



# Essays in Asset Pricing and Macroeconomics

## Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:40046568>

## Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

## Share Your Story

The Harvard community has made this article openly available.  
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

# Essays in Asset Pricing and Macroeconomics

A dissertation presented

by

Thummim Cho

to

The Department of Economics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Economics

Harvard University

Cambridge, Massachusetts

May 2017

© 2017 Thummim Cho

All rights reserved.

**Dissertation Advisors:**

John Campbell

Jeremy Stein

**Author:**

Thummim Cho

**Essays in Asset Pricing and Macroeconomics**

This dissertation presents three essays. The first essay finds that investment strategies which generate “alphas” become endogenously risky by acquiring “betas” with respect to shocks that institutional arbitrageurs are exposed to. This essay provides both theoretical and empirical arguments. The second essay finds that exogenous shocks to liquidity demand cause a variation in the reward for aggregate liquidity provision. To draw this conclusion, this essay uses the daily temperature variation within the summers of the late 19th to early 20th century as a novel proxy for shocks to liquidity demand. The third essay finds that the GDP growth following an exogenous tax change that barely failed to become law is zero. This finding supports the narrative approach to the estimation of tax multipliers taken by Romer and Romer (2010).

## Contents

<b>I. Turning Alphas into Betas: Arbitrage and Endogenous Risk</b>	<b>2</b>
1. Introduction	3
2. A model of arbitrageurs trading multiple assets	10
3. Empirical test of the model	27
4. Auxiliary evidence for endogenous arbitrage risk	66
5. Conclusion	73
<b>II. Does Liquidity Cause Market Return Reversals? A Natural Experiment</b>	<b>75</b>
6. Introduction	76
7. Historical background, data, and measurement	80
8. Temperature effect on aggregate liquidity demand	95
9. Aggregate liquidity demand and market return reversals	102
10. Conclusion	118
<b>III. The Unlegislated Tax Multiplier</b>	<b>120</b>
11. Introduction	121
12. Why are barely failed tax bills useful?	123
13. Data on unlegislated tax changes	124

14. The legislated tax multiplier	130
15. The unlegislated tax multiplier	133
16. Conclusion	136
<b>IV. Appendix to Chapter I</b>	<b>137</b>
A. Model appendix	138
B. Empirical appendix	145

## Acknowledgments

I am deeply indebted to my committee members John Campbell, Jeremy Stein, Samuel Hanson, and Adi Sunderam. I am also grateful to Andrei Shleifer for his guidance and support and to my Ph.D. colleagues Will Diamond, Yosub Jung, Dyvia Kirti, Seunghyup Lee, Jing Li, Gordon Liao, Jeff Picel, Gleb Romanyuk, Geahyun Shin, Argyris Tsiaras, Fernando Yu, and many others for their friendship. I also thank the faculty at the Korea Military Academy for allowing me to conduct part of this research as a faculty officer.

To my parents, Dr. Kyungkeun Cho and Meehee Shin, my fiancé, Na Eun Kim,  
and God.

*"Who shall separate us from the love of Christ?" (Romans 8:35)*



# **Chapter I.**

## **Turning Alphas into Betas: Arbitrage and Endogenous Risk**

# 1. Introduction

Asset pricing “anomalies” are investment strategies with high expected returns but low identifiable risks. These anomalies—such as value and momentum—first gained widespread recognition among finance academics and investment managers in the early 1990s.<sup>1</sup> Since then, arbitrageurs such as hedge funds have allocated growing amounts of capital to these anomalies. As a result, the abnormal returns on these anomalies have fallen, but have not completely disappeared.<sup>2</sup>

What prevents arbitrageurs from completely eliminating anomaly returns? Are anomalies commonly exposed to hidden fundamental risks, so that the remaining anomaly returns represent fair compensation for these hard-to-measure risks? Or have anomalies become increasingly exposed to “endogenous risks” because of the very fact that many arbitrageurs are attempting to exploit them?

In this paper, I argue both theoretically and empirically that arbitrage activity exposes asset pricing anomalies to endogenous risks associated with the act of arbitrage. The emergence of these endogenous risks means that anomaly returns may survive in equilibrium even when the amount of arbitrage capital becomes large.

My key contribution is to draw out the implications of this endogenous risk view for the cross-section of “anomaly assets,” long-short portfolios that exploit asset pricing anomalies. Specifically, if demand curves for anomaly assets slope downward, the prices of anomaly assets comove with shocks to arbitrageur capital. This endogenous comovement is especially large for an anomaly with a large latent mispricing—abnormal return in the absence of arbitrageurs—since it at-

---

<sup>1</sup>Fama and French (1992, 1993, 1996) and Jegadeesh and Titman (1993) ignited this interest. However, anomalies such as size (Banz, 1981) and value (Rosenberg, Reid, and Lanstein, 1985) were documented earlier.

<sup>2</sup>McLean and Pontiff (2016) find an average 32% decline in the returns of 97 anomalies after their publication. Chordia, Subrahmanyam, and Tong (2014) also find that anomaly returns have not completely disappeared.

tracts correspondingly more arbitrage capital. This way, arbitrage turns assets with high “alphas” into assets with high endogenous “betas.” This key insight allows me to develop cross-sectional tests that have more power than pure time-series tests.

To formalize my argument, I develop a model in which arbitrageurs can exploit many anomalies that differ solely in the degree of the latent mispricing that would prevail without arbitrageurs. In my three-period model, there is a continuum of anomaly assets with the same expected cash flow at the final date (time 2). A set of behavioral investors have a downward-sloping demand curve for each anomaly asset at times 0 and 1. Critically, behavioral investors undervalue the anomaly assets’ cash flow, and the degree of undervaluation—the latent mispricing—differs across anomalies.

Arbitrageurs in my model are risk-neutral but face a stochastic funding constraint at time 1 that generates exogenous variation in the capital that they can deploy. As arbitrageur funding at time 1 improves, arbitrageurs devote greater capital to anomaly assets, raising their equilibrium prices. From the perspective of arbitrageurs at time 0, the existence of the stochastic funding constraint means that anomaly prices comove with their capital at time 1. And, crucially, this comovement is stronger for anomalies with higher degrees of latent mispricing. Since arbitrageurs want to hedge their time-1 capital shocks, this makes the more mispriced anomalies endogenously riskier for arbitrageurs to hold at time 0 (Merton, 1973). As a result, in equilibrium, anomalies with greater latent mispricing must offer higher endogenous risk compensation from time 0 to time 1.

In summary, in my model, arbitrage activity necessarily exposes anomaly assets to endogenous risks. These endogenous risks mean that anomaly returns persist in equilibrium. And these endogenous risks imply that an “intermediary asset pricing” model can explain the cross-section of anomaly expected returns: they line up with the anomaly’s exposure to arbitrageur funding shocks even though

the anomaly asset has no fundamental link to those shocks.

The model makes three key predictions about the cross-section of anomaly assets. First, anomalies with greater latent mispricing become more exposed to endogenous risk—i.e., larger  $\alpha$ s turn into larger  $\beta$ s with respect to the funding conditions of arbitrageurs (Proposition 1). Second, this endogenous risk of an anomaly is explained by the amount of arbitrageur capital dedicated to the anomaly—i.e., arbitrageur funding  $\beta$ s line up with anomaly-level measures of arbitrage activity (Proposition 2). Third, an anomaly’s exposure to endogenous risk explains its expected return in equilibrium—i.e., arbitrageur funding  $\beta$ s can “price” the cross-section of anomalies (Proposition 3).

I test the model’s three main predictions using data on 34 equity anomaly assets from 1972 to 2015. Splitting the sample period in half, I proceed under the assumption that the pre-1993 period featured little arbitrage on anomalies whereas the post-1993 period features more arbitrage. I measure the funding conditions of arbitrageurs using the leverage of security broker-dealers, similar to the measure of financial intermediary funding conditions used in Adrian, Etula, and Muir (2014). For main empirical analyses, I use the generalized method of moments (GMM) to obtain conservative standard errors for the test parameters.

My empirical tests support the model’s predictions. In the pre-93 sample, anomalies generated large long-short returns but had little exposure to arbitrageur funding shocks. In the post-93 sample, however, as arbitrageur capital has flowed into anomaly assets, anomaly returns have fallen while their endogenous exposures to arbitrageur funding shocks have risen. And, consistent with the cross-sectional prediction of Proposition 1, an anomaly’s latent mispricing—its pre-93 return—predicts its subsequent endogenous risk—its post-93 beta with respect to arbitrageur funding). Furthermore, as predicted by Proposition 2, these post-93 funding betas are explained by anomaly-specific arbitrage capital inferred from short interests.

Consistent with the intermediary asset pricing logic of Proposition 3, the post-93 expected returns of different anomalies line up with their endogenous risks measured using post-93 betas with respect to arbitrageur funding. The intercept of the cross-sectional regression is *not zero* but positive, which is predicted by the model: anomaly assets generate risk-adjusted returns above the risk-free rate whenever arbitrage capital is insufficient to price all anomaly assets correctly. Interestingly, the price of risk estimated in the pooled period of 1972-2015 is larger than that estimated in the post-93 period, when arbitrageurs have become more important. This is because anomalies with large funding betas and large equilibrium returns in the post-93 period had even larger returns in the pre-93 period before arbitrage began. The funding betas of all anomalies, however, were close to zero in the pre-93 period. As a result, pooling the two periods *increases* the spread of returns and *decreases* the spread of betas, generating an upward bias in the estimated price of risk.

Additional empirical tests support auxiliary implications of the model. First, treating the long and short sides of an anomaly as separate assets, I show that a unit of pre-93 abnormal return turns into a larger post-93 endogenous risk on the short side. This is consistent with the view that short sides of anomalies are primarily traded by leveraged arbitrageurs such as hedge funds but long sides of anomalies are accessible to a wider set of investors. Second, I find that the covariation between anomaly returns and arbitrageur funding conditions occurs only when arbitrageurs are likely to be constrained, consistent with the model's prediction that arbitrageurs exert price pressure on anomalies only when they are constrained (Proposition 4). Finally, I show that using the equity market-neutral index return from Hedge Fund Research (HFR) to measure shocks to arbitrageur capital delivers results similar to those obtained using my proxy for arbitrageur funding shocks.

What alternative explanations might account for my main findings? Suppose

that anomalies with high average returns are exposed to some fundamental—as opposed to endogenous—risk factor that, for whatever reason, has become more correlated with arbitrageur funding shocks in recent years. In this case, an anomaly’s pre-93 mean return would appear to predict its post-93 arbitrageur funding beta, generating my key “ $\alpha$ s into  $\beta$ s” result. If anomalies’ returns were driven by fundamental risks of this sort, one would expect the underlying firms’ cash flows to covary with arbitrageurs funding shocks. To examine this possibility, I examine whether anomaly assets’ cash flows covary with arbitrageur funding shocks, using the return on book equity to measure cash flows as in Campbell and Vuolteenaho (2004), and find no evidence that the anomaly assets have fundamental cash-flow exposures to arbitrageur funding shocks.

**Implications for the literature.** This paper tests cross-sectional predictions of the idea that the act of arbitrage makes mispriced assets endogenously risky. First formalized by Shleifer and Vishny (1997),<sup>3</sup> this idea has been a central explanation for the occurrence of apparent arbitrage opportunities and has been extended to show that the act of arbitrage induces various forms of instability in financial markets.<sup>4</sup> Empirical tests of the idea have relied on time-series variation in arbitrage capital and the ability of an arbitrageur-related risk factor to explain anomaly returns. For instance, Frazzini and Pedersen (2014) show that the “betting against market beta” portfolio realizes a low return when funding constraints

---

<sup>3</sup>Although Shleifer and Vishny (1997) use noise traders to generate shocks to arbitrage capital, the specific source of the arbitrage capital shocks is unimportant. As pointed out in Shleifer (2000), the endogenous risk arises whenever arbitrageurs depend on external (debt or equity) capital, which prevents them from raising more capital when their capital level falls and the mispricing that they bet against widens.

<sup>4</sup>Documented examples of apparent arbitrage opportunities include price divergence in Siamese-twin stocks (Rosenthal and Young, 1990; Froot and Dabora, 1999), negative stub values (Mitchell, Pulvino, and Stafford, 2002; Lamont and Thaler, 2003), and on-the-run vs. off-the-run bond spreads (Amihud and Mendelson, 1991; Warga, 1992; Krishnamurthy, 2002). The instabilities include contagion (Kyle and Xiong, 2001), fire sales (Gromb and Vayanos, 2002; Morris and Shin, 2004; Allen and Gale, 2005), liquidity spirals (Brunnermeier and Pedersen, 2009), and crash risks (Stein, 2009).

tighten, highlighting the endogenous link between the amount of arbitrage capital and the prices of arbitrated assets.<sup>5</sup> Drechsler and Drechsler (2016) show that the cheap-minus-expensive-to-short (CME) portfolio, interpreted as the portfolio of arbitrageurs who focus on shorting, explains the returns on eight prominent equity anomalies.<sup>6</sup> My paper complements these findings by testing a set of new cross-sectional implications of the endogenous-arbitrage-risk idea.

This paper's findings suggest that, from arbitrageurs' point of view, the equity anomalies represent mispricings turned into endogenous risks, contributing to the debate on the nature of asset pricing anomalies.<sup>7</sup> This complements the time-series evidence that anomaly returns have decayed due to increased arbitrage activity following improved liquidity (Chordia, Subrahmanyam, and Tong, 2014) and academic publication (McLean and Pontiff, 2016), as well as the evidence that the return correlation between the top and bottom deciles of an anomaly falls after its academic publication (Liu, Lu, Sun, and Yan, 2015). Although my empirical tests focus on equity anomalies, my predictions apply to other asset classes. Interestingly, Brunnermeier, Nagel, and Pedersen (2009) observe that more profitable currency carry trades are subject to higher crash risks because they attract more arbitrage capital.

Finally, this paper proposes the origin of intermediary asset pricing betas. Intermediary asset pricing theories posit that, in the presence of financial frictions, shocks specific to financial intermediaries carry a risk premium (Gertler and Kiy-

---

<sup>5</sup>See Gârleanu and Pedersen (2011), Chordia, Subrahmanyam, and Tong (2014), Akbas, Armstrong, Sorescu, and Subrahmanyam (2015), and Huang, Lou, and Polk (2016) for additional time-series evidence.

<sup>6</sup>As I discuss in Section 2, they also solve a model of arbitrageurs that shares many similarities to mine.

<sup>7</sup>See, e.g., Fama and French (1993); Lakonishok, Shleifer, and Vishny (1994); Daniel and Titman (1997); Davis, Fama, and French (2000); Campbell, Polk, and Vuolteenaho (2010); and Kozak, Nagel, and Santosh (2015). Some papers attribute anomaly returns to transaction costs (e.g., Korajczyk and Sadka, 2004; Novy-Marx and Velikov, 2016), although there is also an opposing view (e.g., Frazzini, Israel, and Moskowitz, 2015).

otaki, 2010; He and Krishnamurthy, 2012, 2013; and Brunnermeier and Sannikov, 2014). Adrian, Etula, and Muir (2014) test this empirically, finding that intermediary funding shocks inferred from the leverage of broker-dealers explain the returns on equity portfolios sorted by size, value, and momentum and bond portfolios sorted by maturity.<sup>8</sup> However, existing work on intermediary asset pricing offers no explanation on the origin of betas—why some assets have larger exposures to financial sector shocks than others. I show that certain assets have high betas with respect to financial sector shocks because those assets have large latent mispricing and attract large arbitrage capital.

**Outline.** The paper proceeds as follows. Section 2 theoretically examines a model of arbitrageurs exploiting differently mispriced anomalies subject to a stochastic funding constraint. Section 3 empirically tests the model’s implications using the cross-section of anomaly assets. Section 4 presents additional empirical analyses. Section 10 concludes.

---

<sup>8</sup>He, Kelly, and Manela (2016) and Kozak, Nagel, and Santosh (2015) also test intermediary asset pricing.



## 2. A model of arbitrageurs trading multiple assets

In this section, I develop a model in which arbitrageurs can exploit many different anomalies that differ solely in the degree of the latent mispricing that would prevail without arbitrageurs. The model shows that arbitrage activity exposes anomaly assets to endogenous risks and that this endogenous risk is higher for an anomaly with greater latent mispricing. The model generates additional testable predictions about the cross-section of anomaly assets.

### 2.1. Model setup

**Time horizon, assets, and investors.** Consider an economy with three time periods,  $t = 0, 1, 2$ . The economy has two types of assets: a risk-free asset and a continuum of risky assets which I call “anomaly assets.” The risk-free asset is in infinite supply with zero interest rate. An anomaly asset, indexed by  $j \in [0, 1]$ , is a claim to an expected time 2 cash flow of

$$v > 0 \tag{1}$$

and is in zero net supply. The assumption of no cash-flow news at  $t = 0, 1$  and the zero risk-free rate normalization imply that the fundamental value of the anomaly assets to risk-neutral investors is always  $v$ .

There are two types of investors: “arbitrageurs” and “behavioral investors.” Behavioral investors generate mispricings in anomaly assets. They require, for an exogenous reason, positive expected returns for holding the anomaly assets, generating a downward price pressure. Risk-neutral arbitrageurs recognize that the anomaly assets have a fundamental value of  $v$  and trade against mispricings.

**Mispricing.** Anomaly assets differ only in the extent of behavioral investors’ mispricing. At each  $t \in \{0, 1\}$ , behavioral investors’ demand for anomaly asset  $j$ ,

in units of wealth, is

$$B_{j,t} = \frac{E_t [r_{j,t+1}]}{\bar{r}} - j, \quad (2)$$

where  $E_t [r_{j,t+1}]$  denotes asset  $j$ 's conditional expected return and  $\bar{r}$  is a positive constant denoting the most-mispriced asset's expected return in the absence of arbitrage.

This demand curve implies that (i) behavioral investors require a positive expected return for holding an anomaly asset and that (ii) an asset's abnormal return, given any fixed amount of counteracting arbitrage position, increases with the index  $j$ . To see (i), to clear the market with just the behavioral investors ( $B_{j,t} = 0$ ), the expected return on any asset  $j$  must be positive:

$$E_t [r_{j,t+1}] | (B_{j,t} = 0) = \bar{r}j > 0 \quad (3)$$

In the rest of the paper, I refer to this expected return in the absence of arbitrage capital as the anomaly's "latent mispricing." To see (ii), suppose now that arbitrageurs take the same wealth position  $x > 0$  on each anomaly asset. Then, to clear the market ( $x + B_{j,t} = 0$ ), asset  $j$ 's expected return is  $\bar{r}(j - x)$ :

$$E_t [r_{j,t+1}] | (x + B_{j,t} = 0) = \bar{r}(j - x) \quad (4)$$

The derivative of  $\bar{r}(j - x)$  with respect to  $j$  is  $\bar{r} > 0$ , implying that an asset's expected return, holding arbitrage position fixed, increases with  $j$ .

Within each anomaly asset, the market-clearing expected return falls as arbitrage position increases. The slope of (4) with respect to  $x$  is  $-\bar{r}$  for all assets, implying that the marginal effect of arbitrage position on the expected return is the same for all anomalies.

**Arbitrageurs.** The economy has a continuum of identical, risk-neutral arbitrageurs with aggregate mass  $\mu$ . They live through all three periods and seek to maxi-

mize their expected wealth at time 2. Arbitrageurs have limited capital. At time  $t \in \{0, 1\}$ , an arbitrageur's deployable capital  $k_t$  is the sum of its own wealth  $w_t$  and a short-term funding  $f_t$ :

$$k_t = w_t + f_t \quad (5)$$

The wealth evolves according to

$$w_t = w_{t-1} + \int_0^1 r_{j,t} x_{j,t-1} dj, \quad (6)$$

where  $r_{j,t}$  denotes the return on asset  $j$  at time  $t$  and  $x_{j,t}$  is the arbitrageur's position on asset  $j$ . I normalize the time-0 wealth of an individual arbitrageur to be  $w_0 = 1$  so that  $\mu$  is the aggregate arbitrageur wealth at time 0.

Short-term (uncollateralized) funding is available to each arbitrageur at the risk-free rate of zero but is capped at a stochastic funding constraint  $f_t$ . Since part of the funding not used for an arbitrage activity can be invested at the zero interest rate, I assume for notational convenience that arbitrageurs always borrow to the limit  $f_t$ . As we will see, the aggregate arbitrage capital  $\mu k_t$  ("arbitrage capital") will be the only state variable in the model.<sup>9</sup>

Arbitrageurs can take long or short positions on anomaly assets. However, they are required to put up a margin of one for each trade, which prevents them from leveraging up through a long-short trade.<sup>10</sup> An arbitrageur's capital constraint is, therefore,

$$\int_0^1 |x_{j,t}| dj \leq k_t \quad (7)$$

Arbitrageurs' time-1 wealth may become negative. In this case, arbitrageurs are

---

<sup>9</sup>To clarify,  $\mu$  captures the mass of arbitrageurs that changes over a long horizon;  $\mu = 0$  indicates the period before extensive arbitrage and  $\mu > 0$  indicates the period with extensive arbitrage. In contrast, variation in  $k_t$  captures the amount of arbitrageur capital that varies over a short horizon during which the mass of arbitrageurs  $\mu$  is fixed. Hence, for instance, I am assuming that  $k_t$  was low during the recent financial crisis although the mass of arbitrageurs  $\mu$  remained constant.

<sup>10</sup>This ensures that the short-term funding is the only channel for leveraging up.

assumed to exit the economy, taking full responsibility for the liability incurred (unlimited liability) and paying any additional costs of default.

**Equilibrium.** Prices are determined in a competitive equilibrium. Arbitrageurs make optimal investment decisions taking current prices and their expectations of future prices as given, and those prices clear all asset markets. Formally, an equilibrium is defined as follows:

**Definition 1.** An *equilibrium* is the price functional  $\mathbf{p}$ , arbitrageur position  $\mathbf{x}$ , and behavioral investor demand  $\mathbf{B}$  such that

(i)  $\mathbf{x}$  is a solution to the arbitrageur's optimization problem given price  $\mathbf{p}$  and capital constraint (7).

(ii) Price  $\mathbf{p}$  clears the market:  $\mu\mathbf{x} + \mathbf{B} = 0$ .

I solve for equilibrium prices iteratively, beginning with time 1 and moving to time 0. I look for a symmetric equilibrium in which all individual arbitrageurs make identical choices.

**Remarks on modeling choices.** This model is a simple way to deliver intuitions on how arbitrageurs trade multiple anomaly assets and what testable predictions this generates. Most of the modeling choices, however, are not crucial, and alternative specifications generate similar results.

The risk neutrality of arbitrageurs is one such assumption. I use risk neutrality for modeling purposes for two reasons: it most clearly highlights the emergence of endogenous risk and it is the framework used in Shleifer and Vishny (1997) and Brunnermeier and Pedersen (2009), important precursors to my model. Under the risk neutrality assumption, the anomaly assets offer pure arbitrage opportunities in the absence of arbitrage capital since they generate expected returns above the risk-free rate. However, once arbitrageurs trade the anomaly assets with a nonnegligible amount of capital, they cause the prices of the assets to comove

with the level of arbitrage capital, and this endogenous comovement becomes risk through arbitrageurs' intertemporal hedging motive (Merton, 1973). Crucially, a more-mispriced anomaly becomes endogenously riskier since its greater exposure to arbitrage capital makes it a worse instrument for hedging (that is, it realizes a worse return than other assets when arbitrage capital falls and investment opportunities improve).<sup>11</sup>

With risk aversion, a more-mispriced anomaly still becomes endogenously riskier, but the mechanism is different. Under log utility, for instance, an asset's risk is measured by its beta with respect to portfolio return. Since a more-mispriced anomaly offers a larger expected return, arbitrageurs assign a larger portfolio weight to the anomaly, which gives it a larger beta with respect to the portfolio return.<sup>12</sup> In this way, different betas arise because of the different portfolio weights arbitrageurs assign to differently mispriced assets, and these betas explain the expected returns the anomaly assets earn in equilibrium. The model in the online appendix of Drechsler and Drechsler (2016) has this feature and shows a positive relationship between the degree of an asset's underpricing and its beta with respect of arbitrageur portfolio return.

The source of variation in arbitrage capital ( $\mu k_t$ ) in this model is the stochastic funding constraint of arbitrageurs ( $f_t$ ). However, any alternative source of shock that generates variation in arbitrageur capital—the key state variable in the model—generates analytically identical results. For instance, instead of a shock to the constraint on uncollateralized borrowing, one may use a shock to an arbitrageur's margin requirement by making it stochastic (Gârleanu and Pedersen,

---

<sup>11</sup>When the investor's relative risk aversion  $\gamma$  is below 1, as in the case of risk-neutrality, a hedge asset is the one whose return covaries positively with investment opportunities. This is because the speculative motive dominates.

<sup>12</sup>In the language of Shleifer and Vishny (1997), I assume that different anomaly assets are subject to different *levels* of pessimistic sentiments but not different *volatilities* of sentiment. This way, all anomaly assets would have the same fundamental risks (inherent volatilities), but they attain different endogenous risks (betas with respect to arbitrageur portfolio return).

2011; Brunnermeier and Pedersen, 2009). Or one may shut down the borrowing channel altogether and use a shock to arbitrageur wealth ( $w_t$ ) owing to interim cash-flow news, noise trades (that is, stochastic behavioral investor sentiment; the current model has constant sentiment), or stochastic investor flows to generate arbitrage capital shocks.<sup>13</sup>

To describe behavioral investors, I use demand curves rather than more primitive preferences. This allows me to abstract from the underlying cause of a mispricing, which is irrelevant for the rest of the analysis.<sup>14</sup> The demand curves may be generalized to have different parameters govern the latent mispricing ( $\bar{r}_j$ ) and the marginal effect of arbitrageur position on expected return ( $\bar{r}$ ). Introducing a new parameter for this purpose does not affect the model's analytical results.<sup>15</sup>

## 2.2. Two benchmark scenarios: No arbitrage and a complete arbitrage

Before considering the more interesting case of limited arbitrage due to endogenous risks, I consider two benchmark scenarios.

---

<sup>13</sup>Shleifer and Vishny (1997) emphasize the wealth channel of arbitrageur capital. Although not emphasized, the same wealth channel exists in this model; a negative shock to funding  $f_t$  also generates a negative wealth shock  $w_t$  by lowering the values of anomaly assets in the arbitrageur's portfolio. The difference is that the source of a negative wealth shock is the funding condition of arbitrageurs rather than the sentiment of noise traders.

<sup>14</sup>Still, in the Appendix, I provide one way to endogenize the demand curves through heterogeneous beliefs.

<sup>15</sup>Furthermore, although the demand curves are stated in terms of required expected returns, they can be restated in a more conventional form with prices on the left-hand side:

$$p_{j,t} = \frac{E_t [p_{j,t+1}]}{1 + \bar{r} (j + D_{j,t})}$$

### **The no-arbitrage case ( $\mu k_t \leq 0$ )**

The first is the “no-arbitrage” case in which arbitrageurs have zero or negative aggregate capital at all times ( $\mu k_0, \mu k_1 \leq 0$ ). As analyzed during the model setup, this induces the behavioral investors alone to price all assets, and the anomaly assets earn expected returns equivalent to their latent mispricings,  $E_t[r_{j,t+1}] = \bar{r}j$ . The anomaly asset prices at time 0 and time 1 are  $p_{j,0} = v / (1 + \bar{r}j)^2$  and  $p_{j,1} = v / (1 + \bar{r}j)$ , respectively. The prices are deterministic and do not depend on the specific realization of arbitrageur capital.

### **The complete-arbitrage case ( $\mu k_t \geq 1/2$ )**

At the other extreme is the “complete-arbitrage” case. Since arbitrageurs are risk-neutral, they competitively push all expected returns to zero when aggregate arbitrageur capital is large. If this is guaranteed to happen at times 0 and 1, the prices of anomaly assets equal their fundamental value  $v$ :  $p_{j,0} = p_{j,1} = v$ . The prices are deterministic and do not depend on the specific realization of arbitrage capital.

A complete arbitrage occurs if aggregate arbitrage capital  $\mu k_t$  is  $1/2$  or above almost surely both at time 0 and time 1. According to (4), the arbitrageur position required to push asset  $j$ 's expected return to zero is  $j$ . Integrating this over all assets,  $\int_0^1 j dj$ , gives  $1/2$  as the aggregate arbitrage capital required to push all assets' expected returns to zero.

The complete-arbitrage case seems to arise in the actual stock market. These are times when arbitrage capital is persistently sufficient to counteract all mispricings. In these times, anomaly assets have no endogenous risks and generate zero risk-adjusted returns. This point is reiterated theoretically in Proposition 4 and analyzed empirically in Section 3.3.

## 2.3. Limited arbitrage of multiple assets and the emergence of betas

Now I consider the more interesting case in which arbitrage capital may not be sufficient for a complete arbitrage at time 1. I first show that anomaly asset prices at time 1 comove with arbitrage capital due to arbitrageur trading. This makes the anomaly assets ex-ante risky from the perspective of arbitrageurs at time 0. This risk is larger for an anomaly asset with a larger latent mispricing, as it is expected to comove more strongly with arbitrage capital at time 1.

### Equilibrium price at time 1 and endogenous risk

I first determine the prices of anomaly assets at time 1. Since arbitrageurs are risk-neutral, the expected return on any asset arbitrageurs hold will be the same, while the expected return on any asset arbitrageurs do not hold will be lower. This implies that there is a marginal asset. This marginal asset is determined as the point where the amount of capital needed to push down the expected return on all exploited anomalies up to the latent mispricing of the marginal asset is the amount of capital arbitrageurs have.

Let  $j_1^* \in [0,1]$  be the marginal asset. Since latent mispricing increases with  $j$ , arbitrageurs hold assets  $(j_1^*, 1]$  and earn expected returns  $\bar{r}j_1^*$  from them. This expected return implies that behavioral investors take a position  $B_{j,1} = j_1^* - j < 0$  on asset  $j \in (j_1^*, 1]$ , meaning arbitrageurs have a position  $x_{j,1} = -B_{j,1} = j - j_1^* > 0$  on  $j \in (j_1^*, 1]$ . Integrating this position of arbitrageurs over all exploited assets gives the amount of capital arbitrageurs must have to make  $j_1^*$  the marginal asset:  $\int_{j_1^*}^1 (j - j_1^*) dj = \frac{1}{2} (1 - j_1^*)^2$ . Equating this with the actual capital of arbitrageurs,  $\mu k_1$ , gives the marginal asset when aggregate arbitrageur capital is in the intermediate region ( $\mu k_1 \in [0, 1/2]$ ). Below this region, no arbitrage occurs, so  $j_1^* = 1$ . Above this, arbitrageur capital has no further correcting role in anomaly assets,



and  $j_1^* = 0$ . The unexploited assets  $[0, j_1^*)$  generate expected returns equal to their latent mispricings.<sup>16</sup>

In summary, an anomaly asset's equilibrium expected return at time 1 is

$$E_1 [r_{j,2}] = \frac{v}{p_{j,1}} - 1 = \begin{cases} \bar{r}j_1^* & \text{if } j \geq j_1^* \\ \bar{r}j & \text{if } j \leq j_1^* \end{cases}, \quad (8)$$

where  $j_1^*$  is the marginal asset given by

$$j_1^* = \begin{cases} 1 & \text{if } \mu k_1 < 0 \\ 1 - \sqrt{2\mu k_1} & \text{if } \mu k_1 \in [0, \frac{1}{2}] \\ 0 & \text{if } \mu k_1 > \frac{1}{2} \end{cases} \quad (9)$$

Since  $E_1 [r_{j,2}] = v/p_{j,1} - 1$ , translating the expected returns into prices gives the following:

**Lemma 1. (Equilibrium price at  $t = 1$ ).** *Equilibrium price of anomaly asset  $j$  at time  $t = 1$  is*

$$p_{j,1} = \begin{cases} \frac{v}{1+rj} & \text{if } j \leq j_1^* \\ \frac{v}{1+\bar{r}j_1^*} & \text{if } j \geq j_1^* \end{cases} \quad (10)$$

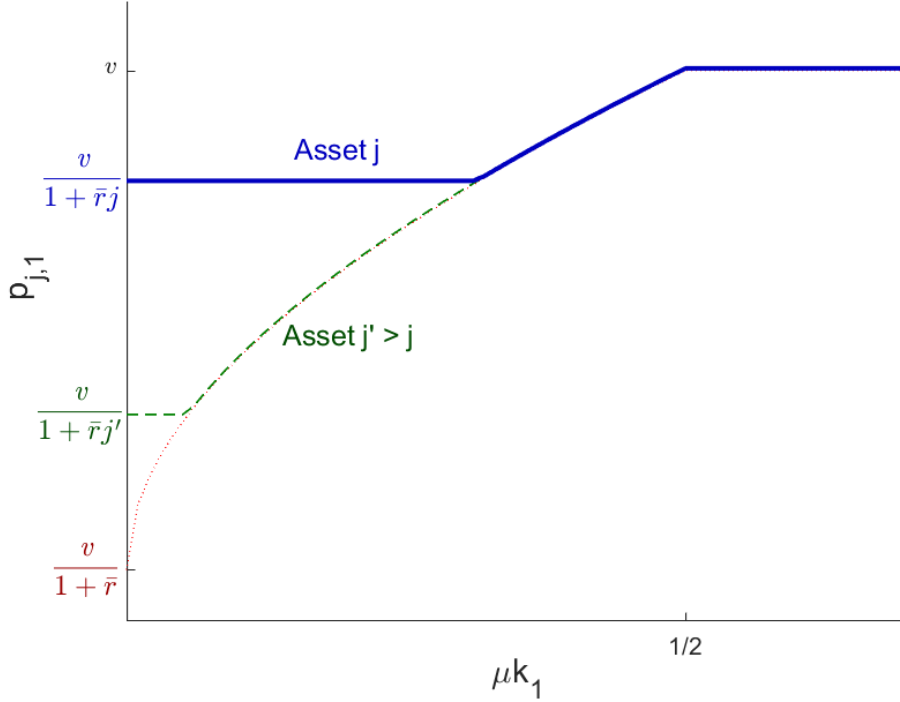
**Proof.** See the Appendix.

Since arbitrageurs equalize expected returns from all exploited assets, the prices of all exploited assets are the same. Figure 1 illustrates the equilibrium time 1 prices of anomaly asset  $j$  and anomaly asset  $j' > j$ .

How does arbitrage capital move the prices of different assets? The *intensive* margin is identical for all assets. When their capital changes, arbitrageurs rebalance their portfolios to ensure that the prices of all exploited assets equal. Hence,

---

<sup>16</sup>That variation in arbitrageur capital has a meaningful effect on asset prices only in the intermediate region of capital is an important feature of Gromb and Vayanos (2009).



**Figure 1: Price of Anomaly Asset  $j$  at Time 1**

The figure plots the time-1 price of anomaly asset  $j$ ,  $p_{j,1}$ , as a function of total arbitrage capital  $\mu k_1$ . The price is given by  $p_{j,1} = \max \left\{ \frac{v}{1+\bar{r}j}, \frac{v}{1+\bar{r}j_1^*} \right\}$ , where  $j_1^* = \min \{0, 1 - \sqrt{2\mu k_1}\}$  is the marginal asset determined by the availability of arbitrageur capital  $\mu k_1$ .

a change in arbitrage capital has the same effect on the prices of assets while they are being exploited. The *extensive* margin, however, applies differently. The larger the latent mispricing, the lower the level of capital from which arbitrageurs begin exploiting the asset. This makes the more-mispriced asset comove with arbitrage capital in a wider region of arbitrage capital. As the reader will see, this will make the more-mispriced asset ex-ante riskier since a larger price covariance with arbitrage capital means a more negative price covariance with the arbitrageur marginal value of wealth.

Since arbitrageurs maximize the expected wealth at time 2, their marginal value of wealth—the value of an additional unit of wealth at time 1—is the gross expected return the extra wealth will generate. This means that the gross expected return earned by exploited assets,  $1 + \bar{r}j_1^*$ , is the arbitrageurs' marginal value of wealth at time 1. However, arbitrageurs' marginal value of wealth is not well-

defined if they have negative realized wealth and exit the financial market (Brunnermeier and Pedersen, 2009). I assume that, in the event of a default, an arbitrageur incurs a marginal bankruptcy cost of  $c$  for each additional dollar of default in addition to taking full responsibility for the negative realized wealth. I then impose a restriction on the value of  $c$  to make an additional unit of wealth more valuable in the default region than in any part of the non-default region.<sup>17</sup> Hence, the marginal value of wealth is as follows:

*Remark 1.* (Arbitrageur's marginal value of wealth at  $t = 1$ ). An arbitrageur's marginal value of wealth at time 1, denoted  $\Lambda_1$ , is

$$\Lambda_1 = \begin{cases} 1 + c & \text{if } k_1 < 0 \\ 1 + \bar{r}j_1^* & \text{if } k_1 \geq 0 \end{cases}, \quad (11)$$

where  $j_1^*$  is the marginal asset specified in (9) and where I assume  $c \geq \bar{r}$  so that the marginal value of wealth is higher in the default region.

Thus, marginal value of wealth decreases as arbitrage capital increases. This means that anomaly assets, which covary positively with arbitrage capital, covary negatively with the arbitrageur marginal value of wealth. This makes anomaly assets risky from the perspective of arbitrageurs at time 0. The risk is larger for a more-mispriced asset (higher  $j$ ) with a larger covariance with arbitrage capital. This is summarized as Lemma 2.

**Lemma 2.** (*Anomaly asset's endogenous risk*). *An anomaly asset is risky as indicated by a negative price covariance with the arbitrageur's marginal value of wealth:*

$$\text{Cov}_0(p_{j,1}, \Lambda_1) \leq 0 \quad \forall j \quad (12)$$

---

<sup>17</sup>Without this assumption of  $c \geq \bar{r}$ , the marginal value of wealth is lower in the default region than in some parts of the non-default region. This could make an asset that pays low in the state of default and pays high in the state of non-default (e.g., the most mispriced asset  $j = 1$ ) safer than an asset that pays the same return in all states (the least-mispriced asset  $j = 0$ ).

Furthermore:

(i) This risk is endogenous, arising only if arbitrageurs have a positive mass in the market so as to generate price pressure:

$$\text{Cov}(p_{j,1}, \Lambda_1) |_{\mu=0} = 0 \quad \forall j \quad (13)$$

(ii) In the cross-section of assets  $j \in [0,1]$ , the riskiness increases with an asset's latent mispricing:

$$\frac{\partial \text{Cov}(p_{j,1}, \Lambda_1)}{\partial (\bar{r}_j)} = \frac{\frac{\partial \text{Cov}(p_{j,1}, \Lambda_1)}{\partial j}}{\frac{\partial (\bar{r}_j)}{\partial j}} \leq 0 \quad (14)$$

**Proof.** See the Appendix.

### Equilibrium price at time 0 and ex-ante pricing of endogenous risk

To find anomaly asset prices at time 0, I first find the arbitrageur's value function. Since each individual arbitrageur is small and risk-neutral, the arbitrageur's value function at time 0 is simply wealth multiplied by the marginal value of wealth,  $\Lambda_0 w_0$ . Since there is no time-0 consumption or discount, this quantity has to equal the time-0 expectation of wealth multiplied by the marginal value of wealth at time 1:

$$\Lambda_0 w_0 = E_0 \left[ \Lambda_1 \left( \int_0^1 \frac{p_{j,1}}{p_{j,0}} x_{j,0} dj + w_0 - \int_0^1 x_{j,0} dj \right) \right] \quad (15)$$

An arbitrageur then maximizes this value function subject to a capital constraint,

$$\int_0^1 |x_{j,0}| dj \leq k_0 \quad (16)$$

I analyze the equilibrium price in the unconstrained and constrained cases separately.

Suppose first that  $k_0$  is large enough to make constraint (16) slack and arbitrageurs unconstrained. Then, taking the derivative of both sides of (15) with respect to  $x_{j,0}$  gives  $E_0[\Lambda_1] = E_0[\Lambda_1 p_{j,1}/p_{j,0}]$ . Furthermore, taking the derivative with respect to  $w_0$  gives  $\Lambda_0 = E_0[\Lambda_1]$ .<sup>18</sup> Hence,

$$p_{j,0} = E_0 \left[ \frac{\Lambda_1}{E_0[\Lambda_1]} p_{j,1} \right] \quad (17)$$

for all assets  $j \in [0,1]$ . Since arbitrageur trading at time 1 makes  $Cov_0(\Lambda_1 p_{j,1}) < 0$  for  $j \in (0,1]$ , even if arbitrageurs are unconstrained at time 0, they do not push the price  $p_{j,0}$  all the way to  $E_0[p_{j,1}]$ .

Suppose now that  $k_0$  is small, in which case the constraint (16) binds and arbitrageurs are constrained. Then, by (5) and (16),  $w_0 = k_0 - f_0 = \int_0^1 x_{j,0} dj - f_0$ , where I use the fact that arbitrageurs have non-negative exposures to anomaly assets in equilibrium. Substituting  $w_0$  with  $\int_0^1 x_{j,0} dj - f_0$  and taking the derivative of the both sides of (15) with respect to  $x_{j,0}$  gives the optimality condition  $\Lambda_0 \geq E_0[\Lambda_1 p_{j,1}/p_{j,0}]$ , which holds with equality if and only if the asset is exploited by arbitrageurs at time 0 and thus has an interior solution in the arbitrageur's optimization problem. Thus, the price of an exploited asset satisfies the fundamental theorem of asset pricing:

$$p_{j,0} = E_0 \left[ \frac{\Lambda_1}{\Lambda_0} p_{j,1} \right] \quad (18)$$

On the other hand, an unexploited asset is priced solely by behavioral investors, who require an expected return of  $\bar{r}_j$ :

$$p_{j,0} = \frac{E_0[p_{j,1}]}{1 + \bar{r}_j} \quad (19)$$

Conditions (18) and (19) imply that  $\Lambda_0$  is pinned down by the expectation of

---

<sup>18</sup>This martingale property  $\Lambda_0 = E_0[\Lambda_1]$  is a consequence of the zero risk-free rate assumption.

returns on exploited assets multiplied by the marginal value of wealth at time 1:

$$\Lambda_0 = \max_{j \in [0,1]} E_0 [\Lambda_1 (1 + r_{j,1})] \quad (20)$$

Hence, arbitrageurs are constrained; they are not the marginal investor of all assets. Instead, arbitrageur's stochastic discount factor  $m_1 = \Lambda_1/\Lambda_0$  prices assets only if they are held by arbitrageurs. Thus, unlike conventional pricing models, arbitrageur pricing is expected to work only on assets that are traded by arbitrageurs.

The equilibrium conditions in both the unconstrained and constrained cases imply that anomaly asset prices at time 0 decrease with  $j$ . If  $j' < j''$ , anomaly  $j'$  is not only subject to a larger mispricing in the absence of arbitrageurs, but is also exposed to a larger endogenous risk. Thus, anomaly  $j''$  must be valued less than anomaly  $j'$ .

This monotonicity of prices at time 0 makes the endogenous risk results in Lemma 2 hold analogously with returns. Consider two assets  $j' < j''$  so that  $j''$  has a larger latent mispricing than  $j'$ . Then, asset  $j''$  not only has a more-negative price covariance with the marginal value of wealth at  $t = 1$ , but also has a lower price at time 0. Since gross return is  $1 + r_{j,1} = p_{j,1}/p_{j,0}$ , this necessarily means that asset  $j''$  has a more negative return covariance with the marginal value of wealth at time 1. This gives Proposition 1, which states Lemma 2 in terms of returns:

**Proposition 1.** (*“Alphas” turn into “betas”*). *In the cross-section of anomaly assets, an anomaly asset's latent mispricing,*

$$\alpha_j \equiv \bar{r}j \quad (21)$$

*predicts its endogenous risk measured as the negative of the beta with the arbitrageur*

stochastic discount factor (SDF)  $m_1 \equiv \Lambda_1/\Lambda_0$ :

$$\beta_j \equiv -\frac{\text{Cov}_0(m_1, r_{j,1})}{\text{Var}_0(m_1)} \quad (22)$$

That is,

$$\frac{\partial \beta_j}{\partial \alpha_j} > 0 \quad (23)$$

**Proof.** See the Appendix.

Hence, anomaly assets' risks—their betas with respect to SDF—arise endogenously in this model. That asset betas arise endogenously is not surprising, given that most nontrivial economies with multiple assets would imply different equilibrium risks of the assets.<sup>19</sup> The difference in this model, however, is that the different risks are generated by arbitrage trading.<sup>20</sup> To emphasize this point, I show that the amount of arbitrage capital devoted to an asset is expected to be larger for an asset with a larger  $\beta_j$ . This is presented as Proposition 2.

**Proposition 2.** *(Beta is explained by anomaly-specific arbitrage capital).  $\beta_j$  increases with the expected arbitrageur position in the asset:*

$$\frac{\partial \beta_j}{\partial E_0[x_{j,1}]} \geq 0 \quad (24)$$

**Proof.** See the Appendix.

This suggests that anomaly-specific measures of arbitrage activity should explain different amounts of endogenous risks in different anomaly assets.

---

<sup>19</sup>For instance, Zhang (2005) shows that value firms can have returns covarying more with the SDF than growth firms if the adjustments in the investment-capital ratio are higher for value firms, especially in bad times. Brunnermeier and Pedersen (2009) show that fundamentally more volatile assets covary more with the speculator's SDF since their market liquidity drops more quickly in times of low liquidity.

<sup>20</sup>The emergence of the beta is perhaps most similar to high-margin securities attaining high funding-liquidity risks owing to their large sensitivities to funding liquidity events (Gârleanu and Pedersen, 2011).

Once arbitrageurs generate endogenous risks, they require a compensation for these risks. This implies that an “intermediary asset pricing” should work on anomaly assets if risk is measured by beta with respect to arbitrageur’s SDF. However, because limited capital can constrain arbitrageurs, the model makes a few nonconventional predictions about pricing assets with the arbitrageur SDF. These are summarized as Proposition 3.

**Proposition 3.** (*“Intermediary asset pricing” with respect to arbitrageur’s SDF*). Suppose asset  $j$  is exploited by arbitrageurs at  $t = 0$ . Then, the asset’s beta with arbitrageurs’ stochastic discount factor  $\Lambda_1/\Lambda_0$  explains its expected return:

$$E_0 [r_{j,1}] = \begin{cases} \bar{r}_j & \text{if } j \text{ is not exploited} \\ \underbrace{\frac{1}{E_0 [\Lambda_1/\Lambda_0]} - 1}_{\text{zero-beta rate}} + \underbrace{\lambda \beta_j}_{\text{risk premium}} & \text{if } j \text{ is exploited} \end{cases} \quad (25)$$

where

$$\lambda = \frac{\text{Var}_0 (\Lambda_1/\Lambda_0)}{E_0 [\Lambda_1/\Lambda_0]} \quad (26)$$

$$\beta_j = -\frac{\text{Cov}_0 (r_{j,1}, \Lambda_1/\Lambda_0)}{\text{Var}_0 (\Lambda_1/\Lambda_0)} \quad (27)$$

with  $E_0 [\Lambda_1/\Lambda_0] = 1$  if arbitrageurs are unconstrained ( $k_0$  large) and  $E_0 [\Lambda_1/\Lambda_0] > 1$  if they are constrained ( $k_0$  small).

**Proof.** Rearranging (18) implies that exploited assets are priced according to  $1 = E_0 [\Lambda_1 \Lambda_0^{-1} (1 + r_{j,1})]$ , so that  $1 = E_0 [\Lambda_1/\Lambda_0] E_0 [1 + r_{j,1}] + \text{Cov}_0 (\Lambda_1/\Lambda_0, r_{j,1})$ . This gives

$$E_0 [r_{j,1}] = \frac{1}{E_0 [\Lambda_1/\Lambda_0]} - 1 + \lambda \beta_j$$

If arbitrageurs are unconstrained,  $\Lambda_0 = E_0 [\Lambda_1]$  so that the zero-beta rate drops out. If arbitrageurs are constrained,  $\Lambda_0 = \max_{j \in [0,1]} E_0 [\Lambda_1 (1 + r_{j,1})] >$



$E_0[\Lambda_1]$  since otherwise, arbitrageurs are not optimally choosing the exploited assets.

If arbitrageurs are constrained at time 0, some anomaly assets are exploited by arbitrageurs while others are not. Hence, arbitrageurs are not the marginal investor of all assets, and choosing the exploited assets is important when estimating the asset pricing model; this is in contrast to Adrian, Etula, and Muir (2014), who essentially assume that financial intermediaries are the marginal investor of all assets. The expected return on an unexploited asset is its latent mispricing  $\bar{r}_j$ , the expected return required by behavioral investors. The expected return on an exploited asset has two components: a zero-beta rate that is common across all exploited assets and a risk premium that is different for each exploited asset.

The zero-beta rate is above the risk-free rate of zero at time 0 if arbitrageur capital is insufficient to “price” all anomaly assets correctly. This is because, in the constrained case, arbitrageurs earn expected returns higher than the levels that would just compensate for the risks that they face; that is, their risk-adjusted returns are positive. Hence, unlike a conventional cross-sectional asset pricing model that tests for a zero intercept, this model predicts that if arbitrageurs are sometimes constrained and expose the prices of anomaly assets to comove with their capital—if they form positive  $\beta_j$ —we should also expect to see a positive intercept in the cross-sectional regression.

What is the interpretation of the zero-beta rate? By definition, the difference between the zero-beta rate and the prevailing risk-free rate represents the arbitrage profit arbitrageurs are generating from their investments. Although not pursued in this paper, inferring the zero-beta rate from a cross-sectional regression of anomaly assets and relating it to the estimate of the price of risk may be an interesting route for an arbitrageur-based asset pricing model to take.

Now I introduce Proposition 4, the last proposition of the model. The endogenous risks generated by arbitrageurs arise only if arbitrageurs are “constrained”

at time 1. This is when aggregate arbitrageur capital at time 1 is in the intermediate region  $\mu k_t \in [0, 1/2]$  so that the anomaly assets are mispriced and variation in arbitrage capital generates price pressure on the anomaly assets. In particular, if  $\mu k_t > 1/2$ , all anomaly assets are already correctly priced so that variation in arbitrageur no longer generates price pressure.

**Proposition 4. (*Betas arise during constrained times*).** *Endogenous risk arises only during constrained times of  $t = 1$ . That is,*

$$\begin{aligned} \text{Cov}_0(\Lambda_1, r_{j,1} | \mu k_1 > 1/2) &= 0 \\ \text{Cov}_0(\Lambda_1, r_{j,1} | \mu k_1 < 1/2) &> 0 \end{aligned} \tag{28}$$

for all  $j \in (0, 1]$ . For this reason, if the funding condition follows a process  $f_t^*$  such that  $f_0^* > 1/2$  and  $f_1^* > 1/2$  almost surely, then neither beta nor abnormal return arises:

$$\beta_j = 0 \text{ and } E_0[r_{j,1}] = 0 \text{ for all } j \in [0, 1] \tag{29}$$

**Proof.** Follows from the price equation in Lemma 1 and the analysis in Section 2.2.

Empirically, I should observe that anomaly assets have zero endogenous risks and zero abnormal returns in times when arbitrageurs have persistently large capital.

### 3. Empirical test of the model

In this section, I test the model's predictions about the cross-section of anomaly assets, using equity anomalies as the empirical counterparts of the differently mispriced anomaly assets in the model. Equity anomalies provide a convenient laboratory because they are easier to construct using publicly available data and straightforward to compare to one another.<sup>21</sup>

---

<sup>21</sup>For instance, fixed-income arbitrage portfolios generated by Duarte, Longstaff, and Yu (2007) use proprietary data and require a separate, nontrivial valuation model for each anomaly asset.

### 3.1. Test environment

**Thirty-four equity anomalies as anomaly assets.** The empirical counterpart of the anomaly assets in the model are 34 equity anomaly assets. For each one, I compute the time-series of quarterly value-weighted (VW) returns on a long-short self-financed portfolio over the period 1972 to 2015. The required data are downloaded from CRSP and Compustat.

I compute the long-short returns on each anomaly asset as follows. At the end of each month from 1972 to 2015, I allocate all domestic common shares trading on NYSE, AMEX, and NASDAQ into deciles based on an anomaly signal, such as the book-to-market ratio, with decile breakpoints determined by NYSE-listed stocks alone.<sup>22</sup> Then, I calculate monthly value-weighted long-short return as the difference between the VW returns on the top and bottom deciles of stocks.<sup>23</sup> I aggregate the monthly returns to the quarterly frequency to match the frequency of the arbitrageur funding shock variable discussed below.

I use 34 distinct anomaly signals to construct the anomaly assets. These comprise 25 standard anomaly signals used in Novy-Marx and Velikov (2016), 6 industry-adjusted signals, and 3 “behavioral” signals meant to exploit investors’ behavioral biases.<sup>24</sup> The construction of the signals is similar to that of Novy-Marx and

---

In contrast, equity anomaly portfolios can be readily constructed once the anomaly signals are calculated.

<sup>22</sup>This ensures that the decile portfolios have comparable market capitalizations. For CRSP, domestic common shares on NYSE, AMEX, and NASDAQ are stocks with share code 10 or 11 and exchange code 1, 2, or 3.

<sup>23</sup>Two exceptions are beta arbitrage and idiosyncratic volatility strategies, for which the plain-vanilla long-short portfolios have large negative exposures to the market portfolio. For these two strategies, I compute returns that go long  $\min\{5, \max\{0, \beta^{\text{Bottom Decile}} / \beta^{\text{Top Decile}}\}\}$  dollar of the top decile and short one dollar of the bottom decile, where  $\beta^{\text{Top Decile}}$  and  $\beta^{\text{Bottom Decile}}$  are the value-weighted market betas of the top and bottom deciles. The market beta used here is calculated at the end of each month using weekly returns in the previous one to three years, depending on data availability. The market factor is downloaded from Kenneth French’s website on June 25, 2016.

<sup>24</sup>Out of 32 signals used in Novy-Marx and Velikov (2016), I exclude 7 for redundancy. For example, I exclude the “ValMomProf” signal since it is simply the sum of a stock’s decile numbers

Velikov (2016) and of Green, Hand, and Zhang (2016); where I deviate, it is so that the signals resemble the actual signals arbitrageurs observe at the end of each month. The online appendix provides more details on how I construct the anomaly signals.

Table 1 lists the 34 anomaly assets along with their mean returns, volatilities, and arbitrageur funding betas (which I will come back to once I discuss my proxy for arbitrageur funding shocks) during the first-half of the sample period (1972Q1-1993Q4, “pre-93”) and the second-half (1994Q1-2015Q4, “post-93”). Twenty-nine of these have positive mean returns in the pre-93 sample, which I later use to measure an asset’s latent mispricing—the abnormal expected return that would prevail in the absence of arbitrageurs. For the five anomaly assets with negative mean returns, I will assume that arbitrageurs flip the direction of their trades to earn positive mean returns. There is some variation in the return volatility of the anomalies, and the ones with larger mean returns tend to have larger volatilities. Hence, it may be important to control for volatility in a regression that proxies for an anomaly’s latent mispricing using the pre-93 mean return. I postpone the discussion of the funding betas.

---

in the three univariate sorts based on value, momentum, and profitability.

**Table 1: Summary Statistics of Anomaly Assets by Sample Period**

This table summarizes the 34 equity anomaly assets used in the empirical section. The abbreviations for the categories are: NMV=Novy-Marx and Velikov (2016); Ind.Adj.=industry-adjusted signals; Behavi.=meant to exploit behavioral biases. Sign indicates whether a higher level of the signal indicates a higher (1) or lower (-1) expected return. Return  $\sigma$  indicates the standard deviation of the time-series of returns. All returns except for those of beta arbitrage and idiosyncratic volatility are value-weighted (VW) long-short returns calculated by subtracting bottom-decile VW return from top-decile VW return. Beta arbitrage and idiosyncratic volatility strategies are hedged for their market exposures based on the CAPM betas of the top and bottom deciles estimated from weekly returns in the previous three years.

No	Anomaly	Category	Sign	1972-1993 ("pre-93")			1994-2015 ("post-93")		
				Mean Return	Return $\sigma$	Funding $\beta$	Mean Return	Return $\sigma$	Funding $\beta$
1	Beta arbitrage	NMV	-1	4.15	24.10	11.70	5.17	28.79	22.54
2	Ohlson's O-score	NMV	-1	5.31	37.00	3.94	1.69	34.53	7.58
3	Size	NMV	-1	1.90	43.04	-12.28	2.17	35.51	-9.57
4	PEAD(SUE)	NMV	1	12.60	25.53	-0.88	5.57	25.12	2.98
5	Value	NMV	1	11.53	39.42	0.88	3.92	37.41	11.20
6	36-month momentum	Behavi.	-1	4.45	45.55	-9.56	2.96	37.80	-5.29
7	Long-run reversals	NMV	-1	5.38	44.51	-9.84	2.74	36.49	-7.49
8	Short-term reversals	NMV	-1	-5.13	23.88	-1.39	-3.72	40.20	-26.75
9	Momentum	NMV	1	20.93	45.69	6.31	8.76	52.91	12.86
10	Annual sales growth Ind-adj change in	Behavi.	-1	3.35	27.12	-4.88	1.33	24.48	11.31
11	employees	Ind.Adj.	-1	-1.49	21.74	2.38	4.23	20.49	3.14
12	Accruals	NMV	-1	6.16	19.68	0.07	4.16	24.28	-6.06
13	Ind-adj book-to-market	Ind.Adj.	1	7.69	31.79	-0.95	1.99	26.91	-14.57
14	Industry momentum	NMV	1	2.98	37.07	4.97	1.03	38.02	1.83
15	Ind-adj firm size Ind-adj cash-flow-to-	Ind.Adj.	-1	4.49	21.83	-1.18	1.73	20.91	-0.34
16	price ratio	Ind.Adj.	1	2.17	23.24	-8.13	-2.93	31.99	-4.96
17	Piotroski's F-score	NMV	1	2.05	16.19	2.58	3.62	19.67	-0.11
18	Idiosyncratic volatility	NMV	-1	3.97	26.96	9.37	3.66	27.65	13.85
19	Price delay	Behavi.	1	-1.71	21.88	3.56	-3.48	17.48	3.55
20	Failure probability	NMV	-1	9.70	48.19	7.49	7.66	56.63	31.74
21	Asset growth	NMV	-1	6.27	24.27	-3.06	5.83	25.47	13.80
22	Net issuance	NMV	-1	5.46	18.74	-3.52	8.13	29.96	4.23
23	Seasonality Ind-adj change in profit	NMV	1	12.43	28.82	-0.16	6.59	27.84	2.22
24	margin Ind-adj change in asset	Ind.Adj.	1	0.21	21.99	-5.58	-3.00	23.64	-1.05
25	turnover	Ind.Adj.	1	1.57	18.53	-3.48	5.35	23.10	10.43
26	PEAD(CAR3)	NMV	1	12.09	17.16	4.49	8.14	22.72	-3.60
27	Investment	NMV	-1	7.76	22.37	2.03	4.81	22.20	6.45
28	Return on market equity	NMV	1	16.25	30.84	-4.84	11.12	46.46	22.14
29	Return on book equity	NMV	1	9.30	34.29	-4.44	8.25	42.22	22.21
30	Return on assets	NMV	1	7.47	31.69	2.14	6.24	40.84	17.34
31	Asset turnover	NMV	1	4.04	26.93	-1.42	4.32	32.00	-10.21
32	Gross margins	NMV	1	-2.40	21.71	-3.77	2.39	20.88	-5.80
33	Gross profitability	NMV	1	0.45	26.34	-0.58	5.08	30.10	-9.93
34	Ind-adj reversals	NMV	-1	-3.36	19.81	-1.66	-4.21	33.92	-26.44
Average across anomalies:				5.24	28.47	-0.58	3.57	31.14	2.62
Standard deviation across anomalies:				5.66	9.15	5.41	3.79	9.50	13.20

**Broker-dealer leverage as arbitrageur funding condition.** The correct measure of risk of an anomaly asset is its beta with respect to arbitrageur’s stochastic discount factor (SDF). Since the SDF is unobserved, I look for an empirical proxy for the arbitrageur funding condition  $f_t$ , the variable underlying the variation in arbitrageur’s SDF in the model.<sup>25</sup>

To measure arbitrageur funding shocks, I use shocks to the book leverage of broker-dealers,

$$f_t = \ln \left( \text{Leverage}_t^{BD} \right) - \ln \left( \text{Leverage}_{t-1}^{BD} \right), \quad (30)$$

which Adrian, Etula, and Muir (2014) use to proxy for financial intermediary funding shocks. Here, a high  $f_t$  or a high leverage shock indicates a favorable funding shock for arbitrageurs.<sup>26</sup> The book leverage of broker-dealers is defined as total financial assets net of repo assets divided by the difference between total financial assets and total liabilities.<sup>27</sup> Quarterly data are available from the flow-of-funds data published by the Federal Reserve.<sup>28</sup> In actual analyses, I annualize the funding shock (30) by multiplying by four and winsorize the series at the 1% and 99% levels to mitigate the effects of outliers (this removes the smallest and largest values, both occurring around the recent financial crisis period).<sup>29</sup> Figure 2 plots the original log leverage series and the leverage shock series  $f_t$ .

---

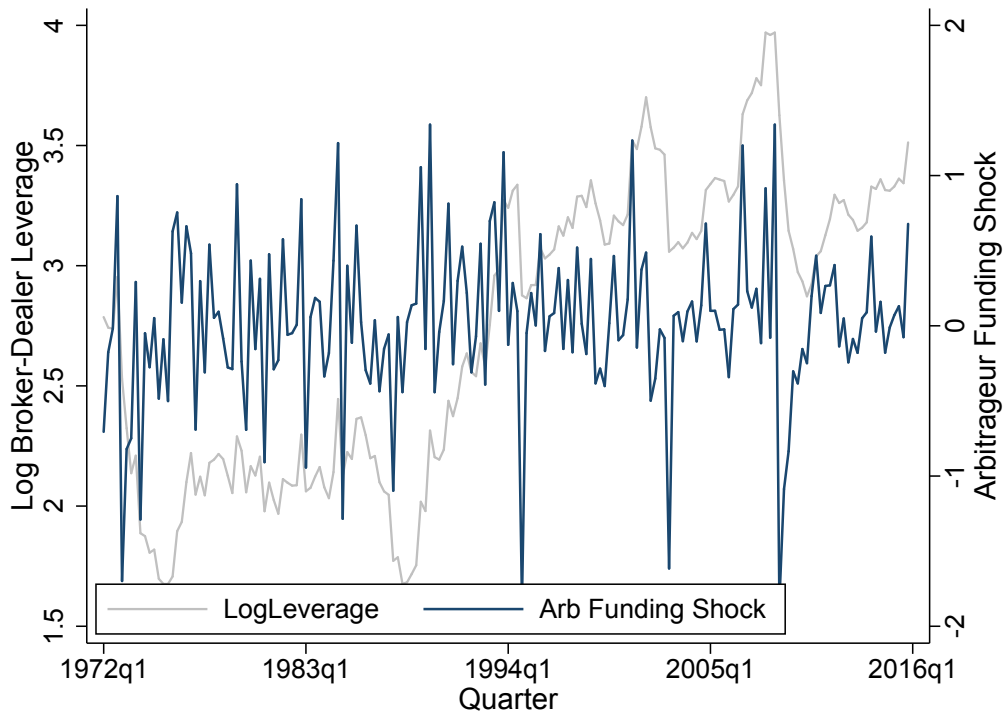
<sup>25</sup>If arbitrageur’s SDF  $m_1$  were approximately linear in the arbitrageur funding condition  $f_1$ , the model’s propositions could be identically stated with the beta with respect to the funding condition  $f_1$ . The approximation would be justified in a conditional model if the arbitrageur funding conditions were expected to vary over a small interval.

<sup>26</sup>There is a slight abuse of notation since  $f_t$  in the model is the *level* of arbitrageur funding, whereas  $f_t$  here in the empirical analysis measures a *shock* to the arbitrageur funding condition.

<sup>27</sup>Hence, the reverse repo (lending money through a repo) is not part of total assets. Instead, the difference between repo borrowing and repo lending (“net repo”) enters into total liabilities. This amounts to assuming that only the relative increase in repo lending (equivalently, a fall in net repo) is taken as a positive funding shock to arbitrageurs.

<sup>28</sup>I use the 2016Q1 release, available at <https://www.federalreserve.gov/releases/z1/20160310/>.

<sup>29</sup>Adrian, Etula, and Muir (2014) also make seasonal adjustments, which I do not do. Making seasonal adjustments does not affect my results. Asl and Etula (2012) and Adrian, Moench, and Shin (2013) also use this series to measure financial sector funding conditions.



**Figure 2: Log Leverage and Arbitrageur Funding Shock**

The figure plots the log of broker-dealer leverage and the arbitrageur funding shock over the period of 1972 to 2015. Broker-dealer leverage is the book leverage of broker-dealers defined as total financial assets net of repo assets divided by the difference between total financial assets and total liabilities. Arbitrageur funding shocks are measured by quarterly shocks to the leverage of broker-dealers. The funding shock is annualized.

Adrian, Etula, and Muir (2014) show that the funding shock is procyclical and has expected signs of correlation with market volatility, Aaa-Baa spread, and financial stocks return. However, among different financial intermediaries, this measure is especially relevant for levered arbitrageurs such as hedge funds since a major part of security broker-dealers' business is prime brokerage for hedge funds. As prime brokers, they provide their hedge fund clients with various types of financing, intermediate securities lending, and serve as custodians of the cash and stocks owned by hedge funds.<sup>30</sup>

In the online appendix, I repeat my main analyses with stochastically detrended

---

<sup>30</sup>Aragon and Strahan (2012) empirically study the financial dependence of hedge funds on prime brokers, using the Lehman bankruptcy.

leverage series,

$$f_t = \ln \left( \text{Leverage}_t^{BD} \right) - \frac{1}{N} \sum_{s=1}^N \ln \left( \text{Leverage}_{t-s}^{BD} \right), \quad (31)$$

with  $N = 4$ ,  $N = 8$ , and  $N = 12$ . I do these robustness checks because a leverage shock to prime brokers may affect their hedge fund clients with a lag. Hedge funds, especially larger ones, often arrange with their prime brokers to “lock in” the margin and collateral requirements for an agreed period. This margin lock-up is typically 90 days, but it can range from 30 to 120 days. Using the stochastically detrended leverage addresses this issue, since it assumes that an innovation to the arbitrageur funding condition at time  $t$  is a weighted average of the innovations in broker-dealer leverage growth in the last four quarters. To see this, simply rewrite (30) as

$$f_t = \sum_{s=0}^{N-1} \frac{N-s}{N} \left[ \Delta \ln \left( \text{Leverage}^{BD} \right) \right]_{t-s}, \quad (32)$$

where  $\left[ \Delta \ln \left( \text{Leverage}^{BD} \right) \right]_t$  indicates the growth rate of leverage from  $t - 1$  to  $t$ . Hence, the beta estimated using the factor resembles the Scholes-Williams beta (Scholes and Williams, 1977) and the Dimson beta (Dimson, 1979), which account for nonsynchronous data. Most test results are similar or stronger when I use  $N = 4$ ,  $N = 8$ , or  $N = 12$ .

For the rest of the paper, I use  $\beta_j$  to denote anomaly asset  $j$ 's beta with respect to the funding condition  $f_t$ , and I refer to it as the anomaly's “funding beta.”<sup>31</sup> I will sometimes refer to the funding beta as “endogenous risk” to highlight that the beta represents an endogenous risk exposure. When necessary, I will add a superscript to denote the sample period in which the beta is estimated.

**Identification through a sample split.** The model's propositions rely on being

---

<sup>31</sup>Adrian, Etula, and Muir (2014) refer to this factor as a “leverage factor.” I refer to it as an “arbitrageur funding shock” to emphasize that it is the empirical counterpart of the arbitrageur funding condition in the model.



able to observe the anomaly assets' abnormal returns in the absence of arbitrage capital ( $\mu = 0$ ) and endogenous risks in the presence of arbitrage capital ( $\mu > 0$ ). Although no single year represents a clear jump in the mass of arbitrageurs in the anomaly assets, I argue that the first-half (pre-93) and the second-half (post-93) of my original sample of 1972Q1-2015Q4 are reasonable proxies for times when arbitrageurs have a negligible mass ( $\mu = 0$ ) in the anomalies and for times when arbitrageurs have a positive mass ( $\mu > 0$ ) in the anomalies, respectively.

There are three justifications for using 1993 as the cutoff year. First, arbitrage capital grew rapidly in the 1990s, with hedge fund assets under management expanding from \$39 billion in 1990 to \$1.73 trillion in 2008 (Stein, 2009). Second, 1993 is the year when some of the most influential papers in equity anomalies were published: Fama and French (1993) popularized the size and value anomalies by rationalizing them with a multifactor model, and Jegadeesh and Titman (1993) introduced the momentum anomaly. These papers spurred the search for new equity anomalies whose abnormal returns are not explained by exposures to size, value, and momentum factors. Third, Chordia, Subrahmanyam, and Tong (2014) also use years prior to 1993 as the period when the trading technology and liquidity had not sufficiently developed to allow for extensive arbitrage at reasonable costs. Main test results, however, are similar when I use different cutoff years within the early 1990s.<sup>32</sup>

An alternative to using a sample split is to use the anomalies' publication years to study the effect of arbitrage trade. This approach is used by both McLean and Pontiff (2016) and Liu, Lu, Sun, and Yan (2015) (LLSY) in their studies of equity anomalies. However, LLSY find that strong arbitrage activities on size and value anomalies began around 1992, although some arbitrage activities occurred following their original discoveries by Banz (1981) and Rosenberg, Reid, and Lanstein

---

<sup>32</sup>The online appendix repeats the main regressions using 1991, 1992, 1994, and 1995 as the cutoff year.

(1985).<sup>33</sup> This suggests that using the 1993 cutoff is a reasonable alternative to using publication years.

Given my sample split approach, I measure the latent mispricing of an anomaly by its mean long-short return in the pre-93 sample ( $\bar{r}_j^{\text{pre93}}$ ) and the endogenous risk of an anomaly by its beta with respect to the arbitrageur funding conditions in the post-93 sample ( $\beta_j^{\text{post93}}$ ). Since the anomaly assets are not strongly exposed to market excess returns or arbitrageur funding shocks in the pre-93 sample, using pre-93 CAPM alphas, pre-93 Fama-French three-factor alpha, or pre-93 arbitrageur funding alphas to measure latent mispricings does not substantially change the paper's main results (see the online appendix).

**Time-series evidence of endogenous risk exposures.** This paper's main contribution is to use the *cross-section* of anomaly assets to test the idea that arbitrage generates endogenous risk by turning  $\alpha$ s into  $\beta$ s. However, I briefly highlight that *time-series* evidence is also consistent with the endogenous risk idea.

First, the returns on anomaly assets have fallen. Table 1 shows that the annualized long-short returns on anomaly assets have fallen from 5.24% in the pre-93 period to 3.57% in the post-93 period. This implies a 32% decline in expected returns as a result of increased arbitrage, similar to the 32% fall in expected returns that McLean and Pontiff (2016) find in anomaly assets after an academic publication. If I assume that arbitrageurs reverse the direction of the trades for the anomalies with negative pre-93 mean returns, the fall in expected returns is from 6.04% to 3.85%, which is a 37% decline.

Second, as the anomalies' expected returns have fallen, their betas with respect to arbitrageur funding shocks have risen. Table 1 shows that the cross-sectional average of arbitrageur funding betas increases from -0.58 to 2.62 between the two sample periods. If I again assume that arbitrageurs reverse the direction of the trades for the anomalies with negative pre-93 mean returns, the change in beta is

---

<sup>33</sup>They attribute this to Fama and French (1992).

from -0.53 to 5.70. Although not reported in the table, in the post-93 period, 8.1% of the time-series variation in the return on an equal-weighted (EW) index of the 34 anomaly assets is explained by the variation in arbitrageur funding shock.<sup>34</sup> In contrast, in the pre-93 period, the EW index return has an  $R^2$  of only 0.3% in the same regression.

Next, I move on to cross-sectional tests. There, I show that cross-sectional evidence strongly points to anomaly assets becoming endogenously riskier due to arbitrage trade.

### 3.2. Mispricing turns into endogenous risk

I first test how well an anomaly's latent mispricing (pre-93 mean return  $\bar{r}_j^{pre}$ ) predicts its endogenous risk (post-93 beta with respect to arbitrageur funding shocks  $\beta_j^{post}$ ) (Proposition 1). The simple intuition is that an anomaly with a larger latent mispricing attracts correspondingly more arbitrage capital, which generates greater endogenous risk. The empirical test is to run the following regression in the cross-section of 34 anomaly assets

$$\beta_j^{post} = b_0 + b_1 \bar{r}_j^{pre} + \eta_j \quad (33)$$

and test if  $b_1 = 0$ .

A complication arises because pre-93 mean return is estimated. If the estimated mean  $\bar{r}_j^{pre}$  is a noisy signal of the actual latent mispricing, and if arbitrageurs observe the true latent mispricing whereas the econometrician does not, then the standard errors for the cross-sectional regression (33) need to be adjusted for the fact that the regressor is generated. Hence, I jointly estimate the coefficients in the cross-sectional regression (33), pre-93 mean long-short returns, and post-93 funding betas using the generalized method of moments (GMM). Since arbitrageurs

---

<sup>34</sup>When constructing this index, I reverse the direction of the trade for anomalies with negative pre-93 long-short returns.

may actually have the identical information set that the econometrician has, using realized returns in the past to gauge an anomaly asset's latent mispricing, I also report ordinary least squares (OLS) t-statistics.

To use GMM, consider the following data-generating process. In the pre-93 period, a mean return is a noisy realization of the latent mispricing:

$$r_{j,t} = \bar{r}_j^{pre} + \epsilon_{j,t} \quad (34)$$

This latent mispricing determines an anomaly's endogenous exposure to funding shocks:

$$\beta_j^{post} = b_0 + b_1 \bar{r}_j^{pre} + \eta_j \quad (35)$$

Here,  $\eta_j$  has a cross-sectional mean of zero. This funding beta then determines the equilibrium expected return in the post-93 period:

$$r_{j,t} = a_j^{post} + \beta_j^{post} f_t + \epsilon_{j,t} \quad (36)$$

These conditions imply the following  $4J$  moment conditions:

$$g_{4J \times 1}(\mathbf{b}) = \begin{bmatrix} E \left[ \left( r_{j,t} - \bar{r}_j^{pre} \right) \mathbf{1}(t \in Pre) \right] \\ E \left[ \left( r_{j,t} - a_j^{post} - \beta_j^{post} f_t \right) \mathbf{1}(t \in Post) \right] \\ E \left[ \left( r_{j,t} - a_j^{post} - \beta_j^{post} f_t \right) f_t \mathbf{1}(t \in Post) \right] \\ E \left[ \left( \beta_j^{post} - b_0 - b_1 r_{j,t} \right) \mathbf{1}(t \in Pre) \right] \end{bmatrix}, \quad (37)$$

I then use a selection matrix to ensure that a cross-sectional expectation is taken over the last set of  $J$  moments,

$$A_{(3J+2) \times 4J} = \begin{bmatrix} \mathbf{I}_{3J \times 3J} & \mathbf{0}_{3J \times J} \\ \mathbf{0}_{1 \times 3J} & \mathbf{1}_{1 \times J} \\ \mathbf{0}_{1 \times 3J} & \bar{\mathbf{r}}'_{1 \times J} \end{bmatrix} \quad (38)$$

where  $\bar{r}$  is a  $J \times 1$  vector of pre-93 mean returns. Then, I find parameter estimates  $\hat{\mathbf{b}}$  such that

$$Ag_t(\hat{\mathbf{b}}) = 0_{(3J+2) \times 1} \quad (39)$$

The chosen selection matrix generates the identical set of moment conditions as in sequential OLS estimations of (34), (36), and then (35).

The first column of Table 2 reports the parameter estimates, OLS t-statistics, and GMM t-statistics from estimating the effect of latent mispricing on funding beta specified in (33). Each percentage of pre-93 mean return turns into a post-93 funding beta of 1.24. To interpret this number, since returns are in percentages whereas the funding shocks are not, a beta of 1.24 means that a 100% increase in the leverage of broker-dealers leads to a 1.24% increase in the anomaly asset return. This “turning alphas into betas” effect is statistically significant based on both GMM and OLS standard errors. The large  $R^2$  implies that the pre-93 return is a strong predictor of an anomaly’s endogenous exposure to arbitrageur funding. The intercept is statistically insignificant, implying that there is no strong trend in the anomalies’ funding betas apart from the endogenous effect coming from arbitrage activity.

Additional control variables have little influence on the results. In the second column, I add pre-93 funding beta as an additional regressor to show that it is not the persistence or magnification of pre-93 beta that drives the large post-93 funding betas. The coefficient of 0.75 on the pre-93 funding beta implies some persistence in the beta, although the effect is significant only based on t-OLS and not based on t-GMM. Despite this persistence, since pre-93 funding beta was small in magnitude, including it in the regression has little effect on the coefficient on pre-93 mean return.

In the third column, I add pre-93 return volatility as an additional regressor, thus addressing the concern that the anomalies with larger pre-93 mean returns also tend to have large pre-93 volatilities; that is, pre-93 mean return may proxy not

**Table 2: Mispricing Turns into Funding Beta**

$$\text{Baseline: } \beta_j^{post} = b_0 + b_1 \bar{r}_j^{pre} + \eta_j$$

This table reports the results from the cross-sectional regression predicting an anomaly asset's post-1993 funding beta using pre-1993 mean long-short return. The dependent variable in the first three columns is post-1993 funding beta  $\beta_j^{post}$ , calculated under the assumption that the beta is constant during the sample period:  $r_{j,t} = a_0 + \beta_j^{post} f_t + \epsilon_t$ . The dependent variable in the fourth column is the post-1993 rate of increase in beta  $\beta_{j,1}^{post}$ , calculated under the assumption that anomaly return attains an increasingly large exposure to arbitrageur funding during the post-1993 sample:  $r_{j,t} = a_0 + (\beta_{j,0}^{post} + \beta_{j,1}^{post} \ln(t)) f_t + \epsilon_t$ , where  $t$  is the number of quarters into the sample ( $t = 1$  for 1994Q1). The dependent variable in the last column is the post-1993 funding correlation  $\rho_j^{post} = \beta_j^{post} \sigma_f^{post} (\sigma_j^{post})^{-1}$  in percentage (%), calculated under the assumption that the correlation is constant during the post-1993 sample. t-OLS is the t-statistic calculated using only the residuals from the cross-sectional regression and accounts for a possible heteroskedasticity of residuals across anomaly assets. t-GMM refers to a t-statistic obtained from the GMM estimation procedure and accounts for the effects of generated regressors and cross-anomaly correlations. Arbitrageur funding shocks are measured by quarterly shocks to the leverage of broker-dealers.

	Post-93 Funding Beta			Post-93 Funding	
	Level $\beta_j^{post}$	Incr. Rate $\beta_{j,1}^{post}$		Corr $\rho_j^{post}$	
Pre-93 Mean Long-Short Return $\bar{r}_j^{pre}$	1.24	1.10	1.19	0.34	1.42
(t-OLS)	(3.51)	(2.82)	(3.19)	(2.49)	(2.80)
(t-GMM)	(2.03)	(1.98)	(2.15)	(0.81)	(2.41)
Pre-93 Funding Beta $\beta_j^{pre}$		0.75			
(t-OLS)		(2.50)			
(t-GMM)		(1.17)			
Pre-93 Return Volatility $\sigma_j^{pre}$			0.07		
(t-OLS)			(0.26)		
(t-GMM)			(0.10)		
Intercept	-3.84	-2.69	-5.49	-1.04	-3.89
(t-OLS)	(-1.44)	(-1.00)	(-0.80)	(-1.00)	(-0.89)
(t-GMM)	(-1.68)	(-1.20)	(-2.46)	(-1.01)	(-0.94)
Anomalies	34	34	34	34	34
Adjusted $R^2$	0.26	0.33	0.24	0.14	0.15

only for latent mispricing, but also for volatility. We see, however, that controlling for pre-93 volatility has little effect on the slope of pre-93 mean return and that pre-93 volatility is a highly insignificant predictor of post-93 funding beta. This shows that it is the latent mispricing rather than volatility that predicts an anomaly asset's endogenous risk.

Thus far I have treated the year 1993 as a “jump” in the mass of arbitrageurs in the anomaly assets. In reality, the increase in the mass of arbitrageurs in the anomalies ( $\mu$  in the model) and hence the anomalies' endogenous exposures to arbitrageur funding shocks would have been gradual even within the post-93 period. Therefore, I allow the arbitrageur funding beta to be an increasing and concave function of time  $t$  within the post-93 period. In particular, I assume the following data-generating process in which an exposure to arbitrageur funding grows at the rate  $\beta_{j,1}^{post} t^{-1}$ , where  $t$  here is the number of quarters into the post-93 sample (1994Q1 being  $t = 1$ ):<sup>35</sup>

$$\begin{aligned}
\text{Pre-93 return:} & & r_{j,t} &= \bar{r}_j^{pre} + \epsilon_{j,t} \\
\text{Post-93 return:} & & r_{j,t} &= a_j^{post} + \left( \beta_{j,0}^{post} + \beta_{j,1}^{post} \ln(t) \right) f_t + \epsilon_{j,t} \\
\text{Beta determination:} & & \beta_{j,1}^{post} &= b_0 + b_1 \bar{r}_j^{pre} + \eta_j
\end{aligned} \tag{40}$$

In this case, the latent mispricing of an anomaly measured by the pre-93 mean return should predict the rate of increase in beta  $\beta_{j,1}^{post}$ .<sup>36</sup>

The fourth column of Table 2 reports the “turning alphas into betas” effect estimated under the assumption of a gradual endogenous risk exposure. Again, the latent mispricing measured by the pre-93 mean return predicts an anomaly's endogenous risk measured by  $\beta_{j,1}^{post}$ . Although the coefficient is not statistically

---

<sup>35</sup>When implementing the GMM, I set  $\ln(t) = 0$  for all  $t \in \text{Pre}$ . The values of  $\ln(t)$  in the pre-93 sample do not affect the results, as all moment conditions involving  $\ln(t)$  are multiplied by the post-93 dummy. Admittedly, the specific function through which I introduce concavity in  $t$  is arbitrary. However, I find that the specific way of introducing concavity in  $t$  does not lead to large changes in the parameter estimate, unless the function is too “linear” over  $t \in \text{Post}$ .

<sup>36</sup>The GMM implementation of this model is explained further in the Appendix.

significant based on t-GMM because of large standard errors, the estimated coefficient is similar in magnitude to the one from the first column. Since there are 88 quarters in the post-93 period (1994Q1-2015Q4), it follows that each percentage of pre-93 mean return leads to a funding beta increase of  $0.34 \times \ln(88) = 1.52$  by the end of the post-93 period. The intercept is small, implying that there is no strong trend in the funding beta other than through arbitrage activity.

Although I have used funding beta to measure an anomaly asset's exposure to arbitrageur funding, beta can change spuriously because of volatility rather than correlation. For instance, to elaborate on the point made earlier, suppose that anomalies with large pre-93 mean returns are precisely the ones with large return volatilities. If such anomalies' large volatilities persist through the post-93 period, then even if anomalies have equal post-93 correlations with arbitrageur funding, pre-93 return would appear to predict post-93 funding beta:

$$\beta_j^{post} = \rho_j^{post} \frac{\sigma_j^{post}}{\sigma_f^{post}} \propto \bar{r}_j^{pre} \quad (41)$$

To rule out this possibility, I repeat the baseline regression using funding correlation  $\rho_j^{post}$  as the dependent variable.<sup>37</sup>

The last column of Table 2 reports the results from predicting an anomaly's post-93 funding correlation using its pre-93 mean return. The results are strong both based on the t-statistics and on  $R^2$ . This suggests that anomalies' large post-93 funding betas are driven by correlations, not volatilities. In terms of magnitude, each percentage of pre-93 mean return raises the anomaly's funding correlation by 1.42 percentage points (%p). Since the anomaly assets' pre-93 mean returns range from  $-5\%$  to  $21\%$ , the predicted post-93 correlation with arbitrageur funding shock ranges from  $-7\%$  to  $30\%$ .

The results presented here are robust to alternative measures of latent mispric-

---

<sup>37</sup>I leave more detailed discussions of the GMM implementation to the Appendix.



ing. The online appendix shows that using pre-93 CAPM alpha, pre-93 Fama-French three-factor alpha, or pre-93 arbitrageur funding alpha generates similar results. The same appendix also shows that using a volatility-neutral measure of latent mispricing generates similar results.

The results here suggest that anomaly assets' exposures to the arbitrageur funding conditions are an endogenous outcome of arbitrage trading. In Section 3.5, I show that these exposures help explain the anomaly assets' equilibrium returns in the post-93 period. Hence, the endogenous risk theory explains the origin of betas before using the betas to explain equilibrium expected returns—a response to the question, “Why are betas exogenous?” (Cochrane, 2011: p.1063). Next, I provide further evidence that the betas are generated by arbitrage activity.

### **3.3. Endogenous risk is explained by anomaly-specific arbitrage capital**

Here, I test the prediction that, if anomaly assets' funding betas are indeed a byproduct of arbitrage activity, then the funding betas must be explained by anomaly-specific measures of arbitrage activity (Proposition 2). To test this, I find measures of arbitrage activity specific to an anomaly. Then, I run a cross-sectional regression to test if an anomaly with greater arbitrage activity has a larger post-93 funding beta during the same period.

I use three measures of anomaly-specific arbitrage capital based on how much shorting there is in the bottom decile of the anomaly relative to the top decile: the difference in the short interest ratio in the bottom and top deciles of an anomaly (“short interest ratio difference”); percentage (log) difference in the short interest ratio in the bottom and top deciles (“log short interest ratio difference”); and difference in the days to cover in the bottom and top deciles (“days to cover difference”). Net shorting is a relatively clean way to measure arbitrage activity since

most shorting is done by hedge funds.<sup>38</sup>

Specifically, short interest ratio, previously used by Hanson and Sunderam (2014), is defined as the number of shares being shorted (short interest) divided by the number of shares outstanding at a given time. I compute the short interest ratio of each stock in each month by dividing its mid-month short interest amount (from Compustat) by shares outstanding on the same day (from CRSP). Then, to obtain the short interest ratio difference measure, I take the post-93 time-series average of the monthly difference between the VW average short interest ratios in the bottom decile and the top decile:

$$DSIR_j^{post} = 100 \times T^{-1} \sum_{t=1}^T \left( SIR_{j,t}^{bottom\ decile} - SIR_{j,t}^{top\ decile} \right) \quad (42)$$

where the value-weighted average short interest ratio of a decile is computed as

$$SIR_{j,t}^{decile} = \sum_{i \in decile_{j,t}} \omega_{i,t} \frac{Short\ Interest_{i,t}}{Shares\ Outstanding_{i,t}} \quad (43)$$

with  $\omega_i$  denoting the weight of stock  $i$  in the relevant extreme decile and  $t$  denoting month.

The second measure of arbitrage activity, log short interest ratio difference, is a variant of the short interest ratio difference measure. Here, I take a percentage difference rather than a level difference in the short interest ratio in the bottom and top deciles because the level difference has a strong positive trend over time; that is, taking a time-series average of the level difference puts disproportionate weights on recent years. Hence, the log short interest ratio difference measure is the following:

---

<sup>38</sup>A report by Goldman Sachs (2014) estimates that 85% of the short interest is held by hedge funds. Similarly, Boehmer, Jones, and Zhang (2013) and Ben-David, Frazoni, and Moussawi (2012) both argue that hedge funds are responsible for most of the short interest. Dechow et al. (2001), Hirshleifer, Teoh, and Yu (2011), and Cao et al. (2012) also interpret short activity on an anomaly as arbitrage activity.

$$DLSIR_j^{post} = T^{-1} \sum_{t=1}^T \left( \ln \left( SIR_{j,t}^{bottom\ decile} \right) - \ln \left( SIR_{j,t}^{top\ decile} \right) \right) \quad (44)$$

where  $t$  is month.

As for the third measure, days to cover (DTC) of a stock is defined as its short interest divided by average daily trade volume. Thus, it measures the expected number of days required to recover all shorted stocks from the market, also interpreted as the liquidity cost of exiting the short positions. Since DTC normalizes short interest by the stock's liquidity, Hong et al. (2015) argues that DTC is a more accurate measure of arbitrage intensity than short interest ratio. To measure an anomaly's average DTC during the post-93 period, I first compute a stock's DTC each month by dividing its monthly short interest ratio by the average share turnover (trade volume divided by shares outstanding) during the same month. I subtract the VW average DTC in the top decile of an anomaly from that in the bottom decile to obtain the DTC difference between the two deciles. I then compute the time-series average of this monthly DTC measure during the post-93 sample period to obtain the average DTC difference between the two deciles of an anomaly:

$$DDTC_j^{post} = T^{-1} \sum_{t=1}^T \left( DTC_{j,t}^{bottom\ decile} - DTC_{j,t}^{top\ decile} \right) \quad (45)$$

where the value-weighted average days to cover of a decile is computed as

$$DTC_{j,t}^{decile} = \sum_{i \in decile_{j,t}} \omega_{i,t} \frac{Short\ Interest_{i,t}}{Avg\ Daily\ Trade\ Volume_{i,t}} \quad (46)$$

with  $\omega_i$  denoting the weight of stock  $i$  in the relevant extreme decile. For all three measures, I compute the same measure for the pre-93 period, although the short interest data starts from 1973 rather than 1972, when my pre-93 period begins.

For each of the three measures of anomaly-specific arbitrage capital, I ask if an anomaly with more arbitrage activity has a larger funding beta. This amounts to running the following regression, where  $ArbCapital_j^{post}$  is an anomaly-specific arbitrage capital measure:

$$\beta_j^{post} = b_0 + b_1 ArbCapital_j^{post} + \eta_j \quad (47)$$

I estimate the parameters using OLS and report t-statistics based on heteroskedasticity-consistent standard errors. Although the explanatory variables in this regression are estimated, I do not need to correct for their variances because it is the *realized* arbitrage activity, not the unobserved “true mean,” that matters for generating the observed betas.<sup>39</sup>

Table 3 reports the test results. During the post-93 period, funding betas are explained by anomaly-specific arbitrage capital. The highly significant slope coefficients, large  $R^2$ , and small intercepts imply that once arbitrage activity has been controlled for, there is little idiosyncratic or common variation in the anomalies’ funding betas. Interpretations of the slope coefficients are as follow. First, the coefficient on the short interest ratio difference implies that, in the cross-section of anomalies, the funding beta increases by 4.94 for an increase of 1/100 in the difference in the short interest ratio between the bottom and the top deciles. Next, the coefficient on the log short interest ratio difference implies that, again in the cross-section of anomalies, the funding beta increases by 0.14 for each 1% point increase in the short-interest-ratio difference between the bottom and the top deciles. Finally, the coefficient on DTC implies that, in the cross-section of anomalies, the funding beta increases by 4.11 for each increase in the number of days it takes to recover the short interest in the bottom decile relative to the top decile.

Interestingly, the ability of anomaly-specific arbitrage capital to explain the

---

<sup>39</sup>However, GMM t-statistics would be useful to control for the cross-asset correlations.

**Table 3: Funding Betas Are Explained by Arbitrage Activity**

$$\text{Baseline: } \beta_j^{post} = b_0 + b_1 DSIR_j^{post} + \eta_j$$

This table reports results from the cross-sectional regressions explaining an anomaly asset's funding beta using an anomaly-specific measure of arbitrage capital. The first three columns report results for the post-1993 period, and the last three columns report results for the pre-1993 period. The dependent variable is the funding beta of an anomaly asset in the pertaining sample period, when arbitrageur funding shocks are measured by quarterly shocks to the leverage of broker-dealers. Short interest ratio difference, log short interest ratio difference, and days to cover difference are the time-series averages of the difference in the value-weighted (VW) average of each measure between the bottom and top deciles of the anomaly asset. Short interest ratio difference uses 100 times the level difference in the VW average short interest ratio (short interest ÷ shares outstanding) in the bottom and the top deciles. Log short interest ratio difference uses the log difference in the VW average short interest ratio in the bottom and the top deciles. Days to cover difference uses the level difference in the VW average days to cover (short interest ÷ average daily trade volume) in the bottom and the top deciles.

	Dependent Variable: Funding Beta					
	Post-93 (1994-2015)			Pre-93 (1972-1993)		
Short Interest Ratio Difference	4.94 (3.98)			11.33 (5.39)		
Log Short Interest Ratio Difference		14.46 (3.81)			6.37 (5.72)	
Days to Cover Difference			4.11 (3.05)			4.14 (4.22)
Intercept	-0.79 (-0.39)	-0.42 (-0.20)	0.98 (0.48)	-2.00 (-2.45)	-1.95 (-2.23)	-1.73 (-2.18)
Anomalies	34	34	34	34	34	34
Adjusted $R^2$	(0.37)	(0.32)	(0.18)	(0.43)	(0.39)	(0.36)

Note: In the parantheses are OLS t-statistics calculated with heteroskedasticity-consistent standard errors. The average of the arbitrage activity measures across all anomalies are 0.69 (*DSIR*), 0.21 (*DLSIR*), and 0.40 (*DDTC*) in the post-93 period. The average of the arbitrage activity measures across all anomalies are 0.13 (*DSIR*), 0.22 (*DLSIR*), and 0.28 (*DDTC*) in the pre-93 period.

cross-section of funding betas is strong in the pre-93 period as well, implying that the pre-93 funding betas were also an outcome of arbitrage activity. However, as noted in the table, the magnitudes of the arbitrage activities were substantially smaller in the pre-93 period, causing the pre-93 funding betas to be smaller in magnitudes than post-93 funding betas.

### 3.4. Mispricing turns into a larger endogenous risk on the short side

Before testing Proposition 3, I carry out a test that is outside the scope of the model but provides further evidence for the endogenous arbitrage risk view. This test separates out the long-short returns on anomaly assets into long (top decile) and short (bottom decile) portfolios and asks whether the same amount of abnormal return turns into a larger endogenous risk on the short side than on the long side of the anomaly. Intuitively, the long side of an anomaly can be exploited by a large class of investors including mutual funds, pension funds, and individual investors who are not exposed to arbitrageur funding shocks, whereas the short side of an anomaly is primarily exploited by arbitrageurs, such as hedge funds, as argued in Section 3.3. Hence, if an anomaly's endogenous risk is a byproduct of arbitrage activity, we should expect a larger "turning alphas into betas" effect on the short side.

To test this, I repeat the test of Proposition 1 in Section 3.2 using 68 portfolios representing long and short sides of the 34 anomaly assets. Since the test portfolios now have significant market exposures, I take the robust approach of measuring the latent mispricing using pre-93 CAPM alpha (instead of simple mean return) and endogenous risk using post-93 funding beta net of market exposure. Hence, the baseline regression is

$$\beta_j^{post} = b_0 + b_1 Short_j + b_2 \alpha_j^{pre,CAPM} + b_3 \alpha_j^{pre,CAPM} \times Short_j + \eta_j, \quad (48)$$

where  $j$  now indexes one of the 68 long- and short-side anomaly portfolios, and  $Short_j$  is a dummy variable for short-side portfolios. I test if  $b_3$  is positive and statistically different from zero.

Table 4 and Figure 3 report the test result. Consistent with the hypothesis, the turning-alphas-into-betas effect is much stronger on the short side than on the

**Table 4: Mispricing Turns into a Larger Funding Beta on the Short Side**

$$\text{Baseline: } \beta_j^{post} = b_0 + b_1 \text{Short}_j + b_2 \alpha_j^{pre,CAPM} + b_3 \alpha_j^{pre,CAPM} \times \text{Short}_j + \eta_j$$

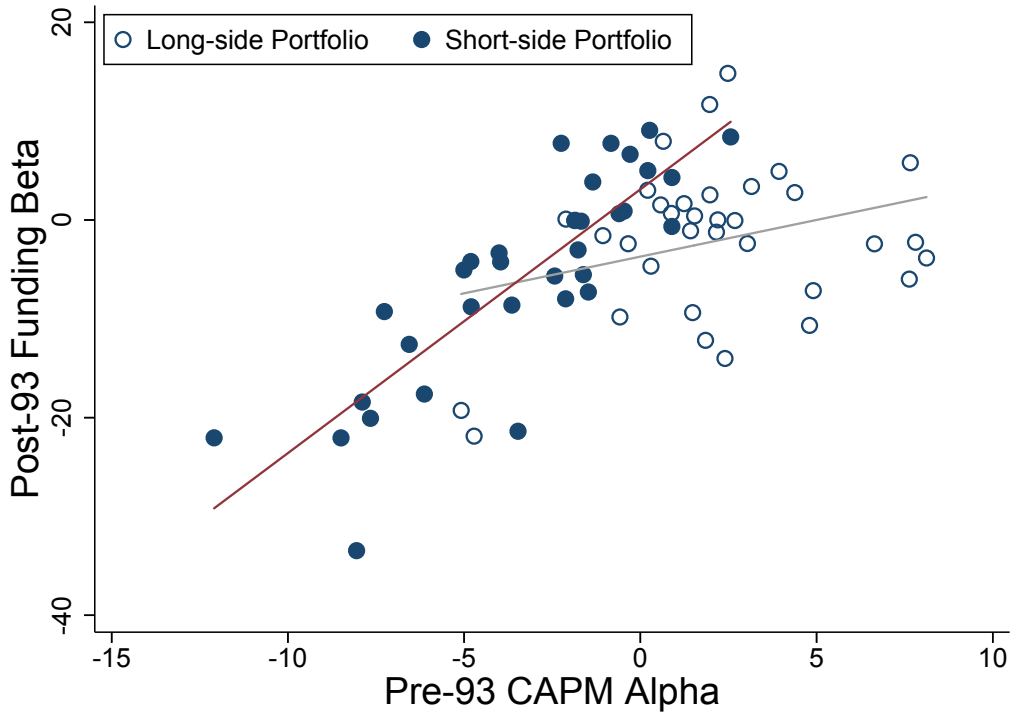
This table reports results from the cross-sectional regressions explaining an anomaly asset's post-1993 funding beta net of market exposure using its pre-1993 CAPM alpha. An anomaly asset is either the long or short side of 34 anomalies. The dependent variable is post-1993 funding beta  $\beta_j^{post}$ , calculated under the assumption that the funding beta net of market exposure is constant during the post-1993 sample:  $r_{j,t} - r_t^f = a_0 + \beta_j^{post} f_t + \beta_j^{post,mkt} (r_t^m - r_t^f) + \epsilon_t$ , where  $r_t^f$  and  $r_t^m$  denote risk-free rate and market return. The independent variable is pre-1993 CAPM alpha:  $r_{j,t} - r_t^f = \alpha_j^{pre,CAPM} + \beta_j^{pre,CAPM} (r_t^m - r_t^f) + \epsilon_t$ . Arbitrageur funding shocks are measured by quarterly shocks to the leverage of broker-dealers.

	Dependent Variable: Post-93 Funding Beta <sup>†</sup>		
	Long Side	Short Side	Combined
Pre-93 CAPM Alpha	0.74 (1.51)	2.67 (8.18)	0.74 (1.51)
Pre-93 CAPM Alpha × 1(Short)			1.93 (3.26)
1(Short)			6.81 (3.07)
Intercept	-3.71 (-1.91)	3.10 (1.16)	-3.71 (-1.97)
Anomalies	34	34	68
Adjusted R <sup>2</sup>	0.06	0.71	0.50

Notes: † This funding beta is net of market exposure. In the parentheses are OLS t-statistics calculated with heteroskedasticity-consistent standard errors.

long side. When the long and short sides of the anomalies are separately analyzed (the first two columns), the slope of the coefficients are significant based on OLS t-statistics. However, the magnitude of the slope and the  $R^2$  of the regression are much larger on the short side. This difference in the slope is statistically significant, as reported in the last column. This result is consistent with arbitrageurs generating larger endogenous risks on the short sides of the anomalies where they have a larger relative presence.

The result here is difficult to reconcile with the idea that a hidden fundamental risk explains the large returns earned by anomaly assets. To see this, suppose that the anomaly assets have always been commonly exposed to single latent risk factor  $L_t$ :  $r_{j,t} = \delta_j L_t + \epsilon_{j,t}$ . Suppose also that, for whatever reason, this latent



**Figure 3: Endogenous Risk Is Larger on the Short Side**

The figure plots the pre-1993 CAPM alphas and post-1993 funding betas of excess returns on long-side and short-side portfolios of anomaly assets. The solid circles are short-side portfolios, and hollow circles are long-side portfolios. The figure shows that mispricing measured by pre-93 CAPM alpha transforms into a larger post-93 funding beta on the short side of the anomalies than on the long side. Funding beta is measured as the beta with respect to arbitrageur funding shocks proxied by quarterly shocks to the leverage of broker-dealers.

risk factor has become more correlated with the arbitrageur funding conditions in recent years:  $\rho(L_t, f_t) = 0$  in pre-93 but  $f_t = \gamma^{post} L_t + \eta_t$  in post-93 for some noise  $\eta_t$ . This would at least explain why pre-93 mean returns appear to explain anomalies' post-93 betas with respect to the arbitrageur funding conditions and why post-93 funding betas explain post-93 expected returns, consistent with the empirical results in Section 3.2 and Section 3.5. In this case, however, the short and long sides of the anomaly assets are expected to have the same coefficient in the “turning alphas into betas” regression. This is because, in this scenario, the coefficient would simply measure how well the arbitrageur funding shock proxies for the latent risk factor during the post-93 period. Analytically, since pre-93 expected return is  $\bar{r}_j^{pre} = \delta_j \bar{L}^{pre}$  and post-93 beta is  $\beta_j^{post} = \delta_j \gamma^{post}$ , the ratio between



the two is a constant:

$$\beta_j^{post} / \bar{r}_j^{pre} = \gamma^{post} / \bar{L}^{pre} \quad (49)$$

Hence, the result here cannot be rationalized by the hidden fundamental risk view, at least under the assumption that there is one latent risk factor to which the anomalies are commonly exposed. Rationalizing it with multiple latent risk factors would require a much more elaborate story in which the long and short sides of the anomalies are exposed to different latent risk factors and those latent risk factors have come to attain different post-93 correlations with the arbitrageur funding conditions.

### **3.5. “Intermediary asset pricing” of anomaly assets based on endogenous risks**

The empirical tests up to this point have focused on showing that funding betas arise as an endogenous outcome of different arbitrage activities on differently mis-priced anomaly assets. An empirical question, however, is whether arbitrageurs take these funding betas into consideration when assessing risks of the anomaly assets. Here, I present some evidence that arbitrageurs are mindful of being exposed to funding betas, the endogenous risks they themselves have generated. However, I also find that jointly explaining the expected returns of 34 different equity anomalies with single factor is a challenging task.

The objective is to test Proposition 3, which predicts that the anomaly assets' endogenous exposures to arbitrage risk explain the expected returns that survive in equilibrium. To do this, in the post-93 period, I run a cross-sectional asset pricing regression using beta with respect to arbitrageur funding. I then test if the price of risk is positive and if the variation in expected returns is explained by the variation in funding betas. In addition, I test the model's nonconventional asset pricing predictions.

Before I delve into the test, I highlight this exercise's close connection to the intermediary asset pricing test of Adrian, Etula, and Muir (2014) (AEM). Proxying for the financial intermediary funding condition using the broker-dealer leverage shocks, they show that the single factor explains the cross-section of returns on 25 equity portfolios sorted by size and value, 10 equity portfolios sorted by momentum, and 6 bond portfolios sorted by maturity. They do this over the years 1968 to 2009, and they test the usual restriction that the intercept of the cross-sectional regression must be zero.

At the most basic level, the exercise here can be viewed as extending the AEM result on size, value, and momentum anomalies to a much wider set of equity anomalies. More importantly, however, I use this exercise to show that the endogenous risks generated by arbitrageurs prevent the initial anomaly returns from disappearing completely. Furthermore, I use this exercise to highlight the endogenous arbitrage risk model's nonconventional asset pricing predictions.

The endogenous-arbitrage-risk model makes two nonconventional predictions in asset pricing. First, if anomaly assets tend to have positive funding betas, we must also see a positive intercept in a cross-sectional regression. For arbitrageurs to generate positive  $\beta$ s in anomalies through price pressure, the anomaly assets must sometimes be mispriced and generate abnormal returns. This means that the zero-beta rate—the risk-adjusted return that arbitrageurs are earning from anomaly assets—must be higher than the risk-free rate, on average, during the sample period. Second, the price of risk is biased upward if the cross-sectional regression is run in a sample that includes both the “pre-arbitrage” and “post-arbitrage” periods. This is because such a regression would try to explain the large returns during the “pre-arbitrage” period driven by mispricing using large betas during the “post-arbitrage” period, when the actual causality flows in the opposite direction from the “pre-arbitrage” return to “post-arbitrage” beta. Although this second point is not formalized as part of Proposition 3, I do examine the issue

empirically.

I now use a cross-sectional asset pricing regression to explain anomaly assets' returns through their exposures to arbitrageur funding shocks. To do this, I jointly estimate the anomaly assets' betas in the time-series of returns and the price of risk in the cross-section of returns in a GMM framework:

$$\begin{aligned} \text{Time-series regressions:} \quad & r_{j,t} = a_j + \beta_j f_t + \epsilon_{j,t} \text{ for } j = 1, \dots, J \\ \text{Cross-sectional regression:} \quad & \bar{r}_j = \lambda_0 + \lambda_1 \beta_j + e_j \end{aligned} \tag{50}$$

where the cross-sectional regression is an OLS regression that puts an equal weight on all anomaly assets. As articulated by Cochrane (2005), mapping these regressions into GMM allows me to obtain standard errors that account for both the fact that  $\beta$ s are estimated and the fact that returns can be correlated across anomaly assets. However, the baseline regression uses only 88 quarters in the post-93 sample to test if exposure to arbitrageur funding has a positive price of risk within this period of large arbitrageur presence. Hence, the estimates of the full variance-covariance matrix of errors  $\epsilon$  may be subject to noise. Given this consideration, I also report t-statistics that adjust the standard errors for the fact that betas are estimated but use the conventional heteroskedasticity-consistent matrix of residuals  $\epsilon$  that restrict the cross-anomaly correlations to be zero ("t-GenReg"). Furthermore, I compare the arbitrageur funding shock's ability to explain the cross-section of anomaly mean returns to that of conventional multifactor models.

Table 5 presents the results. In terms of  $R^2$ , during the post-93 sample, the single arbitrageur funding shock explains 37% of the cross-sectional variation in the long-short returns on 34 anomaly assets. This  $R^2$  is somewhat lower than the  $R^2$ s I obtain from the Fama-French three-factor model (Fama and French, 1993), the Carhart four-factor model (Carhart, 1997), and the Fama-French five-factor model (Fama and French, 2008). However, that the market and size factors exhibit negative prices of risk in most of these multifactor models speaks of the difficulty

of jointly explaining the cross-section of anomaly expected returns.

The estimated price of risk is 0.18 so that each additional unit of funding beta is compensated by an annualized return of 0.18%. This estimated slope, however, is not unequivocally significant. When cross-anomaly correlations are restricted to be zero, the standard error implies a t-statistic of 2.30. However, based on the most conservative GMM standard errors where cross-anomaly correlations are freely estimated, the t-statistic is 1.30. This implies that, in the data, some of the anomaly assets are highly correlated with other anomaly assets, which increases the standard errors.

**Table 5: “Intermediary Asset Pricing” of Anomaly Assets**

$$\text{Baseline: } \bar{r}_j^{post} = \lambda_0 + \lambda_1 \beta_j^{post} + e_j$$

This table reports the risk prices of factors and intercepts estimated in the cross-section of anomaly assets. Returns are long-short returns expressed in annualized percentages. Betas are estimated in the time-series regression  $r_{j,t} = a_j + \beta_j f_t + \epsilon_{j,t}$  for each anomaly. t-GenReg refers to t-statistic corrected for generated regressors *but not* for cross-anomaly correlations. That is, to obtain the standard errors accounting for generated regressors, I allow for heteroskedastic residuals  $\epsilon_j$  for the mean returns and do the correction derived by Shanken (1992), but under the assumption of  $Cov(\epsilon_{j'}, \epsilon_{j''}) = 0$  for  $j' \neq j''$ . t-GMM refers to GMM t-statistic that additionally corrects for correlations across anomaly assets. Arbitrageur funding shocks are measured by quarterly shocks to the leverage of broker-dealers.

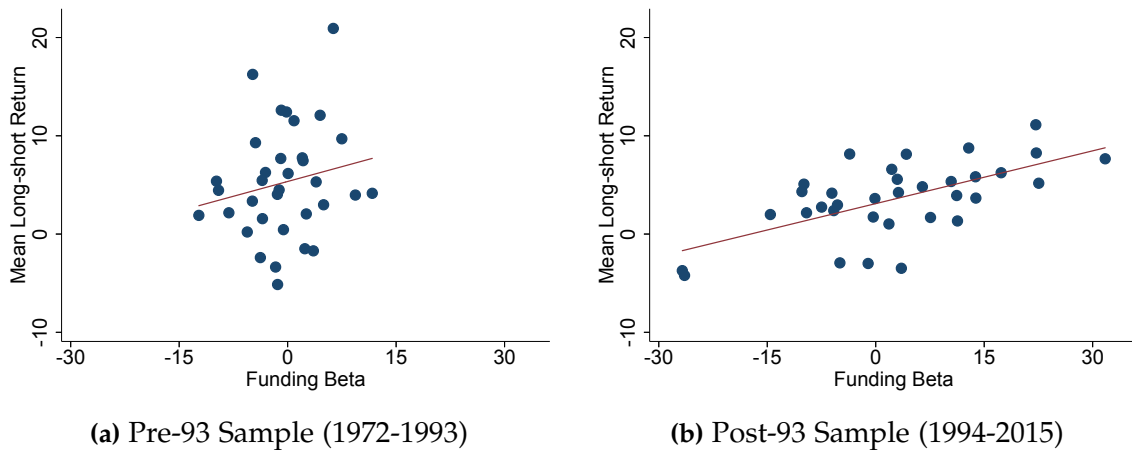
	Post-93 (1994-2015)					Pre-93 (1972-1993)	Pooled (1972-2015)
	Arb Fund- ing	Arb Fund- ing Select 25 <sup>†</sup>	Fama- French 3 Factors	Carhart 4 Factors	Fama- French 5 Factors	Arb Fund- ing	Arb Fund- ing
Arb Funding	0.18	0.20				0.20	0.32
(t-GenReg)	(2.30)	(2.29)				(1.43)	(3.84)
(t-GMM)	(1.30)	(1.74)				(0.67)	(1.62)
Market			-9.30	-6.96	-10.55		
(t-GenReg)			(-2.16)	(-1.60)	(-2.20)		
(t-GMM)			(-1.89)	(-1.52)	(-1.77)		
SMB			-0.57	0.19	-1.68		
(t-GenReg)			(-0.25)	(0.08)	(-0.69)		
(t-GMM)			(-0.22)	(0.07)	(-0.57)		
HML			1.50	2.02	1.28		
(t-GenReg)			(0.53)	(0.72)	(0.43)		
(t-GMM)			(0.52)	(0.72)	(0.44)		
MOM				5.25			
(t-GenReg)				(1.31)			
(t-GMM)				(1.33)			
CMA					1.70		
(t-GenReg)					(0.71)		
(t-GMM)					(0.72)		
RMW					2.93		
(t-GenReg)					(1.30)		
(t-GMM)					(1.40)		
Intercept	3.10	3.14	2.06	1.97	1.91	5.35	4.14
(t-GenReg)	(5.13)	(4.67)	(3.62)	(3.73)	(2.95)	(9.34)	(8.97)
(t-GMM)	(3.15)	(2.94)	(3.02)	(2.80)	(3.01)	(7.93)	(6.11)
Anomalies	34	25	34	34	34	34	34
Quarters	88	88	88	88	88	88	176
Adjusted $R^2$	0.37	0.29	0.49	0.54	0.53	0.01	0.28

Note: † denotes 25 anomaly assets chosen out of 34 by iteratively eliminating anomaly assets that are similar to a linear combination of the other anomaly assets until all anomalies have an  $R^2$  of less than 50% when linearly projected to all the other anomalies.

To see if this is the case, I try restricting my attention to a smaller set of anomaly assets that “span” my 34 anomalies. Starting with the 34 anomalies, I iteratively eliminate an anomaly asset that has the largest  $R^2$  when projected onto the other anomalies until the remaining anomalies have  $R^2$ s of less than 50% in such an exercise. Using these more “independent” anomalies shows a similar price of risk estimate (0.20 rather than 0.18) but a substantially smaller GMM standard error, making the price of risk statistically significant at the 10% level. There are other avenues for dealing with cross-anomaly correlations. One possibility is to use the generalized least squares (GLS) to penalize the anomalies whose residuals are volatile or highly correlated with the residuals of other anomalies, but this still relies on correctly estimating the cross-anomaly correlations of 34 anomaly assets using 88 time-series observations. Another possibility would be to estimate cross-anomaly correlations of the residuals in the entire sample rather than in the post-93 sample, an approach taken by Greenwood (2005b) to deal with the short time-series. I do not explore this approach in this paper.

To summarize the discussion on the price of risk, there is some evidence that arbitrageurs demand compensation for the funding risks that arbitrageurs themselves have created in the anomalies, causing the anomaly returns to survive in equilibrium. Although the GMM t-statistics for the price of risk is not large in the baseline regression, both the large  $R^2$  and the improved performance using the more independent anomaly assets suggests that focusing on the risks that arbitrageurs face is a fruitful path to take in order to jointly explain the cross-section of expected returns of a large set of asset pricing anomalies.

The intercept of the regression is positive and significant, consistent with arbitrageurs being constrained and anomaly assets being mispriced in some parts of the sample period. The intercept of 3.10% measures the average investment opportunity faced by arbitrageurs: arbitrageurs generate a risk-adjusted return that is on average 3.10% above the risk-free rate.

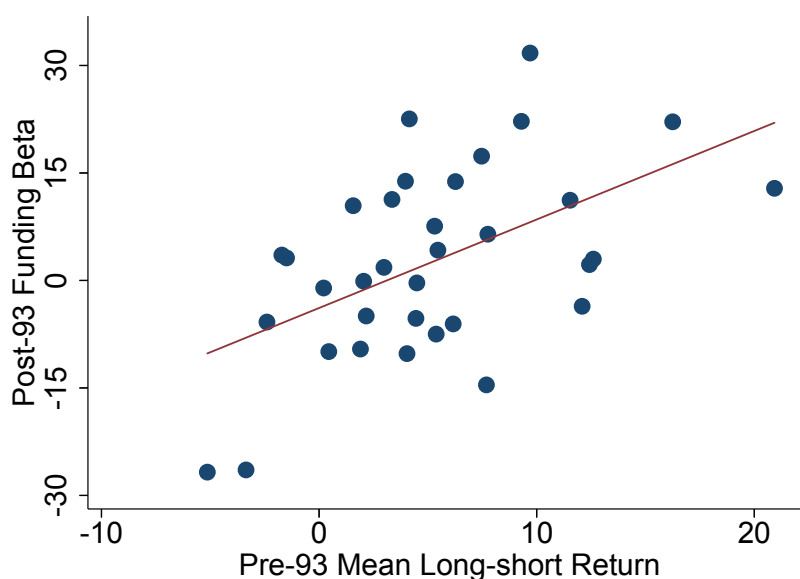


**Figure 4: Arbitrageur Funding Beta and Mean Long-short Return by Sample Period**

These figures plot the mean long-short returns and arbitrageur funding betas of 34 equity anomaly assets in the pre-93 (left) and post-93 (right) samples, proxying for pre-arbitrage and post-arbitrage periods. Return is in annualized percentage. Returns roughly line up with funding betas in the post-93 sample, but not in the pre-93 sample. Funding beta is measured by beta with respect to arbitrageur funding shocks measured by quarterly shocks to the leverage of broker-dealers.

How well does the pricing work in other sample periods? In the pre-93 sample, the estimated price of risk is actually similar in magnitude to that of the post-93 sample, but the low t-statistics suggest that the estimate is unreliable. This suggests that, during this period, arbitrageurs were still small ( $\mu \approx 0$ ) and did not have large presence in the 34 anomaly assets.

Combining the two sample periods, however, generates a much larger estimated price of risk than that in either of the two sample periods. The reason is that large returns come from the pre-arbitrage period and large betas come from post-arbitrage period. Hence, pooling them together inflates the price of risk. This point can be better illustrated using graphs. Figure 4b shows that, during the post-93 period, the anomalies with larger funding betas command larger expected returns. However, from the “alphas to betas” regression (Figure 5), we know that the anomalies with larger post-93 funding betas had even larger expected returns in the pre-93 period because the anomaly’s pre-93 return was what turned into post-93 endogenous risk. Furthermore, Figure 4a shows that pre-93 funding



**Figure 5: Pre-93 Long-short Return Predicts Post-93 Arbitrageur Funding Beta**

This figure plots the relationship between pre-93 mean long-short return and post-93 funding beta. The pre-93 (1972-1993) and post-93 (1994-2015) samples proxy for pre-arbitrage and post-arbitrage periods, respectively. Return is in annualized percentage. Funding beta is measured by beta with respect to arbitrageur funding shocks measured by quarterly shocks to the leverage of broker-dealers.

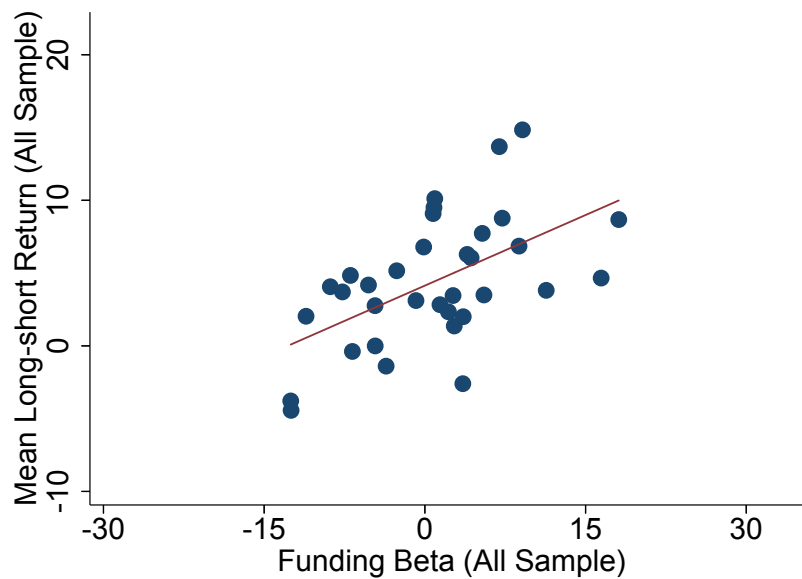
betas were centered around zero, which means that pooling the two samples leads to an attenuation of funding betas. Putting all this together, pooling the pre-arbitrage and post-arbitrage periods means that the betas are attenuated while the expected returns are inflated, which results in an inflated price of risk in the entire sample (Table 5 and Figure 6). The  $R^2$  is slightly below that of the post-93 sample regression since funding betas in the entire sample are essentially white noise added to post-93 funding betas.

In Adrian, Etula, and Muir (2014) too, the estimated price of risk in the entire sample is larger than that in the subsamples. There, the estimated price of risk is (with appropriate scaling to match the price of risk in this paper) is 0.62 in the pooled sample period of 1968-2009 but 0.21 and 0.18, respectively, in the subsamples 1968-1988 and 1989-2009.<sup>40</sup> This is still true even when the price of risk of the

---

<sup>40</sup>I take the last two numbers (0.21 and 0.18) from the online appendix to Adrian, Etula, and Muir (2014) available on Tyler Muir's website. Their leverage factor is expressed as a percentage whereas my funding shock is expressed as the original number, so I apply a scaling factor of 100.





**Figure 6: Arbitrageur Funding Beta and Mean Return in the Pooled Sample (1972-2015)**

This figure plots the mean long-short returns and arbitrageur funding betas of 34 equity anomaly assets over the entire sample period of 1972 to 2015. Return is in annualized percentage. Funding beta is measured by beta with respect to arbitrageur funding shocks measured by quarterly shocks to the leverage of broker-dealers.

leverage factor is estimated in a two-factor setting with the market excess return as the additional factor.<sup>41</sup>

### 3.6. Endogenous risk is generated during constrained times

I now test Proposition 4, which predicts that the endogenous covariation between anomaly asset returns and arbitrageur funding occurs only in times when arbitrageurs are constrained. At such times, a favorable funding shock leads them to dedicate more capital to anomaly assets, increasing their valuations and realized returns. When they are unconstrained, all anomaly assets are already being fully exploited, so a variation in arbitrage capital *does not* lead to variation in the prices of anomaly assets. Furthermore, during a period in which arbitrageurs are persis-

---

<sup>41</sup>This criticism nonetheless applies to this study's post-93 cross-sectional asset pricing as well. Because the mass of arbitrageurs  $\mu$  grew gradually over time even within the post-93 period, by the same logic as above, the prices of risk estimated from splitting the post-93 period into two are lower than that estimated from the entire post-93 period.

tently unconstrained, neither endogenous risks nor abnormal returns should arise in anomaly assets.

To test these predictions, I first need to identify times when arbitrageurs are more likely or less likely to be constrained. To proxy for the *level* (as opposed to the growth) of the arbitrageur funding condition, I take the four-year moving average of the arbitrageur funding shock  $f_t$ ,

$$f_t^{MA} \equiv \frac{1}{17} \sum_{s=-8}^8 f_{t+s}, \quad (51)$$

where  $s$  indexes quarter. This proxy for the level of arbitrageur funding is preferred to a more intuitive measure that simply removes a constant time trend from the original log leverage series, since such a measure would be exposed to medium-term changes in the leverage of broker-dealer leverage unrelated to the funding conditions of arbitrageurs. To mitigate concerns about data mining, I also try measuring the level of arbitrageur funding condition based on the quarterly series of average month-end VIX obtained from the Chicago Board Options Exchange (CBOE).<sup>42</sup> This generates similar results.

Given the measure of arbitrageur funding level, I define “unconstrained” (“constrained”) times within the post-93 period as the quarters in which the level of arbitrageur funding is above (below) the post-93 median. Figure 7 plots the constrained and unconstrained quarters based on the baseline moving-average measure  $f_t^{MA}$ . For comparison, I also plot the log leverage series with constant detrending and quarterly VIX. Based on either the moving-average of funding condition or VIX (top and bottom figures), the constrained quarters approximately fall into (i) the late 1990s to early 2000s and (ii) the financial crisis period. The late-1990s to early-2000s period includes the fall of long-term capital management (LTCM), the dot-com bubble and crash, and the 2003 mutual fund scandal, which

---

<sup>42</sup>The monthly series was downloaded from the CBOE website on August 10, 2016.

also affected hedge funds. These three events would have contributed to persistently low arbitrageur funding conditions during the period. In contrast, the log leverage series with constant detrending (middle figure) treats the mid-1990s and early-to-mid-2010s as constrained quarters. Including these periods as constrained quarters weakens the results.

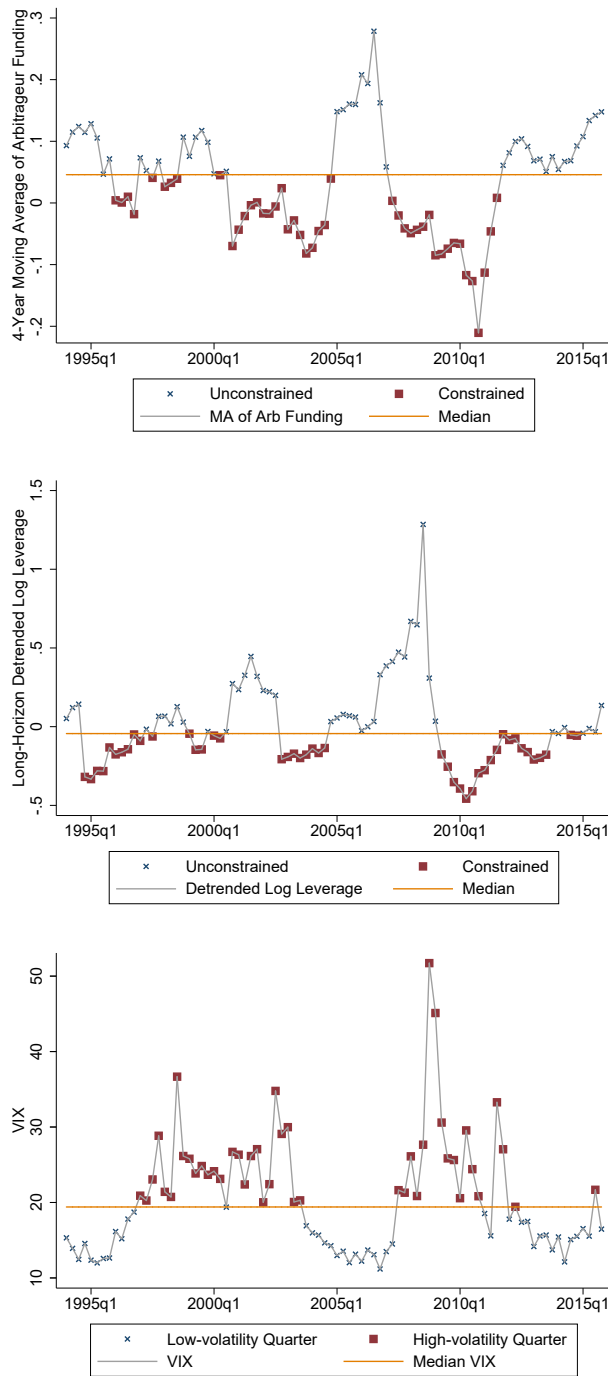
I proceed with the moving-average arbitrageur funding  $f_t^{MA}$  as the measure of arbitrageur funding level. With this, I first test whether funding betas arise only during constrained quarters. Defining a dummy variable  $\mathbf{1}(t \in \text{Constrained})$  to indicate the above-median  $f_t^{MA}$  quarters within the post-93 sample, I estimate

$$r_{j,t} = \left( a_j^{uncon} + \beta_j^{uncon} f_t \right) \mathbf{1}(t \in \text{Unconstrained}) + \left( a_j^{const} + \beta_j^{const} f_t \right) \mathbf{1}(t \in \text{Constrained}) + \varepsilon_{j,t} \quad (52)$$

in the time series for individual anomaly assets as seemingly unrelated regressions. I then test the following joint hypotheses: funding betas are jointly zero during the unconstrained time period ( $\beta_j^{uncon} = 0$ ); funding betas are jointly zero during the constrained time period ( $\beta_j^{const} = 0$ ); and change in returns from unconstrained to constrained times net of funding exposure are jointly zero ( $a_j^{const} - a_j^{uconst}$ ). Theory predicts the first hypothesis but rejects the latter two.

Panel A of Table 6 presents test results. The first two columns show that funding betas arise only during the constrained period, consistent with the prediction that arbitrageurs do not generate endogenous risks when they are unconstrained and all anomalies are fully exploited. The average funding betas are 0.83 and 4.65, respectively, during the unconstrained and constrained time periods. The last column tests if anomaly asset returns increase from unconstrained to constrained times. If constrained times are indeed when arbitrage capital is insufficient to eliminate anomaly returns, I would expect to see an increase in the zero-beta rate and hence positive  $a_j^{const} - a_j^{uconst}$ . Indeed, the changes  $a_j^{const} - a_j^{uconst}$  are jointly different from zero with a cross-sectional average of 4.26, which loosely implies

that the zero-beta rate rises by 4.26% from unconstrained to constrained periods.



**Figure 7: Proxies for Constrained and Unconstrained Quarters, Post-93**

Each of the three figures plots, for the post-1993 period (1994 to 2015), a series representing the level of arbitrageur funding condition as well as the “unconstrained” and “constrained” quarters defined as the quarters in which the level of arbitrageur funding condition is above and below the median, respectively. A high value of the series indicates a good funding condition for the first two figures and bad funding condition for the last figure. The level of arbitrageur funding condition is proxied by the 4-year moving average of the arbitrageur funding shock (top), the log leverage of broker-dealers with a constant detrending (middle), and quarterly average of month-end VIX obtained from CBOE (bottom).

Panel B of Table 6 asks how well an anomaly's latent mispricing—its pre-93 long-short return  $\bar{r}_j^{pre}$ —predicts the cross-section of unconstrained time betas, of constrained time betas, and of the change in  $a_j^{const} - a_j^{uconst}$ . For instance, the second column of Panel B is based on the regression

$$\beta_j^{const} = b_0 + b_1 \bar{r}_j^{pre93} + \eta_j, \quad (53)$$

estimated through GMM, as explained in the Appendix. Theory implies that pre-93 return only predicts the constrained time betas. It should not strongly predict  $a_j^{const} - a_j^{uconst}$  for the following reason. In the model, an anomaly's abnormal return is zero if arbitrageurs are unconstrained and  $\min\{\bar{r}_j, 1/E_0[\Lambda_1/\Lambda_0] - 1\}$  if arbitrageurs are constrained at time 0. That is, in the constrained case, all anomalies should have the same abnormal returns unless the anomaly's latent mispricing  $\bar{r}_j$  is smaller than the abnormal returns arbitrageurs are earning from exploited assets. This means that, in the cross-section of anomalies, pre-93 return should predict the change in abnormal return from unconstrained to constrained times during the post-93 period only for the small subset of anomaly assets that are not exploited by arbitrageurs during constrained times.

The results in Panel B of Table 6 are consistent with the prediction. Although the slope coefficients are positive in all columns, pre-93 mean return strongly predicts only the post-93 constrained time funding betas. The regressions in the first two columns are illustrated in Figure 8. Anomalies' funding betas during unconstrained time are scattered around zero and show no strong relationship with their latent mispricings measured by pre-93 mean returns. In contrast, funding betas during constrained times tend to be positive and show a strong positive relationship with pre-93 mean returns.

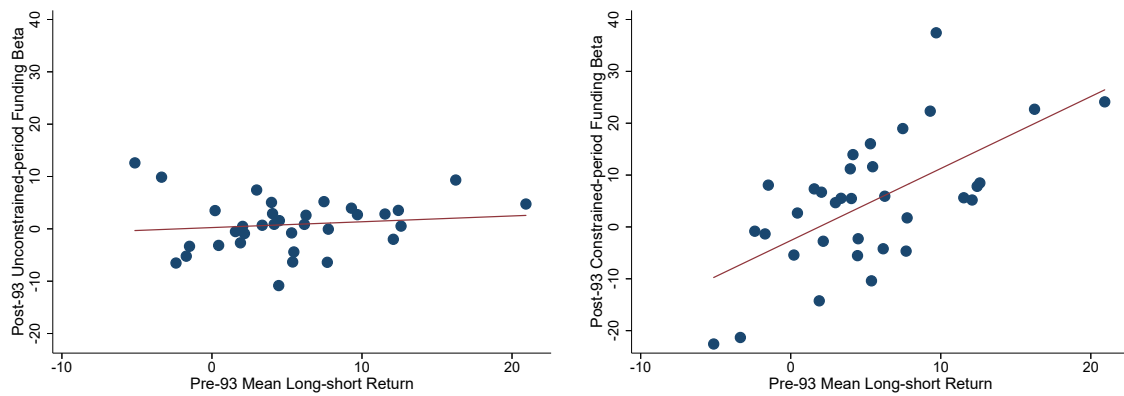
**Table 6: Funding Betas Are Generated during Constrained Times**

Data-generating process :

$$r_{j,t} = \left( a_j^{uncon} + \beta_j^{uncon} f_t \right) \mathbf{1}(t \in Unconstrained) + \left( a_j^{const} + \beta_j^{const} f_t \right) \mathbf{1}(t \in Constrained) + \varepsilon_{j,t}$$

This table analyzes the data-generating process in which anomaly assets' exposures to arbitrageur funding as well as their expected returns change from unconstrained to constrained times within the post-1993 period. The unconstrained and constrained times are proxied by quarters in which the 4-year moving average of arbitrageur funding shock  $f_t$  is above the post-1993 median and below the median, respectively. Panel A reports the mean of the estimated parameters ( $\beta_j^{uncon}$ ,  $\beta_j^{const}$ , and  $a_j^{const} - a_j^{uncon}$ ) in the cross-section of anomaly assets and jointly tests the hypothesis that the actual parameter values are zero. Here, the residual return  $\varepsilon_t$  is assumed to have zero serial correlations. Panel B reports the results from the regression that explains an anomaly asset's estimated parameters ( $\beta_j^{uncon}$ ,  $\beta_j^{const}$ , and  $a_j^{const} - a_j^{uncon}$ ), using its pre-1993 mean long-short return  $\bar{r}_j^{pre}$ . Here, t-OLS is the t-statistic calculated using only the residuals from the cross-sectional regression and accounts for a possible heteroskedasticity of residuals across anomaly assets. t-GMM refers to t-statistic obtained from the GMM estimation procedure and accounts for the effects of generated regressors and cross-anomaly correlations. Arbitrageur funding shocks are measured by quarterly shocks to the leverage of broker-dealers.

	$\beta_j^{uncon}$	$\beta_j^{const}$	$a_j^{const} - a_j^{uncon}$
A. Joint Hypothesis Test (e.g., $\beta_1^{const} = \dots = \beta_j^{const} = 0$ )			
Cross-sectional Average	0.83	4.65	4.26
$p$ (jointly zero)	1.00	0.00	0.01
B. Cross-sectional Prediction (e.g., $\beta_j^{const} = b_0 + b_1 \bar{r}_j^{pre} + \eta_j$ )			
Pre-93 Mean Long-short Return $\bar{r}_j^{pre}$	0.04	1.86	-0.14
(t-OLS)	(0.17)	(3.13)	(-0.57)
(t-GMM)	(0.14)	(2.14)	(-0.35)
Intercept	-0.48	-5.25	4.44
(t-OLS)	(-0.29)	(-1.19)	(3.03)
(t-GMM)	(-0.25)	(-1.82)	(2.53)
Anomalies	34	34	34
Adjusted $R^2$	-0.03	0.27	-0.02



(a) Predicting Unconstrained-time Funding  $\beta$       (b) Predicting Constrained-time Funding  $\beta$

### Figure 8: Mispricing Turns into Endogenous Risk Only in the Constrained Times

The left figure plots anomaly assets' betas with respect to arbitrageur funding condition during the unconstrained quarters of the post-1993 period on the y-axis and the mean long-short returns during the pre-1993 period on the x-axis. The right figure plots the anomaly assets' betas with respect to the arbitrageur funding conditions during the constrained quarters of the post-1993 period. For each anomaly asset, the betas in the unconstrained and constrained periods are estimated using the time-series regression  $r_{j,t} = (a_j^{uncon} + \beta_j^{uncon} f_t) \mathbf{1}(t \in Unconstrained) + (a_j^{const} + \beta_j^{const} f_t) \mathbf{1}(t \in Constrained) + \varepsilon_t$ . The unconstrained and constrained quarters are defined as the quarters in which the level of arbitrageur funding condition—the 4-year moving average of the arbitrageur funding shock  $f_t$ —is above and below the median, respectively. Arbitrageur funding shock is measured by quarterly shocks to the leverage of broker-dealers.



## 4. Auxiliary evidence for endogenous arbitrage risk

In this section, I carry out two additional tests of the endogenous arbitrage risk model. In Section 4.1, I show that the anomaly assets' post-93 exposures to the arbitrageur funding conditions *are not* explained by fundamental cash-flow exposures. In Section 4.2, I show that two main empirical tests show similar results if I measure an anomaly's endogenous risk as the beta with respect to arbitrageur's portfolio returns, which I proxy using the portfolio returns of equity market-neutral hedge funds.

### 4.1. Funding beta is not a fundamental cash-flow risk

In Section 3, I argued that anomaly assets have become endogenously risky due to arbitrage activity. This implies that the anomalies' risks are driven by arbitrageurs generating discount-rate news in the anomalies. Here, I test an alternative explanation, which is that the anomalies' betas with respect to arbitrageur funding are driven by their cash-flow exposure to the arbitrageur funding conditions.

To do this, I first obtain cash-flow news of an anomaly asset. Building on the Campbell-Shiller decomposition (Campbell and Shiller, 1988), Vuolteenaho (2002) show that a firm's price-to-book ratio can be decomposed into a discounted sum of future return-on-equity (ROE) and that of future returns:<sup>43</sup>

$$\ln\left(\frac{ME_{t-1}}{BE_{t-1}}\right) = \sum_{j=0}^{\infty} \rho^j \ln(1 + ROE_{t+j}) - \sum_{j=0}^{\infty} \rho^j \ln(1 + R_{t+j}), \quad (54)$$

where  $ROE$  is defined as the ratio of clean surplus earnings ( $X_t = BE_t - BE_{t-1} + D_t^{gross}$ ) to the firm's beginning-of-the-period book equity ( $BE_{t-1}$ ), and where  $R$  denotes the net return on the firm's stock. Rearranging (54) gives an expression

---

<sup>43</sup>An analogous decomposition of market betas appears in Campbell and Vuolteenaho (2004), Cohen, Polk, and Vuolteenaho (2009), and Campbell, Polk, and Vuolteenaho (2010).

that decomposes the firm's return into cash-flow news and discount-rate news components:

$$\begin{aligned}
r_{i,t} - E_{t-1}r_{i,t} &= \underbrace{\sum_{j=0}^{\infty} \rho^j (roe_{i,t+j} - E_{t-1}roe_{i,t+j})}_{\text{cash-flow news}} - \underbrace{\sum_{j=1}^{\infty} \rho^j (r_{i,t+j} - E_{t-1}r_{i,t+j})}_{\text{discount-rate news}} \\
&\equiv CF_{i,t} + DR_{i,t},
\end{aligned} \tag{55}$$

where  $roe$  and  $r$  respectively denote  $\ln(1 + ROE)$  and  $\ln(1 + R)$ . This means that anomaly  $j$ 's conditional beta with respect to the arbitrageur funding condition  $f_t$ , denoted  $\beta_{j,t}$ , can be decomposed into cash-flow and discount-rate betas:

$$\begin{aligned}
\beta_{j,t} &= \sum_{i \in I_j} w_{i,t}^j \frac{\text{Cov}(CF_{i,t+1}, f_{t+1})}{\text{Var}(f_{t+1})} + \sum_{i \in I_j} w_{i,t}^j \frac{\text{Cov}(DR_{i,t+1}, f_{t+1})}{\text{Var}(f_{t+1})} \\
&= \underbrace{\frac{\text{Cov}(N_{j,t}^{CF}, f_{t+1})}{\text{Var}(f_{t+1})}}_{\text{cash-flow beta}} + \underbrace{\frac{\text{Cov}(N_{j,t}^{DR}, f_{t+1})}{\text{Var}(f_{t+1})}}_{\text{discount-rate beta}} \\
&\equiv \beta_{j,t}^{CF} + \beta_{j,t}^{DR},
\end{aligned} \tag{56}$$

where  $I_j$  represents the constituents of anomaly asset  $j$ ,  $w_{i,t}^j$  is the portfolio weight of  $i$  (which can be negative), and  $N_{j,t}^{CF}$  and  $N_{j,t}^{DR}$  respectively denote cash-flow and discount-rate news about anomaly  $j$ . In summary, an anomaly's funding beta can be decomposed into cash-flow beta  $\beta_{j,t}^{CF}$  and discount-rate beta  $\beta_{j,t}^{DR}$ , and the cash-flow beta of an anomaly can be computed using the underlying firms' ROEs.

In implementing the test, I follow Campbell, Polk, and Vuolteenaho (2010) in proxying for a stock's cash-flow news with its  $ROE$  adjusted for inflation:  $roe_{i,t} = \ln(1 + ROE_t) - 0.4\ln(1 + r_{f,t})$ , where  $r_f$  is the risk-free rate. To prevent shorter-term trends in profitability from driving cash-flow betas, I follow Campbell and Vuolteenaho (2004), Cohen, Polk, and Vuolteenaho (2009), and Campbell, Polk, and Vuolteenaho (2010) in proxying for a stock's cash-flow news using a

**Table 7: Cash-flow Exposure Does Not Explain Funding Beta**

Panel A:  $\beta_1^{CF,post} = \dots = \beta_J^{CF,post} = 0$

Panel B:  $\beta_j^{post} = b_0 + b_1 \beta_j^{CF,post} + \eta_j$

This table asks if anomaly assets' cash-flow funding betas explain their exposure to arbitrageur funding shocks in the post-93 period. Panel A tests the joint hypothesis that the cash-flow funding betas are jointly zero. Panel B tests how well the cash-flow funding beta explains the cross-sectional variation in total funding beta. Cash-flow beta of an anomaly is obtained as the beta of a discounted sum of future cash-flow news with respect to the analogous discounted sum of arbitrageur funding shocks. An anomaly's cash-flow news is calculated by following the procedure of Campbell, Polk, and Vuolteenaho (2010). Arbitrageur funding shocks are measured by quarterly shocks to the leverage of broker-dealers.

	K = 1	K = 4	K = 8	K = 12
A. Joint Hypothesis Test (e.g., $\beta_1^{CF,post} = \dots = \beta_J^{CF,post} = 0$ )				
Cross-sectional Average	-0.52	-1.43	-3.10	-4.85
$p$ (jointly zero)	0.84	0.01	0.00	0.00
B. How much of funding beta is explained by CF beta?				
CF Funding Beta $\beta_j^{CF}$	1.17 (1.76)	0.33 (1.13)	0.07 (0.38)	0.01 (0.05)
Intercept	3.24 (1.59)	3.10 (1.41)	2.84 (1.26)	2.65 (1.18)
Anomalies	34	34	34	34
Adjusted $R^2$	0.09	0.03	0.00	0.00

Notes: In the parantheses are OLS t-statistics calculated with heteroskedasticity-consistent standard errors.

discounted sum of realized return on equities:

$$CF_t \approx \sum_{j=0}^{K-1} \rho^j \ln(1 + ROE_{t+j}), \quad (57)$$

where  $\rho = 0.975^{1/4}$  is the quarterly discount rate and  $K$  is the number of quarters. I use  $K = 1, 4, 8,$  and  $12$  quarters. When using a discounted sum of ROEs, I use the analogously discounted sum of arbitrageur funding shocks  $f_t$  as the factor.

I run two kinds of test on the cash-flow betas of 34 anomaly assets in the post-93 period. First, I test if the anomalies' cash-flow funding betas are positive and jointly different from zero. Second, I test how much of anomalies' funding betas are explained by cash-flow betas.

Table 7 reports the test results. In Panel A, I find that the anomalies' cash-flow funding betas are on average slightly negative. This suggests that anomaly assets do not have positive fundamental exposure to the variation in the arbitrageur funding conditions. In Panel B, I find that the cash-flow betas are only a small part of funding betas. Although the anomalies with higher cash-flow funding betas do tend to have higher total betas, the relationship is not statistically significant, and both the large intercept and the small  $R^2$ s imply that most of the total betas remain unexplained.

A related question is whether the anomalies are linked through the discount-rate components of their returns. Both the theoretical and empirical results of this paper suggests that this must occur in the post-93 period with more arbitrage activities. Lochstoer and Tetlock (2016) find that, between 1964 and 2015, anomalies have little commonality in the discount-rate or cash-flow components of their returns. It would be interesting to see whether this continues to be true in the more recent years with heightened arbitrage activities and with more extreme deciles of the anomalies. If so, it would be strong evidence against my results.

## **4.2. Equity market-neutral hedge fund return as an alternative proxy for shocks to arbitrage capital**

Thus far, the empirical tests have used funding beta to measure an anomaly's endogenous risk. This is consistent with a model in which shocks to arbitrageur capital  $k_t = w_t + f_t$  come entirely from funding shocks  $f_t$ , but in reality, shocks to arbitrageur wealth  $w_t$  orthogonal to funding shocks may generate large variation in arbitrage capital. For example, some of the anomaly assets held by arbitrageurs may underperform for reasons other than the arbitrageur funding conditions, and such a shock to the portfolio may lead the arbitrageurs to de-lever their other positions (e.g., Shleifer and Vishny, 1997). This channel of arbitrageur capital shock would be better measured by the portfolio return of the arbitrageurs.

In Table 8, I report results from testing Proposition 1 and Proposition 3 using beta with respect to equity market-neutral hedge fund return (“HF return beta”), from Hedge Fund Research, as the measure of endogenous risk.<sup>44</sup> Consistent with Proposition 1, an anomaly asset with a larger pre-93 long-short return attains a larger post-93 hedge fund return beta, although the result is not statistically significant at the 5% level based on GMM t-statistics. The second column shows that the result is stronger when I use post-93 hedge fund return correlation as the dependent variable, since it controls for changes in anomaly assets’ volatilities unrelated to arbitrageur trading, which would add noise to the estimated post-93 betas. The last column examines the cross-section of anomaly asset returns with their hedge fund return betas. The estimated price of risk is not large enough to completely overcome large GMM standard errors, but the  $R^2$  of 21% suggests that hedge fund return betas still help explain the cross-sectional variation in anomaly asset returns.

Figure 9 visualizes the regression results in the first and last columns of Table 8. Compared to Figure 5 and Figure 4b, the figures show slightly larger residuals, suggesting that arbitrageur funding shock is a better measure of shocks to arbitrageur capital  $k_t$  than hedge fund portfolio returns.

---

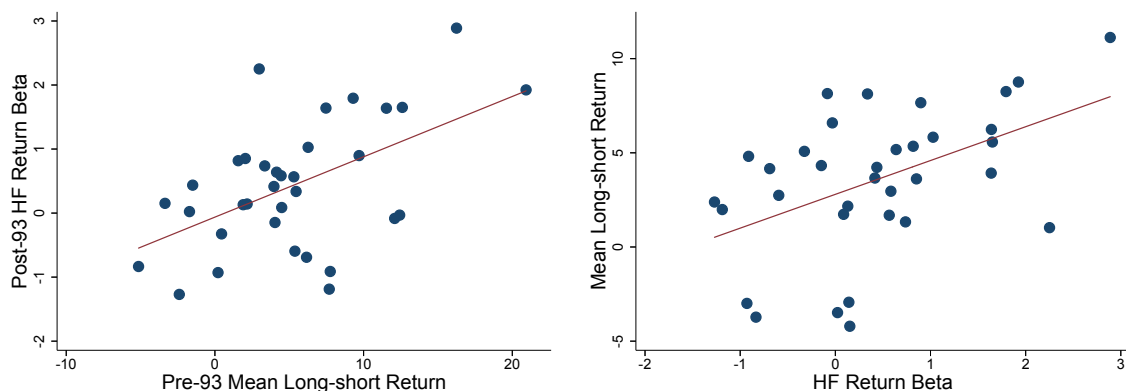
<sup>44</sup>The monthly series was downloaded on June 27, 2016 and converted to quarterly frequency.

**Table 8: Tests Using an Alternative Measure of Shocks to Arbitrage Capital**

$$\begin{aligned} \text{Mispricing to endogenous risk:} & \quad \beta_j^{HF,post} = b_0 + b_1 \bar{r}_j^{pre} + \eta_j \\ \text{Endogenous risk to expected return:} & \quad \bar{r}_j^{post} = \lambda_0 + \lambda_1 \beta_j^{HF,post} + e_j \end{aligned}$$

This table repeats the baseline regressions in Table 2 and Table 5 using the equity market-neutral hedge fund return as the measure of arbitrage capital shocks. The equity market-neutral hedge fund return is obtained by adjusting the market-neutral hedge fund return from Hedge Fund Research (HFR) by removing a small but significant exposure to market excess returns. t-OLS in the first two columns is the t-statistic calculated using only the residuals from the cross-sectional regression and accounts for a possible heteroskedasticity of residuals across anomaly assets. For the last column, t-OLS is the t-GenReg used in Table 5. Specifically, it refers to t-statistic corrected for generated regressors *but not* for cross-anomaly correlations. That is, to obtain the standard errors accounting for generated regressors, I allow for heteroskedastic residuals  $\varepsilon_j$  for the mean returns and do the correction derived by Shanken (1992), but under the assumption of  $Cov(\varepsilon_{j'}, \varepsilon_{j''}) = 0$  for  $j' \neq j''$ . t-GMM refers to t-statistic obtained from the GMM estimation procedure and accounts for the effects of generated regressors and cross-anomaly correlations. The correlation is reported in percentage (%).

	Mispricing Turns into Endogenous Risk (Proposition 1)		Endogenous Risk Explains Expected Return (Proposition 3)
	Post-93 HF Beta $\beta_j^{HF,post}$	Post-93 HF Correlation $\rho_j^{HF,post}$	Post-93 Mean Return $\bar{r}_j^{post}$
Pre-93 Mean Long-short Return $\bar{r}_j^{pre}$	0.09	1.27	
(t-OLS)	(3.84)	(2.76)	
(t-GMM)	(1.85)	(2.15)	
Post-93 HF Beta $\beta_j^{HF,post}$			1.80
(t-OLS)			(1.98)
(t-GMM)			(1.41)
Intercept	-0.06	-0.47	2.80
(t-OLS)	(-0.38)	(-0.13)	(4.66)
(t-GMM)	(-0.26)	(-0.10)	(3.42)
Anomalies	34	34	34
Adjusted $R^2$	0.25	0.13	0.21



(a) Mispricing Turns into  $\beta$  wrt Hedge Fund Return (b)  $\beta$  wrt Hedge Fund Return Explains Post-93 Mean Return

**Figure 9: “Alphas to Betas” and “Intermediary Asset Pricing” Using Equity Market-Neutral Hedge Fund (HF) Return as an Alternative Measure of Arbitrage Capital Shock**

The figures show that the main results of this paper are robust to using the equity market-neutral hedge fund (HF) return to measure arbitrage capital shock. The left figure shows that, in the cross-section of anomaly assets, a large pre-1993 mean long-short return predicts a large post-1993 beta with respect to HF return. The right figure shows that post-1993 mean long-short returns roughly line up with post-1993 betas with respect to HF return. HF return is measured by the return on the equity market-neutral hedge fund index provided by Hedge Fund Research (HFR). The market-neutral hedge fund return is adjusted by removing a small but significant exposure to market excess returns.

## 5. Conclusion

This paper shows that the act of arbitrage generates endogenous risk by turning assets' alphas into betas. The act of arbitrage causes the prices of anomaly assets to comove with shocks to arbitrage capital, and the strength of this comovement—the beta—depends on the amount of arbitrage capital devoted to each anomaly asset. Once these betas arise, they explain the anomaly returns that survive in equilibrium, since arbitrageurs require a compensation for the risk they themselves have created.

This paper links two seemingly disparate points of view: limits to arbitrage and intermediary-based asset pricing. From the limits-to-arbitrage point of view, this paper represents an extension of the idea that arbitrage activity generates endogenous risk to a cross-section of differently mispriced anomaly assets. From the intermediary-based asset pricing point of view, measuring this endogenous risk as the beta with respect to arbitrage capital shocks allows for a cross-sectional pricing of anomaly assets.

There are at least three avenues for future research. First, although I use equity anomalies in the U.S. market as a convenient test laboratory, the model's implications apply to any asset pricing anomalies. It would be interesting to empirically examine if arbitrage has turned mispricings into endogenous risks in other asset classes or other markets. Second, by a mechanism similar to the one explored in this paper, arbitrage activity may cause differently mispriced assets to be differently exposed to crash risks or tail risks.<sup>45</sup> One may formalize this conjecture in a model and take its cross-sectional predictions to data, thereby endogenizing the cross-sectional exposures to higher-order risks that may be priced in equilibrium (e.g., Bali, Cakici, and Whitelaw, 2014; Kelly and Jiang, 2014; Amaya, Christof-

---

<sup>45</sup>Stein (2009) provides such a framework in a single-asset environment. An extension would require embedding differently mispriced assets into such a model.



fersen, Jacobs, and Vasquez, 2015; Bollerslev, Todorov, and Xu, 2015). Third, this paper implies that arbitrageurs such as hedge funds have strong preferences over the characteristics of securities that are known to generate abnormal returns. Hence, in the spirit of Berry, Levinsohn, and Pakes (1995) and Koijen and Yogo (2016), one can estimate arbitrageurs' preferences over those characteristics using institutional holdings data and short interest data. From this, one can infer how arbitrage trading—through both long and short positions—shapes the liquidity of the securities with anomaly characteristics.

# **Chapter II.**

**Does Liquidity Cause Market Return Reversals? A**

**Natural Experiment**

## 6. Introduction

A common explanation for the stock market's short-horizon reversal is aggregate liquidity shocks.<sup>46</sup> The story is that the market sometimes experiences an excessive demand for buying or selling stocks immediately. When this happens, liquidity suppliers require a larger compensation for absorbing the liquidity demand, increasing the expected return from betting against the market (Campbell, Grossman, and Wang 1993) (CGW).<sup>47</sup> This means that the aggregate price movement on that day is more likely to reverse on subsequent days. The causality from an aggregate liquidity shock to market return reversal, albeit conceptually simple, has not been tested using an exogenous shock to aggregate liquidity since such a shock is rarely observed.

This study uses daily temperature variation during the summers of late 19th to early 20th century Manhattan to identify exogenous variation in aggregate liquidity demand. In the absence of advanced communications and transportation technologies, trading stocks in this period involved a substantial physical effort. Naturally, extremely hot summer weather discouraged investors from participating in the daily stock market, causing a 1.2% decline in New York Stock Exchange (NYSE) trade volume for every 1 degree Celsius rise in Manhattan temperature in the summers of 1889-1902.<sup>48</sup> Summer heat, however, had little effect on suppliers of liquidity since they traded stocks using telephones and automatic ticker machines (Donnan 2011), were located at or near the exchange (Longcore and Rees 1996), and traded stocks for a living. This allows me to interpret the decline in

---

<sup>46</sup>From 1926 to 2014, the Standard and Poor's 500 (S&P 500) index had an average daily return of 0.04% but an average absolute value of the return of 0.74%, implying large short-horizon reversals that cannot be explained by daily information flow alone.

<sup>47</sup>Also see Grossman and Miller (1988) for an earlier analysis of supply and demand for market liquidity.

<sup>48</sup>See section 7 for Wall Street Journal articles attributing low trade activities on extreme summer heat.

trade activity on a hot summer day in Manhattan as due primarily to a decline in liquidity demand.

Using temperature variation in the summers of 1889 to 1902 to measure exogenous variation in liquidity demand, I find that liquidity demand has a large effect on market return reversal over a short horizon. A fall in liquidity demand that generates a 1% drop in daily trade volume increases the daily autocorrelation of market returns by 0.018. Other measures of market return reversal—the probability of a market return reversal on the next day and the expected return from a simple reversal strategy on the market index—consistently decline with the exogenous fall in aggregate liquidity demand. The effect is larger for weekly returns; a fall in liquidity associated with a 1% drop in weekly trade volume increases the weekly autocorrelation of market returns by 0.046. Although years 1889-1902 are chosen as the sample period for reasons discussed in section 7, the results are robust to using alternative sample periods.

I back my findings on aggregate market returns using individual stock prices collected from the Commercial and Financial Chronicle. I use the collected data to compute the cross-sectional average reversal profitability of individual stocks that may better capture the return from aggregate liquidity provision than a reversal in the market index. This measure of reward from aggregate liquidity provision as well as the fraction of stocks experiencing a reversal on the next day decrease as aggregate liquidity demand thins out.

This paper reaffirms that idea that shocks to aggregate liquidity change the required reward from aggregate liquidity provision and cause short-horizon reversals in a stock market index. This is consistent with the evidence from individual stock return reversal studies that trades due to an index redefinition—an arguably exogenous liquidity shocks on applicable stocks—cause price changes that shortly reverse (Harris and Gurel 1986; Greenwood 2005a).<sup>49</sup> To study market index re-

---

<sup>49</sup>Related evidence exists on individual stocks. Reversal profitability is higher in illiquid stocks

turn reversal or the cross-sectional average of individual stock reversals, however, the literature has relied on aggregate trade volume (CGW) or aggregate volatility (Nagel 2012) as measures of aggregate liquidity shocks. In contrast, this study uses exogenous shocks to aggregate liquidity to confirm an effect that is consistent in direction but larger in magnitude.

The identified effect of aggregate liquidity demand complements but is distinct from the effect of aggregate liquidity supply. Nagel (2012) shows that reversal profitability increases with volatility, as liquidity suppliers' balance sheet capacity deteriorates. This supply-side effect on aggregate market liquidity, however, moves relatively slowly and is likely to matter only on limited occasions when the capacity constraint binds for liquidity suppliers. I find this to be true in my data; focusing on moderate-return events by winsorizing returns at 5% removes the volatility effect on daily autocorrelation.<sup>50</sup> This suggests that day-to-day changes in aggregate liquidity imbalance and market return autocorrelation are more likely to be governed by time-varying demand for, rather than supply of, liquidity.

This paper strengthens the case that demand curves for stocks slope down. First, the estimated effect of liquidity on market returns implies that the demand for the aggregate stock market slopes down, just as it does for individual stocks (Shleifer 1986; Avramov, Chordia, and Goyal 2006a). That is, even at the aggregate stock market level, where neither asymmetric information nor limited attention plays an important role, arbitrageur capital does not immediately counteract non-informational trades. Put differently, liquidity trades exert a significant price pressure on the market index, providing an empirical ground for the role played by noise traders in various asset pricing models (e.g., Grossman and Stiglitz 1976;

---

(Pastor and Stambaugh 2003; Avramov, Chordia, and Goyal 2006a) and stocks with high trading activity, a proxy for non-informational trades (Conrad, Hameed, and Niden 1994). Studies have also shown that uninformed trades create non-fundamental volatility over a short horizon (Hellwig 1980; Wang 1994; Avramov, Chordia, and Goyal 2006b; Koudijs 2015).

<sup>50</sup>Winsorization has little effect on the relation between volatility and weekly return autocorrelation.

Verrecchia 1982; De Long et al. 1990). Second, the result on the cross-sectional average of individual stock reversals implies that reversals due to non-informational trades is a general phenomenon that occurs in a wider set of stocks than stocks experiencing an index redefinition.

An unanswered question is, what causes the variation in aggregate liquidity demand in the first place? Although there is no direct way to answer this question, I find that a reduction in today's trade volume today due to hot weather *does not* lead to an increased trade volume on the next day—discouraged investors do not substitute their trades intertemporally. This suggests that short-lived motives such as sentiments, rather than true liquidity needs, underlie the identified demand for liquidity. This is consistent with speculative trades by retail investors inducing short-horizon return reversals (Barber, Odean, and Zhu 2009; Foucault, Sraer, and Thesmar 2011).

This study is related to other strands of literature. First, several studies use weather to study the impact of investor psychology on stock returns (Saunders 1993; Hirshleifer and Shumway 2003; Kamstra, Kramer, and Levi 2003; Cao and Wei 2005; Bassi, Colacito, and Fulghieri 2013; Goetzmann, Kim, Kumar, and Wang 2014). I find that weather affected stock market participants' willingness to trade prior to technological advancement in the 20th century. Secondly, outside of finance, studies have documented that hot weather retards economic activity (Dell, Jones, and Olken 2014 review this growing literature). My study complements these findings by presenting a high-frequency (daily) relationship between temperature and stock trading activity. Finally, this paper is part of the growing literature in finance exploiting natural experiments. Two examples are Giroud, Mueller, Stomper, and Westerkamp (2011), who use unexpected snow to instrument for ski resorts' operating performance, and Koudjis (2015), who use weather-induced disruption of news arrival from London to Amsterdam to study the effect of information on individual stock price volatility.

The paper proceeds as follows. In section 7, I present historical background, choose the test sample period, and discuss the data and variables used. In section 8, I quantify the effect of summer temperature on aggregate liquidity by estimating the temperature effect on aggregate trade volume as well as on the fraction of idle stocks during the sample period. I present my main empirical results in section 9. There, I use temperature as an exogenous proxy for aggregate liquidity demand to estimate its effect on various measures of market return reversal. The final section concludes.

## **7. Historical background, data, and measurement**

Wall Street sweltered in the recurrence of extreme heat at the start of the new week. ... [T]he high altitude to which the temperature soared perceptibly reduced speculative participation in the stock market . . . (“Heat Cuts Down Volume: Soaring temperature reduces speculative participation in market,” Wall Street Journal, 16 Jun 1925)

### **7.1. Hot weather as a proxy for a reduction in liquidity trades**

The quote above suggests that extreme summer heat “perceptibly” reduced speculative trades in the daily stock market. If it is true, summer temperature variation in Manhattan during this period represents exogenous variation in aggregate liquidity demand. How exactly, however, did liquidity demand fall on hotter days? In this section, I use narrative records to illuminate the mechanism through which heat reduced speculative trades and to choose an appropriate sample period during which that effect was strong.

Three groups of individuals interacted in the late 19th to early 20th century stock market: ‘outside customers’, ‘brokers’, and ‘traders’. Outside customers were retail investors and small institutional investors without in-house brokers or traders on the exchange. They traded stocks through a broker at a ‘commission house’ (brokerage house) and communicated with the broker in-person by

traveling to the commission house or through telegraphs, letters, and if available, telephones.<sup>51</sup> Some institutional investors had access to automatic ticker machines and received continued updates on the market without calling or visiting the commission house.<sup>52</sup> Those without automatic ticker machines could observe stock price movements at the commission house:

Most such houses provided a “customers’ room” where quotations from all exchanges of which the house is a member are posted on a blackboard as fast as they come out on the tickers, and the principal newspapers and news services are kept on file (p.108 of Selden 1919).

Brokers acted as agents for the outside customers and held membership at the exchange. Upon receiving customer orders from their office partners, usually via a messenger boy, they went to appropriate posts to execute the trades.

The last group—traders—comprised ‘capitalists’, ‘room traders’, and ‘specialists’ who spent the entirety of their working hours at the exchange. Capitalists were employees of large institutional investors that wished to make large purchases and sales without paying a commission to brokers.<sup>53</sup> Room traders made a living by betting on short-term fluctuations and essentially provided liquidity by “standing ready to buy a little below the current market or to sell a little above” (p.85 of Selden 1919). Specialists were the principal providers of liquidity. They specialized in dealing one or more securities and worked as the auctioneer for brokers in addition to trading on their own.<sup>54</sup> Room traders and specialists together made up the liquidity suppliers at the time.<sup>55</sup>

---

<sup>51</sup>See Selden (1919) and Beckert (2003).

<sup>52</sup>Approximately 1000 automatic ticker machines existed in bank and brokerage offices in New York as of 1880 (Donnan 2011).

<sup>53</sup>The annual cost of membership was over \$3,000 a year as of 1917, equivalent to \$56,000 in 2014.

<sup>54</sup>The last four sentences, including the cost of membership in the exchange, are based on Selden (1919).

<sup>55</sup>It is unclear whether capitalists also provided liquidity by holding mispriced securities tem-



Hot weather reduced the demand for trading by outside customers. The reduction in outside customer orders led to a low attendance in commission houses and reduced activity of brokers on the exchange, both of which are cited as responsible for low trading volume:

A slim attendance in commission houses, and the absence of many brokers on the floor owing to the excessively high temperature, accounted for substantial reduction in the volume of dealings yesterday ("Score Further Gains in Market Generally Restrained by Heat," Wall Street Journal, 20 Jun 1929).

To be fair, heat also affected traders:

The heat had a good deal to do with keeping the market inactive, for it thinned out the attendance on the floor and left the traders who remained more inclined to make themselves comfortable on the seats surrounding the posts . . . ("Trading restricted," Wall Street Journal, 23 Jun 1909).

However, while traders may have retreated to seats at the posts or left the posts entirely without brokers to trade with, they were still required to have physical presence at the exchange, ready to execute trades when brokers start appearing again.

This helps clarify the channel through which heat reduced liquidity demand. On normal days, outside customers put trade orders through brokers, who then executed the trades by finding liquidity suppliers. On these days, liquidity suppliers had to absorb a large amount of trade volume from outside customers, and so charged a premium on the liquidity service in the form of a higher likelihood of reversal. However, on hotter days, outside customers became relatively inactive, and so did the brokers who acted upon outside customer orders. Without much demand for their service, liquidity suppliers charged a smaller premium on their liquidity service, which resulted in a lower likelihood of reversal.

---

porarily on their balance sheet. Brokers and the three types of traders comprised almost the entirety of 1,100 members of the NYSE in 1917.

## 7.2. Empirical implementation

Although the stock market in the late 19th to early 20th century allows us to observe exogenous variation in aggregate liquidity demand, executing the idea empirically requires two considerations. The first is the choice of temperature that was most important to investors at the NYSE. Manhattan temperature is a natural choice, but to what extent were NYSE's outside customers concentrated near Manhattan? The second is the choice of sample period. Although narrative records suggest that heat affected stock trading even in late 1920s, technological progress in the 20th century would have made temperature variation an increasingly noisy measure of liquidity. This implies a trade-off between having a longer sample and using stronger variation in liquidity demand. I choose 1889 to 1902 as the default sample period for reasons discussed below.

Manhattan temperature is a natural choice for temperature felt by NYSE investors, as circumstances in the late 19th to early 20th century dictate that a bulk of NYSE trades would have originated from New York City. In this period, the underdevelopment of long-distance communications impaired the competitiveness of trading from another city. Long-distance automatic ticker service was unavailable until 1905, when the service was first set up between New York and Philadelphia. Before then, it took brokerage houses in other cities an additional 15 minutes to receive market updates from New York, because the employees had to hand-copy the stock quotations received by Morse on manifold sheets (Tilghman 1961). Furthermore, major cities like Boston and Philadelphia had regional exchanges that local traders could use, giving investors in these cities less reason to trade at the NYSE.<sup>56</sup> Indeed, I show in section 8 that, controlling for Manhattan temperature, temperatures in other cities have little adverse effect on NYSE trading volume.

---

<sup>56</sup>It is also worth noting that banks and financial services have long had high concentration near NYSE on Wall Street (Longcore and Rees 1996).

What is the appropriate sample period? Trade volume data begins in the year 1889, which restricts me to begin the sample then, but no natural choice exists for the end year. To make a choice, I briefly review technological advancements in the early 1900s that would have weakened the temperature effect on trading. Large-scale air conditioning was first introduced to New York in 1903 (Buchanan 2013). After 1903, as brokerage houses started installing air conditioning, hot weather would no longer have discouraged outside customers from visiting their brokers. Transportation and communication technologies underwent significant improvements in the early 1900s. The advent of the New York City subway system in 1904 would have made it easier for outside customers to visit brokerage houses on hot days. In the late 1800s, telephones and automatic ticker machines were rare, so a typical investor wishing to make two-way communications or receive continuous updates on the market would have had to travel to a nearby telephone exchange or visit the customer's room at a brokerage house.<sup>57</sup> Nevertheless, telephones became more widespread in the early 1900s, with the number of telephones growing at an annual rate of approximately 20% from 1900 to 1910.<sup>58</sup>

Based on these facts, I choose to end the sample period in 1902, which is the last year prior to the aforementioned events that would have weakened the temperature effect. Since the major developments in long-distance communication occurred after 1902, this choice also ensures that Manhattan temperature is a good proxy temperature felt by the majority of NYSE investors within the sample period. Introduction of various technologies, however, was a gradual process, and no single year—including 1903—represents a clear jump in the technological progress. I therefore repeat my main analyses with longer sample periods and show that expanding the sample as far as 1889-1920 has little effect on the sig-

---

<sup>57</sup>One in every 300 persons had a telephone in 1890 (U.S. Census Bureau 1975), and approximately 1000 automatic ticker machines existed in bank and brokerage offices in New York as of 1880 (Donnan 2011).

<sup>58</sup>The calculation is made based on numbers provided by U.S. Census Bureau (1975).

nificance of the estimated effects. Within the selected sample period, I only use data from the summer season to mitigate any seasonal effects and to ensure that temperature has a monotonically negative effect on outside customer trades.

### **7.3. Data and measurement**

I use NYSE index returns measured by the Dow Jones Industrial Average (DJIA) and S&P 500, NYSE individual stock prices hand-collected from digital copies of the Commercial and Financial Chronicle, NYSE total trading volume taken from the NYSE website, and temperature data obtained from the National Climatic Data Center (NCDC), all at the daily frequency.

#### **NYSE index returns**

Daily NYSE index return is measured by DJIA return for years 1889-1925 and S&P 500 return for 1926-2014. I obtain the DJIA returns from 5/26/1896 to 12/31/1925 from the Dow Jones website, and the prior years' cumulative-dividend returns are kindly provided by Schwert (1990).<sup>59</sup> I obtain S&P 500 returns from the Center for Research in Security Prices (CRSP). A notable difference between the two indices is that DJIA is price-weighted while S&P 500 is value-weighted, but the difference is likely to be small since larger stocks tended to have higher prices during the sample period.

#### **Unique data of pre-1903 NYSE individual stock daily returns**

Since no daily individual stock data are available prior to 1926, I collect stock prices and records of all NYSE-listed common stocks in the summers (Jun 11-Sep 10) of 1889 to 1902. This information is hand-collected from the Commercial and

---

<sup>59</sup>The Dow Jones data was retrieved in November 2015 from <http://www.djaverages.com/?go=industrial-index-data&report=performance>.

Financial Chronicle (the Chronicle).<sup>60</sup> The collected information includes daily minimum and maximum prices as well as the information on idle stocks (no buy or sell trade executed) on each day. I then compute daily individual stock returns as the percentage change in the midpoint price. The Chronicle denotes whether there was a dividend on a particular day but do not specify the dividend amount. Since I cannot compute gross-dividend returns without information about the dividend amount, I simply exclude these observations with dividend payments. The result is an unbalanced panel of around 90,000 daily individual stock return observations.

### NYSE trading volume

Daily NYSE trading volume over the years 1888-2014 is obtained from the NYSE website.<sup>61</sup> I compute detrended trading volume,  $Tradvol_t$ , as the log deviation of daily trade volume per trading hour from its trailing 1-year moving average:

$$Tradvol_t = \ln \left( \frac{\text{trade volume}_t}{\text{trade hours}_t} \right) - \frac{1}{N_t} \sum_{s=1}^{N_t} \ln \left( \frac{\text{trade volume}_{t-s}}{\text{trade hours}_{t-s}} \right) \quad (58)$$

where  $N_t$  is the number of trading days in the previous 365 days. This measure of trading volume is similar to those in CGW and Chen, Hong, and Stein (2001), except that it uses trading volume instead of share turnover and that it adjusts for trading hours to make weekday and Saturday trade volumes comparable.<sup>62</sup> Without the trading hour adjustment, trade volume is consistently lower on Sat-

---

<sup>60</sup>Trading volume is available weekly, and I do not use this information in the paper. The hard copy of the Chronicle is not available for June 1894, so that month is excluded from my analyses. The supplementary online appendix to this paper has an example page from the Chronicle.

<sup>61</sup>The data was retrieved in October 2015 from <https://www.nyse.com/data/transactions-statistics-data-library>. The data contains 4 instances in which two volume observations have the same date, and in those cases, I average the two numbers. None of the dates, however, are in the period of interest (1889-1902). Analysts from the NYSE kindly went over and removed all other instances of duplicates.

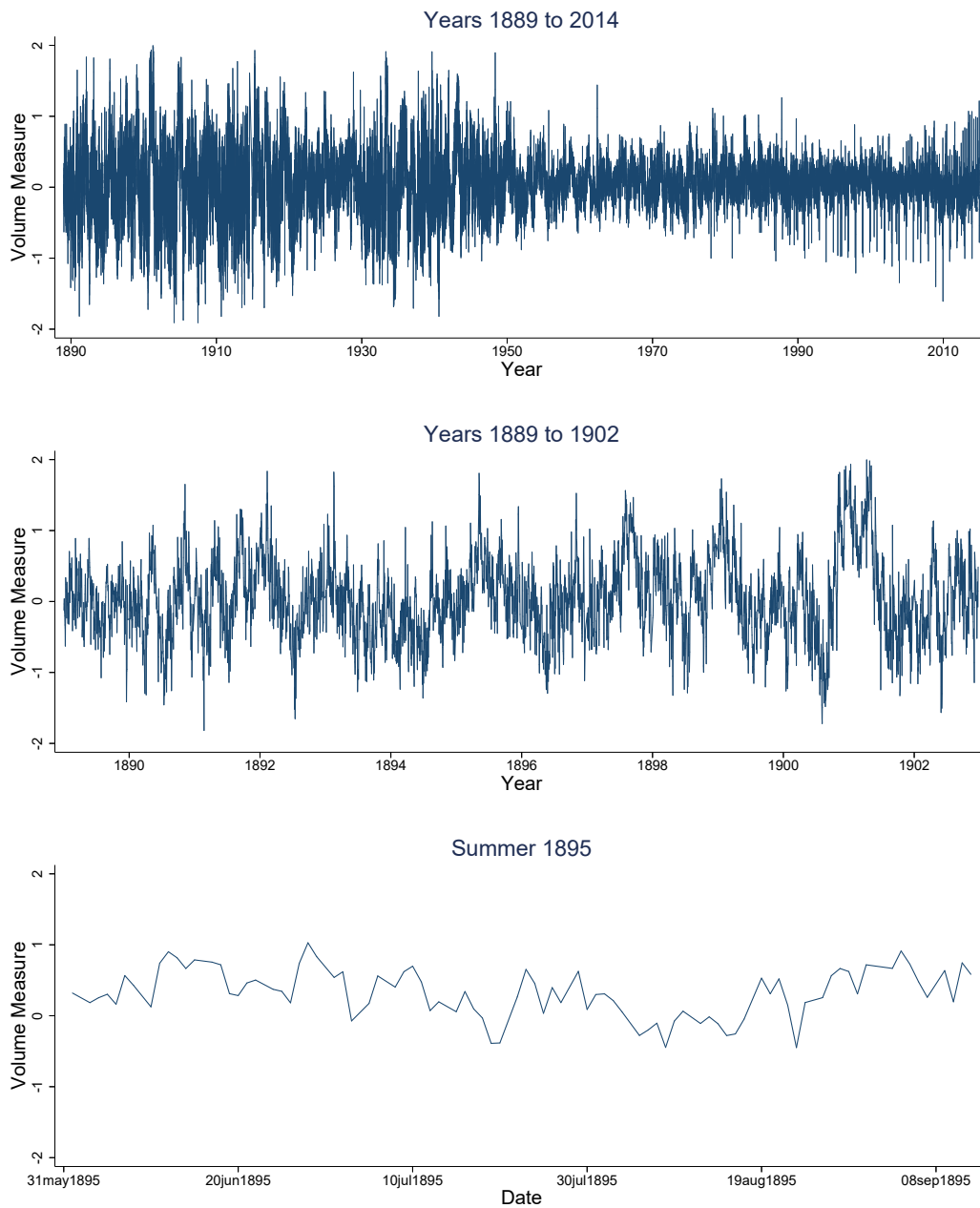
<sup>62</sup>Records of the number of total shares outstanding at NYSE are not available from the 19th century, but using trading volume is equally valid in this study exploiting daily variations.

urdays (which had shorter trading hours), which causes trade volume to function as a Saturday dummy. The results in this study are robust to using trade volume unadjusted for trading hours.

The result is the daily volume series from 1/2/1889 to 12/31/2014, plotted in Figure 10 for the entire sample, the sample of 1889-1902, and the summer of 1895 (representing the middle of the test sample).<sup>63</sup> The detrended volume process is still fairly persistent with the average daily autocorrelation of 0.73 in the 14 summers of 1889-1902. Although this is not considered near unit root, I do take the possibility of a spurious regression seriously, conducting placebo tests with temperatures in other cities and data from later time periods.

---

<sup>63</sup>The drop in the volatility of trade volume around 1950 roughly coincides with the end of Saturday trading. This does not affect my analysis, which focuses on the 1889-1902.



**Figure 10: Detrended Aggregate Trade Volume**

The figure plots, for three different time periods, daily detrended NYSE total trading volume. The detrended volume is computed as the deviation of log total trading volume per hour from its 1-year moving average.

## Manhattan temperature

I obtain daily temperatures in Manhattan Central Park during 1879-2014 from the NCDC website.<sup>64</sup> For additional tests, I also collect daily temperature of Cambridge in Massachusetts (MA), Jacksonville in Florida (FL), San Diego in California (CA), Oxford in United Kingdom (UK), and Sydney in Australia (AU) during 1879-2014 (except that the records of temperature in Cambridge MA start in 1885).<sup>65</sup> The collected data consist of daily maximum and minimum temperatures in degrees Celsius ( $^{\circ}\text{C}$ ), which I use to predict average trading hour temperature  $T$  on each day. I compute average trading hour temperature under the assumption that an intraday variation in temperature reaches the daily maximum at 15:00 and minimum at 5:00 (Lonnqvist 1962) and that the temperature change within a day is linear in time.

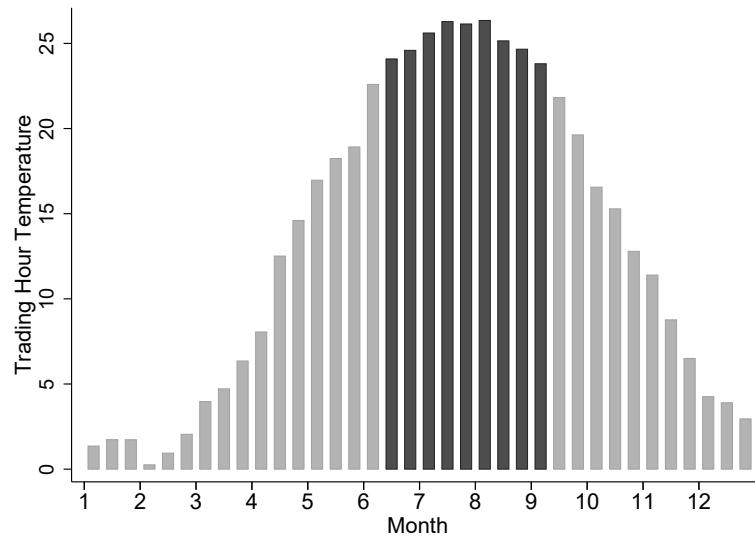
I focus my analyses on the summer season. In particular, I use June 11th to September 10th, the hottest 90 days of a year when average temperature is computed in 10 days' interval (Figure 11). Restricting my analyses to the summer season serves two purposes. First, it prevents seasonality in trade volume as well as in other variables from affecting the empirical results. Stock trade volume is known to be lower in the summer (Gallant, Rossi, and Tauchen 1992; Bouman and Jacobsen 2002; and Hong and Yu 2009), which may be a combination of temperature and seasonal effects (e.g. investors take holidays in hottest parts of the year). Second, using only the summer season ensures that temperature has a monotonic effect on human physiology. Sepannen, Fisk, and Lei (2006) show that task performance in an office environment peaks around 20-22 $^{\circ}\text{C}$  and decreases as temperature increases or decreases from this point. Summer temperature in

---

<sup>64</sup>The data was retrieved in July 2013 from <http://www.ncdc.noaa.gov/data-access/land-based-station-data>.

<sup>65</sup>Although the NCDC interface does not allow me to choose five cities in a systematic manner, I tried to obtain the longest time-series data from scattered geographical locations.





**Figure 11: Manhattan Temperature by 10-Days' Interval (1889-1902)**

The figure plots the average trading hour temperature in Manhattan during the years 1889 to 1902 by 10-days' interval. The summer period used in the paper (Jun 11-Sep 10) is highlighted in dark gray. The trading hour average temperature  $T$  is estimated from the maximum and minimum temperatures assuming a linear intraday variation in temperature with the maximum at 15:00 and minimum at 5:00 (Lonnqvist 1962).

Manhattan is almost always above 22°C, so a rise in temperature is expected to have only an adverse effect on outside customer trades during the summer season.

Even within the summer, however, temperature exhibits seasonality. To isolate the temperature effect from the seasonal effect within the summer, I compute seasonal temperature  $\bar{T}$  as the average trading hour temperature on the same day for the previous 10 years and include it as an additional control variable. This is a simple way to control for seasonality within the summer, but I show that controlling for seasonality through fixed effects produces similar results. In some analyses, I also use deseasonalized temperature  $\tilde{T}$ , defined as temperature minus seasonal temperature.

### Other variables and merged data

Control variables used in this study are conditional volatility of returns ( $\sigma_t^2$  or *volatility<sub>t</sub>*), dummy variables for Friday and Saturday, three dummy variables indicating single trading holiday, two trading holidays, and three or more trading holidays until the next trading day. Gallant, Rossi, and Tauchen (1992) use day of the week dummies to control for the predicted patterns in trading volume within a week and a variant of trading days' gap dummies to control for the effect of having different number of holidays between two consecutive trading days. While Gallant, Rossi, and Tauchen (1992) construct dummy variables for trading holidays based on the number of holidays since the previous trading day, I find that dummy variables based the number of holidays until the *next* trading day have a greater explanatory power. Using the alternative specification has little effect on results.

Return volatility is an important determinant of return autocorrelation. High volatility reduces the inventory-absorption capacity of market makers, leading them to require a high compensation for liquidity provision in the form of a low return autocorrelation (Nagel 2012). Put differently, volatility represents a shock to liquidity supply, whereas temperature represents a shock to liquidity demand during the sample period.

I estimate daily return volatility separately in each of the five sample periods to be defined shortly. I do this using a generalized autoregressive heteroskedasticity (GARCH) model with two ARCH terms and four GARCH terms, which is a model selected based on BIC criterion on the pre-1903 data:

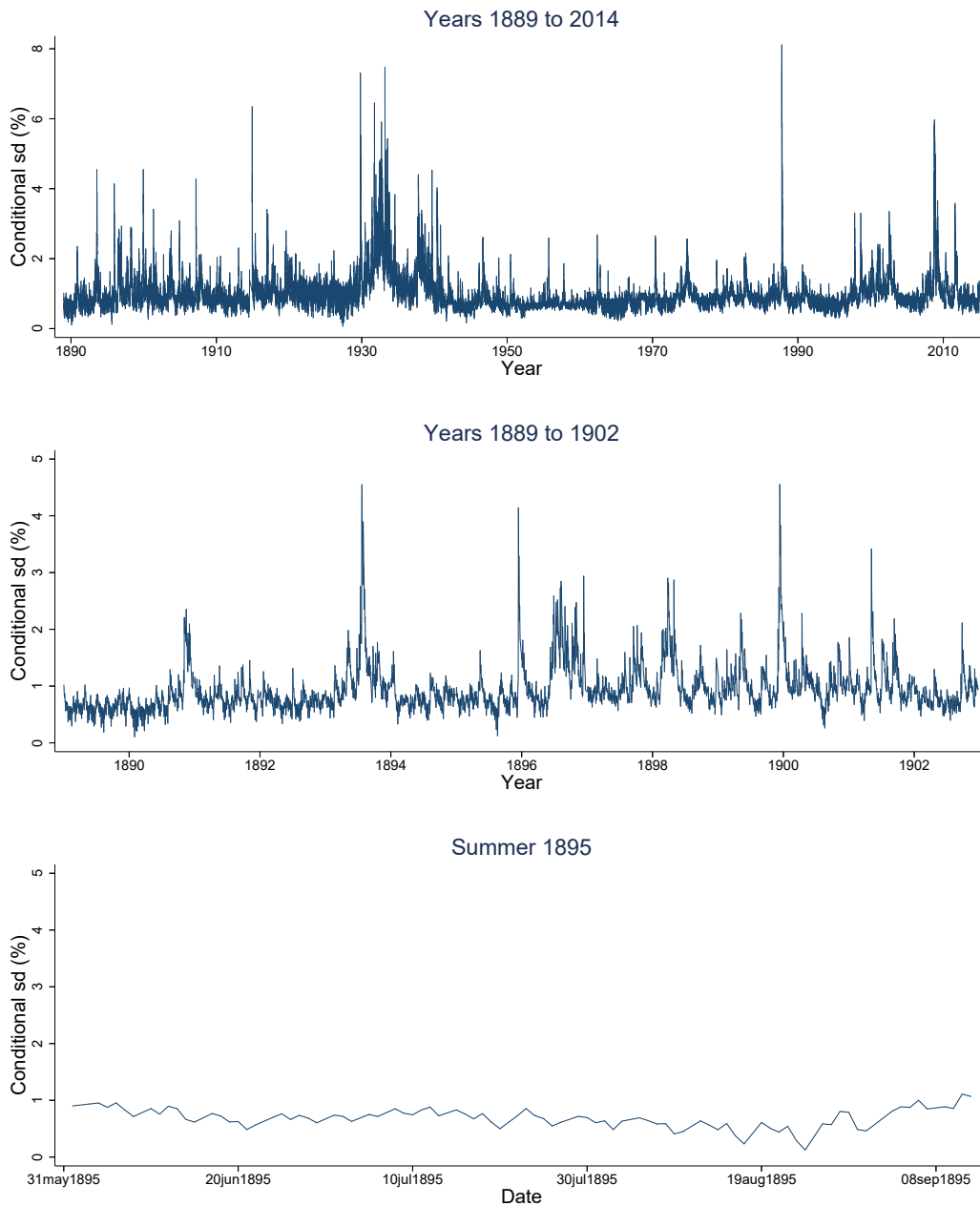
$$r_{m,t} = a_0 + a_1 r_{m,t-1} + a_2 r_{m,t-2} + \varepsilon_t \quad (59)$$

where  $\varepsilon_t \sim N(0, \sigma_t^2)$  and

$$\sigma_t^2 = \gamma_0 + \sum_{i=1}^4 \gamma_i \sigma_{t-i}^2 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 \quad (60)$$

The time series of daily return volatility for the entire sample, the sample of 1889-1902, and the summer of 1895 (representing the middle of the test sample) are plotted in Figure 12. For weekly return volatility, a GARCH model with one ARCH term and one GARCH term is selected based on BIC.

Merging all data gives individual stock and index returns, detrended trade volume, temperature, and other controls over June 11st to September 10th of years 1889 to 2014. I divide this into 5 periods (1889-1902, 1903-1920, 1921-1940, 1941-1960, 1961-2014) and use the earliest period as the test sample. As discussed at the end of section 9, however, the main results are robust to extending the test sample period to include all years from 1889 to 1920. The results disappear in later periods, when temperature has little effect on investor trades. Table 9 presents descriptive statistics by sample period, which I occasionally refer to in the rest of the paper.



**Figure 12: Conditional Volatility of Market Returns**

The figure plots, for three different time periods, daily conditional volatility of market returns. Conditional volatility is the conditional standard deviation obtained from a GARCH model with 4 autoregressive and 2 moving average terms. Market returns are proxied by DJIA for years 1889 to 1925 and by S&P 500 for years 1889 to 2014. The values have been multiplied by 100 to be in % unit.

**Table 9: Descriptive Statistics by Sample Period**

Summer is defined as June 11-September 10, the hottest 3 months of the year (see Figure 11). NYSE index return is measured by the DJIA return during 1889-1925 and S&P 500 return during 1926-2014. When calculating daily autocorrelations, returns are winsorized at the 5% level within each group (e.g. sample and season) by replacing the top and bottom 5% values by the cutoff values.

	(1) 1889-1902	(2) 1903-1920	(3) 1921-1940	(4) 1941-1960	(4) 1961-2014
<b>Daily Manhattan Temperature</b>					
<i>Mean (sd) in degrees Celsius</i>					
Summer	25.2 (3.5)	24.9 (3.6)	25.4 (3.5)	26.1 (3.5)	26.1 (3.5)
All	13.5 (10.1)	13.6 (10.1)	14.1 (10)	14.7 (10)	14.9 (9.9)
<b>Daily Detrended NYSE Trade Volume</b>					
<i>Mean (sd) of log deviation</i>					
Summer	-0.16 (0.57)	-0.17 (0.59)	-0.16 (0.58)	-0.07 (0.36)	-0.01 (0.24)
All	0.03 (0.56)	0.01 (0.58)	0 (0.51)	0.04 (0.38)	0.05 (0.24)
<b>Daily NYSE Index Returns</b>					
<i>Mean (sd) in %</i>					
Summer	0.05 (1.02)	0.01 (0.95)	0.13 (1.51)	0.04 (0.78)	0.02 (0.94)
All	0.02 (1.02)	0.01 (1.06)	0.03 (1.58)	0.05 (0.76)	0.04 (1.01)
<i>Mean (sd) conditional volatility in %</i>					
Summer	0.92 (0.49)	0.93 (0.31)	1.35 (0.69)	0.73 (0.19)	0.92 (0.33)
All	0.94 (0.41)	0.98 (0.35)	1.38 (0.7)	0.74 (0.2)	0.95 (0.43)
<i>Autocorrelation coefficient</i>					
Summer	0.078	0.041	0.014	0.16	0.087
All	0.03	0.05	0	0.167	0.068
<i>Probability of next-day reversal</i>					
Summer	0.47	0.49	0.47	0.43	0.47
All	0.49	0.48	0.48	0.43	0.47

## 8. Temperature effect on aggregate liquidity demand

In section 7, I presented anecdotal evidence that hot weather discouraged demand for liquidity in the sample period. Here, I quantify the size of this effect using aggregate trade volume and the fraction of idle stocks with no buy or sell order.

### 8.1. Hot summer reduces trading activity in the summer: An inter-year regression

Before looking at the temperature effect on trade volume at a daily frequency, I first show that the effect exists even at an annual frequency. I measure summer trading activity using total trade volume over the summer as a percentage of total trade volume over the year (the volumes are before detrending). Then, to test whether summer trading was lower in hotter years, I regress this quantity on the average Manhattan summer temperature in the same year:

$$\frac{\text{Summer volume}_y}{\text{Total volume}_y} = b_0 + b_1 \text{Summer temperature}_y + \varepsilon_y \quad (61)$$

where  $y$  indexes a year. For comparison, I repeat this regression in three other periods.

Table 10 presents the results. Although the small sample size generates a large standard error, column 1 suggests that summer trading was lower in years with hotter summers; for each increase in average summer temperature by 1 degree Celsius, summer trade volume as a fraction of yearly total falls by 2.7%p. The effect can be observed in a graph. Figure 13 visualizes the regression in column 1 and shows an arguably clear negative relationship between temperature and summer trading. Since the temperature effect on trading disappears only gradually over time, I do observe a negative relationship—although not significant at the 10% level—between temperature and volume in the 1903-1920 sample. In the last

**Table 10: High Temperature Reduces Summer Trade Volume (Inter-year)**

$$\frac{\text{Summer volume}_y}{\text{Total volume}_y} = b_0 + b_1 \text{Summer temperature}_y + \varepsilon_y$$

This table reports the inter-year regressions relating NYSE trade volume in the summer to average Manhattan temperature in the summer. Dependent variable is total trade volume over Jun 11-Sep 10 as a percentage of total trade volume in the year. Explanatory variable is the average trading hour temperature in Manhattan over Jun 11-Sep 10. Standard errors corrected for heteroskedasticity are reported in parentheses. \*, \*\*, and \*\*\* indicate significance at 10, 5, and 1 percent, respectively.

	(1)	(2)	(3)	(4)	(5)
	1889-1902	1903-1920	1921-1940	1941-1960	1961-2014
<i>T</i>	-0.027* (0.015)	-0.011 (0.008)	0.009 (0.013)	0.006 (0.011)	-0.000 (0.002)
Observations	14	18	20	20	54
<i>R</i> <sup>2</sup>	0.14	0.06	0.01	0.03	0.00

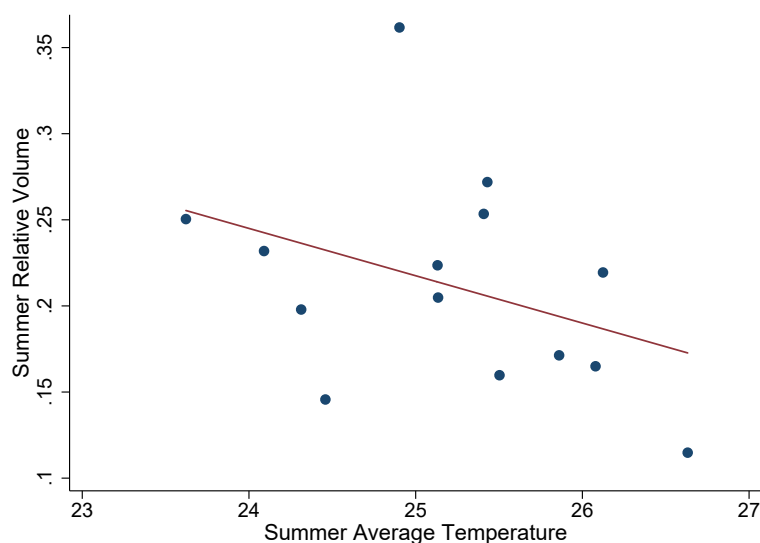
An intercept is included but not reported.

three samples, the low *R*<sup>2</sup>'s imply that temperature has little ability to explain the variation in summer trading.

## 8.2. High temperature reduces trade volume: An inter-day regression

Temperature and trade volume are more strongly related at a daily frequency. To estimate the daily temperature effect on volume, I regress the trading volume variable (*Tradvol*<sub>*t*</sub>) on Manhattan temperature (*T*<sub>*t*</sub>), seasonal temperature ( $\bar{T}_t$ ), volatility ( $\sigma_t$ ), and other controls. The inter-year analysis above shows that the inter-year variation in summer trading can be attributed in part to temperature variation across different years. However, when analyzing the daily temperature effect of volume, I take the conservative approach of exploiting only the within-year variation in trading volume by including year fixed effects. This implies the following regression specification:

$$\text{Tradvol}_t = b_0 + b_1 T_t + b_2 \bar{T}_t + b_3 \sigma_t + \text{other controls} + \text{year fixed effect} + \varepsilon_t \quad (62)$$



**Figure 13: Summer Temperature and Summer Trade Volume (1889-1902)**

This figure plots the average summer temperature in Manhattan on the horizontal axis and total trade volume in the summer relative to total trade volume over a year in the vertical axis. The summer period used in the paper (Jun 11-Sep 10) is highlighted in dark gray. The trading hour average temperature  $T$  is estimated from the maximum and minimum temperatures assuming a linear intraday variation in temperature with the maximum at 15:00 and minimum at 5:00 (Lonnqvist 1962).

Table 11 presents the regression results. The effect of temperature ( $T$ ) is economically and statistically significant in the test sample, with a rise in temperature by 1 degree Celsius (1.8 degrees Fahrenheit) reducing aggregate NYSE trade volume by 1.2%.<sup>66</sup> When year fixed effect is excluded, the reduction in trade volume is larger (2.0%), consistent with the previous finding that year-to-year variation in summer temperature contributes to year-to-year variation in total trading activity in the summer. This temperature effect on trade volume gradually disappears over time. In the 1903-1920 sample, the temperature has a negative but statistically insignificant effect on trade volume. In the latter periods, the magnitudes of the estimates also get smaller.

What do other parameter estimates imply? Seasonal temperature has a larger

---

<sup>66</sup>Although not reported here, I also find that trading volume falls on snowy days in the first sample period but that low winter temperature does not affect trading volume.



**Table 11: High Temperature Reduces Daily Trade Volume**

$$Tradvol_t = b_0 + b_1 T_t + b_2 \bar{T}_t + b_3 \sigma_t + \text{other controls} + \text{year fixed effect} + \epsilon_t$$

OLS regression in the summer 1889-1902 sample (columns 1 to 4) and in the later samples (columns 5 to 8). Dependent variable is daily detrended NYSE trading volume, computed as the deviation of log total trading volume per hour from its 1-year moving average.  $T_t$  and  $\bar{T}_t$  are temperature and seasonal temperature in Manhattan on day  $t$ , respectively. Newey-West HAC standard errors with 20 lags are reported in parentheses. \*, \*\*, and \*\*\* indicate significance at 10, 5, and 1 percent, respectively.

	1889-1902				1903-20	1921-40	1941-60	1961-2014
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$T$	-0.012** (0.006)	-0.012** (0.006)	-0.020*** (0.007)	-0.013** (0.006)	-0.009* (0.005)	-0.001 (0.004)	0.002 (0.003)	-0.000 (0.001)
$\bar{T}$	-0.073*** (0.015)	-0.071*** (0.017)	-0.055* (0.033)	-0.075*** (0.014)	0.003 (0.018)	-0.012 (0.016)	0.011 (0.010)	0.008** (0.003)
Volatility	0.404*** (0.086)		0.163** (0.080)	0.411*** (0.087)	0.436*** (0.117)	0.370*** (0.083)	0.208 (0.162)	0.141*** (0.029)
Year FE	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
N	1,056	1,056	1,056	1,056	1,321	1,518	1,324	3,424

Controls are a constant and dummy variables for Friday, Saturday, and the number of days until next trading day.

negative effect on trade volume than temperature itself, which may be due to a combination of a seasonal effect (e.g. Hong and Yu 2009) and a temperature effect. As expected, volatility is positively related to trade volume in all sample periods.<sup>67</sup>

### A horse race against temperature in other cities

In section 7, I argued using qualitative evidence that Manhattan temperature is a good proxy for temperature felt by a large fraction of NYSE investors. If this claim were correct, I would expect the negative effect of Manhattan temperature on NYSE volume to remain unaffected when temperature in another city is added to

<sup>67</sup>Although not reported in the table, day of the week dummies are significant in some periods. For instance, volume per trading hour is lower on Friday in the 1889-1902 sample but higher on Friday and Saturday in later samples, when trading technology has improved. Among the dummy variables for trading holidays, the dummy variable for trading holidays of 3 days or more is consistently negative, indicating that trade volume falls prior to having a large number of trading holidays.

the regression. Therefore, I add temperatures in other cities to the baseline volume regression, effectively running a horse race between Manhattan temperature and temperature in other cities.

Table 12 presents the results. Adding temperature and seasonal temperature of another city has little effect on the coefficient on Manhattan temperature. For example, temperature in Cambridge, despite its strong correlation with temperature in Manhattan, is not found to reduce NYSE temperature. Interestingly, temperature in San Diego is positively related to NYSE trade volume, which may reflect that temperature in the West Coast was informative about the productivity of agricultural and railroad companies.<sup>68</sup>

---

<sup>68</sup>If San Diego weather indeed carried news about a group of stocks, one may be able to use it to proxy for an exogenous news event.

**Table 12: Temperatures in Other Cities Do Not Reduce NYSE Trade Volume**  
A Horse Race Between Temperature in Manhattan and Temperatures in Other Cities

$$Tradvol_t = b_0 + b_1 T_t^{NY} + b_2 \bar{T}_t^{NY} + b_3 T_t^{Other} + b_4 \bar{T}_t^{Other} + b_5 \sigma_t + \text{other controls} + \text{year fixed effect} + \epsilon_t$$

OLS regression in the summer 1889-1902 sample. Dependent variable is daily detrended NYSE total trading volume, computed as the deviation of log total trading volume per hour from its 1-year moving average. "NY  $T_t$ " and "NY  $\bar{T}_t$ " are temperature and seasonal temperature in Manhattan on day  $t$ , respectively. "Other  $T_t$ " and "Other  $\bar{T}_t$ " are temperature and seasonal temperature of the specified city on day  $t$ , respectively. Newey-West HAC standard errors with 20 lags are reported in parentheses. \*, \*\*, and \*\*\* indicate significance at 10, 5, and 1 percent, respectively.

	(1) Cambridge MA	(2) Jacksonville FL	(3) San Diego CA	(4) Oxford UK	(5) Sydney AU
NY $T$	-0.016* (0.008)	-0.010* (0.006)	-0.012** (0.006)	-0.012** (0.006)	-0.011* (0.006)
NY $\bar{T}$	-0.048** (0.022)	-0.043*** (0.014)	-0.072*** (0.015)	-0.066*** (0.013)	-0.049*** (0.017)
Other $T$	0.004 (0.007)	-0.005 (0.011)	0.026** (0.012)	0.003 (0.008)	-0.002 (0.008)
Other $\bar{T}$	-0.036** (0.018)	-0.079*** (0.027)	0.002 (0.029)	-0.035 (0.022)	0.048** (0.023)
Volatility	0.428*** (0.095)	0.407*** (0.082)	0.386*** (0.086)	0.407*** (0.084)	0.401*** (0.084)
$\rho(\tilde{T}, NY \tilde{T})$	.75	.08	-.07	.01	.02
$\rho(\bar{T}, NY \bar{T})$	.85	.54	.2	.48	-.06
Year FE	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Observations	905	1,056	1,056	1,056	1,056

Controls are a constant and dummy variables for Friday, Saturday, and the number of days until next trading day.

### 8.3. Temperature and the fraction of idle stocks

I use the fraction of idle stocks on a given day as an alternative measure of aggregate liquidity and study the effect of temperature on this measure. For 1889-1902, I use the information collected from the Commercial and Financial Chronicle. I do not have the daily information on idle stocks during the period 1903-1925, but for the period 1926-2014, I collect this information from the daily CRSP data.

The definition of an idle stock is as follows. For the 1889-1902 collected data, a stock is defined as idle if either a buy or sell transaction did not occur for the stock on that day. For the 1926-2014 CRSP data, a stock is defined as idle if its closing price is a bid/ask average (price is marked with a negative symbol) or is unavailable (price is marked zero). When using the CRSP data, I restrict my attention to common stocks (share code 10 and 11) traded on the NYSE (exchange code 1).

The estimated effect of temperature on the fraction of idle stocks, controlling a long-term trend in the fraction through year fixed effects, is presented in Table 13. The contrast between the sample period and the other periods is stark. During 1889-1902, a rise in temperature by 1 degree Celsius increases the fraction of idle stocks by 0.2%p, consistent with the notion that fewer outside customers show up to trade on hotter days. In contrast, temperature has no significant effect on the fraction of idle stocks in the later periods.

The negative coefficients on volatility imply that the fraction of idle stocks tends to fall in times of high volatility. Although high volatility may reduce the supply of liquidity, there is no reason to expect more idle stocks when liquidity supply has deteriorated. Rather, a heightened volatility seems to make investors more attentive to even the smaller stocks which may be less likely to be traded in normal times. The volatility effect falls in magnitude over time as the improvement in trading technology allows even the smallest stocks to be traded relatively easily, reducing the average fraction of idle stocks on any given day.

**Table 13: High Temperature Increases the Fraction of Idle Stocks**

$$\frac{\text{Number of idle stocks}_t}{\text{Total number of stocks}_t} = b_0 + b_1 T_t + b_2 \bar{T}_t + b_3 \sigma_t + \text{other controls} + \text{year fixed effect} + \epsilon_t$$

OLS regression in the summer 1889-1902 sample (column 1) and in the later samples (columns 2 to 4). Dependent variable is the percentage of idle common stocks in the daily market at the NYSE. Coefficients are multiplied by 100 to be in % unit.  $T_t$  and  $\bar{T}_t$  are temperature and seasonal temperature in Manhattan on day  $t$ , respectively. Newey-West HAC standard errors with 20 lags are reported in parentheses. \*, \*\*, and \*\*\* indicate significance at 10, 5, and 1 percent, respectively.

	(1) 1889-1902	(2) 1926-1940	(3) 1941-1960	(4) 1961-2014
$T$	0.215** (0.106)	-0.039 (0.080)	-0.014 (0.048)	0.003 (0.004)
$\bar{T}$	1.219*** (0.285)	-0.111 (0.282)	-0.080 (0.149)	0.013 (0.010)
Volatility	-6.566*** (1.158)	-6.494*** (1.341)	-2.174 (1.633)	-0.293*** (0.113)
Constant	-4.806 (7.802)	52.569*** (8.338)	10.080** (4.026)	0.168 (0.245)
Year FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	1,018	1,138	1,324	3,424

Controls are dummy variables for Friday, Saturday, and the number of days until next trading day.

## 9. Aggregate liquidity demand and market return reversals

Previous sections provided anecdotal and quantitative evidence that temperature variation in the sample period generates exogenous variation in liquidity demand. This section uses this aggregate liquidity demand shock to test the effect of aggregate liquidity demand on market return reversals.

## 9.1. Daily market return reversal

The primary measure of market return reversal is the daily autocorrelation of market returns studied in CGW. As I show below, however, this measure is fairly sensitive to the treatment of outliers. Therefore, I study two additional measures of market return reversal: return on a simple reversal strategy on the market index and the probability of a next-day reversal in the market index.

### The effect of liquidity demand on the autocorrelation of market returns

I study daily autocorrelation of market index returns by allowing temperature and other controls to affect the daily autocorrelation coefficient:

$$r_{m,t+1} = a_0 + (b_0 + b_1 T_t + b_2 \bar{T}_t + b_3 \sigma_t + \text{other controls} + \text{year fixed effect}) r_{m,t} + \varepsilon_{t+1} \quad (63)$$

This specification with daily returns on both sides of the equation is exposed to influence of outliers. To curtail the influence of outliers, I winsorize returns at the 5% and 95% levels, which correspond to cutoff values of  $-1.38\%$  and  $1.47\%$ , respectively, during the sample period.<sup>69</sup>

The outcome of the regression is reported in Table 14. A fall in liquidity demand associated with a 1 degree Celsius rise in temperature increases the daily autocorrelation coefficient by 0.022. To put this in context, this means that a fall in liquidity demand equivalent to a 1% fall in aggregate trade volume (Table 11) increases the autocorrelation of daily market returns by 0.018.<sup>70</sup> Also, since temperature has a standard deviation of 3.5 during the sample period, it follows that a one standard deviation fall in liquidity demand increases the return autocorrelation by 0.077 (see Table 9 for the standard deviation). The coefficients on  $\bar{T}$  tell us

---

<sup>69</sup>See, e.g., Pinegar 2002 for the importance of taming the outliers.

<sup>70</sup>Given by  $0.022/1.2 = 0.018$ , where 1.2% is the fall in aggregate trade volume as a result of a 1 degree Celsius rise in temperature (Table 11).

**Table 14: Liquidity Demand Lowers Autocorrelation in Market Returns**

$$r_{m,t+1} = a_0 + (b_0 + b_1 T_t + b_2 \bar{T}_t + b_3 \sigma_t + \text{other controls} + \text{year fixed effect}) r_{m,t} + \varepsilon_{t+1}$$

OLS regression in the summer 1889-1902 sample. Dependent variable is daily NYSE index return measured by the DJIA return.  $T_t$  and  $\bar{T}_t$  are temperature and seasonal temperature in Manhattan on day  $t$ , respectively.  $T_t$  serves as an exogenous proxy for a *reduction* in aggregate liquidity demand. Winsorization replaces values outside the chosen percentile cutoffs with values at the cutoffs. Newey-West HAC standard errors with 20 lags are reported in parentheses. \*, \*\*, and \*\*\* indicate significance at 10, 5, and 1 percent, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$T$	2.2** (0.9)	2.0** (0.9)	2.5*** (0.9)	2.1** (0.9)		1.9** (0.9)	1.7* (0.9)	1.2 (1.3)	-1.1 (2.3)
$\bar{T}$	-0.6 (2.2)	-1.9 (2.4)	-0.8 (2.1)	-0.9 (2.1)		-0.8 (2.1)	-3.4 (2.2)	2.2 (2.8)	2.4 (2.9)
Volatility	-10.8 (8.2)	-5.2 (9.0)	-11.8 (7.5)		-5.2 (8.8)	-1.2 (4.6)	-10.1 (8.4)	-19.7** (8.3)	-18.5*** (6.8)
Tradvol		-22.4*** (8.4)			-22.6*** (8.2)				
$\tilde{T}^2$			-0.4 (0.2)						
Year FE	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
Winsorize	5%	5%	5%	5%	5%	5%	5%	1%	No
N					1,056				

Excludes controls. Coefficients expressed in 1/100.

that seasonal variation in temperature has little influence on market return reversals. The negative coefficient on the squared deseasonalized temperature implies that the marginal effect of temperature on autocorrelation falls as temperature rises; this may be due to a decreasing marginal effect in the temperature-liquidity relation or in the liquidity-autocorrelation relation.

Comparing columns 1, 8, and 9 of the table shows that the effects of temperature and volatility are both sensitive to the treatment of outliers. With winsorization at 5%, liquidity demand (temperature) significantly affects the return autocorrelation, but volatility does not. In contrast, the opposite happens with winsorization at 1% or with no winsorization; the estimated effect of liquidity demand becomes insignificant, but the effect of volatility becomes significant. This implies that the

volatility effect is concentrated in a small number of large return events. In terms of magnitude, a one standard deviation increase in volatility (0.49; see Table 9) leads to a fall in autocorrelation coefficient by 0.091.

Winsorization, however, has little influence on the effect of trade volume. Even with the 5% winsorization, trade volume retains negative relation to autocorrelation. The coefficient of  $-0.226$  is similar in magnitude to that in CGW ( $-0.212$  to  $-0.427$ , depending on sample period), implying that the estimated effect of trade volume is similar to that in CGW (sample period 1962 to 1988) is part of the evidence that the market environment in the past was similar to what it is now. The magnitude of the estimated effect implies that a one standard deviation rise in trade volume (0.64; see Table 9) leads to a fall in autocorrelation coefficient by 0.143.

The results indicate that day-to-day variation in liquidity demand shapes the reversal behavior of market index returns. Although the link between aggregate liquidity and market return reversal has been established in previous studies, the exogeneity of temperature, and its effect on liquidity demand, are strong evidence of causality. When demand for liquidity falls, as it does on hotter days in the sample period, market returns exhibit a higher autocorrelation and thus a decreased likelihood of a reversal.

The estimated effect of liquidity demand on autocorrelation using temperature is approximately 8 times the magnitude estimated using trade volume, implying a large attenuation bias when trade volume is used to proxy for aggregate liquidity demand. The coefficient on temperature implies that a 1 degree Celsius rise in temperature that leads to a decrease in total trade volume by about 1% causes the autocorrelation to increase by 0.018. On the other hand, when volume falls by 1% for a unspecified reason, return autocorrelation is expected to increase by just 0.0023 ( $0.226 \times 0.01$ ), about 1/8 of the effect estimated using temperature. A corollary is that only about 1/8 of the variation in daily trade volume can be



attributed to changes in liquidity demand. Attenuation bias is therefore large, and the magnitude of the coefficient on trade volume does not allow for an easy economic interpretation.

This autocorrelation study also suggests that the variation in volatility generates a meaningful variation in liquidity supply only during substantially volatile times with large absolute values of returns. Although one could argue that the estimated effect without winsorization is spurious, models of financial intermediation with a VaR constraint (e.g. Brunnermeier and Pedersen 2009; Adrian and Shin 2010) provides an alternative interpretation. A small swing in the market leads to only a small increase in the conditional volatility, in which case the market maker's balance sheet would not reach the VaR constraint. A series of large movements in the market, however, can lead to a substantial increase in the conditional volatility to make the market maker's VaR constraint bind, limiting the market maker's ability to absorb liquidity demand.

What one cannot conclude from this study is whether the disappearance of the temperature effect under no winsorization implies that liquidity demand is unimportant in volatile times or that temperature variation does not lead to liquidity demand variation in such times. Even if liquidity demand affects autocorrelation in all times, if large swings in market returns alert even the outside customers and induce them to trade despite hot weather, the temperature-autocorrelation relation would disappear in such times. This point is similar to that made earlier about the quadratic effect of temperature on market return autocorrelation.

The estimated impact of liquidity demand on market return autocorrelation is strong but hinges on winsorization of returns. Thus, I consider two alternative measures of short-horizon reversal in market returns—return on a simple reversal strategy on the market and the likelihood of reversal in the market return—which I analyze next.

## The effect of liquidity demand on a simple reversal strategy return on the market

A return autocorrelation regression can be understood as using past return information to forecast future returns. However, one may try predicting future returns without past return information, relying solely on the hypothesis that market return is more likely to reverse following a heightened demand for liquidity.

I implement this idea by studying returns on the simple strategy that goes long 1 dollar of the market index if the market goes down today and short 1 dollar of the market index if the market goes up today:

$$r_{t+1}^{rev} \equiv r_{m,t+1} \mathbf{1}(r_{m,t} < 0) \quad (64)$$

If liquidity demand indeed helps predict reversal on the next trading day, temperature today should negatively predict expected return on this strategy. Thus, I regress return on this strategy tomorrow on today's temperature and other controls,

$$r_{t+1}^{rev} = b_0 + b_1 T_t + b_2 \bar{T}_t + b_3 \sigma_t + \text{other controls} + \text{year fixed effect} + \epsilon_{t+1} \quad (65)$$

and test if  $b_1 = 0$  as a usual. I do this without winsorizing the return data because return now only appears in the left-hand side of the equation, mitigating the outlier effect that occurred in the previous specification with return on both sides of the equation.

Table 15 reports the results. Even if the information about today's return is omitted, variation in liquidity demand associated with temperature change helps predict reversal of market return on the next trading day. In particular, a 1 degree Celsius rise in temperature reduces the expected return from betting against today's market index movement by 2.2 basis points, or 2.2% of the standard devi-

**Table 15: Liquidity Demand Increases Return on A Simple Reversal Strategy on the Market**

$$r_{t+1}^{rev} = b_0 + b_1 T_t + b_2 \bar{T}_t + b_3 \sigma_t + \text{other controls} + \text{year fixed effect} + \epsilon_{t+1}$$

OLS regression in the summer 1889-1902 sample. Dependent variable is  $r_{t+1}^{rev} = r_{m,t+1} \times \mathbf{1}(r_{m,t} < 0)$ , where  $r_{m,t}$  is the NYSE index return on day  $t$  as measured by the DJIA return. Coefficients are multiplied by 100 to be in % unit.  $T_t$  and  $\bar{T}_t$  are temperature and seasonal temperature in Manhattan on day  $t$ , respectively.  $T_t$  serves as an exogenous proxy for a *reduction* in aggregate liquidity demand. No winsorization is done on the data. Newey-West HAC standard errors with 20 lags are reported in parentheses. \*, \*\*, and \*\*\* indicate significance at 10, 5, and 1 percent, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$T$	-0.022*** (0.009)	-0.021** (0.009)	-0.026*** (0.009)	-0.022*** (0.008)		-0.021** (0.008)	-0.019** (0.008)
$\bar{T}$	0.007 (0.022)	0.016 (0.023)	0.010 (0.021)	0.007 (0.021)		0.008 (0.022)	0.026 (0.021)
Volatility	0.199 (0.156)	0.149 (0.164)	0.212 (0.151)		0.144 (0.163)	-0.157* (0.086)	0.207 (0.152)
Tradvol		0.124* (0.071)			0.133** (0.068)		
$\tilde{T}^2$			0.004** (0.002)				
Year FE	Yes	Yes	Yes	Yes	Yes	No	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	No
Observations	1,056	1,056	1,056	1,056	1,056	1,056	1,056

Controls are a constant and dummy variables for Friday, Saturday, and the number of days until next trading day.

ation.<sup>71</sup> The effects of volatility and trade volume, however, are not strong enough to overcome the increased noise in the prediction exercise.

### Liquidity demand and the likelihood of a next-day market return reversal

The last measure of daily reversal is the likelihood of a next-day return reversal in the market index. Despite the information loss associated with discretizing the market return behavior into a binary event of reversal and non-reversal, this approach has the advantage of being less subject to the influence of outlier returns. For this, I define  $reversal_{t+1} \equiv \mathbf{1}[r_{m,t} r_{m,t+1} < 0]$  as the event that the market return

<sup>71</sup>The standard deviation of this simple reversal strategy is 1.02% during the sample period.

$r_{m,t}$  has a reversal on the next trading day. I then use the logit specification to study how liquidity demand influences the probability of a reversal:

$$\ln\left(\frac{\Pr[\text{reversal}_{t+1} = 1]}{\Pr[\text{reversal}_{t+1} = 0]}\right) = b_0 + b_1 T_t + b_2 \bar{T}_t + b_3 \sigma_t + \text{other controls} + \text{year fixed effect} + \epsilon_{t+1} \quad (66)$$

Table 16 reports the results in terms of marginal effects from the mean values of the covariates. The fall in demand associated with a 1 degree Celsius rise in temperature reduces the probability of market return reversal by 1.5%p. On the other hand, neither volatility nor trade volume significantly affects the likelihood of a reversal. This again suggests that the effects of volatility and trade volume on return autocorrelation are concentrated in a small number of large return events, consistent with the effect of volatility on return autocorrelation disappearing with the 5% winsorization. Seasonal temperature shows that a reversal becomes more likely during seasonably hotter days in the summer, suggesting that the supply of liquidity may be low on these days for seasonal reasons. Finally, including the squared deseasonalized temperature term has little effect on estimated coefficients.

## 9.2. Weekly market return reversal

Thus far I have focused on daily market returns. The results, however, extend to the autocorrelation of weekly market returns.

To conduct weekly regressions, I redefine variables to have a weekly horizon. Market return is still the DJIA return but is measured over a week. Weekly temperature, seasonal temperature, and deseasonalized temperature are defined as their average daily values over a week. Volatility, as explained in section 7, is the conditional volatility of weekly market return. Trade volume is defined as the total NYSE trade volume over a week, stochastically detrended using the prior 52 weeks' moving average. Day of the week dummies and trading days' gap dum-

**Table 16: Liquidity Demand Increases Likelihood of Market Return Reversal**

$$\ln \left( \frac{Pr[reversal_{t+1} = 1]}{Pr[reversal_{t+1} = 0]} \right) = b_0 + b_1 T_t + b_2 \bar{T}_t + b_3 \sigma_t + \text{other controls} + \text{year fixed effect} + \epsilon_{t+1}$$

Logistic regression in the summer 1889-1902 sample. Dependent variable is the binary variable indicating the event that the NYSE index return switches the sign on the next trading day (1 = reversal, 0 = no reversal). NYSE index return is measured by the DJIA return.  $T_t$  and  $\bar{T}_t$  are temperature and seasonal temperature in Manhattan on day  $t$ , respectively.  $T_t$  serves as an exogenous proxy for a *reduction* in aggregate liquidity demand. The table reports the marginal effects from the mean values of the covariates. Standard errors corrected for heteroskedasticity are reported in parentheses. \*, \*\*, and \*\*\* indicate significance at 10, 5, and 1 percent, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$T$	-0.015*** (0.005)	-0.015*** (0.005)	-0.016*** (0.005)	-0.015*** (0.005)		-0.014*** (0.005)	-0.014*** (0.005)
$\bar{T}$	0.021* (0.012)	0.022* (0.012)	0.022* (0.012)	0.022* (0.012)		0.016 (0.011)	0.030*** (0.011)
Volatility	0.084* (0.051)	0.078 (0.053)	0.087* (0.051)		0.078 (0.052)	-0.005 (0.032)	0.087* (0.050)
Tradvol		0.013 (0.039)			0.011 (0.037)		
$\tilde{T}^2$			0.001 (0.001)				
Year FE	Yes	Yes	Yes	Yes	Yes	No	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	No
Observations	1,056	1,056	1,056	1,056	1,056	1,056	1,056

Controls are a constant and dummy variables for Friday, Saturday, and the number of days until next trading day.

mies are no longer necessary.

Using the sample period (summers of 1889-1902), I estimate the weekly market return autocorrelation via specification (63) and report the results in Table 17. I find that a rise in weekly temperature by 1 degree Celsius increases the weekly return autocorrelation by 0.17. This effect is larger than that for daily returns. Since a rise in weekly temperature by 1 degree Celsius decreases the weekly aggregate trade volume by 3.7% (not reported in tables), it follows that a fall in liquidity demand generating a 1% drop in aggregate trade volume lowers the weekly return autocorrelation by 0.046 (compared to 0.018 for daily returns). As weekly returns

**Table 17: Liquidity Demand Lowers Autocorrelation in Weekly Market Returns**

$$r_{m,t+1} = a_0 + (b_0 + b_1 T_t + b_2 \bar{T}_t + b_3 \sigma_t + \text{other controls} + \text{year fixed effect}) r_{m,t} + \varepsilon_{t+1}$$

OLS regression in the summer 1889-1902 sample. Dependent variable is weekly NYSE index return measured by the DJIA return.  $T_t$  and  $\bar{T}_t$  are average daily temperature and seasonal temperature in Manhattan in week  $t$ , respectively.  $T_t$  serves as an exogenous proxy for a *reduction* in aggregate liquidity demand. Winsorization replaces values outside the chosen percentile cut-offs with values at the cutoffs. Newey-West HAC standard errors with 4 lags are reported in parentheses. \*, \*\*, and \*\*\* indicate significance at 10, 5, and 1 percent, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$T$	17.2** (6.8)	13.1** (6.3)	21.7*** (7.2)	16.3** (8.2)		14.5** (6.4)	17.1** (6.7)	11.1** (5.0)
$\bar{T}$	-18.9** (9.2)	-13.6 (8.4)	-23.5** (10.0)	-12.8 (11.0)		-10.1 (9.3)	-19.0** (9.1)	-11.3 (7.8)
Volatility	-55.4*** (17.7)	-53.8*** (17.3)	-60.3*** (15.7)		-49.7*** (18.4)	-14.0 (10.6)	-55.0*** (17.5)	-52.0*** (14.3)
Tradvol		-78.4*** (23.3)			-94.2*** (27.1)			
$\tilde{T}^2$			-3.4** (1.6)					
Year FE	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
Winsor	No	No	No	No	No	No	1%	5%
N	196	196	196	196	196	196	196	196

An intercept is included but not reported. Coefficients expressed in 1/100.

tend to have less outliers than daily returns, the estimated effect of temperature is statistically significant with or without winsorization.

Both volatility and volume have large effects on weekly return autocorrelation. Conditional volatility of weekly returns has a standard deviation of 0.79 in the sample period, so a one standard deviation rise in volatility leads to a large fall in return autocorrelation of 0.44 (0.79 times  $-0.554$  in column 1). Trading volume has a standard deviation of 0.51 in the sample period, implying that a one standard deviation rise in volume leads to a fall in return autocorrelation by 0.40 (0.51 times  $-0.784$  in column 2). The effects of temperature, volatility, and volume on return autocorrelation are all larger for weekly return than for daily returns, suggesting that a weekly liquidity event may be a more accurate measure of risk that market

makers care about.

### 9.3. Reward from aggregate liquidity provision

Although the impact of aggregate liquidity demand on market return reversal has been explained in the context of liquidity suppliers requiring a premium for betting against the market index, the exact positions held by liquidity suppliers may differ from the composition of the market index. That is, liquidity suppliers are likely to have a large position on stocks with large short-horizon movements, not just on stocks with larger market capitalizations.

For this reason, researchers have looked at a more direct proxy for liquidity suppliers' positions to study the impact of aggregate liquidity on reward from aggregate liquidity provision. (Lehman 1990; Lo and MacKinlay 1990; Nagel 2012). Following this line of work, I use individual stock price data from 1889-1902 to test whether the exogenous variation in liquidity demand affects liquidity suppliers' expected returns. I also look at the fraction of individual stocks having reversal on the next trading day as another measure of reward from liquidity provision.

Lehman (1990) and Nagel (2012) suggest proxying for return on aggregate liquidity provision using

$$L_{t+1} = \sum_{i=1}^N \omega_{i,t} r_{i,t+1} \quad (67)$$

with the time- $t$  position on stock  $i$  given as

$$\omega_{i,t} = -\frac{r_{i,t} - r_{m,t}^{EW}}{\frac{1}{2} \sum_{i=1}^N |r_{i,t} - r_{m,t}^{EW}|} \quad (68)$$

where  $i = 1, \dots, N$  indexes the individual stocks, and  $r_{m,t}^{EW} = N^{-1} \sum_{i=1}^N r_{i,t}$  is the equal-weighted market index return between  $t - 1$  and  $t$ . The numerator of portfolio weight  $\omega$  assumes that market makers have negative position on stocks that outperform the equal-weighted market index and positive position on stocks that underperform the index. The denominator adjusts the portfolio weight based on

the assumption that a dollar of either long or short position requires 50 cents—that the required margin is a constant 1/2 for all stocks and all times (Nagel 2012). The denominator adjustment also implies that the market maker’s asset position in the strategy is constant over time, unaffected by time-varying volatility.<sup>72</sup>

If  $L_{t+1}$  is an accurate proxy for return on liquidity provision, I would expect it to be lower following a hotter day (high  $T_t$ ), when aggregate liquidity demand is lower and market makers thus require lower compensation for providing liquidity; i.e.,  $\partial E_t[L_{t+1}]/\partial T_t < 0$ . To test this, I regress  $L_{t+1}$  on temperature  $T_t$  and other controls:

$$L_{t+1} = b_0 + b_1 T_t + b_2 \bar{T}_t + b_3 \sigma_t + \text{other controls} + \text{year fixed effect} + \epsilon_{t+1} \quad (69)$$

Since daily returns on individual stocks are unavailable in the late 19th to early 20th century, I use the hand-collected data discussed earlier to compute  $\{L_t\}$  (except for June 1894, when data are unavailable).

The regression results are presented in columns 1 through 6 of Table 18 (coefficients are multiplied by 100 to be in % unit). Consistent with the hypothesis, an increase in temperature lowers the expected return from liquidity provision. An increase in Manhattan temperature by 1 degree Celsius lowers the expected return on liquidity provision by 4.7 basis points, or 1.7% of the standard deviation.<sup>73</sup> This temperature effect remains largely unchanged under different specifications except that the effect becomes insignificant when year fixed effects are removed. This suggests that the return on aggregate liquidity provision undergoes large changes over different years for reasons other than temperature (e.g., increased competition in market making).

---

<sup>72</sup>Nagel (2012) discusses two other proxies for return on liquidity provision, and these use different denominator adjustment from the one this proxy uses. Using these other proxies generates similar results.

<sup>73</sup>The standard deviation of this return over the first sample is 2.9%.



**Table 18: Liquidity Demand Increases Reward from Aggregate Liquidity Provision**

OLS regression in the summer 1889-1902 sample. Dependent variable is the market maker's daily return from aggregate liquidity provision in columns 1 to 6 (see section 9.3) and the fraction of stocks having reversal on the next trading day in column 7. Coefficients are multiplied by 100 to be in % unit.  $T_t$  and  $\bar{T}_t$  are temperature and seasonal temperature in Manhattan on day  $t$ , respectively.  $T_t$  serves as an exogenous proxy for a *reduction* in aggregate liquidity demand. Newey-West HAC standard errors with 20 lags are reported in parentheses. \*, \*\*, and \*\*\* indicate significance at 10, 5, and 1 percent, respectively.

	Reward from aggregate liquidity provision						% Reversed
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$T$	-0.047** (0.024)	-0.051** (0.024)	-0.045* (0.024)		-0.031 (0.024)	-0.048** (0.022)	-0.281** (0.143)
$\bar{T}$	0.163* (0.083)	0.143* (0.084)	0.166** (0.083)		0.139* (0.084)	0.194** (0.090)	0.082 (0.301)
Volatility	1.075*** (0.248)	1.190*** (0.232)		1.212*** (0.238)	0.680*** (0.185)	1.057*** (0.261)	3.813** (1.817)
Tradvol		-0.276* (0.145)		-0.333** (0.155)			
Year FE	Yes	Yes	Yes	Yes	No	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	No	Yes
Observations	1,016	1,016	1,016	1,016	1,016	1,016	1,016

Controls are a constant and dummy variables for Friday, Saturday, and the number of days until next trading day.

Expected return on liquidity provision tends to fall as seasonal temperature rises ( $\bar{T}$ ). One way to rationalize this is that the number of liquidity suppliers tends to fall in predictably hotter parts of the summer. Although this would be consistent with the earlier observation that the fall in aggregate trade volume on seasonably hot days is larger than what the temperature effect alone would predict, I refrain from extrapolating too much based on this result.

As noted in Nagel (2012), high volatility increases the expected return from aggregate liquidity provision. The interpretation is that the supply of liquidity falls with volatility because volatility reduces market makers' ability to hold additional inventory on their balance sheets. Finally, controlling for temperature, aggregate

trade volume negatively predicts return on liquidity provision.<sup>74</sup> This would be true if an unusual trade volume were a proxy for private information flow for individual stocks with large movements (Wang 1994), but conflicts with the notion that an unusual trade volume indicates a flow of speculative trades (CGW). One way to make sense of this puzzling result is that, since the return on aggregate liquidity provision here treats each stock equally regardless of its size (equation 68 does not use information about the size of each stock), it overrepresents the private information channel of small stocks.

Column 7 of Table 18 shows the impact of liquidity demand on the fraction of stocks with a next-day reversal. The result is again consistent with a fall in liquidity demand lowering the market maker's expected return: a rise in temperature by 1 degree Celsius reduces the fraction of individual stocks with a next-day reversal by 2.9%p.

#### **9.4. Further discussions and robustness checks**

The results in this section suggest that that a fall in liquidity demand associated with a rise in temperature leads to an increased autocorrelation of short-horizon market returns and thus a lower likelihood of a market return reversal. An unanswered question is, what is the nature of this aggregate liquidity demand captured by the variation in temperature? Are these trades motivated by correlated liquidity needs or by correlated sentiments?

To tease out the liquidity story from the sentiment story, I test for the presence of pent-up liquidity. If investors trade less on hotter days because they are postponing to resolve their liquidity needs, the unresolved need for liquidity would be manifested as a heightened level of trade on the next trading day. On the other hand, if investors wanted to trade for short-lived motives such as sentiments, the

---

<sup>74</sup>Cooper (1999) also finds that return on reversal strategy on NYSE-AMEX stocks declines with trading activity.

unresolved need for liquidity is likely to disappear by the next trading day.

As a simple test, I regress day- $t$  trade volume on the previous trading day's temperature and test if the coefficient is positive; i.e., I test if hot weather on the previous trading day leads to more pent-up liquidity today. I also try regressing day- $t$  trade volume on today's temperature interacted with temperature on the previous trading day; i.e., I test if hot weather on the previous trading day increases the potential demand for liquidity today so that the same increase in temperature leads to a greater reduction in trade volume. I measure the previous day's weather with either the dummy variable for deseasonalized temperature higher than the one standard deviation or simply with the original deseasonalized temperature variable.

The results, reported in Table 19, are *inconsistent* with investors postponing to resolve their liquidity needs. Previous day's hot weather tends to *decrease* the trade volume next day. Also, the contemporaneous impact of temperature on trade volume does not significantly increase if the previous day's temperature is higher. This suggests that the liquidity demand identified in this study through temperature variation is driven by short-lived motives like sentiments.

This paper relies on the identification assumption that Manhattan temperature variation in the summer determined demand for liquidity at the NYSE in the period prior to relevant technological advancement. I then determine the appropriate test sample period to be 1889-1902 based on evidence that this marks the period prior to the introduction of most technologies that eased stock trading on hot summer days. Although the year 1902 is chosen as the last year of the sample period based on the introduction of large-scale air conditioning in 1903, the choice is admittedly arbitrary. Here, I show that the results are not sensitive to how far into the 20th century the test sample period is extended.

This paper's online appendix repeats the baseline return autocorrelation regression for daily and weekly returns over four expanding sample periods: 1889-1905,

**Table 19: No Evidence of Pent-up Liquidity**

OLS regression in the summer 1889-1902 sample. Dependent variable is daily detrended NYSE total trading volume, computed as the deviation of log total trading volume per hour from its 1-year moving average..  $T_t$  and  $\bar{T}_t$  are temperature and seasonal temperature in Manhattan on day  $t$ , respectively.  $T_t$  serves as an exogenous proxy for a *reduction* in aggregate liquidity demand on day  $t$ .  $\tilde{T}_t \equiv T_t - \bar{T}_t$  indicates unseasonal temperature. Dummy variable for hot weather on previous trading day indicates that the value of unseasonal temperature is above one standard deviation:  $Hot_{t-1} \equiv \mathbf{1}(\tilde{T}_{t-1} > \sigma(\tilde{T}))$ . Newey-West HAC standard errors with 20 lags are reported in parentheses. \*, \*\*, and \*\*\* indicate significance at 10, 5, and 1 percent, respectively.

	(1)	(2)	(3)	(4)
$T$	-0.010* (0.005)	-0.014** (0.006)	-0.012** (0.005)	-0.012** (0.005)
$\bar{T}$	-0.075*** (0.015)	-0.073*** (0.015)	-0.073*** (0.015)	-0.073*** (0.015)
Volatility	0.402*** (0.086)	0.406*** (0.086)	0.403*** (0.086)	0.403*** (0.086)
$Hot_{t-1}$	-0.060* (0.035)	-0.097** (0.040)		
$T \times Hot_{t-1}$		0.016 (0.011)		
$\tilde{T}_{t-1}$			-0.000 (0.004)	-0.000 (0.004)
$T \times \tilde{T}_{t-1}$				0.000 (0.001)
Year FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	1,052	1,052	1,052	1,052

Controls are a constant and dummy variables for Friday, Saturday, and the number of days until next trading day.

1889-1910, 1889-1915, and 1889-1920. The temperature effect on market return autocorrelation remains statistically significant as the sample end year increases, although the magnitudes of the coefficients gradually fall as the temperature effect on liquidity demand diminishes over time.

To measure variation in liquidity demand using temperature variation, it is important to control for the seasonal variation in temperature within a summer. This is done using the single seasonal temperature variable, but I demonstrate here that using seasonal fixed effects produces similar results. The online appendix shows

that excluding the seasonal temperature variable and instead introducing increasingly fine seasonal fixed effects—from an interval of 30 days to an interval of 3 days—generates similar results on the return autocorrelation regression. Neither the magnitude nor significance of the estimated effects depends importantly on the choice of seasonal fixed effects.

The reliance on historical data is both a strength and weakness of this study. Focusing on the past years allows for the natural experiment exploited by this study—i.e., aggregate demand for liquidity falls as summer temperature exogenously rises. Nevertheless, the further we go back into time, the less relevant the results may be to today's market. Still, there are reasons to believe that studying the stock market in the late 19th to early 20th century generates insights relevant to today's market. First, the microstructure of NYSE attained the modern form by this time. One critical change occurred further back in 1871, when the trading mechanism changed from call market to continuous market, allowing trades to occur asynchronously. Furthermore, as explained in section 7, the classification of traders into market makers and outside customers as well as the market makers' special role in providing liquidity were true then as they are now. Finally, the estimated effect of trade volume on daily return autocorrelation during the sample period is similar to that in CGW, implying that liquidity demand per unit of aggregate trade volume was similar in this study's (1889-1902) and CGW's (1962-1988) sample periods.

## 10. Conclusion

In this paper, I use a natural experiment to study the impact of aggregate liquidity demand shocks on short-horizon reversals in a stock market index. The natural experiment I exploit is that, prior to the advancement in temperature control, communication, and transportation technologies, hot summer weather discouraged a sizable quantity of outside customers from participating in the stock market. I

find that day-to-day changes in the identified demand for liquidity—likely the outcome of sentiment or other short-lived motives—lead to changes in the short-horizon behavior of market returns.

The suggested natural experiment and the unique records of individual stock returns in 1889-1902 may be used to explore other topics. For example, combining the individual stock price data based on daily high and low prices with the temperature proxy for outside investors, one can test theories of bid-ask spreads or volatilities. This would be possible using the range-based measures of bid-ask spreads (Corwin and Schultz 2012) and volatility (Alizadeh, Brandt, and Diebold 2002) but is left to future work.

# **Chapter III.**

## **The Unlegislated Tax Multiplier**

## 11. Introduction

The tax multiplier study of Romer and Romer (2010) (RR) based on the narrative approach is influential among researchers and policymakers alike.<sup>75</sup> Using presidential and congressional documents to identify “exogenous” tax bills that were enacted irrespective of the prevailing or anticipated economic conditions, they estimate a tax multiplier of three: an exogenous fall in tax by 1 percent of the GDP raises the GDP in the next three years by approximately 3 percent.

Since a tax multiplier of three is larger than that estimated by most other studies (e.g., Blanchard and Perotti, 2002; Barro and Redlick, 2011; and Favero and Giavazzi, 2012), the RR multiplier has come under scrutiny.<sup>76</sup> For instance, Calda and Kamps (2012), Charhour, Schmitt-Grohe and Uribe (2012), Perotti (2012), and Mertens and Ravn (2014) point to different ways in which the RR estimate may be biased. We focus on another source of bias, namely that the period in which exogenous bills are enacted are not truly similar to other times.

Our main contribution is to gauge the reliability of the RR estimate without having to identify the specific sources of a potential bias a priori. The idea is to estimate the RR tax multiplier using *unlegislated* tax bills that barely failed to pass as indicated by having passed at least one chamber of the Congress but subsequently failing to become law (“unlegislated” or “barely failed” from hereon). Since these unlegislated tax bills should not have a causal effect on the GDP, estimating the unlegislated tax multiplier allows us to isolate the changes in the GDP due only to the prevailing or anticipated economic conditions surrounding the passage of an exogenous tax bill if an eventual passage of a bill that has passed at least one chamber hinges on a random course of political events.

---

<sup>75</sup>For instance, the study was cited in the 2014 Economic Report to the President (CEA, 2014).

<sup>76</sup>Mountford and Uhlig (2009), however, also find a relatively large multiplier.



### 11.1. The kind of biases are we concerned about?

We claimed earlier that even the exogenous tax changes identified by RR may occur under current and anticipated economic conditions that are different from other times. Why is this possible?

This happens, for instance, if the popularity of free market policies lead to both tax cuts and industry deregulation, as it did during the Reagan era. The Reagan tax cut in 1981 is one of the largest exogenous tax cuts identified by RR, but the tax cut was complemented by the deregulation of telecommunication, transportation, finance, and agriculture.<sup>77</sup> If both tax cuts and industry deregulation contribute to GDP growth in the short term, attributing the GDP growth solely to the tax cut can bias the tax multiplier estimated from this time period upward.<sup>78</sup>

This bias differs from the one identified in RR and cannot be addressed by their narrative approach alone. In the motivating example above, even if the tax cut were exogenous to the economic circumstance, the presence of free market policies complementing the tax cut would still generate a positive bias caused by the omitted policy variables. One can deal with this bias by explicitly including relevant policy variables, but quantifying the magnitudes of policy changes is a challenge.<sup>79</sup>

Our study nonetheless is meaningful also as a sanity check for the RR classification of tax bills. If following the same classification approach we find normal GDP movements following proposed tax changes that barely failed to become law, it

---

<sup>77</sup>One example is the Federal Communications Commission's decision to end the de facto monopoly rights of Western Electric in telecommunication equipment industry in the 1970's and 1980's. Olley and Pakes (1996) documents the consequent competition in equipment industry increased the productivity of the equipment industry. This time period coincides with the passage of a series of tax reforms aimed at long-term economic growth during the Reagan administration.

<sup>78</sup>A bias in the same direction can also occur based on the expectation that the bill would pass and affect future GDP growth. We thank Robert Barro for this point.

<sup>79</sup>Mulligan and Shleifer (2005), Coffrey et al. (2012), Dawson and Seater (2013), and Omar Al-Ubaydli and Patrick A. McLaughlin (2015) have made progress in quantifying the magnitudes of regulations.

would increase the reliability of the classification approach.

## 11.2. What we find

We find a unlegislated tax multiplier of close to zero, suggesting that tax multiplier estimates based on legislated tax bills suffer from a selection bias. This result is fairly robust to the specification of our test.

## 12. Why are barely failed tax bills useful?

Our methodology is best explained in the potential outcome framework of applied microeconomics. The output growth is the outcome variable  $\Delta Y$ , the tax revenue change is the treatment variable  $\Delta T$  (for expositional purposes, we suppose  $\Delta T$  takes the value of either 1 or 0.), and the passage of the tax revenue change is the selection variable  $P$ . Formally, we can write:

$$\Delta Y_i = \Delta Y_i(\Delta T_i, P_i) + \epsilon_i \quad (70)$$

We are interested in how random tax changes affect the outcome growth. Formally, this is captured by

$$\gamma = \mathbb{E}[\Delta Y_i(\Delta T_i = 1, P_i) - \Delta Y_i(\Delta T_i = 0, P_i)] \quad (71)$$

Romer and Romer (2010) treats both the passage and the tax change as treatment variable and compares the outcome growth when both the treatment and selection happen to the output growth when either treatment or selection does not happen. Formally, this is equivalent to

$$\gamma^{RR} = \mathbb{E}[\Delta Y_i(\Delta T_i = 1, P_i = 1) - \Delta Y_i((\Delta T_i, P_i) \in \{(1,0), (0,1), (0,0)\})] \quad (72)$$

This estimate is generally not  $\gamma$ . This paper attempts to remedy this problem.

To that end, we construct a matching variable  $\tilde{P}_i$  such that it makes  $P$  irrelevant in expectation:  $\mathbb{E}[\Delta Y_i(\Delta T_i = 1, P_i) | \tilde{P}_i] = \mathbb{E}[\Delta Y_i(\Delta T_i = 1) | \tilde{P}_i]$ . Then,

$$\begin{aligned} \gamma^{CJ} &= \mathbb{E}[\Delta Y_i(\Delta T_i = 1, P_i) - \Delta Y_i(\Delta T_i = 0, P_i) | \tilde{P}_i] \\ &= \mathbb{E}[\Delta Y_i(\Delta T_i = 1) - \Delta Y_i(\Delta T_i = 0) | \tilde{P}_i] \end{aligned} \quad (73)$$

Econometrically, we need  $\tilde{P}$  that subsumes all the effects from  $P$  to  $\Delta Y$ . For instance, agents can foresee the passage of tax bills as signals of the broad pro-market stance of the Congress and increase investment. We need a signal that subsumes this effect. We believe the passage of one chamber of the Congress be a good candidate for  $\tilde{P}$ .

### 13. Data on unlegislated tax changes

To do this study, we need to first identify all tax-related bills that pass at least one chamber of the Congress but do not eventually become law. Then, for each of these unlegislated tax bills, we need to find the proposed tax revenue changes.

To identify tax-related bills, we rely on the fact that most tax-related bills seriously considered by the Congress have tax revenue estimates provided by the Joint Committee on Taxation (JCT).<sup>80</sup> Hence, we start with the universe of 3,435 documents available on the JCT website from 1975 to 2017. Although these documents all contain potentially useful information, some of these documents are

---

<sup>80</sup>Exceptions are social security tax bills, which may not be considered by the JCT. For these, RR rely on the Social Security Bulletin and the Annual Report of the Board of Trustees of the Federal Old Age and Survivors Insurance Trust Fund to obtain the bill information as well as revenue estimates. We exclude social security tax bills for now.

tax comparison studies and technical studies of tax rules that contain no information about tax revenue changes. Since it is impractical to go through all 3,435 documents to identify tax-related bills with revenue estimates, we take a short cut.

Our short cut is to read only the documents associated with bills classified as tax bills by Congress.gov, the website for U.S. federal legislative information. We do this based on the assumption that the bills that are not classified as primarily a tax bill by Congress.gov are associated with smaller tax revenue changes. Congress.gov has the universe of over million bills from 1975 to 2017, and it categorizes 25544 bills (of which, 144 are enacted) as tax bills. We then match the 25544 taxation bills to the titles of the JCT documents to obtain 575 JCT documents related to the bills. From the 575 JCT documents, we manually collect the revenue estimates for 188 tax bills.

The JCT revenue estimates contain the estimated changes in the tax revenues for 1 to 8 years following the tax change. We then extract the “surprise” component of the tax revenue changes by taking the component of the tax revenue change that is different from the 1-year future value of the prior year’s tax revenue change, where future value is calculated using the 1-year Treasury nominal rate. For instance, suppose the JCT document states that the proposed tax cut in year  $t$  is expected to reduce tax revenues by \$5 billion in year  $t + 1$ , \$8 billion in year  $t + 2$ , and \$9 billion in year  $t + 3$ . If the realized Treasury rates are indicated by  $i$ , then the tax revenue surprises are taken as  $-\$5$  billion in year  $t + 1$ ,  $-\$8 + (1 + i_{t+1})\$5$  in year  $t + 2$ , and  $-\$9 + (1 + i_{t+2})\$8$  in year  $t + 3$ .<sup>81</sup>

To replicate the RR procedure, in addition to collecting tax revenue estimates, we also need to determine whether each bill is exogenous or endogenous to the prevailing economic conditions. To do this, we read the bill texts and the Economic Report of the President to determine whether each of those 188 bills are exogenous

---

<sup>81</sup>The 1-year rate is downloaded from the Board of Governors of the Federal Reserve System, series H15/H15/RIFLGFCY01\_N.B, on 5/4/2017.

in the RR sense.<sup>82</sup> Out of those 188, we found 74 exogenous bills that pass one chamber of the Congress but fail to be enacted. Five of them get re-introduced in later periods, so we drop them. In total, we have the revenue estimates for 69 exogenous tax bills that barely fail.

For analyses, we need economic variables. Since the GDP growth rate will be the dependent variable in most regressions, we download the nominal GDP and price index data from the Bureau of Labor Statistics.<sup>83</sup> Using GDP and price index  $P$ , we compute the GDP growth rate  $Y$  as

$$\Delta Y_t = \ln\left(\frac{GDP_t}{P_t}\right) - \ln\left(\frac{GDP_{t-1}}{P_{t-1}}\right) \quad (74)$$

We also normalize all of our tax revenue surprises through a division by GDP. That is, after computing the total exogenously proposed tax revenue surprise in year  $t$ ,  $\Delta TR_t$ , we compute our proposed tax change variable  $\Delta\tau_t$  as

$$\Delta\tau_t = \frac{\Delta TR_t}{GDP_t}$$

which is valid since both  $\Delta TR_t$  and  $GDP_t$  are nominal quantities. The time series of unlegislated exogenous tax changes,  $\Delta\tau_t$ , is plotted in Figure 1a.

For comparison, we also plot the time series of *legislated* exogenous tax changes, which we denote by  $\Delta T_t$  in the rest of the paper, in Figure 1b. For years 1978-2007, this series is simply taken from RR. For the years 2008-2016, we repeat the RR classification for the legislated tax bills to identify exogenous tax changes. The two series have a correlation of 0.21.

---

<sup>82</sup>RR categorize the motivation for tax changes into four: long-term economic growth, inherited deficit concerns, match increase government spending, and countercyclical tax changes. The first two are taken as exogenous motivations, and the last two are endogenous motivations. So far, the authors divided up this task, but we plan to cross-check our classifications by first individually classifying all tax bills and then comparing our classifications with each other. We will then jointly determine the classifications of the bills for which we classified differently.

<sup>83</sup>GDP and price index data are respectively downloaded from the National Income and Product Accounts, Table 1.1.5 and 1.1.4, on 5/2/2017.

**Table 20: Determinants of the Enactment of an Exogenous Tax Bill**

	Point estimate ( $\beta$ )	Robust Standard Error
Contentious House	0.53	(0.35)
Contentious Senate	-0.77	(0.31)
Party in Power	0.04	(0.09)
$\Delta Y_{t-1}$ (%)	-0.02	(0.04)
$\Delta Y_{t-2}$ (%)	0.02	(0.03)
Intercept	0.47	(0.18)
$N$		103
Adjusted $R^2$		0.05

### 13.1. Do the legislated and unlegislated tax changes occur in similar political and economic conditions?

Before using exogenous but unlegislated tax changes as the control group in a tax multiplier study, it is useful to know whether the chance of an exogenous tax bill becoming law conditional on having passed one chamber of the Congress is close to random. Hence, we use a linear discrete choice model to study what determines the chance of a bill becoming a law:

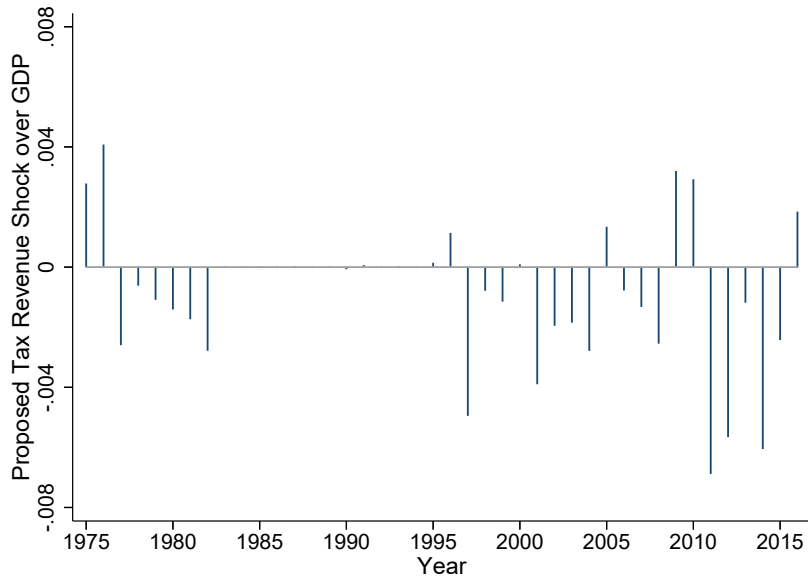
$$1(\text{bill } i \text{ becomes a law} | \text{passed one chamber}) = X_i \beta$$

where  $X_i$  comprises the following political and economic variables: contentious House (margin of majority in the House less than 5%), contentious Senate (margin of majority in the Senate less than 5%), party in power (House majority, Senate majority, and the President are from the same party), GDP growth in the previous year, and GDP growth two years prior. We use OLS to run the regression.<sup>84</sup>

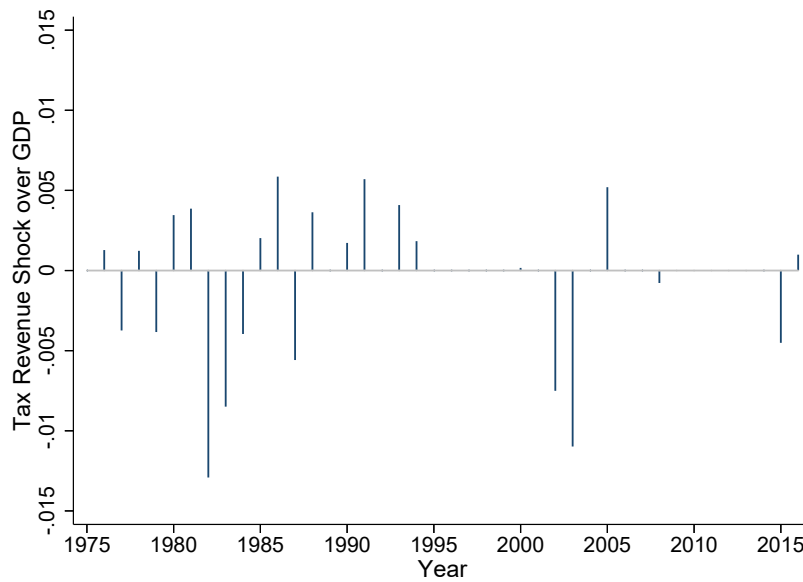
Table 1 reports the results. To interpret the point estimates, a more contentious Senate has a marginally significant effect in lowering the chance of a bill passing. It

---

<sup>84</sup>Using a logit model gives similar results.

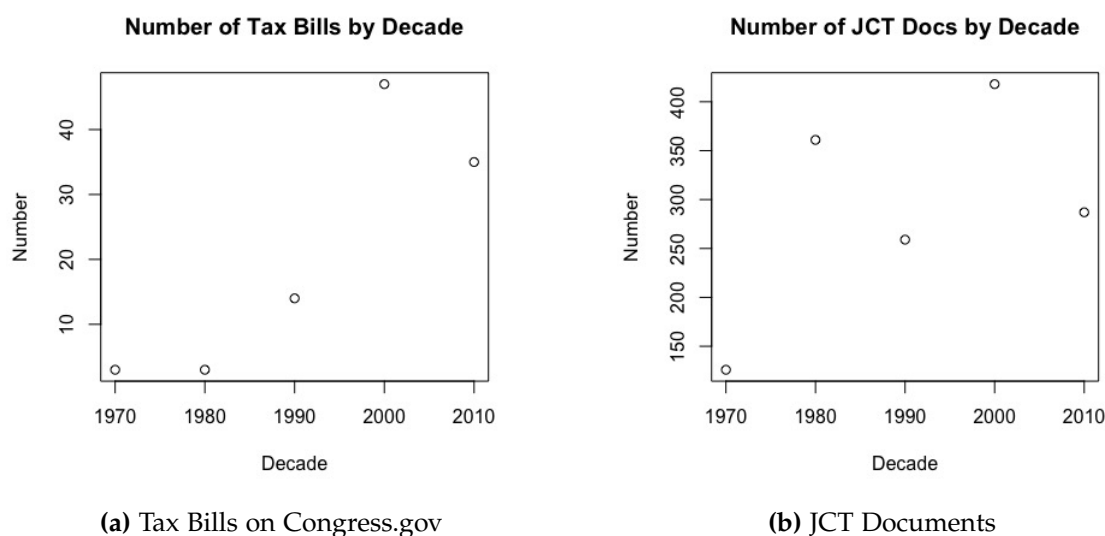


(a) Exogenous Tax Changes That Barely Failed (“Unlegislated”)



(b) Exogenous Tax Changes That Become Law (“Legislated”)

**Figure 14: Unlegislated vs. Legislated Exogenous Tax Changes, 1975-2016**



**Figure 15: Number of Tax Bills on Congress.gov vs. Number of Joint Committee on Taxation (JCT) Documents by Decade**

is, however, counterintuitive that a more contentious House increases the chance of a bill passing, although the effect is not significant. When single party is in power, the chance of a bill passing increases but not meaningfully. Lagged GDP growth rates do not seem to have meaningful effects.

### 13.2. A note about the incomplete data collection process

A prominent feature of the time-series of tax changes, plotted in Figure 1a, is that tax changes are completely missing in the 1980s and early 1990s. This is not because of a lack of tax bills in the 1980s but because of a classification issue on the Congress.gov website.

In Figure 2, we plot the number of bills classified as tax bills on Congress.gov as well as the number of Joint Committee on Taxation documents. Looking at Figure 2a, we see that very few bills were classified as tax bills in the 1970s and 1980s. This is not because there were few tax-related bills legislated. Figure 2b shows that 1970s had a large number of bills with tax revenue consequences. The problem seems to be that the Congress.gov website has classified a large fraction



of tax-related bills in the 1980s as “economics and public finance” bills.

This problem of missing 1980s data will be resolved in our future work as we plan to manually go through all JCT documents. The current draft, however, will suffer from the lack of tax bills from the period. Later on, we try excluding the pre-1995 years altogether to assess the sensitivity of our results to sample selection.

## 14. The legislated tax multiplier

RR estimates the tax multiplier using quarterly data over 1950 to 2007. In comparison, the unlegislated tax change data we collect are annual over 1978 to 2016. Hence, for the purpose of comparison, we replicate the results in RR using the same annual frequency over the period 1978 to 2016.

### 14.1. Using quarterly data

To do this, we use the sample period 1950-2007 to estimate

$$\Delta Y_t = a + \sum_{i=0}^M b_i \Delta T_{t-i} + trend_t + e_i, \quad (75)$$

where  $M=12$ .<sup>85</sup> Following RR, we also try including lagged GDP growth rates to run

$$\Delta Y_t = a + \sum_{i=0}^M b_i \Delta T_{t-i} + \sum_{j=1}^N c_j \Delta Y_{t-j} + trend_t + e_i, \quad (76)$$

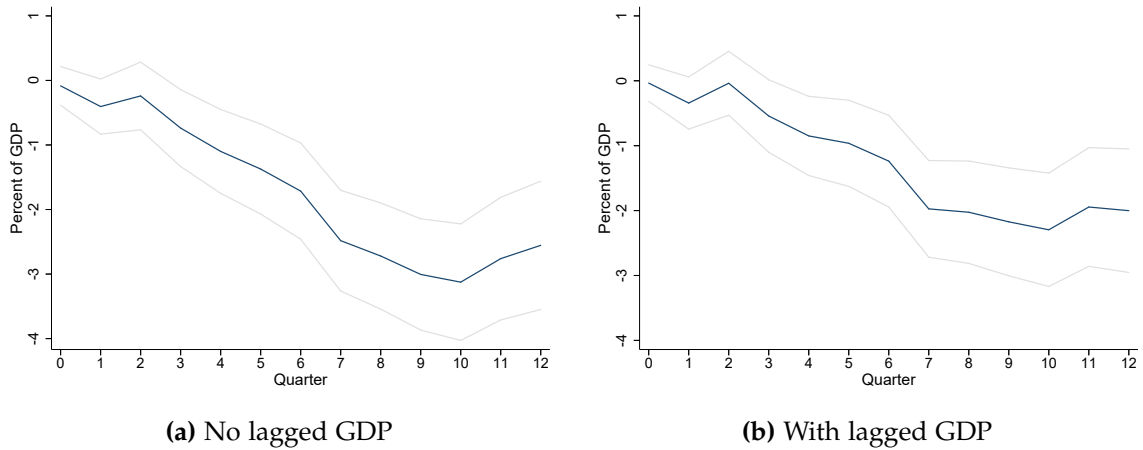
where  $N = 11$  to keep the same sample period.<sup>86</sup>

The resulting impulse responses of GDP following a 1 percentage point increase in the tax revenue as a fraction of GDP along with the one-standard-deviation con-

---

<sup>85</sup>Although RR do not include a time trend, but we include it to control for the fact that GDP growth slows down slowly over the long horizon.

<sup>86</sup>There is also a third specification based on VAR, which we do not do.



**Figure 16: Estimated Impact of A Legislated, “Exogenous” Tax Increase of 1 Percentage of GDP (quarterly over 1950.1-2007.4)**

confidence intervals are plotted in Figure 3.<sup>87</sup> The implied tax multiplier is between 2 and 3.

## 14.2. Using annual data

How do the results change if we use the original sample period but use annualized data? To do this, we sum across the tax revenue changes over different quarters within the same year. Then, we estimate the annualized versions of (75) and (76):

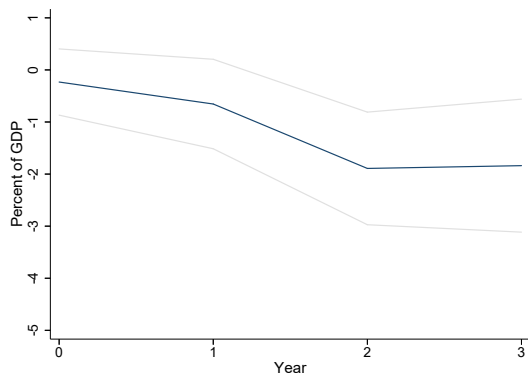
$$\begin{aligned} \Delta Y_t &= a + \sum_{i=0}^N b_i \Delta T_{t-i} + trend_t + e_i, \\ \Delta Y_t &= a + \sum_{i=0}^N b_i \Delta T_{t-i} + \sum_{j=1}^N c_j \Delta Y_{t-j} + trend_t + e_i \end{aligned} \tag{77}$$

over the sample period 1951-2007 where  $N = 3$ .

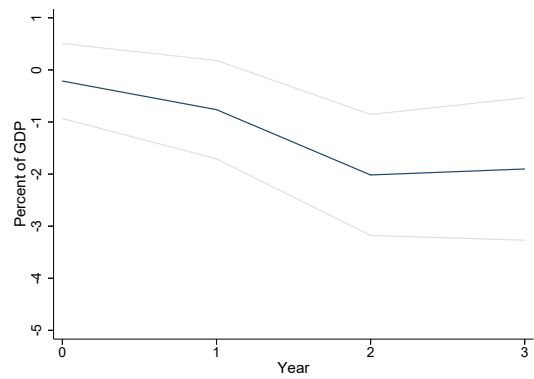
The impulse responses are plotted in Figures 4a and 4b. Both figures show that the tax multiplier identified with the annualized data is around 2.

---

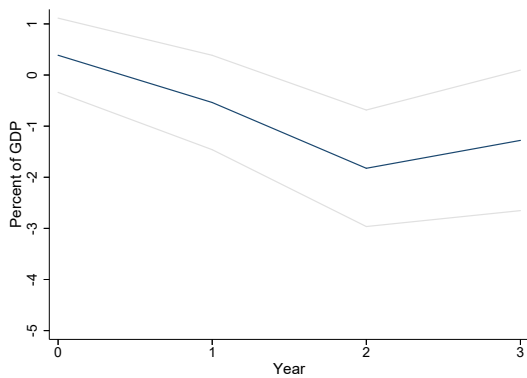
<sup>87</sup>The variance of the impulse response is calculated based on 10,000 random draws from a multivariate normal distribution with the mean and variance equal to the point estimates and the variance-covariance matrix of the parameter estimates. We do not assume serial correlation in the residuals.



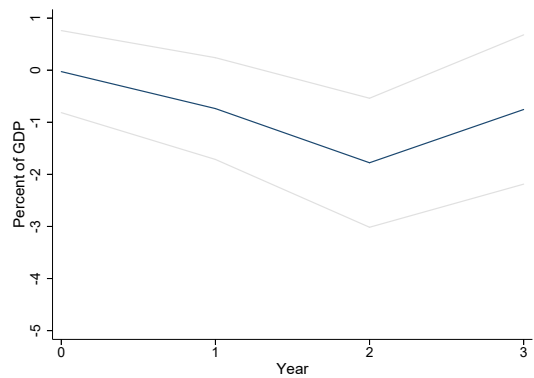
(a) No lagged GDP (1951-2007)



(b) With lagged GDP (1951-2007)



(c) No lagged GDP (1978-2016)



(d) With lagged GDP (1978-2016)

**Figure 17: Estimated Impact of A Legislated, “Exogenous” Tax Increase of 1 Percentage of GDP (annual over different periods)**

Our collected data have tax revenue information beginning 1975. Hence, to allow for three lagged values, we need to use the annual sample of 1978 to 2016. Here, for comparison, we use the same sample period and the annual frequency to estimate the RR tax multiplier.

Figures 4c and 4d report the impulse responses estimated using the two models. As before, the cumulative GDP response is the largest by the second year, although the rebound is more prominent starting the third year. The bottom line, however, is that the RR tax multiplier is around 2 when using the annual frequency.

## 15. The unlegislated tax multiplier

Now we estimate the unlegislated tax multiplier using tax changes that would have occurred had the tax bills passed. Since the bills did not actually pass, the resulting multiplier should reflect only the prevailing and anticipated economic conditions surrounding those bills without the interference of the causal effect of tax changes on the GDP.

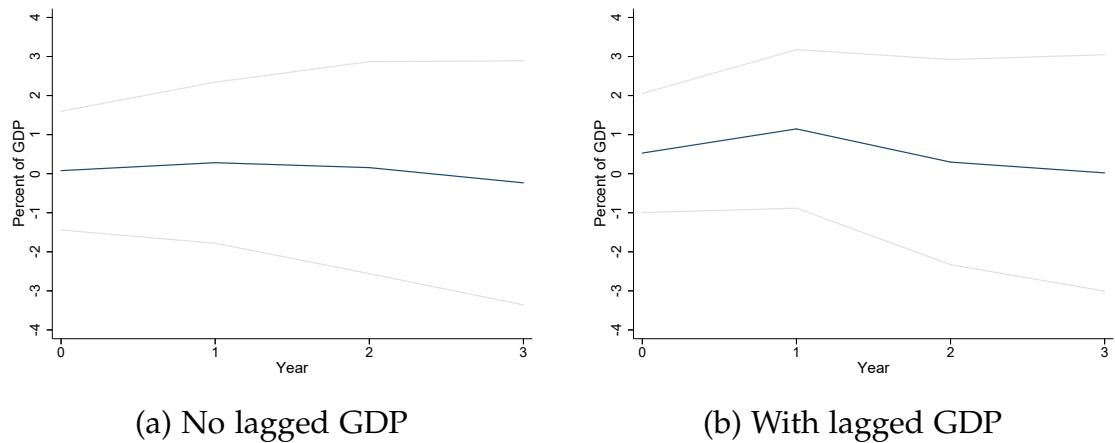
### 15.1. Baseline estimate

Specifically, we use the sample period 1978-2016 to estimate the two models,

$$\begin{aligned}\Delta Y_t &= a + \sum_{i=0}^N b_i \Delta \tau_{t-i} + trend_t + e_i, \\ \Delta Y_t &= a + \sum_{i=0}^N b_i \Delta \tau_{t-i} + \sum_{j=1}^N c_j \Delta Y_{t-j} + trend_t + e_i\end{aligned}\tag{78}$$

where  $\Delta \tau$  measures the unlegislated tax changes. Again, the lag  $N$  is 3 years.

The estimates under both assumptions (Figures 5a and 5b) show that we cannot reject the null that the unlegislated tax multiplier is zero. When the lagged GDP growth rates are controlled for, however, the point estimate of the unlegislated tax multiplier is positive and slightly above one.



**Figure 18: Estimated Impact of A Unlegislated, “Exogenous” Tax Increase of 1 Percentage of GDP (annual over 1978-2016)**

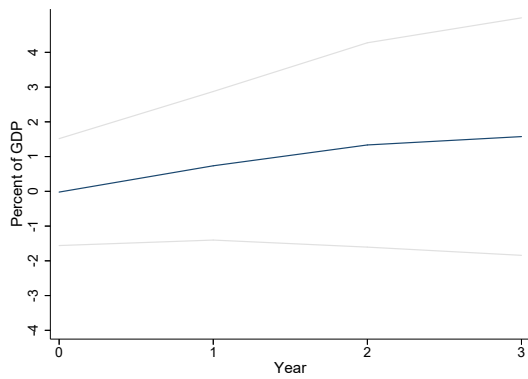
## 15.2. Robustness checks

Can we reliably conclude that the unlegislated tax multiplier is zero? Given that our data collection has focused mainly on 1995-2016, it may be premature to conclude one way or the other. Here, however, I check whether the result is driven by other factors. Hence, we check whether including current and lagged (up to 3 years) legislated tax changes, including current and lagged (up to 3 years) government spending, excluding the two crisis years of 2008 and 2009 (with the use of a dummy variable), or using only the years with reliable data (1995-2016, time trend excluded since the sample period is now short) changes anything.<sup>88</sup>

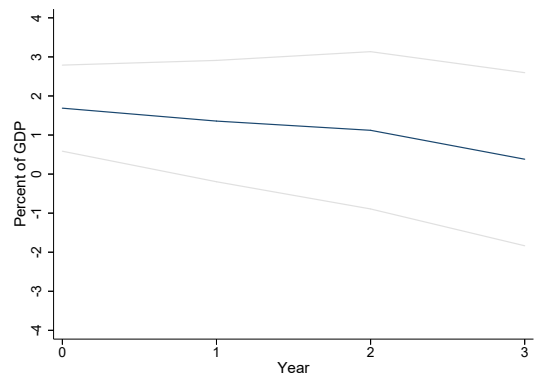
Figure 6 reports the results. Controlling for legislated tax changes or excluding the crisis years do not strongly affect the result. Controlling for government spending associates a unlegislated tax increase with an *increase* in the GDP growth rate. In contrast, excluding the years 1978-1994 associates a unlegislated tax increase with a slight decrease in the GDP growth rate. There is, however, no conclusive evidence that the unlegislated tax multiplier is either positive or negative.

---

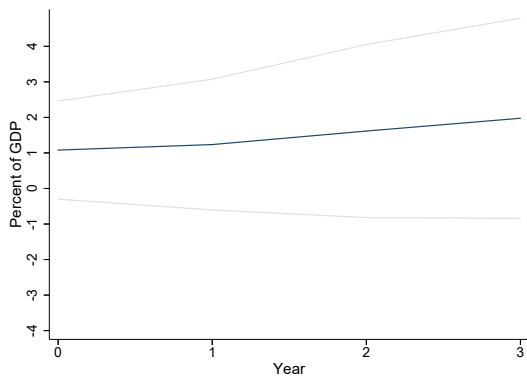
<sup>88</sup>Government spending is computed as the federal total gross expenditure minus interest payments. The data are downloaded from the National Income and Product Accounts (Table 3.2) on 5/3/2017. The government spending is then normalized by the nominal GDP.



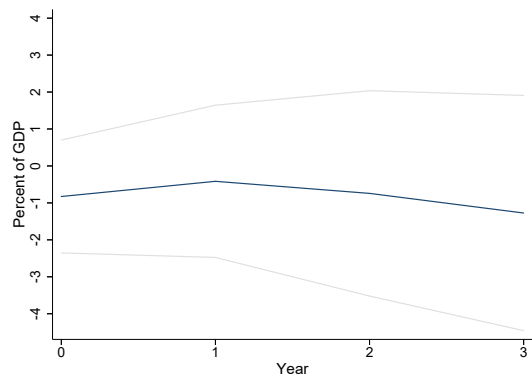
(a) With legislated tax changes



(b) With government spending



(c) Exclude crisis (2008 and 2009)



(d) Exclude 1978-1994

**Figure 19: Estimated Impact of A Unlegislated, “Exogenous” Tax Increase of 1 Percentage of GDP (annual over 1978-2016)**

## 16. Conclusion

In this paper, we examined the growth of GDP following proposed tax changes that barely failed to become law. Although our data collection is incomplete, we at this point find that those tax changes are not associated with GDP movements substantially different from normal times. This increases the reliability of the tax multipliers estimated using legislated tax bills.

Although we focus on the Romer-Romer (2010) study of tax multiplier, our method is also relevant to studies that use legislated bills to assess the government spending multiplier (e.g. Ramey and Shapiro, 1998; Ramey, 2011; and Barro and Redlick, 2011). A priori, we do not know whether the bias will be zero in government spending study. Future studies with our method of using “barely failed” bills as the counterfactual is warranted to assess the potential bias.

# **Chapter IV.**

## **Appendix to Chapter I**



## A. Model appendix

### A.1. Proofs of lemmas and propositions

Before proving Lemma 1, I establish the following argument made in the body of the text:

*Remark 2.* The following is true about anomaly assets at  $t = 1$ :

- (i) All anomaly assets “exploited” by arbitrageurs have the same returns.
- (ii) There exists a marginal anomaly asset  $j_1^* \in [0, 1]$  such that asset  $j$  is exploited if and only if  $j \in [j_1^*, 1]$ .
- (iii) All exploited assets generate return  $\bar{r}j_1^*$ , the expected return earned by asset  $j_1^*$  in the absence of arbitrageurs.

**Proof.** (i) Suppose otherwise. Then we can find two exploited assets  $j'$  and  $j''$  such that

$$E_1 r_{j',2} < E_1 r_{j'',2}$$

But in this case, a risk-neutral arbitrageur can increase its expected portfolio return by reducing its dollar position on  $j''$  by  $dx_{j'',1}$  and using it to increase its position on  $j'$  by  $dx_{j',1} = p_{j',1} p_{j'',1}^{-1} dx_{j'',1}$ .

To see (ii) and (iii), note that there must be some  $j^* \in [0, 1]$  such that the equal expected return earned by all exploited assets is  $\bar{r}j_1^*$ . It suffices to show that asset  $j$  is exploited if and only if  $j \in [j_1^*, 1]$ . Suppose first that  $j \in (j_1^*, 1]$  but is not exploited. Then, since  $j$  earns an expected return,

$$E_1 r_{j,2} = \bar{r}j > \bar{r}j_1^*$$

so that arbitrageurs are not optimizing. Now suppose  $j \in [0, j_1^*)$ . Then since

its demand by behavioral investors, given expected return  $\bar{r}j_1^*$ , is

$$B_{j,t} = j_1^* - j < 0,$$

the market does not clear if arbitrageurs take a long position on the asset.

**Proof of Lemma 1 (Equilibrium price at time  $t = 1$ ).** Trivially, if  $k_1 \leq 0$ , behavioral investors price all assets to ensure

$$E_1 [r_{j,2}] = \frac{v}{p_{j,1}} = 1 + \bar{r}j,$$

which implies

$$p_{j,1} = \frac{v}{1 + \bar{r}j}$$

Next, suppose  $k_1 \geq 0$  but arbitrageurs cannot remove all mispricings. Then, by Lemma 1, there exists  $j_1^* \in (0,1)$  such that arbitrageurs exploit assets if and only if  $j \in [j_1^*, 1]$  and earn the expected return

$$E_1 r_{j,2} = \bar{r}j_1^*$$

from them. Then, behavioral investor demand for each asset  $j \in [j_1^*, 1]$  is

$$B_{j,t} = j_1^* - j$$

in dollar position. Thus, for the market to clear, arbitrageur's demand for the asset needs to be  $x_{j,1} = j_1^* - j$ . Integrating this over  $[j_1^*, 1]$  should equal total arbitrageur capital, so that

$$\int_{j_1^*}^1 (j - j_1^*) dj = \mu k_1$$

This gives<sup>89</sup>

$$j_1^* = 1 - \sqrt{2\mu k_1}$$

and

$$p_{j,1} = \frac{v}{1 + \bar{r}j_1^*}$$

On the other hand, unexploited assets are priced by behavioral investors so that for all  $j \in [0, j_1^*]$ ,

$$p_{j,1} = \frac{v}{1 + \bar{r}j}$$

Finally, suppose arbitrageurs are unconstrained; that is,  $\mu k_1 \geq 1/2$ . Then, all anomaly assets are fully exploited so that

$$p_{j,1} = v$$

for all  $j \in [0, 1]$ .

**Proof of Lemma 2 (Anomaly asset's endogenous risk).** The negative covariance follows from the fact that

$$\frac{\partial \Lambda_1}{\partial p_{j,1}} = \frac{\partial \Lambda_1 / \partial k_1}{\partial p_{j,1} / \partial k_1} = \frac{- \text{ or } 0}{+ \text{ or } 0} \leq 0$$

(i) Suppose  $\mu \rightarrow 0^+$ . Then,  $k_1 \rightarrow 0^+$  so that  $\psi_1 = 1 + \bar{r}$  and  $p_{j,1} = v / (1 + \bar{r}j)$ .

Since both are deterministic,

$$\lim_{\mu \rightarrow 0} \text{Cov}_0(p_{j,1}, \Lambda_1) = 0$$

---

<sup>89</sup>Note that the other root is ruled out since it is always greater than 1, the largest possible value of  $j_1^*$ .

(ii) Note that

$$\frac{\partial \text{Cov}_0(p_{j,1}, \Lambda_1)}{\partial(\bar{r}_j)} = \frac{\frac{\partial \text{Cov}_0(p_{j,1}, \Lambda_1)}{\partial j}}{\frac{\partial(\bar{r}_j)}{\partial j}} = \frac{1}{\bar{r}} \times \frac{\partial \text{Cov}_0(p_{j,1}, \Lambda_1)}{\partial j}$$

Now, since  $\text{Cov}_0(p_{j,1}, \Lambda_1) = E_0[\Lambda_1 p_{j,1}] - E_0[\Lambda_1] E_0[p_{j,1}]$ ,

$$\begin{aligned} \text{Cov}_0(p_{j,1}, \Lambda_1) &= v \int_{-\infty}^0 \frac{1+c}{1+\bar{r}_j} dF(k_1) + v \int_0^{k_1(j)} \frac{1+\bar{r}_1^*}{1+\bar{r}_j} dF(k_1) + v \int_{k_1(j)}^{\infty} dF(k_1) \\ &- v E_0[\Lambda_1] \left( \int_{-\infty}^0 \frac{1}{1+\bar{r}_j} dF(k_1) + v \int_0^{k_1(j)} \frac{1}{1+\bar{r}_j} dF(k_1) + v \int_{k_1(j)}^{1/2} \frac{1}{1+\bar{r}_1^*} dF(k_1) \right. \\ &\quad \left. + v \int_{1/2}^{\infty} dF(k_1) \right), \end{aligned}$$

where  $k_1(j)$  is the value of  $k_1$  that gives  $j$  as the marginal asset, and  $F$  is the conditional cumulative density function of  $k_1$ . Thus, the derivative of the covariance with respect to  $j$  gives

$$\begin{aligned} \frac{\partial \text{Cov}_0(p_{j,1}, \Lambda_1)}{\partial j} &= -v \left( \int_{-\infty}^0 \frac{(1+c)\bar{r}}{(1+\bar{r}_j)^2} dF(k_1) + \int_0^{k_1(j)} \frac{(1+\bar{r}_1^*)\bar{r}}{(1+\bar{r}_j)^2} dF(k_1) \right) \\ &+ E_0[\Lambda_1] v \left( \int_{-\infty}^0 \frac{\bar{r}}{(1+\bar{r}_j)^2} dF(k_1) + \int_0^{k_1(j)} \frac{\bar{r}}{(1+\bar{r}_j)^2} dF(k_1) \right), \end{aligned}$$

where the Leibniz terms cancel out by the fact that  $j_1^*(K_1(j)) = j$ . Rearranging terms gives

$$\begin{aligned} \frac{\partial \text{Cov}_0(p_{j,1}, \Lambda_1)}{\partial j} &= -\frac{v\bar{r}}{(1+\bar{r}_j)^2} \left( \int_{-\infty}^{k_1(j)} \Lambda_1 dF(k_1) - E_0[\Lambda_1] \int_{-\infty}^{k_1(j)} dF(k_1) \right) \\ &= -\frac{v\bar{r}}{(1+\bar{r}_j)^2} (E_0[\Lambda_1 | j < j_1^*] - E_0[\Lambda_1]) F(k_1(j)) \\ &< 0, \end{aligned}$$

since  $E_0[\Lambda_1 | j < j_1^*] > E_0[\Lambda_1]$ .

**Lemma 3. (Monotonicity of prices at  $t = 0$ ).** For any  $j' < j''$  such that  $j', j'' \in [0, 1]$ ,

$$p_{j',0} \geq p_{j'',0}$$

**Proof.** Suppose for a contradiction that  $j' < j''$  but  $p_{j',0} < p_{j'',0}$ . Suppose also that  $j''$  is priced by arbitrageurs so that

$$p_{j'',0} = E_0 \left[ \frac{\Lambda_1}{\Lambda_0} p_{j'',1} \right]$$

Since  $p_{j',1} \geq p_{j'',1}$  in all states of  $t = 1$ , it must be that

$$p_{j',0} \geq E_0 \left[ \frac{\Lambda_1}{\Lambda_0} p_{j',1} \right] \geq E_0 \left[ \frac{\Lambda_1}{\Lambda_0} p_{j'',1} \right]$$

which is a contradiction. Now suppose that  $j''$  is priced by behavioral investors so that

$$p_{j'',0} = \frac{1}{1 + \bar{r}_{j''}} E_0 \left[ p_{j'',1} \right]$$

Since  $p_{j',1} \geq p_{j'',1}$  in all states of  $t = 1$ , it must be that

$$p_{j',0} \geq \frac{1}{1 + \bar{r}_{j'}} E_0 \left[ p_{j',1} \right] \geq \frac{1}{1 + \bar{r}_{j''}} E_0 \left[ p_{j'',1} \right]$$

which is also a contradiction.

**Proof of Proposition 1 (“Alphas” turn into “betas”).** Suppose  $j < j'$ . Then, by Lemma 2,  $-Cov_0(p_{j,1}, \Lambda_1) > -Cov_0(p_{j',1}, \Lambda_1)$ . Furthermore, by Lemma 3 in the Appendix,  $p_{j,0} < p_{j',0}$ . Thus,  $\beta_j = -Cov_0(\Lambda_1/\Lambda_0, r_{j,1})/Var_0(\Lambda_1/\Lambda_0) > -Cov_0(\Lambda_1/\Lambda_0, r_{j',1})/Var_0(\Lambda_1/\Lambda_0) = \beta_{j'}$ .

**Proof of Proposition 2** (*Beta is explained by anomaly-specific arbitrage capital*).

Recall that arbitrageur position on  $j$  is  $x_{j,1} = j - j_1^*$ . Hence, the unconditional expectation of arbitrageur position is  $E_0 x_{j,1} = j - E_0 [j_1^*]$ . Thus,  $\partial \beta_j / \partial E_0 [x_{j,1}] = (\partial \beta_j / \partial j) \times (\partial j / \partial E_0 [x_{j,1}]) = \partial \beta_j / \partial j > 0$ .

## A.2. Endogenizing the demand curves of behavioral investors

Suppose that each behavioral investor  $i$  believes the state variable underlying the cash flow at  $t = 2$  follows

$$v_{j,t+1} = \frac{1}{\phi_i} v_{j,t} + \varepsilon_{j,t+1} \quad (79)$$

Asset  $j$  has unique set of behavioral investors whose total mass equals one and whose beliefs are distributed according to

$$F_j(\phi_i) \sim U \left[ 1 + \bar{r} \left( j - \frac{1}{\theta} \right), 1 + \bar{r} \left( j + \frac{1}{\theta} \right) \right] \quad (80)$$

These behavioral investors, like arbitrageurs, cannot trade on margin. Then, given price  $p_{j,t}$  and the expectation that price at  $t + 1$  is  $p_{j,t+1} = a + v_{j,t+1}$ , the net demand for asset  $j$  by behavioral investors is

$$D_{j,t} = \frac{\theta}{2\bar{r}} \left( \int_{1+\bar{r}(j-\frac{1}{\theta})}^{v_{j,t}/p_{j,t}} di - \int_{v_{j,t}/p_{j,t}}^{1+\bar{r}(j+\frac{1}{\theta})} di \right) \quad (81)$$

which implies the demand of

$$D_{j,t} = \theta \left( \frac{E_t r_{j,t+1}}{\bar{r}} - j \right) \quad (82)$$

For behavioral investors at  $t = 1$ , this demand curve is the exact dollar position on  $j$  demanded by them, since they expect  $p_{j,2} = v_{j,2} = v_{j,1} + \varepsilon_{j,2}$ . For behavioral

investors at  $t = 0$ , however, price at  $t = 1$  is not necessarily expected to be linear in  $v_{j,1}$ . Still, if they perceive the price at  $t = 1$  to be approximately linear in  $v_{j,1}$  so that  $p_{j,1} \approx a + bv_{j,1}$ , then the demand curve is

$$D_{j,0} = \theta \left( b \frac{E_t r_{j,t+1}}{\bar{r}} - j + \frac{b}{\bar{r}} - 1 \right) \quad (83)$$

which is analytically identical to (82) for the purpose of this paper's model.

## B. Empirical appendix

### B.1. GMM formulations

#### Latent mispricing predicts funding beta

Consider the following data-generating process. In the pre-93 period, long-short return is a noisy realization of latent mispricing:

$$r_{j,t} = \bar{r}_j^{pre} + \epsilon_{j,t} \quad (84)$$

This latent mispricing determines the exposure of an anomaly asset to a funding shock:

$$\beta_j^{post} = b_0 + b_1 \bar{r}_j^{pre} + \eta_j \quad (85)$$

where  $\eta_j$  has a cross-sectional mean of zero. This beta then determines the long-short return in the post-93 period:

$$r_{j,t} = a_j^{post} + \beta_j^{post} f_t + \epsilon_{j,t} \quad (86)$$

These conditions imply the following  $4J$  moment conditions:

$$\mathcal{g}_{4J \times 1}(\mathbf{b}) = \begin{bmatrix} E \left[ \left( r_{j,t} - \bar{r}_j^{pre} \right) \mathbf{1}(t \in Pre) \right] \\ E \left[ \left( r_{j,t} - a_j^{post} - \beta_j^{post} f_t \right) \mathbf{1}(t \in Post) \right] \\ E \left[ \left( r_{j,t} - a_j^{post} - \beta_j^{post} f_t \right) f_t \mathbf{1}(t \in Post) \right] \\ E \left[ \left( \beta_j^{post} - b_0 - b_1 r_{j,t} \right) \mathbf{1}(t \in Pre) \right] \end{bmatrix}, \quad (87)$$



where the parameter vector is  $\mathbf{b} = [\bar{r}_1^{pre} \dots \bar{r}_J^{pre} a_1^{post} \dots a_J^{post} \beta_1^{post} \dots \beta_J^{post} \dots \beta_J^{post} b_0 b_1]'$ . At each  $t$ , the errors are,

$$g_t(\mathbf{b}) = \begin{bmatrix} \epsilon_{1,t} \mathbf{1}(t \in Pre) \\ \vdots \\ \epsilon_{J,t} \mathbf{1}(t \in Pre) \\ \epsilon_{1,t} \mathbf{1}(t \in Post) \\ \vdots \\ \epsilon_{J,t} \mathbf{1}(t \in Post) \\ \epsilon_{1,t} f_t \mathbf{1}(t \in Post) \\ \vdots \\ \epsilon_{J,t} f_t \mathbf{1}(t \in Post) \\ (\epsilon_{1,t} + \eta_1) \mathbf{1}(t \in Pre) \\ \vdots \\ (\epsilon_{J,t} + \eta_J) \mathbf{1}(t \in Pre) \end{bmatrix} \quad (88)$$

Note that the errors  $\epsilon_{j,t}$  represent the error in estimating the true latent mispricing  $\bar{r}^{pre}$  and the errors  $\eta_j$  represent the error in predicting post-93 beta using true  $\bar{r}^{pre}$ . Since the errors  $\epsilon_{j,t}$  are explicitly included in the last set of  $J$  moments, this GMM formulation takes into account the generated regressor problem for  $\bar{r}_j^{pre}$ .

Since the last  $J$  moments represent errors in the cross-sectional regression (33), they require an expectation taken in the cross-section of anomaly assets. Hence, I use the following selection matrix to take a cross-sectional expectation:

$$A_{(3J+2) \times 4J} = \begin{bmatrix} \mathbf{I}_{3J \times 3J} & \mathbf{0}_{3J \times J} \\ \mathbf{0}_{1 \times 3J} & \mathbf{1}_{1 \times J} \\ \mathbf{0}_{1 \times 3J} & \bar{\mathbf{r}}'_{1 \times J} \end{bmatrix} \quad (89)$$

where  $\bar{\mathbf{r}}$  is a  $J \times 1$  vector of pre-93 long-short returns. Then, the estimation problem

is to choose  $\hat{\mathbf{b}}$  to set

$$Ag_t(\hat{\mathbf{b}}) = 0_{(3J+2) \times 1} \quad (90)$$

or, equivalently, to minimize the sum of squared moments:

$$\hat{\mathbf{b}}_{(3J+2) \times 1} = \arg \min \left\{ (Ag_t)'_{1 \times (3J+2)} (Ag_t)_{(3J+2) \times 1} \right\} \quad (91)$$

Under this formulation, the parameter estimates will be identical to those derived from sequential OLS estimates of (84), (85), and (86). In particular, the last two rows of the selection matrix generate two moment conditions of the OLS estimation of (33). To see this, note that the vector of 1s in the right middle block of the selection matrix gives

$$\begin{aligned} 0 &= E_T [\epsilon_{1,t} + \eta_1] + \dots + E_T [\epsilon_{J,t} + \eta_J] \\ &= E_T \left[ \beta_1^{\text{post}} - b_0 - b_1 r_{1,t} \right] + \dots + E_T \left[ \beta_J^{\text{post}} - b_0 - b_1 r_{J,t} \right] \\ &\Leftrightarrow 0 = \sum_{j=1}^J \left( \beta_j^{\text{post}} - b_0 - b_1 E_T [r_{j,t}] \right) \\ &\Leftrightarrow 0 = E_J \left[ \beta_j^{\text{post}} - b_0 - b_1 E_T [r_{j,t}] \right] \end{aligned} \quad (92)$$

Analogously, the right lower block of the selection matrix implies

$$0 = E_J \left[ \left( \beta_j^{\text{post}} - b_0 - b_1 E_T [r_{j,t}] \right) E_T [r_{j,t}] \right] \quad (93)$$

This shows that the last two rows of the selection matrix ensures that the cross-sectional regression has the same moments implied by the OLS implementation of it.

The spectral density for the moments is

$$S_{4J \times 4J} = \sum_{\tau=-\infty}^{\infty} E [g_t(\mathbf{b}) g_{t-\tau}(\mathbf{b})'], \quad (94)$$

which I estimate assuming no serial correlation:

$$S_T = E_T [g_t(\mathbf{b}) g_t(\mathbf{b})'] \quad (95)$$

Hansen (1982) shows that

$$\sqrt{T}(\hat{\mathbf{b}} - \mathbf{b}) \rightarrow N \left[ 0, (Ad)^{-1} ASA' \left( (Ad)^{-1} \right)' \right], \quad (96)$$

where  $d$  is a matrix representing the sensitivity of moments with respect to parameter values:

$$d_{4J \times (3J+2)} = \frac{\partial g_T(\mathbf{b})}{\partial \mathbf{b}'} \quad (97)$$

Hence,

$$var(\hat{\mathbf{b}}) = \frac{1}{T} (Ad)^{-1} ASA' \left( (Ad)^{-1} \right)' \quad (98)$$

### Adding pre-93 beta as another regressor

These conditions imply the following  $6J$  moment conditions:

$$g_{6J \times 1}(\mathbf{b}) = \begin{bmatrix} E \left[ \left( r_{j,t} - \bar{r}_j^{pre} \right) \mathbf{1}(t \in Pre) \right] \\ E \left[ \left( r_{j,t} - a_j^{post} - \beta_j^{post} f_t \right) \mathbf{1}(t \in Post) \right] \\ E \left[ \left( r_{j,t} - a_j^{post} - \beta_j^{post} f_t \right) f_t \mathbf{1}(t \in Post) \right] \\ E \left[ \left( r_{j,t} - a_j^{pre} - \beta_j^{pre} f_t \right) \mathbf{1}(t \in Pre) \right] \\ E \left[ \left( r_{j,t} - a_j^{pre} - \beta_j^{pre} f_t \right) f_t \mathbf{1}(t \in Pre) \right] \\ E \left[ \left( \beta_j^{post} - b_0 - b_1 r_{j,t} - b_2 \beta_j^{pre} \right) \mathbf{1}(t \in Pre) \right] \end{bmatrix} \quad (99)$$

The selection matrix is

$$A_{(5J+3) \times 6J} = \begin{bmatrix} \mathbf{I}_{5J \times 5J} & \mathbf{0}_{5J \times J} \\ \mathbf{0}_{1 \times 5J} & \mathbf{1}_{1 \times J} \\ \mathbf{0}_{1 \times 5J} & \bar{\mathbf{r}}'_{1 \times J} \\ \mathbf{0}_{1 \times 5J} & \boldsymbol{\beta}^{pre}_{1 \times J} \end{bmatrix} \quad (100)$$

The formulae for spectral density  $S$ , the matrix  $d$ , and the covariance of estimates are analogous to those in the previous analysis.

### Adding pre-93 volatility as another regressor

These conditions imply the following  $6J$  moment conditions:

$$g_{5J \times 1}(\mathbf{b}) = \begin{bmatrix} E \left[ \left( r_{j,t} - \bar{r}_j^{pre} \right) \mathbf{1}(t \in Pre) \right] \\ E \left[ \left( r_{j,t} - a_j^{post} - \beta_j^{post} f_t \right) \mathbf{1}(t \in Post) \right] \\ E \left[ \left( r_{j,t} - a_j^{post} - \beta_j^{post} f_t \right) f_t \mathbf{1}(t \in Post) \right] \\ E \left[ \left( \frac{T/2}{T/2-1} \left( r_{j,t} - \bar{r}_j^{pre} \right)^2 - \left( \sigma_j^{pre} \right)^2 \right) \mathbf{1}(t \in Pre) \right] \\ E \left[ \left( \beta_j^{post} - b_0 - b_1 r_{j,t} - b_2 \sigma_j^{pre} \right) \mathbf{1}(t \in Pre) \right] \end{bmatrix} \quad (101)$$

The selection matrix is

$$A_{(4J+3) \times 5J} = \begin{bmatrix} \mathbf{I}_{4J \times 4J} & \mathbf{0}_{4J \times J} \\ \mathbf{0}_{1 \times 4J} & \mathbf{1}_{1 \times J} \\ \mathbf{0}_{1 \times 4J} & \bar{\mathbf{r}}'_{1 \times J} \\ \mathbf{0}_{1 \times 4J} & \boldsymbol{\sigma}^{pre}_{1 \times J} \end{bmatrix} \quad (102)$$

The formulae for spectral density  $S$ , the matrix  $d$ , and the covariance of estimates are analogous to those in the previous analysis.

## Latent mispricing predicts the rate of increase in funding beta

I assume the following data-generating process in which an exposure to arbitrageur funding grows at the rate  $\beta_{j,1}^{post} t^{-1}$ , where  $t$  is the number of quarters into the post-93 sample (1994Q1 being  $t = 1$ ):

$$\begin{aligned}
 \text{Pre-93 return:} & & r_{j,t} &= \bar{r}_j^{pre} + \epsilon_{j,t} \\
 \text{Post-93 return:} & & r_{j,t} &= a_j^{post} + \left( \beta_{j,0}^{post} + \beta_{j,1}^{post} \ln(t) \right) f_t + \epsilon_{j,t} \\
 \text{Beta determination:} & & \beta_{j,1}^{post} &= b_0 + b_1 \bar{r}_j^{pre} + \eta_j
 \end{aligned} \tag{103}$$

Here,  $\beta_{j,1}^{post}$  indicates the growth of funding beta due to growth of arbitrageur mass over time ( $\mu$  in the model). The moment conditions are now

$$g_{5J \times 1}(\mathbf{b}) = \begin{bmatrix} E \left[ \left( r_{j,t} - \bar{r}_j^{pre} \right) \mathbf{1}(t \in Pre) \right] \\ E \left[ \left( r_{j,t} - a_j^{post} - \beta_{j,0}^{post} f_t - \beta_{j,1}^{post} \ln(t) f_t \right) \mathbf{1}(t \in Post) \right] \\ E \left[ \left( r_{j,t} - a_j^{post} - \beta_{j,0}^{post} f_t - \beta_{j,1}^{post} \ln(t) f_t \right) f_t \mathbf{1}(t \in Post) \right] \\ E \left[ \left( r_{j,t} - a_j^{post} - \beta_{j,0}^{post} f_t - \beta_{j,1}^{post} \ln(t) f_t \right) \ln(t) f_t \mathbf{1}(t \in Post) \right] \\ E \left[ \beta_{j,1}^{post} - b_0 - b_1 r_{j,t} \right] \mathbf{1}(t \in Pre) \end{bmatrix} \tag{104}$$

The parameter vector is

$$\mathbf{b}_{(4J+2) \times 1} = \begin{bmatrix} \bar{r}_1^{pre} \\ \vdots \\ \bar{r}_J^{pre} \\ a_1^{post} \\ \vdots \\ a_J^{post} \\ \beta_{1,0}^{post} \\ \vdots \\ \beta_{J,0}^{post} \\ \beta_{1,1}^{post} \\ \vdots \\ \beta_{J,1}^{post} \\ b_0 \\ b_1 \end{bmatrix} \quad (105)$$

The selection matrix is

$$A_{(4J+2) \times 5J} = \begin{bmatrix} \mathbf{I}_{4J \times 4J} & \mathbf{0}_{4J \times J} \\ \mathbf{0}_{1 \times 4J} & \mathbf{1}_{1 \times J} \\ \mathbf{0}_{1 \times 4J} & \bar{r}'_{1 \times J} \end{bmatrix} \quad (106)$$

The formulae for spectral density  $S$ , the matrix  $d$ , and the covariance of estimates are analogous to those in the previous analysis.

## Latent mispricing predicts funding correlation

The data-generating process here is assumed to be identical to that of the latent mispricing to funding beta regression. The moment conditions are

$$\mathcal{G}_{(5J+2) \times 1}(\mathbf{b}) = \begin{bmatrix} E \left[ \left( r_{j,t} - \bar{r}_j^{pre} \right) \mathbf{1}(t \in Pre) \right] \\ E \left[ \left( r_{j,t} - \bar{r}_j^{post} \right) \mathbf{1}(t \in Post) \right] \\ E \left[ \left( \frac{T/2}{T/2-1} \left( r_{j,t} - \bar{r}_j^{post} \right)^2 - \left( \sigma_j^{post} \right)^2 \right) \mathbf{1}(t \in Post) \right] \\ E \left[ \left( f_t - \bar{f}^{post} \right) \mathbf{1}(t \in Post) \right] \\ E \left[ \left( \frac{T/2}{T/2-1} \left( f_t - \bar{f}^{post} \right)^2 - \left( \sigma_f^{post} \right)^2 \right) \mathbf{1}(t \in Post) \right] \\ E \left[ \left( \frac{T/2}{T/2-1} \left( r_{j,t} - \bar{r}_j^{post} \right) \left( f_t - \bar{f}^{post} \right) - \sigma_{j,f} \right) \mathbf{1}(t \in Post) \right] \\ E \left[ \sigma_{j,f}^{post} \left( \sigma_j^{post} \right)^{-1} \left( \sigma_f^{post} \right)^{-1} - b_0 - b_1 r_{j,t} \right] \mathbf{1}(t \in Pre) \end{bmatrix} \quad (107)$$

where  $T/2$  is the number of periods in each subsample. The parameter vector is

$$\mathbf{b}_{(4J+4) \times 1} = \begin{bmatrix} \bar{r}_1^{pre} \\ \vdots \\ \bar{r}_J^{pre} \\ \bar{r}_1^{post} \\ \vdots \\ \bar{r}_J^{post} \\ \sigma_1^{post} \\ \vdots \\ \sigma_J^{post} \\ \bar{f}^{post} \\ \sigma_f^{post} \\ \sigma_{1,f}^{post} \\ \vdots \\ \sigma_{J,f}^{post} \\ b_0 \\ b_1 \end{bmatrix} \quad (108)$$

The selection matrix is

$$A_{(4J+4) \times (5J+2)} = \begin{bmatrix} \mathbf{I}_{(4J+2) \times (4J+2)} & \mathbf{0}_{(4J+2) \times J} \\ \mathbf{0}_{1 \times (4J+2)} & \mathbf{1}_{1 \times J} \\ \mathbf{0}_{1 \times (4J+2)} & \bar{r}'_{1 \times J} \end{bmatrix} \quad (109)$$



### “Intermediary asset pricing” of anomaly assets based on endogenous risks

The moment conditions for the cross-sectional test are the conditions for estimating  $\beta$ s in the time series and the conditions for estimating  $\lambda$ s in the cross section:

$$g(\mathbf{b}) = \begin{bmatrix} E[r_{j,t} - a_j - \beta_j f_t] \\ E[(r_{j,t} - a_j - \beta_j f_t) f_t] \\ E[r_{j,t} - \lambda_0 - \lambda_1 \beta_j] \end{bmatrix} \quad (110)$$

This vector represents  $(2 + K)J$  moment conditions. To obtain OLS coefficients, I use the selection matrix

$$A = \begin{bmatrix} \mathbf{I}_{(J+JK) \times (J+JK)} & \mathbf{0}_{(J+JK) \times J} \\ \mathbf{0}_{1 \times (J+JK)} & \mathbf{1}_{1 \times J} \\ \mathbf{0}_{K \times (J+JK)} & \boldsymbol{\beta}'_{K \times J} \end{bmatrix} \quad (111)$$

where  $\mathbf{1}_{1 \times J}$  is a vertical vector of ones and  $\boldsymbol{\beta}$  is a  $J$  by  $K$  matrix of  $\beta$ s.

I compute two types of standard errors. I compute Shanken standard errors under the assumption of zero autocorrelations and zero cross-anomaly correlations. Hence, I take the panel of residuals from the time-series regressions for betas,

$$\epsilon_t = r_t - \beta f_t \quad (112)$$

where  $r_t$  and  $\beta$  are vertical vectors of different anomaly assets' returns and betas, respectively. Then, the variance of the price of risk is computed as

$$Var(\hat{\lambda}) = \frac{1}{T} \left[ \Sigma_f + (\beta' \beta)^{-1} \beta' \Sigma \beta (\beta' \beta)^{-1} \left( 1 + \lambda' \Sigma_f^{-1} \lambda \right) \right] \quad (113)$$

where

$$\Sigma = \text{diag} \left( \sum_{t=1}^T \epsilon_t \epsilon_t' \right) \quad (114)$$

GMM standard errors are estimated in the standard way, which allows the errors to be correlated in the cross-section.

### Funding betas are formed during constrained times

Here, I test whether pre-93 mean return predicts post-93 change in beta from unconstrained to constrained times. The moment conditions are,

$$g_{6J \times 1}(\mathbf{b}) = \begin{bmatrix} E \left[ \left( r_{j,t} - \bar{r}_j^{pre} \right) \mathbf{1}(t \in Pre) \right] \\ E \left[ \left( r_{j,t} - \left( a_{j,0}^{post} + \beta_{j,0}^{post} f_t \right) \right) \mathbf{1}(t \in Post, Unconstrained) \right] \\ E \left[ \left( r_{j,t} - \left( a_{j,0}^{post} + \beta_{j,0}^{post} f_t \right) \right) f_t \mathbf{1}(t \in Post, Unconstrained) \right] \\ E \left[ \left( r_{j,t} - \left( a_{j,0}^{post} + \Delta a_j^{post} + \left( \beta_{j,0}^{post} + \Delta \beta_j^{post} \right) f_t \right) \right) \mathbf{1}(t \in Post, Constrained) \right] \\ E \left[ \left( r_{j,t} - \left( a_{j,0}^{post} + \Delta a_j^{post} + \left( \beta_{j,0}^{post} + \Delta \beta_j^{post} \right) f_t \right) \right) f_t \mathbf{1}(t \in Post, Constrained) \right] \\ E \left[ \left( \Delta \beta_j^{post} - b_0 - b_1 r_{j,t} \right) \mathbf{1}(t \in Pre) \right] \end{bmatrix} \quad (115)$$

To obtain OLS coefficients, I use the selection matrix

$$A = \begin{bmatrix} \mathbf{I}_{5J \times 5J} & \mathbf{0}_{5J \times J} \\ \mathbf{0}_{1 \times 5J} & \mathbf{1}_{1 \times J} \\ \mathbf{0}_{1 \times 5J} & \bar{r}'_{1 \times J} \end{bmatrix} \quad (116)$$

Testing if  $\bar{r}_j^{pre}$  predicts  $\beta_{j,0}^{post}$  or  $\Delta a_j^{post}$  just requires changing the last moment condition.

## B.2. Constructing the anomaly assets

See the online appendix.

## References

- [1] Adrian, Tobias, Erkko Etula, and Tyler Muir, 2014, Financial intermediaries and the cross-section of asset returns, *Journal of Finance* 69(6), 2557-2596.
- [2] Adrian, Tobias, Emanuel Moench, and Hyun Song Shin, 2013, Dynamic leverage asset pricing, working paper.
- [3] Adrian, Tobias and Hyun Song Shin, 2010, Liquidity and leverage, *Journal of Financial Intermediation* 19(3), 418-437.
- [4] Akbas, Ferhat, Will J. Armstrong, Sorin Sorescu, and Avanidhar Subrahmanyam, 2015, Smart money, dumb money, and capital market anomalies, *Journal of Financial Economics* 118(2), 355-382.
- [5] Alizadeh, Sassan, Michael W. Brandt, and Francis X. Diebold, 2002, Range-based estimation of stochastic volatility models, *Journal of Finance* 57(3), 1047-1091.
- [6] Allen, Franklin, and Douglas Gale, 2005, From cash-in-the-market pricing to financial fragility, *Journal of the European Economic Association* 3(2-3), 535-546.
- [7] Amaya, Diego, Peter Christoffersen, Kris Jacobs, and Aurelio Vasquez, 2015, Does realized skewness predict the cross-section of equity returns? *Journal of Financial Economics* 118(1), 135-167.
- [8] Amihud, Yakov, and Haim Mendelson, 1991, Liquidity, maturity, and the yields on US Treasury securities, *Journal of Finance* 46(4), 1411-1425.
- [9] Aragon, George O., and Philip E. Strahan, 2012, Hedge funds as liquidity providers: Evidence from the Lehman bankruptcy, *Journal of Financial Economics* 103(3), 570-587.
- [10] Asl, Farshid M., and Erkko Etula, 2012, Advancing strategic asset allocation in a multi-factor world, *Journal of Portfolio Management* 39(1), 59-66.
- [11] Avramov, Doron, Tarun Chordia, and Amit Goyal, 2006a, Liquidity and auto-correlations in individual stock returns, *Journal of Finance* 61(5), 2365-2394.
- [12] Avramov, Doron, Tarun Chordia, and Amit Goyal, 2006b, The impact of trades on daily volatility, *Review of Financial Studies* 19(4), 1241-1277.

- [13] Bali, Turan G., Nusret Cakici, and Robert F. Whitelaw, 2014, Hybrid tail risk and expected stock returns: When does the tail wag the dog? *Review of Asset Pricing Studies* 4(2), 206-246.
- [14] Banz, Rolf W., 1981, The relationship between return and market value of common stocks, *Journal of Financial Economics* 9(1), 3-18.
- [15] Barber, Brad M., Terrance Odean, and Ning Zhu, 2009, Do retail trades move markets?, *Review of Financial Studies* 22(1), 151-186.
- [16] Barro, Robert J. and Charles J. Redlick, 2011, The macroeconomic effects of government purchases and taxes, *Quarterly Journal of Economics* 126(1), 51-102.
- [17] Bassi, Anna, Riccardo Colacito, and Paolo Fulghieri, 2013, 'O sole mio: an experimental analysis of weather and risk attitudes in financial decisions, *Review of Financial Studies* 26(7), 1824-185.
- [18] Beckert, Sven, 2003, *The monied metropolis: New York City and the consolidation of the American bourgeoisie, 1850-1896*, Cambridge: Cambridge University Press.
- [19] Ben-David, Itzhak, Francesco Franzoni, and Rabih Moussawi, 2012, Hedge fund stock trading in the financial crisis of 2007–2009, *Review of Financial Studies* 25(1), 1-54.
- [20] Berry, Steven, James Levinsohn, and Ariel Pakes, 1995, Automobile prices in market equilibrium, *Econometrica* 63(4), 841-890.
- [21] Blanchard, Olivier and Roberto Perotti, 2002, An empirical characterization of the dynamic effects of changes in government spending and taxes on output, *Quarterly Journal of Economics* 117(4), 1329-1368.
- [22] Boehmer, Ekkehart, Charles M. Jones, and Xiaoyan Zhang, 2013, Shackling short sellers: The 2008 shorting ban, *Review of Financial Studies* 26(6), 1363-1400.
- [23] Bollerslev, Tim, Viktor Todorov, and Lai Xu, 2015, Tail risk premia and return predictability, *Journal of Financial Economics* 118(1), 113-134.
- [24] Bouman, Sven, and Ben Jacobsen, 2002, The Halloween indicator, "Sell in May and go away": Another puzzle, *American Economic Review* 92(5), 1618-1635.

- [25] Brunnermeier, Markus, Stefan Nagel, and Lasse Heje Pedersen, 2009, Carry trades and currency crashes, *NBER Macroeconomics Annual 2008* (23), 313–347
- [26] Brunnermeier, Markus K., and Lasse Heje Pedersen, 2009, Market liquidity and funding liquidity, *Review of Financial studies* 22(6), 2201-2238.
- [27] Brunnermeier, Markus K., and Yuliy Sannikov, 2014, A macroeconomic model with a financial sector, *American Economic Review* 104(2), 379-421.
- [28] Buchanan, Matt, 2013, An apparatus for treating air: The modern air conditioner, *The New Yorker*, Jun. 5, 2013. <http://www.newyorker.com/tech/elements/an-apparatus-for-treating-air-the-modern-air-conditioner>.
- [29] Caldara, Dario and Christophe Kamps, 2012, The analytics of SVARs: A unified framework to measure fiscal multipliers, *Review of Economic Studies*, forthcoming.
- [30] Campbell, John Y., Sanford J. Grossman, and Jiang Wang, 1993, Trading volume and serial correlation in stock returns, *Quarterly Journal of Economics* 108(4), 905-939.
- [31] Campbell, John Y., Christopher Polk, and Tuomo Vuolteenaho, 2010, Growth or glamour? Fundamentals and systematic risk in stock returns, *Review of Financial Studies* 23(1), 305-344.
- [32] Campbell, John Y., and Robert J. Shiller, 1988, The dividend-price ratio and expectations of future dividends and discount factors, *Review of financial studies* 1(3), 195-228.
- [33] Campbell, John Y., and Tuomo Vuolteenaho, 2004, Bad beta, good beta, *American Economic Review* 94(5), 1249–1275.
- [34] Cao, Bing, Dan Dhaliwal, Adam C. Kolasinski, and Adam V. Reed, 2012, Bears and numbers: Short sellers exploit and affect mispricing of earnings, working paper.
- [35] Cao, Melanie, and Jason Wei, 2005, Stock market returns: A note on temperature anomaly, *Journal of Banking and Finance* 29(6), 1559-1573.
- [36] Carhart, Mark M., 1997, On persistence in mutual fund performance, *Journal*

of Finance 52(1), 57-82.

- [37] CEA, 2014, Economic Report to the President, United States Government Printing Office.
- [38] Chahrour, Ryan, Stephanie Schmitt-Grohé, and Martín Uribe. A model-based evaluation of the debate on the size of the tax multiplier, *American Economic Journal: Economic Policy* 4(2), 28-45.
- [39] Chen, Joseph, Harrison Hong, and Jeremy C. Stein, 2001, Forecasting crashes: trading volume, past returns, and conditional skewness in stock prices, *Journal of Financial Economics* 61(3), 345-381.
- [40] Chordia, Tarun, Avanidhar Subrahmanyam, and Qing Tong, 2014, Have capital market anomalies attenuated in the recent era of high liquidity and trading activity? *Journal of Accounting and Economics* 58(1), 41-58.
- [41] Cochrane, John H., 2005, *Asset pricing* (revised edition), Princeton University Press.
- [42] Cochrane, John H., 2011, Presidential address: Discount rates, *Journal of Finance* 66(4), 1047-1108.
- [43] Coffey, Bentley, Patrick A. McLaughlin, and Robert D. Tollison, 2012, Regulators and redskins, *Public Choice* 153(1-2), 191-204.
- [44] Cohen, Randolph B., Christopher Polk, and Tuomo Vuolteenaho, 2009, The price is (almost) right, *Journal of Finance* 64(6), 2739-2782.
- [45] Conrad, Jennifer S., Allaudeen Hameed, and Cathy Niden, 1994, Volume and autocovariances in short-horizon individual security returns, *Journal of Finance* 49(4), 1305-1329.
- [46] Cooper, Michael, 1999, Filter rules based on price and volume in individual security overreaction, *Review of Financial Studies* 12(4), 901-935.
- [47] Corwin, Shane A., and Paul Schultz, 2012, A simple way to estimate bid-ask spreads from daily high and low prices, *Journal of Finance* 67(2), 719-760.
- [48] Daniel, Kent, and Sheridan Titman, 1997, Evidence on the characteristics of cross sectional variation in stock returns, *Journal of Finance* 52(1), 1-33.

- [49] Davis, James L., Eugene F. Fama, and Kenneth R. French, 2000, Characteristics, covariances, and average returns: 1929 to 1997, *Journal of Finance* 55(1), 389-406.
- [50] Dawson, John W., and John J. Seater, 2013, Federal regulation and aggregate economic growth, *Journal of Economic Growth* 18(2), 137-177.
- [51] Dechow, Patricia M., Amy P. Hutton, Lisa Meulbroek, and Richard G. Sloan, 2001, Short-sellers, fundamental analysis, and stock returns, *Journal of Financial Economics* 61(1), 77-106.
- [52] Dell, Melissa, Benjamin Jones, and Benjamin Olken, 2014, What do we learn from the weather? The new climate-economy literature, *Journal of Economic Literature* 52(3), 740-798.
- [53] De Long, J. Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann, 1990, Noise trader risk in financial markets, *Journal of Political Economy* 98(4), 703-738.
- [54] Dimson, Elroy, 1979, Risk measurement when shares are subject to infrequent trading, *Journal of Financial Economics* 7(2), 197-226.
- [55] Donnan, Chris, 2011, A brief history of market data, Presentation at the European Financial Information Summit 2011. [http://www.financialinformationsummit.com/digital\\_assets/6006/Chris\\_Donnan.pdf](http://www.financialinformationsummit.com/digital_assets/6006/Chris_Donnan.pdf).
- [56] Drechsler, Itamar, and Qingyi Freda Drechsler, 2016, The shorting premium and asset pricing anomalies, working paper.
- [57] Duarte, Jefferson, Francis A. Longstaff, and Fan Yu, 2007, Risk and return in fixed-income arbitrage: Nickels in front of a steamroller? *Review of Financial Studies* 20(3), 769-811.
- [58] Fama, Eugene F., and Kenneth R. French, 1992, The cross-section of expected stock returns, *Journal of Finance* 47(2), 427-465.
- [59] Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33(1), 3-56.
- [60] Fama, Eugene F., and Kenneth R. French, 1996, Multifactor explanations of asset pricing anomalies, *Journal of Finance* 51(1), 55-84.

- [61] Fama, Eugene F., and Kenneth R. French, 2008, Dissecting anomalies, *Journal of Finance* 63(4), 1653-1678.
- [62] Faucault, Thierry, David Sraer, and David Themar, 2011, Individual investors and volatility, *Journal of Finance* 66(4), 1369-1406.
- [63] Favero, Carlo, and Francesco Giavazzi. Measuring tax multipliers: The narrative method in fiscal VARs, *American Economic Journal: Economic Policy* 4(2), 69-94.
- [64] Frazzini, Andrea, Ronen Israel, and Tobias J. Moskowitz, 2015, Trading costs of asset pricing anomalies, working paper.
- [65] Frazzini, Andrea, and Lasse Heje Pedersen, 2014, Betting against beta, *Journal of Financial Economics* 111(1), 1-25.
- [66] Froot, Kenneth A., and Emil M. Dabora, 1999, How are stock prices affected by the location of trade? *Journal of Financial Economics* 53(2), 189-216.
- [67] Gallant, A. Ronald, Peter E. Rossi, and George Tauchen, 1992, Stock prices and volume, *Review of Financial Studies* 5(2), 199-242.
- [68] Gârleanu, Nicolae, and Lasse Heje Pedersen, 2011, Margin-based asset pricing and deviations from the law of one price, *Review of Financial Studies* 24(6), 1980-2022.
- [69] Gertler, Mark, and Nobuhiro Kiyotaki, 2010, Financial intermediation and credit policy in business cycle analysis, *Handbook of Monetary Economics* 3(3), 547-599.
- [70] Giroud, Xavier, Holger M. Mueller, Alex Stomper, and Arne Westerkamp, 2012, Snow and leverage, *Review of Financial Studies* 25(3), 680-710.
- [71] Goetzmann, William N., Dasol Kim, Alok Kumar, and Qin Wang, 2014, Weather-induced mood, institutional investors, and stock returns, *Review of Financial Studies*, forthcoming.
- [72] Goldman Sachs, 2008, Goldman Sachs hedge fund trend monitor, February 20, 2008.
- [73] Green, Jeremiah, John R.M. Hand, and Frank Zhang, 2016, The Characteristics that Provide Independent Information about Average U.S. Monthly Stock



Returns, *Review of Financial Studies*, forthcoming.

- [74] Greenwood, Robin, 2005a, Excess comovement of stock returns: Evidence from cross-sectional variation in Nikkei 225 weights, *Review of Financial Studies* 21(3), 1153-1186.
- [75] Greenwood, Robin, 2005b, Short-and long-term demand curves for stocks: theory and evidence on the dynamics of arbitrage, *Journal of Financial Economics* 75(3), 607-649.
- [76] Gromb, Denis, and Dimitri Vayanos, 2002, Equilibrium and welfare in markets with financially constrained arbitrageurs, *Journal of Financial Economics* 66(2), 361-407.
- [77] Gromb, Denis, and Dimitri Vayanos, 2009, Financially constrained arbitrage and cross-market contagion, working paper.
- [78] Grossman, Sanford J., and Merton H. Miller, 1988, Liquidity and market structure, *Journal of Finance* 43(3), 617-633.
- [79] Grossman, Sanford J., and Joseph E. Stiglitz, 1976, Information and competitive price systems, *American Economic Review* 66(2), 246-253.
- [80] Hansen, Lars Peter, 1982, Large sample properties of generalized method of moments estimators, *Econometrica* 50(4), 1029-1054.
- [81] Hanson, Samuel G., and Adi Sunderam, 2014, The growth and limits of arbitrage: Evidence from short interest, *Review of Financial Studies* 27(4), 1238-1286.
- [82] Harris, Lawrence, and Eitan Gurel, 1986, Price and volume effects associated with changes in the S&P 500 list: New evidence for the existence of price pressures, *Journal of Finance* 41(4), 815-829.
- [83] He, Zhiguo, and Arvind Krishnamurthy, 2012, A model of capital and crises, *Review of Economic Studies* 79(2), 735-777.
- [84] He, Zhiguo, and Arvind Krishnamurthy, 2013, Intermediary asset pricing, *American Economic Review* 103(2), 732-770.
- [85] He, Zhiguo, Bryan Kelly, and Asaf Manela, 2016, Intermediary asset pricing: New evidence from many asset classes, *National Bureau of Economic*

Research, working paper.

- [86] Hellwig, Martin F., 1980, On the aggregation of information in competitive markets, *Journal of Economic Theory* 22(3), 477-498.
- [87] Hirshleifer, David, and Tyler Shumway, 2003, Good day sunshine: stock returns and the weather, *Journal of Finance* 58(3), 1009-1032.
- [88] Hirshleifer, David, Siew Hong Teoh, and Jeff Jiewei Yu, 2011, Short arbitrage, return asymmetry, and the accrual anomaly, *Review of Financial Studies* 24(7), 2429-2461.
- [89] Hong, Harrison, Weikai Li, Sophie X. Ni, Jose Scheinkman, and Philip Yan, 2015, Days to cover and stock returns, National Bureau of Economic Research, working paper.
- [90] Hong, Harrison, and Jialin Yu, 2009, Gone fishin': Seasonality in trading activity and asset prices, *Journal of Financial Markets* 12(4), 672-702.
- [91] Huang, Shiyang, Dong Lou, and Christopher Polk, 2016, The booms and busts of beta arbitrage, working paper.
- [92] Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *Journal of finance* 48(1), 65-91.
- [93] Kamstra, Mark J., Lisa A. Kramer, and Maurice D. Levi, 2003, Winter blues: A SAD stock market cycle, *American Economic Review* 93(1), 324-343.
- [94] Kelly, Bryan, and Hao Jiang, 2014, Tail risk and asset prices, *Review of Financial Studies* 27(10), 2841-2871.
- [95] Koijen, Ralph SJ, and Motohiro Yogo, 2016, An equilibrium model of institutional demand and asset prices, working paper.
- [96] Korajczyk, Robert A., and Ronnie Sadka, 2004, Are momentum profits robust to trading costs? *Journal of Finance* 59(3), 1039-1082.
- [97] Koudijs, Peter, 2015, The boats that did not sail: Asset price volatility in a natural experiment, *Journal of Finance*, accepted online manuscript.

- [98] Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh, 2015, Interpreting factor models, working paper.
- [99] Krishnamurthy, Arvind, 2002, The bond/old-bond spread, *Journal of Financial Economics* 66(2), 463-506.
- [100] Kyle, Albert S., and Wei Xiong, 2001, Contagion as a wealth effect, *Journal of Finance* 56(4), 1401-1440.
- [101] Lakonishok, Josef, Andrei Shleifer, and Robert W. Vishny, 1994, Contrarian investment, extrapolation, and risk, *Journal of Finance* 49(5), 1541-1578.
- [102] Lamont, Owen A., and Richard H. Thaler, 2003, Anomalies: The law of one price in financial markets, *Journal of Economic Perspectives* 17(4), 191-202.
- [103] Lehman, Bruce N, 1990, Fads, martingales, and market efficiency, *Quarterly Journal of Economics* 105(1), 1-28.
- [104] Liu, Qi, Lei Lu, Bo Sun, and Hongjun Yan, 2015, A model of anomaly discovery, working paper.
- [105] Lo, Andrew W., and A. Craig MacKinlay, 1990, When are contrarian profits due to stock market overreaction?, *Review of Financial studies* 3(2), 175-205.
- [106] Lochstoer, Lars A., and Paul C. Tetlock, 2016, What drives anomaly returns? working paper.
- [107] Longcore, Travis R., and Peter W. Rees, 1996, Information technology and downtown restructuring: The case of New York City's financial district, *Urban Geography* 17(4), 354-372.
- [108] Lonnqvist, Olov, 1962, On the diurnal variation of surface temperature, *Tellus* 14(1), 96-101.
- [109] McLean, R. David, and Jeffrey Pontiff, 2016, Does academic research destroy stock return predictability? *Journal of Finance* 71(1), 5-32.
- [110] Mertens, Karel, and Morten O. Ravn, 2014, A reconciliation of SVAR and narrative estimates of tax multipliers, *Journal of Monetary Economics* 68(S), S1-S19.

- [111] Merton, Robert C., 1973, An intertemporal capital asset pricing model, *Econometrica* 41(5), 867-887.
- [112] Mitchell, Mark, Todd Pulvino, and Erik Stafford, 2002, Limited arbitrage in equity markets, *Journal of Finance* 57(2), 551-584.
- [113] Morris, Stephen, and Hyun Song Shin, 2004, Liquidity black holes, *Review of Finance* 8(1), 1-18.
- [114] Mountford, Andrew, and Harald Uhlig, 2009, What are the effects of fiscal policy Shocks? *Journal of Applied Econometrics* 24(6), 960–992.
- [115] Mulligan, Casey B., and Andrei Shleifer, 2005, The extent of the market and the supply of regulation, *Quarterly Journal of Economics* 120(4), 1445-1473.
- [116] Nagel, Stefan, 2012, Evaporating liquidity, *Review of Financial Studies* 25(7), 2005-2039.
- [117] Novy-Marx, Robert, and Mihail Velikov, 2016, A taxonomy of anomalies and their trading costs, *Review of Financial Studies* 29(1), 104-147.
- [118] Omar Al-Ubaydli and Patrick A. McLaughlin, 2015, RegData: A numerical database on industry-specific regulations for all US industries and federal regulations, 1997–2012, working paper.
- [119] Pastor, Lubos and Robert F. Stambaugh, 2003, Liquidity risk and expected stock returns, *Journal of Political Economy*, 111(3), 642-685.
- [120] Perotti, Roberto. The effects of tax shocks on output: not so large, but not small either, *American Economic Journal: Economic Policy* 4(2), 214-237.
- [121] Pinegar, Michael J., 2002, Losing sleep at the market: Comment, *American Economic Review* 92(4), 1251-1256.
- [122] Ramey, Valerie A, 2011, Can government purchases stimulate the economy? *Journal of Economic Literature* 49(3), 673-685.
- [123] Ramey, Valerie A., and Matthew D. Shapiro, 1998, Costly capital reallocation and the effects of government spending, *Carnegie-Rochester Conference Series on Public Policy* 48, North-Holland.

- [124] Romer, Christina D., and David H. Romer, 2010, The macroeconomic effects of tax changes: estimates based on a new measure of fiscal shocks. *American Economic Review* 100(3), 763-801.
- [125] Rosenberg, Barr, Kenneth Reid, and Ronald Lanstein, 1985, Persuasive evidence of market inefficiency, *Journal of Portfolio Management*, 11(3), 9-16.
- [126] Rosenthal, Leonard, and Colin Young, 1990, The seemingly anomalous price behavior of Royal Dutch/Shell and Unilever NV/PLC, *Journal of Financial Economics* 26(1), 123-141.
- [127] Saunders, Edward M., 1993, Stock prices and wall street weather, *American Economic Review* 83(5), 1337-1345.
- [128] Scholes, Myron, and Joseph Williams, 1977, Estimating betas from nonsynchronous data, *Journal of Financial Economics* 5(3), 309-327.
- [129] Schwert, G. William, 1990, Indexes of United States stock prices from 1802 to 1987, *Journal of Business* 63(3), 399-426. <http://schwert.ssb.rochester.edu/dstock.htm> (accessed July 2013).
- [130] Selden, George Charles, 1919, *The machinery of Wall Street: why it exists, how it works and what it accomplishes*, New York: The Magazine of Wall street.
- [131] Seppanen, Olli, William J. Fisk, and Q. H. Lei, 2006, Effect of temperature on task performance in office environment, Lawrence Berkeley National Laboratory, working paper.
- [132] Shanken, Jay, 1992, On the estimation of beta-pricing models, *Review of Financial studies* 5(1), 1-33.
- [133] Shleifer, Andrei, 1986, Do demand curves for stocks slope down?, *Journal of Finance* 41(3), 579-590.
- [134] Shleifer, Andrei, 2000, *Inefficient markets: An introduction to behavioral finance*, Oxford University Press.
- [135] Shleifer, Andrei, and Robert W. Vishny, 1997. The limits of arbitrage, *Journal of Finance* 52(1), 35-55.
- [136] Stein, Jeremy C., 2009, Presidential address: Sophisticated investors and

market efficiency. *Journal of Finance* 64(4), 1517-1548.

- [137] Tilghman, Charles R., 1961, Telegraph history: Some early days of Western Union's stock ticker service 1871-1910, *Western Union Technical Review* 15(2), 75-79.
- [138] U.S. Census Bureau, 1975, Historical statistics of the United States, colonial times to 1970, Bicentennial Edition, Part 2, Washington, D.C.: U.S. Government Printing Office. <http://www2.census.gov/prod2/statcomp/documents/CT1970p2-05.pdf>.
- [139] Verrecchia, Robert E., 1982, Information acquisition in a noisy rational expectations economy, *Econometrica* 50(6), 1415-1430.
- [140] Vuolteenaho, Tuomo, 2002, What Drives Firm-Level Stock Returns? *Journal of Finance* 57(1), 233-264.
- [141] Wang, Jiang, 1994, A model of competitive stock trading volume, *Journal of Political Economy* 102(1), 127-168.
- [142] Warga, Arthur, 1992, Bond returns, liquidity, and missing data, *Journal of Financial and Quantitative Analysis* 27(4), 605-617.
- [143] Zhang, Lu, 2005, The value premium, *Journal of Finance* 60(1), 67-103.