Short-sale constraints and the market portfolio∗

Adem Dugalic† Diego Torres Patiño‡
Stanford University

February 1, 2018

Abstract

We study how short-sale constraints on stock lending affect asset prices in an equilibrium model with multiple assets. We endow investors with heterogeneous beliefs in order to generate short selling demand. We obtain a CAPM-like equation that links asset-specific excess returns with the market equity premium. In the presence of short selling constraints in the market, the model gives rise to asset-specific alphas that are explained by both asset-specific and market-wide short-sale constraints; unconstrained stocks have higher risk-adjusted expected returns relative to the market portfolio, whereas the opposite holds for constrained stocks. In the absence of short-sale constraints, the model reduces to the standard CAPM. We test the model using extensive data on short interest and borrow fees. The model is able to empirically explain asset prices for 10 borrow-fee-sorted portfolios, as opposed to CAPM and factor models which produce unexplained alphas that are significantly different from zero for some low and high borrow fee portfolios.

∗We are extremely grateful to Jonathan Berk, Shai Bernstein, Nick Bloom, Svetlana Bryzgalova, and Isaac Sorkin for helpful conversations and suggestions, as well as to Kevin Mak for practical insights on the institutional details of the short selling market. All errors are ours.
†Department of Economics, Stanford University, 579 Serra Mall, Stanford, CA 94305. Email: adugalic@stanford.edu. Website: http://www.stanford.edu/~adugalic.
‡Department of Economics, Stanford University, 579 Serra Mall, Stanford, CA 94305. Email: dtorresp@stanford.edu.
1 Introduction

Classical asset pricing models, such as the capital asset pricing model (CAPM), assume a frictionless market with homogeneous investors. Extensions to the model as well as other more recent theories have acknowledged that investors are heterogeneous and that the efficiency of financial markets depends on the ability of investors to short sell securities believed to be overvalued. In a very influential paper, Miller (1977) has argued that in the presence of investors with heterogeneous beliefs about an asset return and with short selling constraints, the asset will be overpriced because it cannot be sold (short) by investors with pessimistic beliefs, and as a consequence the marginal investor will be an optimist. In this paper we study, both theoretically and empirically, the effect of short selling constraints on the cross section of the U.S. stock prices. On the theoretical side, we propose a model that extends the standard CAPM model by incorporating short selling constraints on the lending side of the market. On the empirical side, we use an extensive database on short interest and borrow fees covering over 95% of U.S. public equities to first demonstrate that standard models, such as the CAPM and multi-factor models, underprice short selling constrained stocks (i.e. stocks with high borrow fees should have lower risk-adjusted returns than predicted by the models) and overprice non-constrained stocks (i.e. stocks with borrow fees close to zero should have higher risk-adjusted returns than predicted by the models), and then to assess the performance of our model in explaining the cross section of asset prices. We show that our model, which accounts for short selling constraints, reduces the underpricing of short selling constrained stocks: unexplained residual negative returns are reduced in absolute value by about one half, and also cease to be statistically significant. In addition, the model corrects for the overpricing of non-constrained stocks, which constitute an overwhelming majority of stocks, by reducing the unexplained residual positive returns by about 50 basis points per year.

Our model builds upon a standard CAPM with multiple assets and a continuum of investors by endowing investors with heterogeneous beliefs about future asset payoffs, which can create a demand for short selling by pessimistic investors. In order to short sell an asset, an investor must first borrow the asset from another investor who holds the asset and has not lent it yet, at which point the borrower delivers a cash collateral to the lender in addition to any borrow fees.¹ In the next period the borrower returns the asset and

¹The borrow fee need not be positive since the collateral itself is valuable as it can earn riskless return.
gets back the cash collateral. We model short selling constraints on the securities lending side of the market by assuming that an investor who goes long in a particular asset can only lend up to a fixed fraction of its holdings of the asset, where the fraction depends on and represents various types of short selling constraints specific to that asset at a given point in time. This assumption, for instance, proxies for the decentralized nature of the market for lending securities: the shares offered for lending by asset holders are not always located by potential borrowers, as matching between lenders and borrowers is costly and not instantaneous. In addition, many investors are restricted from lending their shares due to various reasons. If in an equilibrium the number of shares lent by an investor who is long in an asset is smaller than the fraction representing the constraint, it implies that short selling is unconstrained as borrowers are not borrowing the entire supply of the located lendable shares and consequently borrow fees are near zero. If the number of shares lent is such that the constraint is binding, it implies that all the located shares are lent, and borrow fees increase until the market is cleared (i.e. the borrowing demand equals the supply of the shares available for lending). When the model is taken to the data, short interest can be used to identify the magnitude of the constraint for any asset with positive cost of borrowing (net of brokerage fees).

The model gives rise to two equations that jointly determine equilibrium borrow fees and asset prices. It is possible to obtain one equation that links equilibrium borrow fees and asset prices free of the variable representing the amount of disagreement among investors, which is difficult to quantify and measure. The equation is testable as all of its components are quantifiable and can be constructed from the variables in our datasets. We express the equation in the CAPM style, linking asset-specific excess returns with the market equity premium. Short selling constraints give rise to asset-specific alphas that can be decomposed into two parts: first, the market-wide alpha, which takes into account the borrow fees and short interest of all securities in the market and is proportional to the asset-specific beta;\(^2\) and second, the asset-specific alpha, which is negative and decreasing in the asset-specific borrow fee and short interest.\(^3\) The intuition behind the “part 2” component is as follows. First, higher fees reduce short selling demand but also encourage going long in the asset because potential income from lending the asset is higher. Both of these forces put an

\(^2\)We refer to this component as “part 1”.

\(^3\)We refer to this component as “part 2”.

---

The negative of the borrow fee is known as the rebate rate and is the variable of choice in studies focused on the repo markets.
upward pressure on the asset price. Second, for a constrained asset, higher short interest indicates that the asset is easier to lend, which also encourages going long in the asset, again causing a rise in the asset price. The “part 1” component is more subtle and overlooked in the literature: constrained short selling leads to overvaluation of the market portfolio, implying that unconstrained stocks (with positive beta) will produce higher risk-adjusted returns than what is predicted by the standard models.

We confirm the predictions of the model in the data. Following the literature, we begin by sorting stocks by borrow fees and grouping them into 10 portfolios. We show that standard models overprice the highest-borrow-fee portfolios while underpricing the unconstrained portfolios. Next, we estimate the model-implied alphas, and show that they account for most of the mispricing produced by the standard models. Finally, we decompose the alphas into “part 1” and “part 2” and show that each plays the role implied by the model: the “part 1” alpha corrects for the overpricing of the unconstrained portfolios, while “part 2” corrects for the underpricing of the constrained portfolios.

This paper is related to several strands of literature. On the theoretical side, we build upon the capital asset pricing model, introduced in Treynor (1961), Sharpe (1964), Lintner (1965a,b), and Mossin (1966). Short selling constraints are a part of a large literature on limits to arbitrage and various institutional constraints that prevent asset prices from reaching an efficient level, started by Duffie (1996) and Shleifer and Vishny (1997) and recently surveyed in Gromb and Vayanos (2010). The analysis of short selling frictions in models with investors endowed with heterogeneous beliefs is explored in Harrison and Kreps (1978), Scheinkman and Xiong (2003), Hong and Stein (2003), Hong et al. (2006), Geanakoplos (2010), and Simsek (2013), among others. Duffie et al. (2002) study the effect of the over-the-counter nature of the market for lending securities on borrow fees. We complement the theoretical literature by proposing a model that allows for a quantitative assessment of the effect of the short selling frictions on the lending side of the market.

On the empirical side, a majority of studies used datasets obtained from a single institution participating in the stock loan market and are thus limited in terms of their coverage of the market. Instead, this paper uses a proprietary dataset covering most of the market that recently became available. Most relevant to this paper is a working paper by Drechsler and Drechsler (2014); they use this dataset with a sample running from 2004 to 2013 but focus on the frictions on the borrowing side of the market, which limit some of the market
participants from short selling stocks and thus imply that the shorting premium is short sellers’ compensation for the concentrated risk they bear in shorting stocks that they find overpriced. While this story may play some role in explaining the overpricing due to short selling constraints, it does not explain what gives rise to borrow fees, which are taken as exogenous in their model; in addition, they do not test the model itself, as we do in our paper. Among other relevant empirical work, D’avolio (2002) and Usher (2015) describe the market for borrowing and lending U.S. equities, and Geczy et al. (2002); Asquith et al. (2005); Cohen et al. (2007); Blocher et al. (2013), and Porras Prado et al. (2016) show that high borrow fee stocks have lower returns.

Layout. The rest of the paper is organized as follows. Section 2 describes the market for borrowing and lending U.S. equities. Section 3 lays out the theoretical model. Section 4 describes the data sources and provides the descriptive statistics. Section 5 presents the empirical strategy and main results. Section 6 concludes.

2 Institutional Background

In this section, we briefly discuss the role of borrow fees in the short selling market; for a comprehensive coverage on the institutional background regarding securities lending, please see D’avolio (2002) and Usher (2015).

Demand in the securities lending market—securities borrowing—comes from hedge funds and other investment funds. Directional long/short funds and prominent short sellers willing to pay high prices to short specific stocks are usually involved in cases of high borrow fees, but other players with hedging needs (to hedge long positions in their portfolio) or stock delivery obligations also participate in the low borrow fee markets. Supply—securities lending—mostly comes from mutual funds, index funds, pension funds, sovereign wealth funds, and insurers; for retail accounts to be able to lend securities, they must first explicitly inform their brokers of their desire to do so. Lenders earn extra returns from the borrow fees while keeping their long position (Blocher and Whaley (2016) show that ETF managers slant their holdings towards stocks with higher lending fees). Lenders also participate in the market for cash financing purposes, since borrowers must post collateral; however, this is more the case in repo markets. Brokers participate on both sides of the market (for a commission); they manage the inventory of shares available for lending,
locate the requested shares, execute the transactions, and manage the collateral.

Upon origination, a borrower posts collateral and the lender transfers the stock certificate; note that it is the lender (and subsequently anyone who purchases the stock from the lender) who will have voting rights and earn any distributions. While the stock loan is active, in order to replicate true long and short cash positions for the lender and borrower, the borrower must replicate all distributions, including dividends and cash or stock distributions, to the lender, in addition to paying the borrow fees (which can vary on a daily basis but are usually paid monthly). Collateral adjustments are also made as the stock price changes to keep the margin constant. The stock loan can be terminated at any time at the request of either party; the borrower must return the stock certificate to the lender, who returns the collateral plus the rebate (the interest earned on the collateral at the risk-free rate). See Figure 1 for an illustration of a typical transaction.

\[ \text{Figure 1: A Securities Lending Transaction} \]

Note: Based on Usher (2015).

Central to our paper is the question regarding the existence of borrow fees. In the absence of frictions, 100% of the shares are initially available for lending, and any time a share is lent and shorted, it can be lent again by the new owner of the stock certificate. Thus, 100% of the shares should be available for lending at any time. This would result in excess supply and a borrow fee of zero\(^4\)—or negative, since interest (the rebate) must be

\(^4\)Other than tax implications from manufactured dividends taxed at a different rate than actual divi-
paid on the collateral.

In our paper, we focus on frictions in the lending side of the market. Many investors are restricted from lending their shares. This group includes insiders (who can be prohibited from lending their shares by their company’s policy), IPO investors during the lockup period, cash accounts that do not give their brokers explicit permission to do so, and some funds as per their directives. In addition, some insiders and investors are not willing to lend them to short sellers in order to avoid downward pressure on stock prices; Lamont (2012) shows that firms will use a variety of strategies in this regard, which does result in their stock being overpriced, although the overpricing eventually gets corrected. Some investors do not lend them because of counterparty risk (but may lend them for a higher fee); this includes institutions with possibly uncertain but time-sensitive stock delivery obligations who cannot risk delays in receiving their stock back whenever they recall it and any investor worried about the counterparty not returning the borrowed securities at all outweighing the lending income benefit. Brokers themselves may choose to keep some lendable inventory out of the market in case an investor recalls their shares unexpectedly. Finally, small investors cannot lend them even if they want to because the transaction costs of locating their shares are too large: the broker’s time may be more valuable than the small spread they may earn on the borrow fees, and many counterparties are only interested in shorting a large lot of shares from a single or a limited number of parties, since this reduces recall risk (in a binary sense at least, and thus the expected number of transactions needed to keep a short position active). Thus, after some investors lend their shares, 100% of the float may end up being restricted from lending.

Another issue to keep in mind when analyzing sudden changes in borrow fees is that stock transactions settle two days after the transaction date (t+2), so if the lendable supply is used up in less than two days, it can’t be shorted again until after the transactions settle. This delays stock rehypothecation and may cause the supply constraints to bind even in the absence of any of the above circumstances. Before September 2017, transactions settled in three days\textsuperscript{5}, so we can expect this issue to be much less frequent in the future.

A high borrow fee may be the result of any of these constraints combined with high borrowing demand for the stock. The latter may happen if there is a great disparity of opinions regarding how much a stock should be worth. Non-public information can easily

lead to a large disparity of opinions, whether it is negative private or inside information regarding that stock or positive information regarding a comparable stock (so the stock is shorted for hedging purposes). Finally, the existence of arbitrage opportunities, such as merger or acquisition rumors and related securities such as convertible debt and options, is also conducive to high borrowing demand. A high borrow fee will then incentivize long investors to make their shares available for lending and reduce short selling demand until an equilibrium in the securities lending market is reached.

3 Model

In this section we study how short selling constraints affect asset borrow fees and prices in an equilibrium model with multiple assets. We endow investors with heterogeneous beliefs about future asset payoffs, which can create a demand for short selling by pessimistic investors. Short selling constraints are imposed on the security lending side of the market. This means that an investor who goes long in a particular asset can lend up to a fraction \( \phi \) of its holdings of the asset, where \( \phi \) represents various types of exogenous short selling constraints specific to that asset at a given point in time, as described in the previous section. The model implies a CAPM-like equation that links specific asset excess returns with the market equity premium. Short selling constraints give rise to asset-specific alphas that depend on asset-specific and market-wide short-sale constraints.

3.1 Setup

Consider an overlapping-generations (OLG) economy in discrete time with a unit mass of investors of type \( j \in \{ A, B \} \) born at time \( t \) with wealth \( W_{y,j,t} \) (where \( y \) stands for “young”) who live for 2 periods. Investors trade risky securities \( n = 1, ..., N \), assumed without loss of generality to be in unit supply. Let \( \delta_t \) denote the dividend vector, and \( P_t \) denote the ex-post price vector for risky securities at time \( t \). Let \( \mathbb{E}_t(P_{t+1} + \delta_{t+1}) = \mu_t \) and \( \text{Var}_t(P_{t+1} + \delta_{t+1}) = \Sigma_t \). Investors of type \( A \) believe that \( \mathbb{E}_{A,t}(P_{t+1} + \delta_{t+1}) = \mu_t + \epsilon_t \) and investors of type \( B \) believe that \( \mathbb{E}_{B,t}(P_{t+1} + \delta_{t+1}) = \mu_t - \epsilon_t \), where \( \epsilon_t \in \mathbb{R}^N \). Both types of investors agree on \( \text{Var}_t(P_{t+1} + \delta_{t+1}) = \text{Var}_{A,t}(P_{t+1} + \delta_{t+1}) = \text{Var}_{B,t}(P_{t+1} + \delta_{t+1}) = \Sigma_t \). Thus, \( |\epsilon_t^{(n)}| \) denotes the amount of disagreement about asset \( n \) at time \( t \); we impose symmetry in order to be able to test the model in the data, since \( |\epsilon_t^{(n)}| \) is hard to estimate.
There is a market for borrowing and lending risky securities. Let $f_t$ denote the vector of borrow fees at time $t$, meaning that for each unit of asset $n$ borrowed, the borrower pays fraction $f_t^{(n)}$ of the price $P_t^{(n)}$ to the lender at time $t$. In addition, the borrower puts $m P_t^{(n)}$ amount of cash as a collateral for every unit of stock $n$ borrowed.\footnote{In practice, $m = 1.02.$} We further assume that each investor who goes long in asset $n$ at time $t$ can lend to up to $\phi_t^{(n)} \in [0, 1]$ units of asset $n$ held, were $\phi_t^{(n)}$ represents various types of short selling constraints specific to asset $n$ at time $t$. In addition to the risky securities, there is a one-period risk-free asset in infinite supply, a unit of which costs $Q_t$ at time $t$ and pays 1 unit of wealth at time $t + 1$.

A young investor of type $j \in \{A, B\}$ at time $t$ chooses a portfolio consisting of shares of risky assets, $x_j$, risk-free bonds, $b_j$, and shares to lend, $z_j$, in order to maximize the mean-variance utility of its time $t + 1$ wealth $W_{j,t+1}$, taking prices and fees as given. In particular,

$$\max_{x_j, b_j, z_j} \mathbb{E}_j[W_{j,t+1}] - \frac{\gamma}{2} \text{Var}_t[W_{j,t+1}]$$

s.t. $W_{j,t+1} = x_j' (P_{t+1} + \delta_{t+1}) + b_j - m \min(z_j, x_j)' P_t$

$$W_{j,t} = x_j' P_t + b_j Q_t - \min(z_j, x_j)' \text{diag} \left[m + f_t^{(n)}\right] P_t$$

$$0 \leq z_j^{(n)} \leq \max \left(0, \phi_t^{(n)} x_j^{(n)} \right)$$

, for $n = 1, \ldots, N$  \hspace{1cm} (4)

Equation (2) gives the expression for the wealth of an old investor of type $j$ at time $t + 1$ as a function of the investor’s portfolio holding, while Equation (3) is the budget constraint of the same investor at time $t$. The first term on the right-hand side of each equation corresponds to the holdings of risky assets, the second term corresponds to the position in risk-free bonds, and the third term corresponds to the collateral. To understand the third term, first suppose that a young investor of type $j$ at time $t$ shorts asset $n$, i.e. $x_j^{(n)} < 0$. Then the investor must pay the total fee resulting from such a position in asset $n$ in the amount of $x_j^{(n)} f_t^{(n)} P_t^{(n)}$, in addition to posting collateral in the amount of $x_j^{(n)} m P_t^{(n)}$. Each of the two terms reduces the amount of time $t$ wealth spent on the available assets, explaining the term in Equation (3). Furthermore, the investor receives the collateral back at time $t + 1$, which explains the term in Equation (2). Note that in that case $z_j^{(n)} = 0$ since the investor cannot lend the asset that he does not own. If the investor on the other hand goes long in asset $n$, i.e. $x_j^{(n)} > 0$, then he may lend some of the shares. In that case
\( z_j^{(n)} > 0 \), and the cash flows due to the collateral go in the direction opposite to that of the first case. Equation (4) is the collateral constraint: an investor who goes long in asset \( n \), i.e. \( x_j^{(n)} > 0 \), cannot lend more than fraction \( \phi_i^{(n)} \) of his holdings of asset \( n \).

A competitive equilibrium consists of a sequence of prices \( P_t \) and \( Q_t \), fees \( f_t \), and allocations \((x_j,t, b_j,t, z_j,t)\) for \( j \in \{A, B\} \) such that, given prices and fees, allocation \((x_j,t, b_j,t, z_j,t)\) solves the problem of a young investor of type \( j \in \{A, B\} \) at time \( t \) and markets clear:

\[
\frac{1}{2} x_{A,t} + \frac{1}{2} x_{B,t} = 1
\]

\[
\min(z_{A,t}, x_{A,t}) + \min(z_{B,t}, x_{B,t}) = 0
\]

Equation (5) is the market clearing condition for the market for buying and selling risky assets, while Equation (6) is an equivalent condition for the market for lending and borrowing risky assets.

### 3.2 Results

**Equilibrium Characterization**

After substituting out \( b_j \) and \( W_{j,t+1} \), the problem of young investor of type \( j \) at time \( t \) simplifies to

\[
\max_{x_j, z_j} x_j' \left[ \mathbb{E}_{j,t}(P_{t+1} + \delta_{t+1}) - R_{t+1}^f P_t \right] - \frac{\gamma}{2} x_j' \Sigma_t x_j
\]

\[
+ R_{t+1}^f (z_j + \min(0, x_j))' \text{diag} \left[ \left( 1 - \frac{1}{R_{t+1}^f} \right) m + f_t^{(n)} \right] P_t
\]

s.t. \( 0 \leq z_j^{(n)} \leq \max \left( 0, \phi_t^{(n)} x_j^{(n)} \right) \)

where \( R_{t+1}^f \equiv \frac{1}{Q_t} \) is the risk-free return from period \( t \) to \( t + 1 \). The first-order conditions with respect \( x_j \) and \( z_j \) are

\[
x_j : \quad 0 = \mathbb{E}_{j,t}(P_{t+1} + \delta_{t+1}) - R_{t+1}^f P_t - \gamma \Sigma_t x_j + \text{diag} \left[ \left( x_j^{(n)} \geq 0 \right) \phi_t^{(n)} \right] \lambda_{j,t}
\]

\[
+ R_{t+1}^f \text{diag} \left[ 1 \left( x_j^{(n)} < 0 \right) \left( \left( 1 - \frac{1}{R_{t+1}^f} \right) m + f_t^{(n)} \right) \right] P_t
\]

\[
z_j : \quad 0 = R_{t+1}^f \text{diag} \left[ 1 \left( x_j^{(n)} \geq 0 \right) \left( \left( 1 - \frac{1}{R_{t+1}^f} \right) m + f_t^{(n)} \right) \right] P_t - \lambda_{j,t} + \eta_{j,t}
\]
where \( \lambda_{j,t} \) and \( \eta_{j,t} \) are vectors of Lagrange multipliers on the collateral and non-negativity lending constraints, respectively. Observe that if \( \left(1 - \frac{1}{R_{t+1}}\right) m + f_t^{(n)} = 0 \), then Equation (8) implies that \( \lambda_{j,t}^{(n)} = 0 \) and \( \eta_{j,t}^{(n)} = 0 \) and the last two terms on the right-hand side of the \( n \)th row of Equation (7) drop out. If \( \left(1 - \frac{1}{R_{t+1}}\right) m + f_t^{(n)} < 0 \), then Equation (8) implies that \( \lambda_{j,t}^{(n)} = 0 \) and \( \eta_{j,t}^{(n)} > 0 \), implying \( z_j^{(n)} = 0 \). Thus, in equilibrium nobody can short asset \( n \), so the last two terms on the right-hand side of the \( n \)th row of Equation (7) drop out. Thus, the last two terms on the right-hand side of Equation (7) matter if and only if \( \left(1 - \frac{1}{R_{t+1}}\right) m + f_t^{(n)} > 0 \), i.e. \( f^{(n)} > -m \left(1 - \frac{1}{R_{t+1}}\right) \). In that case, some agents must short asset \( n \) in equilibrium because \( z_j^{(n)} > 0 \) whenever \( x_j^{(n)} > 0 \). Note that in that case we can link \( \phi_t^{(n)} \) to the short interest for asset \( n \). In particular, if \( f^{(n)} > -m \left(1 - \frac{1}{R_{t+1}}\right) \) and letting \( j \) denote the type of investor who shorts asset \( n \) in equilibrium, then

\[
\phi_t^{(n)} = -\frac{x_t^{(n)}}{x_t^{(n)}} = -\frac{\frac{1}{2}x_t^{(n)}}{1 - \frac{1}{2}x_t^{(n)}} = \frac{s_t^{(n)}}{1 + s_t^{(n)}}
\]

where \( s_t^{(n)} \equiv -\frac{1}{2}x_t^{(n)} \) is the short interest for asset \( n \).

Let \( \iota_t^{(n)} = 1 \) if \( f^{(n)} > -m \left(1 - \frac{1}{R_{t+1}}\right) \) (i.e. if the short selling constraint on asset \( n \) binds) and \( \iota_t^{(n)} = 0 \) otherwise. Using Equation (8) to substitute the multipliers out in Equation (7) (and using the fact that if \( \lambda_{j,t}^{(n)} > 0 \) then \( \eta_{j,t}^{(n)} = 0 \)), we have that

\[
0 = \mathbb{E}_{j,t}(P_{t+1} + \delta_{t+1}) - R_{t+1}^f P_t - \gamma \Sigma_t \varepsilon_j
+ R_{t+1}^f \text{diag} \left[ \iota_t^{(n)} \left(1 - (1 - \phi_t^{(n)}) \mathbb{1} \left(x_j^{(n)} \geq 0\right)\right) \left(1 - \frac{1}{R_{t+1}^f}\right) m + f_t^{(n)} \right] P_t
\]  

(9)

Taking the difference and sum of Equation (9) between the two types of investors, we obtain the following equations, respectively:

\[
0 = \varepsilon_t - \frac{1}{2} \gamma \Sigma_t \varepsilon_j (x_{A,t} - x_{B,t}) - \frac{1}{2} R_{t+1}^f \text{diag} \left[ \iota_t^{(n)} \left(1 - \phi_t^{(n)}\right) \left(2 \cdot \mathbb{1} \left(x_j^{(n)} \geq 0\right) - 1\right) \left(1 - \frac{1}{R_{t+1}^f}\right) m + f_t^{(n)} \right] P_t
\]

\[
0 = \mathbb{E}_t(P_{t+1} + \delta_{t+1}) - \gamma \Sigma_t 1 - R_{t+1}^f \left(I - \frac{1}{2} \text{diag} \left[ \iota_t^{(n)} \left(1 + \phi_t^{(n)}\right) \left(1 - \frac{1}{R_{t+1}^f}\right) m + f_t^{(n)} \right] \right) P_t
\]

where in the last equation we applied the equilibrium condition \( \frac{1}{2} (x_A + x_B) = 1 \). Observe also that if \( \iota_t^{(n)} > 0 \), then \( \frac{1}{2} \left(x_j^{(n)} - x_j^{(n)}\right) = \frac{1 + \phi_t^{(n)}}{1 - \phi_t^{(n)}} \left(2 \cdot \mathbb{1} \left(x_j^{(n)} \geq 0\right) - 1\right) \). If \( \iota_t^{(n)} = 0 \), then we can without loss of generality assume that inequality \( f_t^{(n)} \leq -m \left(1 - \frac{1}{R_{t+1}^f}\right) \) in fact
holds with equality (since the last term in the two equations above drops out anyways).

Defining \( \hat{x}_t \equiv \frac{1}{2}(x_{A,t} - x_{B,t}) \), we summarize the equilibrium conditions for the prices \( P_t \) and borrow fees \( f_t \) in the following theorem:

**Theorem 1.** The equilibrium prices \( P_t \) and borrow fees \( f_t \) satisfy the following set of conditions:

\[
\mathbb{E}_t(P_{t+1} + \delta_{t+1}) - R_{t+1}^f P_t = \gamma \Sigma_t \mathbf{1} - \frac{1}{2} R_{t+1}^f \text{diag} \left( \left( 1 + \phi_t^{(n)} \right) \left( 1 - \frac{1}{R_{t+1}^f} \right) m + f_t^{(n)} \right) P_t \tag{10}
\]

\[
e_t = \gamma \Sigma_t \hat{x}_t + \frac{1}{2} R_{t+1}^f \text{diag} \left( \left( 1 - \phi_t^{(n)} \right) \left( 2 \cdot 1 \left( x_{A}^{(n)} \geq 0 \right) - 1 \right) \left( 1 - \frac{1}{R_{t+1}^f} \right) m + f_t^{(n)} \right) P_t \tag{11}
\]

\[
f_t^{(n)} \geq - \left( 1 - \frac{1}{R_{t+1}^f} \right) m, \quad \text{for } n = 1, \ldots, N \tag{12}
\]

\[
- \frac{1 + \phi_t^{(n)}}{1 - \phi_t^{(n)}} \leq \hat{x}_t^{(n)} \leq \frac{1 + \phi_t^{(n)}}{1 - \phi_t^{(n)}}, \quad \text{for } n = 1, \ldots, N \tag{13}
\]

\[
\hat{x}_t^{(n)} = \frac{1 + \phi_t^{(n)}}{1 - \phi_t^{(n)}} \left( 2 \cdot 1 \left( y_A^{(n)} \geq 0 \right) - 1 \right), \quad \text{if } f_t^{(n)} > - \left( 1 - \frac{1}{R_{t+1}^f} \right) m, \quad \text{for } n = 1, \ldots, N \tag{14}
\]

Note that for an asset \( n \) for which the short selling constraint does not bind, i.e. \( f_t^{(n)} = - \left( 1 - \frac{1}{R_{t+1}^f} \right) m \), Equation (10) simplifies to the standard asset pricing relationship:

\[
\mathbb{E}_t \left( P_{t+1}^{(n)} + \delta_{t+1}^{(n)} \right) - R_{t+1}^f P_t^{(n)} = \gamma \Sigma_t^{(n)} \mathbf{1}
\]

where \( \Sigma_t^{(n)} \) is the \( n \)th row of the covariance matrix \( \Sigma_t \).

**Modified CAPM**

Equation (10) in Theorem 1 provides a relationship between asset prices and borrow fees. Importantly, the disagreement vector \( e_t \) does not enter Equation (10), but only Equation (11). Consequently, all the variables in Equation (10) are observable in the data, which allows for empirical testing. Focusing on asset \( n \) in Equation (10), we have that

\[
\mathbb{E}_t \left( P_{t+1}^{(n)} + \delta_{t+1}^{(n)} \right) - R_{t+1}^f P_t^{(n)} = \gamma \text{Cov}_t \left( P_{t+1}^{(n)} + \delta_{t+1}^{(n)}, (P_{t+1} + \delta_{t+1})' \mathbf{1} \right) - \frac{1}{2} R_{t+1}^f \left( 1 + \phi_t^{(n)} \right) \left( 1 - \frac{1}{R_{t+1}^f} \right) m + f_t^{(n)} P_t^{(n)}
\]
and after dividing by $P^{(n)}_{t}$, we get

$$
E_t \left(R^{(n)}_{t+1}\right) - R^{f}_{t+1} = \gamma \text{Cov}_t \left(R^{(n)}_{t+1}, R^{m}_{t+1}\right) P^{f}_1 - \frac{1}{2} R^{f}_{t+1} \left(1 + \phi^{(n)}_{t}\right) \left(1 - \frac{1}{R^{(n)}_{t+1}}\right) m + f^{(n)}_{t} \tag{15}
$$

where $R^{(n)}_{t+1} \equiv \frac{R^{(n)}_{t+1} + \delta^{(n)}_{t+1}}{P^{(n)}_{t}}$ and $R^{m}_{t+1} \equiv \frac{(P^{(m)}_{t+1} + f^{(m)}_{t+1})1}{P^{f}_{t}}$ are returns on asset $n$ and on the market portfolio of risky assets. After multiplying Equation (15) by market weight $w^{(n)}_{t} \equiv \frac{P^{(n)}_{t}}{P^{f}_{t}}$ and summing over the equations for all $N$ risky assets, we obtain the following equation:

$$
E_t \left(R^{m}_{t+1}\right) - R^{f}_{t+1} = \gamma \text{Var}_t \left(R^{m}_{t+1}\right) P^{f}_1 - \frac{1}{2} R^{f}_{t+1} w^{(n)}_t \text{diag} \left[1 + \phi^{(n)}_{t}\right] \left(1 - \frac{1}{R^{(n)}_{t+1}}\right) m + f^{(n)}_t \tag{16}
$$

Combining Equations (15) and (16) and defining $r^{(n)}_{t+1} \equiv R^{(n)}_{t+1} - 1$, $r^{m}_{t+1} \equiv R^{m}_{t+1} - 1$, and $r^{f}_{t+1} \equiv R^{f}_{t+1} - 1$ produces the following result:

**Theorem 2.** In any equilibrium, we have that

$$
E_t \left(r^{(n)}_{t+1}\right) - r^{f}_{t+1} = \alpha^{(n)}_t + \beta^{(n)}_t \left(E_t \left(r^{m}_{t+1}\right) - r^{f}_{t+1}\right) \tag{17}
$$

where

$$
\alpha^{(n)}_t \equiv \frac{1}{2} \left(1 + r^{f}_{t+1}\right) \left[\beta^{(n)}_t w^{(n)}_t \text{diag} \left[1 + \phi^{(n)}_{t}\right] \left(1 - \frac{1}{1 + r^{f}_{t+1}}\right) m + f^{(n)}_t\right]
$$

$$
\beta^{(n)}_t \equiv \frac{\text{Cov}_t \left(r^{(n)}_{t+1}, r^{m}_{t+1}\right)}{\text{Var}_t \left(r^{m}_{t+1}\right)}
$$

If short selling is unconstrained for all assets at time $t$, $\alpha_t$ will be $0$, and we get the standard CAPM. Otherwise, $\alpha_t \neq 0$ in general. Note that $\alpha^{(n)}_t$ for asset $n$ depends on the short interest and borrow fees of other assets. $\alpha_t$ is the sum of two components. We first explain the intuition behind the second component, $\alpha^{(n)}_{2,t} \equiv -\frac{R^{(n)}_{t+1}}{2} \left(1 + \phi^{(n)}_{t}\right) \left(1 - \frac{1}{R^{(n)}_{t+1}}\right) m + f^{(n)}_t$, referred to as “part 1”. It represents the direct effect of short-sale constraints on asset $n$ and has a negative effect on $\alpha^{(n)}_t$. All else equal, the higher the borrow fee $f^{(n)}_t$, the more negative this term is. This is intuitive: the higher the borrow fee, the more attractive it is to buy the asset and less attractive it is to sell, which pushes the price up and reduces the expected return. More interestingly, for any borrow fee $f^{(n)}_t$ in excess of $-\left(1 - \frac{1}{R^{(n)}_{t+1}}\right) m$, this term decreases when the short selling constraint gets relaxed (i.e. when $\phi^{(n)}_t$ goes up). Intuitively, when $\phi^{(n)}_t$ goes up and short selling for asset $n$ is constrained, long investors find
the asset more attractive because they can earn more by lending it. If the borrow fee \( f_t^{(n)} \) is not allowed to change, short investors have no desire to change their short position. To keep the market in equilibrium, price \( P_t^{(n)} \) needs to increase to discourage long investors from demanding and lending too much and to encourage short investors to absorb the excess lending supply. In equilibrium, \( f_t^{(n)} \) will respond as well, and the price \( P_t^{(n)} \) could move in either direction. To better understand the effect of short and long investors on the return, observe that this component can be written as the (negative of the) sum of two parts:

\[
\frac{R_{t+1}}{2} \left( 1 - \frac{1}{R_{t+1}} \right) \left( 1 - \frac{1}{R_{t+1}} \right) m + f_t^{(n)} \quad \text{and} \quad \frac{R_{t+1}}{2} \phi_t^{(n)} \left( 1 - \frac{1}{R_{t+1}} \right) m + f_t^{(n)}.
\]

The first part is the cost of short selling a dollar of asset \( n \), weighted by the mass of short sellers in that asset, \( \frac{1}{2} \); the second part is the gain from lending the maximum possible fraction of a dollar of asset \( n \) held, weighted by the mass of long investors in that asset, \( \frac{1}{2} \).

The first component in \( \alpha_t^{(n)} \), \( \alpha_{1,t}^{(n)} \equiv \frac{1}{2} R_{t+1} f_t^{(n)} w_t' \text{diag} \left[ 1 + \phi_t^{(n)} \right] \left( 1 - \frac{1}{R_{t+1}} \right) m + f_t^{(n)} \), or “part 1”, comes from the fact that the market portfolio is “overvalued” due to short selling constraints. Note that this term depends on a specific asset only through the asset beta. If \( \beta_t^{(n)} > 0 \), then this term has a positive effect on \( \alpha_t^{(n)} \): returns for asset \( n \) and the market correlate positively, so if the expected market return goes down (due to short selling constraints), then the expected return for asset \( n \) goes up relative to the market. The effect is opposite if \( \beta_t^{(n)} < 0 \).

4 Data

4.1 Data Sources

We obtain data related to short selling borrow fees and short interest from Markit Securities Finance. Markit collects self-reported daily data on stock loan availability and stock lending transactions from prime brokers, custodians, asset managers, and hedge funds, thus covering both sides of the market. Their proprietary database is the most comprehensive database on short selling. We believe the accuracy of the data is not a reason for concern, since participants have limited incentives to report false information due to Markit’s wide coverage and the anonymity of the data. Markit offers the borrowing and lending samples separately to different types of users, thus further protecting the anonymity of the data. We have access to the borrowing sample, so borrow fee data (based on transactions reported
by borrowers) includes the brokers’ spread. Markit reports the number of shares borrowed both by the borrower sample (that is tied to the reported borrow fees) and by the entire market. We assume the fee data is representative of the entire market and use the short interest measure corresponding to the entire market to calculate the short interest ratio as the number of shares sold short as a percentage of shares outstanding.

We merge the short selling data with the CRSP database on security prices and returns (from Wharton Research Data Services, WRDS) by CUSIP. Following the literature, we keep only stocks classified as “common stock” (CRSP codes 10 and 11) trading on the main US exchanges (NYSE, NASDAQ, and AMEX, exchange codes 11, 14, and 12, respectively), resulting in a sample of over 4,000 stocks. We calculate total returns and number of shares outstanding for each stock using dividend and stock split adjusted prices and drop any weeks (for a given stock) where the stock did not have any trades in three or more regular trading days.

Finally, we obtain two-year CAPM, FF3, and FFC4 weekly beta estimates for each stock (which will be inputs for our alpha calculations) from Beta Suite by WRDS, which we merge to CRSP by PERMNO, and the relevant time series of risk-free rates, market returns, and other FF3 and FFC4 factors from Kenneth French’s website\(^7\). Although all of our data is available at a daily frequency, we choose a weekly frequency instead in order to reduce some of the noise in stock prices. Our sample runs from January 4, 2010 to August 31, 2015.

4.2 Summary Statistics

High borrow fees are relatively uncommon: on an average day, most stocks have a borrow fee of zero plus a small brokers’ spread of 0.25-0.375% per year, and 80% of stocks have a borrow fee (including spread) of less than 1%, also known as “general collateral”. Only 5% of stocks have a borrow fee greater than 10%, and only 1% have a fee greater than 30%.

Table 1 presents summary information on the short interest ratio, median return, and average market cap by borrow fee bin, as well as the average number of stocks in each bin through time. We chose the bin cutoffs arbitrarily using different bin widths with the goal of grouping stocks with comparable short selling characteristics at the expense of

\(^7\)Kenneth French’s Data Library
Table 1: Summary Statistics

<table>
<thead>
<tr>
<th>Portfolio</th>
<th>Min fee (%) yr</th>
<th>Avg fee (%) yr</th>
<th>Max fee (%) yr</th>
<th>SIR (%)</th>
<th>Return (med % wk)</th>
<th>Market cap ($ in million)</th>
<th>n stocks</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.3</td>
<td>.4</td>
<td>0.5</td>
<td>1.5</td>
<td>.33</td>
<td>6560</td>
<td>2470</td>
</tr>
<tr>
<td>2</td>
<td>0.5</td>
<td>.6</td>
<td>1</td>
<td>5.9</td>
<td>.69</td>
<td>1467</td>
<td>329</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1.4</td>
<td>2</td>
<td>5.8</td>
<td>.37</td>
<td>894</td>
<td>155</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>2.5</td>
<td>3</td>
<td>6.8</td>
<td>.37</td>
<td>749</td>
<td>63</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>3.8</td>
<td>5</td>
<td>7.5</td>
<td>.47</td>
<td>663</td>
<td>92</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>5.9</td>
<td>7</td>
<td>8.1</td>
<td>.16</td>
<td>536</td>
<td>44</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>8.3</td>
<td>10</td>
<td>9</td>
<td>.27</td>
<td>458</td>
<td>65</td>
</tr>
<tr>
<td>8</td>
<td>10</td>
<td>13.6</td>
<td>20</td>
<td>10.9</td>
<td>-.04</td>
<td>498</td>
<td>65</td>
</tr>
<tr>
<td>9</td>
<td>20</td>
<td>24.8</td>
<td>30</td>
<td>10.1</td>
<td>-.15</td>
<td>474</td>
<td>16</td>
</tr>
<tr>
<td>10</td>
<td>30</td>
<td>46.2</td>
<td>216.7</td>
<td>8.9</td>
<td>-.01</td>
<td>406</td>
<td>29</td>
</tr>
</tbody>
</table>

Notes: The sample runs from January 4, 2010 to August 31, 2015. Borrow fees are annualized and expressed in terms of percent of the asset price, although they are charged on a monthly basis or upon settlement of the stock loan transaction. SIR stands for short interest ratio, which is the number of shares sold short as a percentage of the number of shares outstanding. Return is the median weekly return expressed in percentage terms. Market cap is the average market capitalization of a company in the bin. “n stocks” stands for the average number of companies included in the bin.

equal bin sizes. For example, a pair of stocks with fees of 21% and 29% are viewed by market participants as much more similar than stocks with fees of 1% and 9%. If we had instead separated our sample into deciles, all relatively constrained stocks would have been grouped in the same bin, which would have not allowed us to properly test our model. Our results are not sensitive to different choices of bin cutoffs.

The average market cap of high borrow fee stocks is an order of magnitude lower than that of low borrow fee stocks. This is expected, as large cap stocks have a large analyst following and represent the bulk of institutional investments, so there is a correspondingly large number of research articles published about them, helping the market to come close to having a consensus view on the firms’ prospects and thus reducing disagreement and short selling demand, while also making it easier for brokers to find lendable shares, thus increasing short selling supply; on the other hand, small cap stocks have a lower following and are thus more susceptible to differences of opinion among investors as well as both intentional and unintentional inaccuracies in their financial statements, which increases short selling demand, while their more fragmented and lower institutional ownership makes it harder for brokers to find lendable shares, thus decreasing short selling supply. Stocks with high borrow fees also have negative median gross (i.e. before lending fees) weekly returns in our sample. The short interest ratio is increasing in the borrow fees for stocks with borrow fees below 20%; after that, it is decreasing. This is intuitive, as short interest is an equilibrium variable: the most highly constrained stocks are the ones with the highest
fees, but one of the reasons they are constrained is their limited availability of lending supply, which mechanically results in a lower short interest ratio.

5 Empirical Analysis

5.1 Methodology

We follow the extensive anomalies literature in finance in testing an asset pricing model’s performance with respect to a specific variable by first sorting stocks according to the variable and grouping them into portfolios and then fitting the model to historical data on stock returns and thus testing whether the model can explain the portfolios’ returns or there is an unexplained residual alpha (constant term in a regression framework) that is statistically significantly different from zero. The baseline models used are CAPM and the Fama-French three-factor (FF3) and Fama-French-Carhart four-factor (FFC4) models. Thus, if an asset pricing model proposes $E(r^{(n)}) = f(X)$ for an asset $n$, \(^8\) after sorting and assigning stocks to portfolios, we test whether the residual $\alpha_{res}^{(n)}$ is equal to 0 in the following regressions:

$$E(r^{(n)}) - f(X) = \alpha_{res}^{(n)} + \epsilon^{(n)}$$ \hspace{2cm}(18)

If some $\alpha_{res}^{(n)} \neq 0$ (generally, but not necessarily, for one or more of the extreme portfolios), the model is considered incapable of explaining asset prices for stocks with a particular characteristic, which is then considered a market anomaly. For instance, in the case of a CAPM test, $X = (r^f, r^m, \beta^{(n)})$, and $f(X) = r^f + \beta^{(n)}(r^m - r^f)$, where $\beta^{(n)} = \frac{\text{Cov}(r^{(n)}, r^m)}{\text{Var}(r^m)}$, as stated in Theorem 2.

We start by sorting stocks by borrow fees and grouping them into the portfolios detailed in Table 1, where portfolio 1 consists of stocks that are easy (and essentially free) to short, with borrow fees below 0.5% per year, and portfolio 10 consists of stocks that are hard and costly to short, with borrow fees above 30% per year. As documented before, \(^9\) standard models such as CAPM, FF3, and FFC4 fail to explain returns for high borrow fee stocks, with negative residual alphas that are different from zero at standard significance levels.

\(^8\)X is the vector of variables specific to a particular test.

\(^9\)See, for example, Drechsler and Drechsler (2014).
In addition, we find that low borrow fee stocks have positive residual alphas that are small but still economically significant.

Next, we test whether the proposed model that incorporates short selling frictions is able to empirically explain asset prices where standard models fail. We first calculate the proposed model’s “short alpha”, \( \alpha^{(n)} \) for asset \( n \), as defined in Theorem 2 (i.e. the alpha predicted by the model as a result of short selling constraints), using the data from Markit and different beta estimates. Next, we perform the test outlined in Equation (18), with \( X = (r_f, r_m, \alpha^{(n)}, \beta^{(n)}) \), and \( f(X) = \alpha^{(n)} + r_f + \beta^{(n)}(r_m - r_f) \). Finally, we compare the residual alphas resulting from this estimation to the residual alphas from the baseline model (i.e. CAPM).

Finally, we separately study the effect of each of the components of \( \alpha \), namely \( \alpha_1 \) (i.e. part 1) and \( \alpha_2 \) (i.e. part 2). We do this because the two components play different roles in the model. In particular, part 1 results from the market portfolio being overvalued, which implies that the unconstrained stocks will produce higher risk-adjusted returns in the future relative to the market portfolio (i.e. implying that standard models overprice individual stocks on average). Part 2, the negative and much larger (in absolute value) term, comes from the fact that constrained stocks underperform their CAPM (or FF3, FFC4) risk-adjusted expected returns. Thus, part 1 should correct for the overpricing of all stocks, whereas part 2 should correct for the underpricing of high borrow fee stocks. To perform the test for part \( i \), for \( i \in \{1, 2\} \), in Equation (18) we take \( X = \left( r_f, r_m, \alpha_i^{(n)}, \beta^{(n)} \right) \) and \( f(X) = \alpha_i^{(n)} + r_f + \beta^{(n)}(r_m - r_f) \).

### 5.2 Results

Table 2 shows residual alphas for various models for portfolios sorted by borrow fees, as detailed in Table 1. Columns (1)-(3) provide residual alphas from the capital asset pricing model (CAPM), Fama-French three-factor (FF3), and Fama-French-Carhart four-factor (FFC4) models, respectively. Column (4) shows the residual alpha from the model proposed in the paper, referred to as “modified CAPM”. Columns (5) and (6) provide residual alphas from equivalent modifications of FF3 and FFC4 models. Since we look at returns at weekly frequencies, residual alphas are computed with respect to weekly returns and expressed in percentage terms.

---

\(^{10}\)for any positive-beta stock
Table 2: Residual Alphas

Notes: Residual alphas are computed with respect to weekly returns and expressed in percentage terms. Columns (1)-(3) provide residual alphas for CAPM, the Fama-French three-factor (FF3), and Fama-French-Carhart four-factor (FFC4) models, respectively. Column (4) shows the residual alpha from the model proposed in the paper, referred to as “modified CAPM”. Columns (5) and (6) provide residual alphas from equivalent modifications of FF3 and FFC4 models. ***, **, * indicate significance at 1, 5, and 10%.

It can be seen from columns (1)-(3) that standard models overpredict the returns of high fee portfolios (i.e. portfolios 5-10), with statistically significant and economically large estimates. In addition, the standard models underpredict the performance of low fee portfolios, although the magnitudes are substantially smaller than in the case of large fee portfolios. Including the model-implied alphas substantially reduces underpricing of high fee portfolios: the estimates cease to be statistically significant, and the magnitudes fall in absolute value by about 50% (i.e. for portfolios 7, 8, and 9). The modified CAPM does not perform substantially differently from the modified FF3 and FFC4 models.

Table 3: Residual Alphas due to Decomposition of Model-Implied Alpha

Notes: Residual alphas are computed with respect to weekly returns and expressed in percentage terms. Column (1) provides residual alphas for CAPM. Columns (2) and (3) modify the standard CAPM by including part 1 and part 2 of the model-implied alpha, respectively. Column (4) shows the residual alpha from the model proposed in the paper. ***, **, * indicate significance at 1, 5, and 10%.

Table 3 shows residual alphas after applying separately each of the components of the
model-implied alpha, $\alpha_1$ and $\alpha_2$ (i.e. columns (2) and (3), respectively), to the CAPM. For comparison, included are also residual alphas from the CAPM (i.e. column (1)) and from the modified CAPM which includes both components of $\alpha$ (i.e. column (4)). It can be seen that $\alpha_2$, and not $\alpha_1$, corrects for the underpricing of high fee portfolios. On the other hand, $\alpha_1$ reduces the residual alpha for all the portfolios by about 1 basis point (i.e. by comparing columns (3) and (4)). Since we are working with weekly returns, this amounts approximately to a 0.5% reduction on an annual basis, which is economically significant. This effect is intuitive, since $\alpha_1$ results from the market portfolio being overvalued, which implies that stocks will produce higher\footnote{for any positive beta stock} risk-adjusted returns in the future relative to the market portfolio. Because low fee portfolios are overpriced by standard models, and most stocks fall into this category (as can be seen in Table 1), stocks are on average overpriced by standard models, and thus the inclusion of $\alpha_1$ works in the desired direction by reducing the overpricing across all stocks. Even though $\alpha_1$ appears to have a much smaller effect than $\alpha_2$, note that $\alpha_1$ affects all stocks, whereas $\alpha_2$ only affects high borrow fee stocks, which constitute a small fraction of the stock market (i.e. see Table 1). Thus, both components are important in explaining asset prices and the market portfolio returns under the presence of short selling constraints.

6 Conclusion

It has been shown that traditional asset pricing models, like CAPM and Fama-French factor models, are unable to explain returns—both gross and net of borrow fees—for portfolios of high borrow fee stocks, i.e. highly short-sale constrained stocks, which have been linked to several anomaly characteristics. In this paper, we directly model the effects of short-sale constraints on asset prices. We use a CAPM setting where investors have heterogeneous beliefs, thus generating short-selling demand; we focus on constraints on the lending side of the market: investors are constrained from lending a fraction of the stock they hold. The resulting asset pricing equilibrium equation is similar to the CAPM equation but has a non-zero asset-specific alpha term composed of two parts: a large (in absolute value) negative term for high borrow fee stocks and a much smaller positive term for all [positive-beta] stocks. The negative term predicts lower risk-adjusted expected returns relative to the market portfolio for high borrow fee stocks, consistent with the existing literature; the
positive term is a result of these high borrow fee stocks being overvalued, which absent this correction would result in the market portfolio being overvalued, which cannot be the case in equilibrium. The latter is, to our knowledge, a novel result.

We then test the model directly using data on borrow fees, sorting assets into 10 portfolios by borrow fee. The results are consistent with the model: unexplained residual negative returns in short-sale constrained stocks are half as large as in the CAPM model and not statistically significantly different from zero, whereas unexplained residual positive returns in unconstrained stocks are lower and also not statistically significant, as opposed to the CAPM model; while this effect is smaller, at 50 basis points per year—which might explain why it has been overlooked in related studies—it is still economically significant.

We identify two areas for future research. First, in this model we take stock lending supply constraints as exogenous; it would be interesting to understand these constraints and how they affect borrow fees. Second, the positive alpha term, $\alpha_1$, is a measure of market overvaluation due to short-sale constraints; it is worth exploring whether this term includes information that can be used to predict market returns or other macroeconomic indicators.
References


