What You See is Not What You Get: The Costs of Trading Market Anomalies

Andrew J. Patton and Brian M. Weller*

Duke University

November 28, 2018

Abstract

Is there a gap between the profitability of a trading strategy "on paper" and that which is achieved in practice? We answer this question by developing a general technique to measure the real-world implementation costs of financial market anomalies. Our method extends Fama-MacBeth regressions to compare the on-paper returns to factor exposures with those achieved by mutual funds. Unlike existing approaches, our approach delivers estimates of all-in implementation costs without relying on parametric microstructure models or explicitly specified factor trading strategies. After accounting for implementation costs, typical mutual funds earn low returns to value and no returns to momentum.

JEL: G12, G14, G23

Keywords: Trading Costs, Performance Evaluation, Mutual Funds, Market Efficiency

*We thank Mark Carhart, Alex Chinco, Andrew Detzel, Campbell Harvey, David Hsieh, Robert Korajczyk, Alex Kostakis, Juhani Linnainmaa, Toby Moskowitz, Ronnie Sadka, and seminar participants at Duke Fuqua, Manchester, Warwick, the Rodney L. White Center Conference on Financial Decisions and Asset Markets, the Yale Junior Finance Conference, the UBC Winter Finance Conference, and John Cochrane's 60th Birthday Conference for valuable comments and discussion. Send correspondence to Brian Weller, Duke University, 213 Social Sciences Building, Box 90097, Durham, NC 27708; telephone: (919) 660-1720. E-mail: brian.weller@duke.edu.

I. Introduction

Empirical asset pricing overflows with explanations for differences in average returns across securities. The proliferation of predictors distracts from genuine market anomalies from which we might draw lessons about risks, preferences, and beliefs. Recent calls to action by Harvey, Liu, and Zhu (2016), Harvey (2017), and Hou, Xue, and Zhang (2017) have focused on high false discovery rates and scurrilous academic practices. Fundamentally they question whether candidate factors in the cross-section of expected returns are real and actionable.

We give on-paper trading strategies the benefit of the doubt and instead investigate whether they are implementable in practice, thereby representing true expected return factors or market anomalies. This line of inquiry originates with Fama (1970), who considers the role of transactions costs in defining market efficiency and departures therefrom.

Despite nearly fifty years of subsequent research, accurately measuring real-world implementation costs for academic factors remains a formidable challenge. Existing approaches generally fall into two categories. The first category entails using proprietary trading data to analyze costs for a single firm (e.g., Keim and Madhavan (1997), Engle, Ferstenberg, and Russell (2012), and Frazzini, Israel, and Moskowitz (2015)). Although selected firms may not be representative of asset managers as a whole, such analyses provide an informative lower bound on the implementation costs of factor strategies. The second approach uses market-wide trading data such as NYSE Trade and Quote (TAQ) to estimate trading costs for individual securities and then cumulate simulated costs of trade implied by dynamic factor strategies (e.g., Lesmond, Schill, and Zhou (2004), Korajczyk and Sadka (2004), and Novy-Marx and Velikov (2016)). Papers in this category typically extrapolate price impact estimates from small trades to large factor portfolios or ignore price impact costs entirely.

Our work complements these approaches with a new cross-sectional technique that combines the best elements of both. Like papers that utilize proprietary trading data, our estimates reflect the all-in costs of implementing factor strategies, and they apply equally well for past and modern market environments (for which Lesmond, Ogden, and Trzcinka (1999)'s zero-return day measure fails, for example). Like papers that estimate transaction cost functions using market data, our baseline methodology captures the costs faced by *representative* practitioners of factor investing rather than single, special investment managers. In contrast with both approaches, our methodology facilitates the evaluation of implementation costs (1) without specifying the precise trades used to implement factor strategies and (2) for arbitrary subsets of the asset management universe trading universe. These innovations are important because existing methods using precisely specified factor strategies with different data sources and sets of firms disagree on the implementability of factor strategies. For example, Lesmond, Schill, and Zhou (2004) find no net-of-costs return to momentum using TAQ data and a representative set of traders, whereas Frazzini, Israel, and Moskowitz (2015) find positive momentum premia for a large hedge fund. Although our approach rests on few assumptions, it has sufficient resolution to reconcile these disparate results in a transparent way.

Our methodology is an extension of the familiar Fama and MacBeth (1973) procedure. In the first stage, time-series regressions estimate factor loadings β_i for each asset i, and in the second-stage, cross-sectional regressions estimate the compensation per unit of factor exposure λ_t at each date t. Standard assets are based on stock portfolios, and the resulting estimates of compensation for factor exposure, denoted $\bar{\lambda}_k^S$, represent the "on-paper" profitability of a given factor strategy. We augment the set of assets to include 4,267 U.S. domestic mutual funds, and we allow the compensation for factor exposure earned by mutual funds, $\bar{\lambda}_k^{MF}$, to differ from that which is available on paper. Unlike stock portfolio returns, (gross) mutual fund returns reflect the real-world implementation costs of factor strategies, thus the difference between mutual fund and stock portfolio compensation $\bar{\lambda}_k^S - \bar{\lambda}_k^{MF}$ delivers an estimate of implementation costs for factor k.[1,2] Because costs per unit of exposure are likely to be negatively correlated with factor exposures—funds that earn greater net returns to a factor are more likely to take greater exposures to it—our estimate of implementation costs represents a *lower bound* on the costs faced by a representative mutual fund.

Our empirical analysis focuses on the implementation costs of mutual funds for the market (MKT), value (HML), size (SMB), and momentum (UMD). We choose these factors because they comprise the dominant empirical models in academic finance (e.g., Fama and French (1992) and Carhart (1997)) and because they serve as the basis for hundreds of billions of dollars in investments. We study mutual funds as our set of asset managers because they collectively manage more than \$16 trillion of capital in the United States,[3] and the mutual fund industry has been better populated for a longer period of time than alternative asset managers such as hedge funds. Our approach is readily extended to other factors and market participants, however.

Our analysis delivers new empirical facts on the all-in implementation costs of anomalies for typical mutual funds. First, momentum strategies suffer extreme underperformance in practice relative to on-paper strategies: our full-sample estimates of all-in implementation costs are in the range of 7.2%–7.6% per year, which eliminates most profits accruing to momentum during the 1970–2016 period. About half of this cost is due to mutual funds' inability to short. Our all-in cost estimates are considerably larger than those typically estimated using bid-ask spreads alone (e.g., Novy-Marx and Velikov (2016)). We conclude—as Lesmond, Schill, and Zhou (2004) do—that momentum strategies are unprofitable for typical asset managers when a broader set of implementation costs are considered. Second, mutual fund implementation costs sharply reduce returns to the value factor; we estimate all-in costs of 2.6%–4.1% per year. In contrast, mutual funds implementation costs for the market and size factors are approximately zero.

Our approach also yields insights into the sources of implementation costs for typical firms. Simple modifications to the set of test portfolios, factors, and slopes considered allow us to at-

[1] We use gross returns to focus on the efficiency of mutual funds' investing technology rather than on the distribution of rents between managers and investors embedded in net returns.

[2] Our more sophisticated approaches account for time- and cross-sectional variation in implementation costs, which we discuss further below.

[3] Per the 2017 Investment Company Fact Book, available at http://www.icifactbook.org/.

tribute costs to three primary sources. First, by excluding microcap stock portfolios, we can gauge the potential shadow costs of investability restrictions faced by real-world investors. Doing so we find that difficulty in investing in the smallest stocks explains reductions in realizable factor compensation of 1% per year for value and momentum. Second, we develop two "long-only" variants of the Carhart factors to assess the role of institutional constraints on shorting. We find that shorting frictions explain roughly half of mutual fund underperformance on momentum, and between one fifth and one third of underperformance on value. Third, we consider the gap attributable to mutual funds tracking alternative variants of the usual academic factors. Sorting funds by their time series R^2 from the four-factor model, we estimate that around half of the the average mutual fund underperformance on value and momentum is associated with uncompensated departures from the academic factors.[4]

As a third empirical contribution, we analyze variation in implementation costs across funds and time and demonstrate the importance of considering such variation in evaluating the implementability of factor strategies. While the typical firm's compensation for momentum is indistinguishable from zero, subsets of the mutual fund universe may achieve positive returns to momentum net of costs. A focused analysis on smaller subsets is also important from an aggregate market efficiency perspective because a violation exists if the *marginal investor* sees anomalous profits, even if a typical investor does not. For this purpose we segment the mutual-fund universe by (lagged) total net assets. Size is a natural sorting dimension because Berk and Green (2004), Pastor, Stambaugh, and Taylor (2015), Berk and van Binsbergen (2015), and others link scale to gross-of-fees performance. We rerun our cross-sectional analysis using each mutual fund size category separately, and we confirm that small and large mutual funds achieve different returns to momentum from "typical" mutual funds. Using this insight we reconcile conflicting evidence on the transactions-cost rationale for the continued existence of the momentum anomaly.[5]

Our approach provides us with an estimate of the gap in factor-mimicking portfolio performance ($\lambda_{kt}^{S} - \lambda_{kt}^{MF}$) for each particular factor and date, and we use this information to study determinants of the time series of average implementation costs. We document that industry inflows are associated with increased strategy costs, which in turn neutralize the secular declines in bid-ask spreads that affect the first dollar traded in factor strategies. As a consequence, bid-ask spread based measures increasingly underestimate the true costs of factor strategies as asset management (and factor investing in particular) grows in scale.

While our new approach delivers simple, nonparametric, estimates of the implementation costs for factor trading strategies, it does face some limitations. First, as mentioned above, our approach delivers lower bounds on implementation costs. In our empirical analysis these bounds do not

[4]This analysis also addresses a potential concern about the strategies that mutual funds actually trade, and it is discussed in detail below.

[5]We also run subsample analyses by quintile of total net assets and four-factor R^2s. Our methodology can accommodate many other splits of interest, e.g., sorting by factor betas sheds light on typical gains to running combined strategies. We leave investigations of other cuts of the mutual fund universe to future work.

greatly limit the economic conclusions we can draw: the estimated costs are already so high as to eliminate or severely attenuate the on-paper profitability of strategies like value and momentum for typical mutual funds. For other strategies, estimates that indicate positive returns net of costs do not necessarily imply that an anomaly can be implemented by typical investors. In this sense our measures can diagnose an implementability problem with a factor, but they cannot deliver a clean bill of health.

Secondly, our technique relies on real-world asset managers to reveal implementation costs through realized returns to their chosen factor exposures. We cannot speak to the costs of new factors that asset managers have not had an opportunity to trade.[6] For the same reason, our approach cannot estimate implementation costs for counterfactual factor exposures to evaluate strategy carrying capacities, unlike approaches that rely on parametric transaction cost models.

Finally, like much of the literature on performance evaluation, our method is susceptible to criticism of the choice of factors included in the analysis. A manager who is following a strategy that does not correspond to an approximate linear combination of those included in the model may appear to have high implementation costs for the included strategies, even though she has low costs for the strategy actually being implemented. We verify that omitted mutual fund strategies do not drive our high implementation cost estimates by replicating large performance gaps for funds with returns almost completely explained by the academic factors (the average R^2 of the four-factor model for these funds' return histories is 94%). For these funds, the scope for omitted strategies is too small to explain the observed real-world performance gaps.

II. Related Literature

The Fama and French three-factor model has been the benchmark for empirical asset pricing since its introduction in 1992. This empirical model supplanted the CAPM, but its new value and size factors had little theoretical motivation.[7] As factors continued to emerge over the next quarter century—most notably, the momentum anomaly of Jegadeesh and Titman (1993)—several strands of literature emerged in an attempt to tame the "factor zoo" (Cochrane (2011)). One active strand investigates the implementation costs of anomalies with a particular focus on size, value, and momentum anomalies. While implementation costs cannot explain why expected return discrepancies come to be in the first place, this literature (reviewed below) seeks to rationalize the continued existence of market anomalies as their byproduct. Our paper advances this line of inquiry by introducing a new and readily generalizable approach for measuring the real-world implementation costs of return factors and anomalies.

[6] This caveat does not apply in the particular case of momentum. Grinblatt, Titman, and Wermers (1995) argue that momentum-like strategies are endemic among mutual funds in their 1975–1984 sample, decades before the publication of Jegadeesh and Titman (1993).

[7] Banz (1981) and Basu (1977) document price-earnings ratios and market capitalization as *characteristics* associated with deviations from the CAPM.

Existing methods for measuring implementation costs take two approaches. The first approach uses specialized trading data to evaluate the costs of trade for large investment managers with the implicit assumption that these managers are representative of sophisticated investment managers more generally. These papers typically assess trading costs using Perold (1988)'s implementation shortfall measure, which captures the difference between realized profits and on-paper profits using a preset decision price. This approach dates back at least to Keim and Madhavan (1997)'s analysis of the transactions costs of a variety of investment styles for $83 billion of trades.[8] In this vein Keim (2003) uses institutional trading data for 33 firms and finds that trading costs likely eliminate profits to on-paper momentum strategies.

A key challenge to this method is that institutional trading is endogenous; traders are particularly aggressive in their trading targets when liquidity is readily available, which in turn imparts a downward bias to estimated cost functions. Frazzini, Israel, and Moskowitz (2015) overcome this challenge by using data from an investment manager whose trading targets are model-generated and selected irrespective of market conditions. Armed with more than $1 trillion of trades, they analyze value, size, and momentum anomalies and find that all of them are implementable and scalable to tens or hundreds of billions of dollars of invested capital. By their reckoning (and by contrast with Keim (2003)'s managers), major anomalies continue to be anomalous if their asset manager's costs are representative of typical investment managers' costs.

The second approach trades off accuracy for representativeness in estimating implementation costs. Rather than using proprietary trading data for a single asset manager to estimate costs directly, other studies derive transactions costs using aggregate price and transaction records and extrapolate estimated price impact functions to factor trading strategies.[9] Much of this literature focuses on the momentum anomaly because of its high turnover, and even the originating article establishing the momentum anomaly considers a trading-costs explanation (Jegadeesh and Titman (1993) and later Jegadeesh and Titman (2001)). Notably none of these papers use precise "all-in" trading cost measures like implementation shortfall because theoretical or "decision-date" prices are not obtainable outside of specialized trading data.

Chen, Stanzl, and Watanabe (2002) estimate separate price impact functions for 5,173 individual stocks and calculate the trading costs accruing to size, value, and momentum strategies. The authors suggest that all factors have break-even carrying capacities on the order of millions of dollars (HML) to hundreds of millions of dollars (SMB). By their calculations, factor strategies are not investable. Lesmond, Schill, and Zhou (2004) suggest that momentum in practice trades in "disproportionately high cost securities" rather than the typical-transactions cost securities Jegadeesh and Titman (1993) use for approximating the momentum factor's trading costs. Using

[8]Other studies use Keim and Madhavan (1997)'s calibrated transaction-cost functions to decompose fund performance for a larger universe of funds. For example, Wermers (2000), like our study, finds that implementation costs meaningfully erode mutual fund returns.

[9]Grundy and Martin (2001) and Barroso and Santa-Clara (2015) invert this logic and calculate the transactions costs that would be required to wipe out the momentum anomaly.

effective spreads from TAQ, commission schedules from a discount brokerage, and "all-in" frictions implied by zero-trading days (Lesmond, Ogden, and Trzcinka (1999)), Lesmond, Schill, and Zhou (2004) argue that trading costs erase the returns to the momentum anomaly.

Korajczyk and Sadka (2004) present more optimistic results on the investability of factor strategies. Korajczyk and Sadka (2004) use TAQ data to estimate effective and quoted spreads, the primary proportional costs studied in the literature, and price impact or "non-proportional trading cost" functions from Glosten and Harris (1988) and Breen, Hodrick, and Korajczyk (2002). In utilizing different non-proportional cost functions from Lesmond, Schill, and Zhou (2004), Korajczyk and Sadka (2004) extrapolate trade-level costs to find positive net-of-cost returns to the momentum anomaly. They invert their cost function estimates to obtain a break-even momentum strategy carrying capacity of \$5 billion. Novy-Marx and Velikov (2016) measure trading costs using effective spreads recovered from Hasbrouck (2009)'s Bayesian Gibbs sampler and tally costs of trading size, value, and momentum strategies, among others. The authors estimate strategy carrying capacities of \$5 billion for momentum (as in Korajczyk and Sadka (2004)), \$170 billion for size, and \$50 billion for value (the latter two of which are comparable to Frazzini, Israel, and Moskowitz (2015)'s estimates). These approaches do not account for the price impact costs of large institutional investors, and they likely overestimate the true strategy carrying capacities as a result.

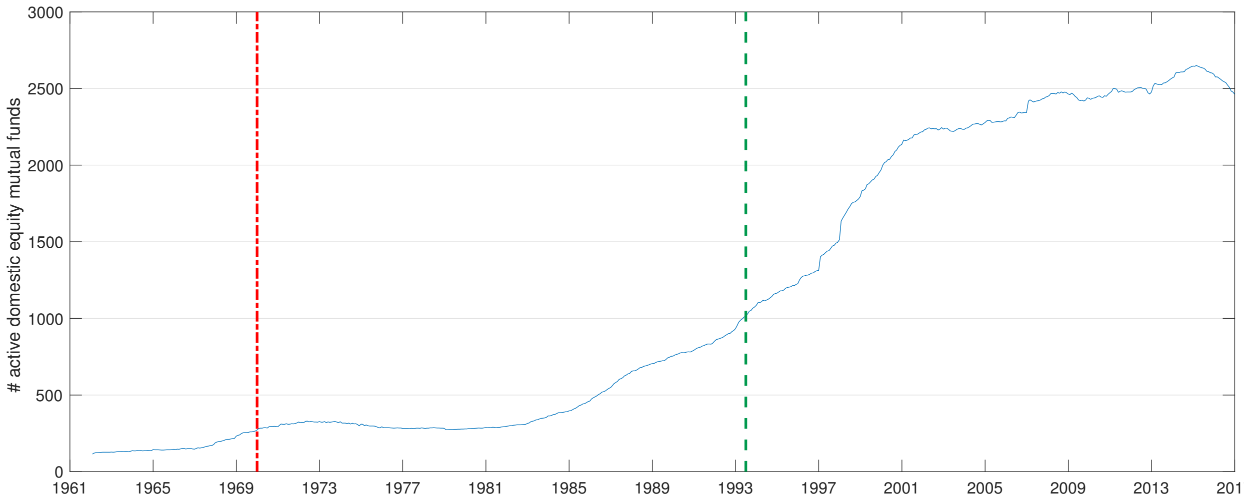
In concurrent work, Arnott, Kalesnik, and Wu (2017) argue, as we do, that mutual funds deliver much lower returns on value and momentum anomalies than on-paper factor counterparts might indicate. Our paper differs from theirs in four key respects. First, we modify standard Fama-MacBeth regressions to develop a cost-estimation procedure that is robust to heterogeneity in implementation costs across both funds and time. Second, we decompose costs to highlight the respective roles of shorting, investability, and liquidity frictions. Third, we slice the cross section of mutual funds to distinguish among funds of different attributes and in so doing reconcile previous work on implementation costs. Finally, our approach compares factor-mimicking portfolio returns for mutual funds and stock portfolios. Arnott, Kalesnik, and Wu (2017)'s use of on-paper factor returns as a benchmark is valid only if investors can frictionlessly trade stock factors.

III. Data

Our mutual fund sample consists of 4,267 United States domestic equity mutual fund groups with at least 24 non-missing monthly gross returns from January 1970 to December 2016. Appendix A details our mutual fund filtering methodology. Therein we describe a number of data cleaning and filtering steps based on the recommendations of Berk and van Binsbergen (2015), Pastor, Stambaugh, and Taylor (2015), and others. One data processing step bears special mention here: we map funds delineated by share class into fund groups. Share classes for funds with identical investments differ in fees charged to investors, but they are not otherwise economically distinct. To aggregate returns within a fund group, we take total-net asset weighted gross-of-fee returns.

Figure I: Count of Active Domestic Equity Mutual Funds by Month

Figure plots the count of non-missing returns by month for United States domestic equity mutual funds. The dashed line at January 1970 marks the starting point of our 1970–2016 sample. The dashed line at July 1993 marks the midpoint of the post-1970 sample as well as the start date for our post-Jegadeesh and Titman (1993) sample.



CRSP provides returns net of management and 12b-1 fees, and we convert these into gross returns by adding expense ratios divided by 12, following Fama and French (2010). We use "fund group" and "fund" interchangeably henceforth.

Significant changes in the count of active mutual funds reflect both a secular growth in the mutual fund industry and continual improvements in data quality.[10] Figure I highlights these changes by plotting the number of non-missing returns for domestic equity mutual funds by month. The number of funds increases from 276 in January 1970 to 979 in July 1993 to 2,463 in December 2016. Because the number and composition of funds varies widely over time, we conduct our analysis both on an extended sample and on a more recent subsample. Our long sample runs from January 1970 to December 2016. We discard the 1962–1969 window during which monthly returns are less consistently provided and during which several of our liquidity proxies are not available. Our recent subsample consists of the second half of the long sample and runs from July 1993 to December 2016. This start date postdates Jegadeesh and Titman (1993)'s documenting of the momentum anomaly, the most recently discovered factor we consider. Table I reports summary statistics for the set of mutual funds used in our analysis. All told the 1970–2016 sample consists of 724,995 fund-month observations and the 1993–2016 sample consists of 597,992 fund-month observations.

Much of our analysis compares mutual funds with similar stocks as measured by loadings on equity risk factors. Our Fama-MacBeth tests of Section IV combine mutual fund data with common

[10] Pages 1–2 of the CRSP mutual fund database guide details the amalgamation of data sources used to construct returns from December 1961 through the present. Page 16 discusses the merge of classifications into CRSP objective or style codes that we use to restrict the set of funds to United States domestic equity funds.

Table I: Domestic Equity Mutual Fund Sample Summary Statistics

Table presents summary statistics for the 1970–2016 sample of 4,267 United States domestic equity mutual funds. The top subtable provides information on the time series of the number of active funds for each date as well as cross-sectional information on fund lifetimes and total net assets (TNA) at sample start, middle, and end. The bottom subtable reports distributional information on fund excess returns. $\bar{\rho}$ is the average pairwise correlation with other mutual funds' returns, and $\rho_{S\&P500}$ is the correlation with the S&P 500.

Unit	Funds #	Lifetime Years	TNA, Jan. 1970 Million USD	TNA, July 1993 Million USD	TNA, Dec. 2016 Million USD
Mean	1286	14.16	128.74	552.87	2590.70
Std. Dev.	917	10.50	302.83	1533.70	13254.00
25%	324	5.75	3.96	37.48	70.93
50%	1023	11.58	23.90	118.36	314.00
75%	2282	19.58	91.18	431.83	1421.30

Unit	Mean Return % / Month	Return Vol. % / Month	Sharpe Ratio Annualized	$\bar{\rho}_{MF}$ %	$\rho_{S\&P500}$ %
Mean	0.46	4.88	0.41	74.10	84.97
Std. Dev.	0.63	1.97	0.41	16.49	18.42
25%	0.32	3.86	0.25	71.77	81.94
50%	0.56	4.66	0.44	77.86	89.55
75%	0.78	5.59	0.60	82.32	94.53

test portfolios. Our first portfolio set consists of the Fama-French 25 size-value double-sorted portfolios plus 25 size-beta portfolios, 25 size-prior return portfolios, and 25 size-Amihud illiquidity portfolios to ensure adequate dispersion in factor loadings to identify risk premia. We supplement this set of test assets with an expanded cross section following the recommendation of Lewellen, Nagel, and Shanken (2010). In our larger portfolio set, we add 49 industry portfolios, 25 size-operating profitability portfolios, 25 size-investment portfolios, 10 market beta-sorted portfolios, 10 market capitalization-sorted portfolios, 10 book-to-market ratio sorted portfolios, 10 Amihud illiquidity-sorted portfolios, 10 operating profitability-sorted portfolios, and 10 investment-sorted portfolios for a total of 269 portfolios. With the exception of the illiquidity-sorted portfolios, all portfolio data are downloaded from Ken French's website. Decile illiquidity portfolios sort stocks by the median daily Amihud illiquidity (daily absolute returns over dollar volume) over the prior calendar year, and stocks are assigned for the following year using deciles the end of June to match the timing convention of the other portfolio data.[11] The 25 size-illiquidity portfolios sort first on

[11]Our monthly stock sample consists of all CRSP stocks (share codes 10 or 11) with at least 24 non-missing monthly returns, for a total of 22,121 unique PERMNOs over the 1970–2016 sample period.

lagged market capitalization and then on Amihud illiquidity quintile within each size bin to ensure that all portfolios are non-empty. Our analysis uses both equal- and value-weighted stock portfolios.

We include several market and funding liquidity variables to proxy for time-varying cost factors that may affect the performance of mutual funds relative to stocks. Our market liquidity variables are Amihud illiquidity (Amihud (2002)), Pastor-Stambaugh liquidity levels (Pastor and Stambaugh (2003)), Corwin and Schultz (2012) NYSE-average bid-ask spreads, and the CBOE S&P 500 Volatility Index (VIX), as motivated by Nagel (2012). We use Corwin and Schultz (2012)'s methodology to estimate bid-ask spreads because it enables measurement of market liquidity before TAQ becomes available in 1993 and because it captures average effective spread levels and innovations better than other pre-TAQ methodologies (see Corwin and Schultz (2012) Table IV).[12] We use the CBOE S&P 100 Volatility Index (VXO) in place of the VIX in the pre-1990 period for which the VIX is not available. We compute Amihud illiquidity using CRSP daily data with values averaged within a month as in Amihud (2002), and we obtain the Pastor-Stambaugh series and CBOE VXO/VIX series from Robert Stambaugh's website and the Federal Reserve of St. Louis's FRED database, respectively.

Our funding liquidity variables are Frazzini and Pedersen (2014)'s "betting against beta" (BAB) factor, He, Kelly, and Manela (2017)'s intermediary capital ratio, the 10-year BAA minus 10-year Treasury spread, and the 3-month LIBOR minus 3-month Treasury yield or "TED" spread. The first two series are expressly designed to capture institutions' funding liquidity constraints, and the latter two series are common proxies in the funding liquidity literature (e.g., Brunnermeier (2009)). We obtain BAB from AQR's website, intermediary capital ratios from Asaf Manela's website, and credit and TED spreads from FRED.

IV. Fama-MacBeth Estimates of Implementation Costs

A. Fama-MacBeth Methodology

In this section, we consider the compensation per unit of risk exposure and investigate whether mutual funds obtain the same risk premium that academics achieve on paper. In our baseline estimation, we assume that mutual funds have a constant per-unit cost for implementing academic anomalies. Investing in a market index with $\beta_{MKT} = 1$ results in a performance gap of η relative to the on-paper performance of a market index, and investing in a levered version of the market more generally results in a performance gap of $\eta\beta_{MKT}$. In this setting, we would expect performance differences between stock and mutual fund portfolios to be linear in factor exposure.

We estimate the "implementation gap" using augmented Fama and MacBeth (1973) two-stage regressions for the Carhart four-factor model (Carhart (1997)). The time-series regression step

[12] Corwin and Schultz make their code available at https://www3.nd.edu/~scorwin/HILOW_Estimator_Sample_002.sas. As in their paper, we compute cross-sectional averages using only NYSE-listed stocks, and we use their variant of estimated spreads in which negative values are set to 0.

is standard except for the choice of test assets. As discussed in the preceding section, we have $N_S = 100$ and $N_S = 269$ stock portfolios for the baseline and extended portfolio sets, respectively. In addition to stock portfolios, we also include $N_{MF} = 4,267$ mutual funds, of which more than a thousand are active in the typical month. As diversified entities spanning a wide range of multifactor risk exposures, mutual funds unlike stocks need not be grouped into portfolios via a characteristic-sorting procedure.

The $N_S + N_{MF}$ first-stage time series regressions are

$$r_{it} = \alpha_i + \sum_k f_{kt}\beta_{ik} + \epsilon_{it},\ i = 1, \dots, N_S, N_{S+1}, \dots, N_S + N_{MF}, \tag{1}$$

where r_{it} is the month t gross return on stock portfolio or mutual fund i net of the contemporaneous risk-free rate and f_{kt} (for $k = 1, ..., K$) is the return on factor k at date t. The usual second-stage cross-sectional regressions are extended to accommodate the possibility of differences in risk pricing for stocks and mutual funds,

$$r_{it} = \sum_k \lambda^S_{kt}\hat{\beta}_{ik}1_{i\in S} + \sum_k \lambda^{MF}_{kt}\hat{\beta}_{ik}1_{i\in MF} + \epsilon_{it},\ t = 1, \dots, T. \tag{2}$$

Regression (2) is equivalent to two separate cross-sectional regressions run on stocks and mutual funds because the indicators partition the set of observations and coefficients. λ^S_{kt} and λ^{MF}_{kt} represent the factor-mimicking portfolio returns for stocks and mutual funds, that is, the hypothetical date-t returns to a stock or mutual fund portfolio with $\beta_k = 1$ and $\beta_j = 0\,\forall\, j \neq k$. If these factors were tradeable by real-world investors, f_{kt}, λ^S_{kt}, and λ^{MF}_{kt} would all be equal. The difference $\hat{\lambda}^\Delta_{kt} \equiv \hat{\lambda}^S_{kt} - \hat{\lambda}^{MF}_{kt}$ is our estimate of the implementation costs for strategy k, and it is the gap between the on-paper returns of a given strategy ("what you see") and the actual returns achieved by an asset manager facing real-world implementation costs ("what you get"). Conceptually this difference captures both direct costs such as spreads and price impact from factor trading as well as indirect costs such as investing in liquid versions of factors to robustify strategies against outflows. Our point estimates are the average of the monthly differences in factor compensation $\bar{\lambda}^\Delta_k$, and we construct Newey and West (1987) standard errors for this difference using three monthly lags to account for serial correlation and heteroskedasticity in the λ-difference series.

Throughout our analysis, we estimate cross-sectional slopes of returns on risk exposures assuming that risk exposures are constant. In making this assumption we prioritize minimizing the errors-in-variables problem arising from using noisy betas as inputs in the second-stage Fama-MacBeth regression. This problem is vitally important because we do not want to find differences in λs simply as a byproduct of higher measurement error in mutual fund betas. Static betas effectively eliminate this issue; empirically, cross-sectional slopes are virtually unaffected by measurement error because the time-series variance of estimated betas is about two orders of magnitude smaller than the cross-sectional variance of estimated betas for both stock and mutual fund portfolios.[13] In using static betas, we trade off against taking on model misspecification arising from time-varying stock portfolio or mutual fund risk exposures. Note, however, that if funds on average have timing ability, then using static betas in place of time-varying betas understates true implementation costs. For example, if funds scale up their betas when λ is high, then cross-sectional slopes for mutual funds $\hat{\lambda}_t$ are biased up, and the average estimated factor compensation $\bar{\hat{\lambda}}$ exceeds its true value.

Following Lettau, Maggiori, and Weber (2014) and others, we omit the constant term in (2) to force cross-sectional average alphas to zero. Economically this omission forces the typical zero-risk security or mutual fund to have zero excess (gross) return at each point in time. We impose this restriction because the slope on β_{MKT} is not otherwise well identified in our stock portfolio sample, namely the time series of the intercept α_t and the estimated market risk premium $\lambda_{MKT,t}$ are strongly negatively correlated and of similar magnitudes. By contrast in the mutual fund sample, market beta has a large and positive risk price regardless of whether a constant is included. Empirically, none of the other factor risk premia are meaningfully affected.

B. Baseline Estimates

Table II presents estimates of Equation (2). The λ^{Δ} value in the upper-left corner indicates that the difference in compensation per unit of market exposure is 0.38% per year greater for risk exposures taken via mutual funds than in (100 value-weighted) on-paper stock portfolios. This difference declines slightly to 0.21% per year when assessed against the full set of 269 portfolios. Neither effect is statistically or economically significant, and the absence of a performance gap is robust to using equal-weighted portfolios (bottom subtable) rather than value-weighted portfolios. This result is unsurprising as mutual funds are expected to be relatively good at implementing the market factor.

Broadening our focus to columns 1–4, we see that mutual funds underperform stocks in isolating factor exposures for two of the other Carhart factors. The average implementation gaps for value (HML) and momentum (UMD) range from of 50%–80% of the total on-paper factor return in stock portfolios. The remaining compensation to mutual funds for HML and UMD are positive ($\lambda^{MF} > 0$), but they are only 1%–3% per year and not statistically distinguishable from zero. Conversely, HML and UMD factors are both highly compensated and statistically robust in value-weighted stock portfolios in this period. On-paper compensation for size factor (SMB) exposure has a smaller positive point estimate, but this value is not reliably different from zero.

Notably the point estimates for the differences λ^{Δ} for HML and UMD are typically more statistically significant than either of the components of the difference λ^{S} or λ^{MF}. This feature

[13]The multiplicative attenuation bias in cross-sectional slopes is proportional to $\overline{\sigma^2_\epsilon}/\sigma^2_{\hat{\beta}} \approx$ 1%–2%, where $\sigma^2_{\hat{\beta}}$ is the cross-sectional variance of estimated betas (including measurement error) and $\overline{\sigma^2_\epsilon}$ is the average variance of the time-series beta estimates.

Table II: Implementation Cost Estimates in Fama-MacBeth Regressions — Baseline Specification

Table reports Fama-MacBeth estimates of the compensation for factor exposure for stock portfolios (second panel), domestic equity mutual funds (third panel), and their difference (top panel). Coefficients are the average cross-sectional slopes $\bar{\lambda}_k$ across monthly regressions of excess returns r_{it} on time-series betas $\hat{\beta}_{ik}$,

$$r_{it} = \sum_k \lambda^S_{kt} \hat{\beta}_{ik} 1_{i \in S} + \sum_k \lambda^{MF}_{kt} \hat{\beta}_{ik} 1_{i \in MF} + \epsilon_{it},\ t = 1, \ldots, T,$$

where k indexes the four Carhart (1997) factors and λ^Δ is defined as $\lambda^S - \lambda^{MF}$. Stock portfolio sets are described in Section III. All coefficients are annualized and reported in percent. Standard errors are Newey-West with three lags. t statistics are reported in parentheses.

(a) Value-Weighted Stock Portfolios

		1970 − 2016				**1993 − 2016**			
	N_S	MKT	HML	SMB	UMD	MKT	HML	SMB	UMD
λ^Δ	100	-0.38	3.81***	0.26	7.18***	-0.11	3.12***	-0.24	4.27***
t-stat		(-1.28)	(5.08)	(0.42)	(5.53)	(-0.32)	(3.83)	(-0.29)	(2.64)
λ^Δ	269	-0.21	2.59***	-0.07	7.30***	0.28	2.09***	-0.97	5.04***
t-stat		(-0.88)	(3.81)	(-0.14)	(5.54)	(1.25)	(3.31)	(-1.39)	(2.89)
λ^S	100	6.60***	6.43***	1.27	8.72***	7.67**	5.43*	1.96	6.01
t-stat		(2.75)	(3.51)	(0.75)	(3.74)	(2.35)	(1.93)	(0.81)	(1.60)
λ^S	269	6.77***	5.20***	0.94	8.85***	8.06**	4.40	1.23	6.78*
t-stat		(2.82)	(2.84)	(0.56)	(3.80)	(2.49)	(1.54)	(0.51)	(1.83)
λ^{MF}	—	6.98***	2.62	1.01	1.54	7.78**	2.31	2.20	1.73
t-stat		(2.86)	(1.51)	(0.59)	(0.63)	(2.38)	(0.83)	(0.92)	(0.45)
T		564	564	564	564	282	282	282	282
$\bar{N}_{MF}$		1286	1286	1286	1286	2123	2123	2123	2123

$^*p < .10$, $^{**}\ p < .05$, $^{***}\ p < .01$

(b) Equal-Weighted Stock Portfolios

		1970 − 2016				**1993 − 2016**			
	N_S	MKT	HML	SMB	UMD	MKT	HML	SMB	UMD
λ^Δ	100	-0.36	4.47***	2.34**	6.83***	0.07	3.16***	2.14	3.71**
t-stat		(-0.76)	(5.57)	(2.41)	(5.21)	(0.12)	(3.29)	(1.55)	(2.21)
λ^Δ	269	0.25	3.31***	2.22**	8.51***	0.95	2.01*	2.05	6.04***
t-stat		(0.5)	(3.58)	(2.05)	(6.19)	(1.45)	(1.96)	(1.34)	(3.13)
λ^S	100	6.62***	7.09***	3.35***	8.37***	7.85**	5.48**	4.34	5.45
t-stat		(2.75)	(3.91)	(1.70)	(3.59)	(2.39)	(1.99)	(1.53)	(1.44)
λ^S	269	7.23***	5.93***	3.23	10.06***	8.73***	4.33	4.25	7.78**
t-stat		(3.02)	(3.03)	(1.56)	(4.17)	(2.69)	(1.47)	(1.43)	(1.98)
λ^{MF}	—	6.98***	2.62	1.01	1.54	7.78**	2.31	2.20	1.73
t-stat		(2.86)	(1.51)	(0.59)	(0.63)	(2.38)	(0.83)	(0.92)	(0.45)
T		564	564	564	564	282	282	282	282
$\bar{N}_{MF}$		1286	1286	1286	1286	2123	2123	2123	2123

$^*p < .10$, $^{**}\ p < .05$, $^{***}\ p < .01$

reflects the netting out of common variation in factor realizations between the λ time series. Ideally the residual variation in λ^{Δ} captures only random variation in trading costs. In practice this residual variation also captures idiosyncratic differences in estimated risk prices associated with using different sets of test assets; the difference between λ^{Δ} estimated from the set of 100 stock portfolios and the set of 269 stock portfolios suggests that the implementation gap depends in part on the stock benchmarks employed.

Columns 5–8 reproduce these tests for the July 1993 to December 2016 sample. Mutual funds achieve lower returns to HML and UMD and higher returns to SMB than in the full sample, and these returns are universally statistically indistinguishable from zero. For stock portfolios, the compensation for HML and UMD (SMB) exposures also decreases (increases) relative to the full sample. The net effect of these changes is a small decrease in the typical implementation gap for HML and a moderate decrease in the implementation gap for UMD. The implementation gap is roughly unchanged for market exposure (effectively zero) and SMB exposure (positive but now statistically insignificant). Focusing on the latter sample with a more broadly representative set of mutual funds does not change our conclusions on the high real-world efficacy of achieving market exposure and size and the low real-world efficacy of implementing value and momentum.

In sum, *no factor other than the market earns reliably positive risk premia for the typical mutual fund*. This finding is our first main result, and below we investigate whether it survives alternative weightings (including equal-weighted returns in the bottom subtable), enriched methodologies, additional controls, and sample splits.[14,15] Section IV.E presents a detailed comparison of our estimates to prior work, and we postpone most discussion of economic mechanisms and implications until that time. However, we understand that implementation cost estimates of this magnitude may surprise some readers, and it is worth mentioning now an intuitive channel by which real-world compensation to these factors may fall to zero even as on-paper returns persist.

Using evidence from mutual fund flows, Berk and van Binsbergen (2016) and Barber, Huang, and Odean (2016) demonstrate that investors appear to use the CAPM to evaluate market risks, and compensation accruing to non-market sources is perceived as skill. By this logic, capital market equilibrium requires that real-world factor compensation *must* be squeezed to zero because these other factors are not seen as risky. Otherwise, if funds were to achieve positive returns to HML and UMD, funds could load on these factors to achieve "alpha" and attract inflows indefinitely. Our findings indicate that equilibrium obtains through greater implementation costs rather than through reduced on-paper factor compensation. Higher implementation costs resulting from greater fund

[14] E.g., Section D below evaluates a purely characteristic-based variant of our Fama-MacBeth regressions with similar results.

[15] Appendix D introduces a complementary, matched-pairs approach to investigating mutual fund implementation costs. In the spirit of Daniel, Grinblatt, Titman, and Wermers (1997), this approach compares returns to high book-to-market ratio, small size, and high prior return stocks and mutual funds with similar risk characteristics. The analysis therein has the ancillary benefit of controlling for differences in the distribution of betas between stock portfolios and mutual funds, which may be important if compensation for factor exposure is earned only in some segments of this distribution.

size (Berk and Green (2004)) or industry scale (Pastor and Stambaugh (2012)) are two established mechanisms by which these costs may adjust until achievable anomalous returns disappear.

C. Estimates When Costs Vary Across Funds and Time

Time-varying implementation costs complicate the comparison of compensation per unit of factor risk. To see why, consider the following augmented model of mutual fund costs. As before, let there be a set of academic factors f, where f_t is a $1 \times K$ vector. Each mutual fund i implements its favored version of academic factors and earns a return of

$$h_{it} = f_t - \eta_{it}, \tag{3}$$

where η_{it} reflects tilts away from the academic factor on account of trading costs or factor optimization. This section differs from the previous one in that we no longer assume that η is constant across funds and time in interpreting λ^{Δ}. The η_{it} term in turn can be decomposed into components,

$$\eta_{it} = \eta_i + \eta_t \gamma_i + \tilde{\eta}_{it}. \tag{4}$$

The first component is the fixed, fund-specific cost of trading a factor. The second component is a set of L time-varying liquidity costs, η_t multiplied by the $L \times K$ loadings of all factors on these liquidity costs, γ_i. Finally, $\tilde{\eta}_{it}$ is a $1 \times K$ set of idiosyncratic costs, e.g., a surprise redemption demand that makes continued investment in factor k more costly for fund i.

In this heterogeneous-cost specification, funds earn returns of

$$\begin{aligned} r_{it} &= \alpha_i + h_{it}\beta_i + \epsilon_{it} \\ &= (\alpha_i - \eta_i\beta_i) + (f_t - \eta_t\gamma_i)\,\beta_i + (\epsilon_{it} - \tilde{\eta}_{it}\beta_i)\,. \end{aligned} \tag{5}$$

An ideal test compares the average compensation f_t for factor exposure for on-paper investment in stocks against the compensation h_{it} for factor exposure for real-world investment. In the constant-cost setting of Section IV.A, we achieve this ideal: η_{it} simplifies to η, and Fama-MacBeth regressions recover consistent estimates of h as the difference in λs in Equation (2).

By contrast, in this general setting we face two key challenges that complicate the comparison of f_t and h_{it}. First, trading costs vary over time, and these costs may covary with factor realizations. For example, during the 2007–2008 Financial Crisis, the aggregate market declines sharply just as funding and market liquidity deteriorate significantly. Omitting relevant liquidity factors thus contributes to an omitted variable bias in time-series estimates of β_i for investment managers, which in turn potentially invalidates simple comparisons of second-stage slope estimates. Second, investment managers select their risk exposures endogenously. An investor who has discovered improvements upon academic factors and another who faces particularly high trading costs are unlikely to select the same factor exposures, all else equal. For this reason we would expect mutual fund-specific trading costs η_i to be correlated with β_i in the cross-section.

We now address these two sources of bias. First, to address the omission of trading-cost factors, we assume that trading costs or optimization gains for mutual funds are spanned by liquidity proxies considered in the literature (and described in Section III). To avoid overfitting by including too many correlated liquidity proxies, we start with two:[16] the first principal component of four market-liquidity variables (Amihud illiquidity, Pastor-Stambaugh liquidity, Corwin-Schultz bid-ask spreads, and the CBOE VIX/VXO) and the first principal component of four funding-liquidity variables (Frazzini and Pedersen (2014)'s "betting against beta" factor, He, Kelly, and Manela (2017)'s intermediary capital risk factor, 10-year BAA minus 10-year Treasury spreads, and 3-month LIBOR minus 3-month Treasury yield or "TED" spreads).[17] We normalize all liquidity variables to have unit standard deviation before taking principal components because liquidity proxies vary widely in their scales. We assign these components an *illiquidity* interpretation by normalizing them to be positively correlated with the VIX/VXO.

We then run Fama-MacBeth regressions as before, but we extend the factor model to include these liquidity proxies in the time-series regressions,

$$r_{it} = \alpha_i + \sum_k f_{kt}\beta_{ik} + \sum_l \tilde{\eta}_{lt}\tilde{\gamma}_{il} + \epsilon_{it},\ \ i = 1, \dots, N_S, N_{S+1}, \dots, N_S + N_{MF}, \tag{6}$$

where $\tilde{\eta}_{lt}$ are the liquidity factor proxies at time t. The second-stage cross-sectional regressions are exactly as in Equation (2).

The mismatch in model specification for the time-series and cross-sectional regressions is intentional, and the decomposition of the resulting second-stage coefficient estimates reveals why the second source of bias—cross-sectional heterogeneity in implementation costs—makes our results conservative. In the time-series regressions, we recover fund exposures to the academic factors, and we need the additional liquidity proxy variables to cleanse the estimated mutual fund factor loadings of omitted illiquidity components. By contrast, in the second stage, we recover the cost per unit exposure to the academic factors and do not want to include the liquidity proxy exposures. Excluding the liquidity factors only in the second stage delivers $\hat{\lambda}_t^S = \lambda_t^S$ and

$$\hat{\lambda}_t^{\Delta} = \lambda_t^S - \frac{cov\left(r_{it}^{MF}, \beta_i\right)}{var\left(\beta_i\right)} = -\frac{cov\left(\alpha_i - \eta_{it}\beta_i, \beta_i\right)}{var\left(\beta_i\right)} = \bar{\eta}_t - \frac{cov\left(\left(\bar{\eta}_t - \eta_{it}\right)\beta_i, \beta_i\right)}{var\left(\beta_i\right)}. \tag{7}$$

[16] Ideally we would use all liquidity variables rather than their principal components because we want time-varying determinants of η_{it} to lie in the span of the liquidity-augmented factor model. To this end we include all proxies in a sparse-regression approach in Appendix C.

[17] The CBOE VXO and the TED spread series start in January 1986. Our principal components procedure accommodates the missing liquidity proxy data using MATLAB's alternating least squares (ALS) algorithm. ALS extracts factors and completes missing data by conjecturing principal components and iteratively estimating principal component loadings ϕ and factor values g until the distance between known and fitted values achieves a local minimum. We run PCA-ALS from 1,000 starting points and select the global distance-minimizing factors and loadings.

The final equality makes the standard assumption that alphas and betas are cross-sectionally uncorrelated. $\bar{\eta}_t$ represents the cross-sectional average per-unit liquidity costs to implementing the factor. The second term is the covariance between deviations from the average costs and βs. Funds with a particular skill in investing in a factor likely have higher exposures to it, β_i is endogenous, so β_i is high when $\bar{\eta}_t - \eta_{it}$ is high, and β_i is close to zero when $\bar{\eta}_t - \eta_{it}$ is negative (negative betas do not reverse the sign on costs). Combining these features, the overall covariance is positive, and the cross-sectional slopes of returns with respect to β_i are biased upward ($\hat{\lambda}_t^{MF} > \lambda_t^{MF}$).[18] Consequently λ_{kt}^{MF} is an upper bound on the realizable gains to factor investing per unit risk exposure, and λ_{kt}^{Δ} is a lower bound on the costs of implementing a factor strategy.

Table III presents results from the liquidity-extended first-stage regression. Results are virtually the same as those of the baseline specification in Table II with one exception. Mutual funds' (already low) annual compensation for UMD exposure decreases from 1.54% to 1.28% in the long sample and from 1.73% to 0.76% in the recent sample, suggesting that liquidity factor exposure at least partly explains mutual funds' compensation for momentum. Asness, Moskowitz, and Pedersen (2013) find that momentum loads positively on liquidity risk, and we find that the same holds for mutual funds' implementation of momentum. We examine this feature in detail in Section VI.B.

D. Cross-Sectional Characteristic Regressions

The Fama-MacBeth regression approach of the preceding sections estimates implementation costs for asset pricing factors under the assumption that factor exposures are the source of risk premia. However, Daniel and Titman (1997) and Daniel, Grinblatt, Titman, and Wermers (1997), among others, argue that characteristics such as book-to-market ratios and market capitalization dominate factors in explaining the cross-section of expected stock returns and mutual fund performance.[19] To address this class of models, we modify our baseline two-stage regression approach to use characteristics rather than factor betas. The resulting cross-sectional slopes are estimates of the compensation to characteristics accruing to on-paper stock portfolios and in real-world mutual funds.

We obtain characteristic prices in the style of Fama-MacBeth regressions by replacing the time-series beta estimates from Equation (1) with stock portfolio or fund characteristics, c_{ikt}, in the cross-sectional regressions,

$$r_{it} = \sum_k \lambda_{kt}^S c_{ikt} 1_{i \in S} + \sum_k \lambda_{kt}^{MF} c_{ikt} 1_{i \in MF} + \epsilon_{it},\ t = 1, \ldots, T. \tag{8}$$

Because characteristics are directly observed rather than estimated, we no longer face an errors-in-

[18] Including liquidity proxies in the second-stage introduces a more opaque omitted variable bias, as we discuss in Appendix B.

[19] Berk (2000) and Davis, Fama, and French (2000) provide other views in the debate on compensation to factors or characteristics. We thank Juhani Linnainmaa and Ronnie Sadka for the suggestion to consider both perspectives.

Table III: Implementation Cost Estimates in Fama-MacBeth Regressions — Liquidity PCs

Table reports Fama-MacBeth estimates of the compensation for factor exposure for stock portfolios (second panel), domestic equity mutual funds (third panel), and their difference (top panel). Coefficients are the average cross-sectional slopes $\bar{\lambda}_k$ across monthly regressions of excess returns r_{it} on time-series betas $\hat{\beta}_{ik}$,

$$r_{it} = \sum_k \lambda_{kt}^S \hat{\beta}_{ik} 1_{i \in S} + \sum_k \lambda_{kt}^{MF} \hat{\beta}_{ik} 1_{i \in MF} + \epsilon_{it},\ t = 1, \ldots, T,$$

where k indexes the four Carhart (1997) factors and λ^Δ is defined as $\lambda^S - \lambda^{MF}$. First-stage regression estimates include these factors, the first principal component of market liquidity proxies, and the first principal component of funding liquidity proxies. Liquidity proxies and stock portfolio sets are described in Section III. All coefficients are annualized and reported in percent. Standard errors are Newey-West with three lags. t statistics are reported in parentheses.

(a) Value-Weighted Stock Portfolios

		1970 − 2016				**1993 − 2016**			
	N_S	MKT	HML	SMB	UMD	MKT	HML	SMB	UMD
λ^Δ	100	-0.44	4.07***	0.35	7.49***	-0.12	3.30***	-0.24	5.23***
t-stat		(-1.45)	(5.17)	(0.57)	(5.71)	(-0.36)	(3.92)	(-0.29)	(3.09)
λ^Δ	269	-0.22	2.83***	-0.02	7.55***	0.27	2.30***	-0.93	5.84***
t-stat		(-0.92)	(3.87)	(-0.03)	(5.70)	(1.22)	(3.54)	(-1.32)	(3.29)
λ^S	100	6.55***	6.71***	1.26	8.77***	7.68**	5.38*	1.98	5.99
t-stat		(2.74)	(3.63)	(0.74)	(3.76)	(2.37)	(1.90)	(0.82)	(1.59)
λ^S	269	6.77***	5.47***	0.89	8.84***	8.08**	4.39	1.30	6.60*
t-stat		(2.83)	(2.94)	(0.53)	(3.78)	(2.51)	(1.51)	(0.54)	(1.78)
λ^{MF}	—	6.99***	2.64	0.90	1.28	7.80**	2.09	2.22	0.76
t-stat		(2.87)	(1.51)	(0.53)	(0.52)	(2.41)	(0.74)	(0.92)	(0.20)
T		564	564	564	564	282	282	282	282
$\bar{N}_{MF}$		1286	1286	1286	1286	2123	2123	2123	2123

$^*p < .10$, $^{**}\ p < .05$, $^{***}\ p < .01$

(b) Equal-Weighted Stock Portfolios

		1970 − 2016				**1993 − 2016**			
	N_S	MKT	HML	SMB	UMD	MKT	HML	SMB	UMD
λ^Δ	100	-0.53	4.48***	2.69***	6.92***	0.00	2.67***	2.32*	4.00**
t-stat		(-1.12)	(5.35)	(2.75)	(5.14)	(0.00)	(2.59)	(1.71)	(2.27)
λ^Δ	269	0.05	3.64***	2.64**	8.54***	0.74	1.92*	2.29	5.73***
t-stat		(0.09)	(3.79)	(2.44)	(5.95)	(1.09)	(1.81)	(1.56)	(2.82)
λ^S	100	6.46***	7.12***	3.60*	8.20***	7.81**	4.76*	4.54	4.76
t-stat		(2.70)	(3.88)	(1.84)	(3.51)	(2.40)	(1.73)	(1.63)	(1.26)
λ^S	269	7.04***	6.28***	3.55*	9.82***	8.55***	4.01	4.51	6.49
t-stat		(2.97)	(3.14)	(1.73)	(4.05)	(2.66)	(1.35)	(1.57)	(1.64)
λ^{MF}	—	6.99***	2.64	0.90	1.28	7.80**	2.09	2.22	0.76
t-stat		(2.87)	(1.51)	(0.53)	(0.52)	(2.41)	(0.74)	(0.92)	(0.20)
T		564	564	564	564	282	282	282	282
$\bar{N}_{MF}$		1286	1286	1286	1286	2123	2123	2123	2123

$^*p < .10$, $^{**}\ p < .05$, $^{***}\ p < .01$

variables problem arising from using estimated betas in the second-stage regression. This feature allows us to use time-varying characteristics c_{ikt} rather than averages over the full time series. Indeed this replacement is critical because while betas are relatively stable across the sample period, characteristics such as average market capitalization have strong time trends.[20]

With the modified methodology in hand, the next step is to specify the set of characteristics and their construction. We follow Daniel and Titman (1997) and Daniel et al. (1997) in using market capitalization, book-to-market ratios, and prior returns as characteristics. We construct these characteristics at the stock-month level using book-to-market ratios from the most recent fiscal year, market capitalization at the end of the current month, and prior 12-month minus 2-month returns. We then lag book-to-market ratios and market capitalization by one month to ensure that all characteristics are available to market participants at the start of month t. To control data errors in the book-to-market ratio, we drop negative values and winsorize at the 1% level within each date.

We then build characteristics at the stock portfolio and mutual fund group levels as value-weighted averages of the characteristics of their constituent stocks. For stock portfolios, we use Ken French breakpoints where available to partition NYSE/AMEX/NASDAQ common stocks (share code 10 or 11). For illiquidity-sorted portfolios, we use quintiles of Amihud illiquidity in univariate-sorted portfolios and conditional quintiles of Amihud illiquidity by market capitalization bin in double-sorted portfolios. The stock portfolio value of each characteristic is the value-weighted average of its constituent stocks' characteristics.

To construct mutual fund characteristics, we first obtain mutual fund holdings using the Thomson Reuters mutual funds holdings database (s12). We match holdings at the fund level using MFLINKS to convert Thomson Reuters identifiers to CRSP mutual fund identifiers.[21] We form fund-level characteristics as the dollar holdings-weighted average of stock-level characteristics and fund-group characteristics as the TNA-weighted average of fund-level characteristics. Finally, we take logs of book-to-market ratios and market capitalization to prevent the regressions from being dominated by outlier firms.[22]

[20]Because of these time trends and because a zero value of a characteristic need not command a zero risk premium (unlike betas), we also include a constant as part of the characteristic set. The inclusion of a constant at each date eliminates the influence of time trends by absorbing shifts in the means of the characteristics. However, by the same token, absorbing time-varying means of the characteristics renders the constant term uninterpretable, and we do not report it in our results.

[21]Details on the merge procedure are available at the Guide for MFLINKS on WRDS; most importantly for our application, the link table matches up to 98% of the domestic equity funds in CRSP for the March 1980 to September 2015 period in which linking data is available.

[22]Characteristic ranks are an alternative transformation sometimes used in characteristic regressions, but they are inappropriate in our setting for two reasons. First, a change in rank has a different meaning for stock portfolios and mutual funds, particularly in light of the flow-performance relationship related to mutual funds' prior returns characteristic. Second, the distribution of characteristics differs for stock portfolios and mutual funds, so ranking must be performed across all entities so as to not destroy information about differences in average characteristics. However, doing so introduces the undesirable feature that stock portfolio characteristics depend on the set of mutual funds considered and vice-versa.

Table IV: Implementation Cost Estimates in Fama-MacBeth Regressions — Characteristic Model

Table reports Fama-MacBeth estimates of the compensation for characteristic exposure for stock portfolios (second panel), domestic equity mutual funds (third panel), and their difference (top panel). Coefficients are the average cross-sectional slopes $\bar{\lambda}_k$ across monthly regressions of excess returns r_{it} on characteristics,

$$r_{it} = \left(\lambda_{0t}^S + \lambda_{Bt}^S BM_{it} + \lambda_{St}^S SIZE_{it} + \lambda_{Pt}^S P212_{it}\right) 1_{i \in S} + \left(\lambda_{0t}^{MF} + \lambda_{Bt}^{MF} BM_{it} + \lambda_{St}^{MF} SIZE_{it} + \lambda_{Pt}^{MF} P212_{it}\right) 1_{i \in MF} + \epsilon_{it},\ t = 1, \ldots, T,$$

where BM denotes lagged log book-to-market ratios, $SIZE$ denotes lagged log market capitalization, and $P212$ denotes prior 2-12 month return. λ^Δ is defined as $\lambda^S - \lambda^{MF}$. Stock portfolio sets are described in Section III. All coefficients are annualized and reported in percent. Standard errors are Newey-West with three lags. t statistics are reported in parentheses.

(a) Value-Weighted Stock Portfolios

		1980 − 2015			**1993 − 2015**		
	N_S	BM	$SIZE$	$P212$	BM	$SIZE$	$P212$
λ^Δ	100	1.86	-0.37	10.64*	2.10	-0.51	9.04
t-stat		(1.07)	(-0.97)	(1.90)	(0.89)	(-1.10)	(1.25)
λ^Δ	269	-0.06	-0.11	7.16*	-1.33	-0.14	7.99*
t-stat		(-0.04)	(-0.42)	(1.92)	(-0.78)	(-0.39)	(1.69)
λ^S	100	2.53	-0.69	16.07**	2.17	-0.98	10.55
t-stat		(1.28)	(-1.21)	(2.36)	(0.82)	(-1.38)	(1.14)
λ^S	269	0.67	-0.43	13.02**	-1.14	-0.60	10.17
t-stat		(0.58)	(-1.07)	(2.49)	(-0.76)	(-1.16)	(1.41)
λ^{MF}	—	0.69	-0.34	5.70*	0.10	-0.49	1.91
t-stat		(0.48)	(-0.90)	(1.82)	(0.05)	(-1.06)	(0.44)
T		429	429	429	267	267	267
$\bar{N}_{MF}$		997	997	997	1405	1405	1405

*$p < .10$, ** $p < .05$, *** $p < .01$

(b) Equal-Weighted Stock Portfolios

		1980 − 2015			**1993 − 2015**		
	N_S	BM	$SIZE$	$P212$	BM	$SIZE$	$P212$
λ^Δ	100	2.78	-0.44	10.48*	3.60	-0.62	10.39
t-stat		(1.52)	(-1.07)	(1.80)	(1.43)	(-1.24)	(1.36)
λ^Δ	269	-0.17	0.18	13.07***	-1.22	0.16	16.20***
t-stat		(-0.12)	(0.64)	(3.71)	(-0.70)	(0.42)	(3.53)
λ^S	100	3.35	-0.76	15.89**	3.51	-1.08	11.87
t-stat		(1.60)	(-1.36)	(2.29)	(1.24)	(-1.59)	(1.24)
λ^S	269	0.47	-0.14	18.82***	-1.19	-0.31	18.21***
t-stat		(0.47)	(-0.49)	(3.93)	(-0.90)	(-0.90)	(2.71)
λ^{MF}	—	0.69	-0.34	5.70*	0.10	-0.49	1.91
t-stat		(0.48)	(-0.90)	(1.82)	(0.05)	(-1.06)	(0.44)
T		429	429	429	267	267	267
$\bar{N}_{MF}$		997	997	997	1405	1405	1405

*$p < .10$, ** $p < .05$, *** $p < .01$

Table IV reports results of Fama-MacBeth style regressions using our characteristic pricing model. By contrast with Daniel and Titman (1997), we do not find strong evidence of compensation for characteristics in stock or mutual fund portfolios. This result is likely due to the sensitivity of characteristic-based pricing models to the choice of functional form, and average return compensation may not be linear in logs. By contrast, the prior returns characteristic is highly compensated in both value- and equal-weighted stock portfolios: a 1% increase in prior return is associated with a 10–19 basis point increase in future returns. Turning to mutual funds, compensation to this characteristic is a far lower 1.9–5.7 basis points, and only the latter value is even marginally statistically significant. Hence the implementation gap on the momentum characteristic remains prohibitively high at 56%–90% of the on-paper stock portfolio compensation. Paralleling our results in Sections IV.B–C, we conclude that mutual funds cannot reliably earn premia on characteristic versions of any of the Carhart anomalies.

E. Comparison with Cost Estimates from Other Work

Table V compares our real-world factor return estimates with estimates from selected works in the literature. Novy-Marx and Velikov (2016) estimate trading costs by summing effective bid-ask spreads of traded securities, and by their reckoning, momentum's trading costs reduce the gross strategy return from 16.0% per year to 8.16% per year (Table 3 of their paper). These positive momentum returns net-of-costs likely significantly overstate achievable returns, however, because their calculation ignores the price impact of trading that is particularly relevant to institutional investors.

Papers that consider price impact costs reach mixed conclusions on the implementability of momentum. Korajczyk and Sadka (2004) suggest that momentum profits exist only at small scales (the table reports only their returns net of proportional costs, and by their reckoning, non-proportional costs quickly overwhelm strategy returns), and Lesmond, Schill, and Zhou (2004) argue that high transaction costs preclude profitable momentum strategies altogether. Because these studies estimate transactions-cost functions using all TAQ transactions, their average implementation cost estimates smooth over heterogeneous investors and over trades unrelated to momentum strategies. Nevertheless, Lesmond, Schill, and Zhou (2004) find that momentum has an economically unimportant premium for the average trader.

Our factor compensation estimates fall on the lower end of the spectrum, and our results are most similar to Lesmond, Schill, and Zhou (2004) in that we find no net-of-cost compensation to momentum. We square our implementation cost estimates with prior work in two ways. First, we decompose implementation costs to better understand what frictions erode mutual funds' ability to capture factor premia. Section V considers the roles of shorting frictions and limitations on funds' investable universe, as well as the trade-off between tracking error and performance more generally. Second, Section VI considers cross-sectional and time-series variation in costs across funds, and

Table V: Comparison with Selected Factor Profitability Estimates from Prior Work

Table presents estimates of factor strategy returns. The top panel reports cross-sectional slopes from Fama-MacBeth regressions as in Table X. For brevity we report only the estimates in which liquidity proxy principal components appear in the time-series step, and we focus on the slopes for the full sample of mutual funds and for small mutual funds (lagged total net assets between $10 million and $50 million). As before, standard errors are Newey-West with three lags. The second panel presents value-weighted momentum strategy returns from Table IV of Korajczyk and Sadka (2004). Alphas are constructed relative to the Fama-French three factors. $\alpha_{net}^{espr.}$ and $\alpha_{net}^{qspr.}$ represent excess momentum returns net of proportional costs as measured by effective spreads and quoted spreads, respectively. The third panel reports equal-weighted strategy returns from Table 3 of Lesmond, Schill, and Zhou (2004) (value-weighted returns are not reported). r_{net}^{LDV} and r_{net}^{direct} are momentum returns net of Lesmond, Ogden, and Trzcinka (1999)-implied costs and "direct" costs (consisting of bid-ask spreads and trading commissions), respectively. The fourth panel tabulates realized strategy returns from Table IV of Frazzini, Israel, and Moskowitz (2015). The final panel reports value-weighted strategy returns net of Hasbrouck (2009)-implied effective spreads from Table 3 of Novy-Marx and Velikov (2016). Throughout returns are annualized and t statistics are reported in parentheses.

		HML	SMB	UMD
Cross-Sectional Slopes w/ PCA	λ^{MF}	2.64	0.90	1.28
1970–2016	t-stat	(1.51)	(0.53)	(0.52)
	λ^{MF}_{small}	2.55	1.37	2.62
	t-stat	(1.37)	(0.82)	(0.97)
Korajczyk and Sadka (2004)	α_{gross}			6.84***
1967–1999	t-stat			(4.54)
	$\alpha_{net}^{espr.}$			5.40***
	t-stat			(3.59)
	$\alpha_{net}^{qspr.}$			4.80***
	t-stat			(3.17)
Lesmond, Schill, and Zhou (2004)	r_{gross}			7.83***
1980–1998	t-stat			(6.22)
	r_{net}^{LDV}			0.13
	t-stat			(0.07)
	r_{net}^{direct}			2.24
	t-stat			(1.22)
Frazzini, Israel, and Moskowitz (2015)	r_{gross}	4.86	7.98***	2.26
1986–2013	t-stat	(1.12)	(3.01)	(0.40)
	r_{net}	3.51	6.52**	-0.77
	t-stat	(0.80)	(2.48)	(-0.14)
Novy-Marx and Velikov (2016)	r_{gross}	5.64***	3.96*	15.96***
1963–2013	t-stat	(2.68)	(1.66)	(4.80)
	r_{net}	5.04**	3.36	8.16**
	t-stat	(2.39)	(1.44)	(2.45)

*$p < .10$, ** $p < .05$, *** $p < .01$

we find substantial heterogeneity. Differences between "average" and "skilled" funds reconcile the lower costs seen in studies of single funds using proprietary trading data and studies of average traders in TAQ.

V. Decomposing Implementation Costs

A. The Role of Mutual Fund Shorting Constraints

The implementation gap we estimate reflects the difference between on-paper and real-world performance for zero-cost factor strategies. Such strategies consist of financing long position by shorting other stocks, for example, selling "growth" stocks to purchase "value" stocks. Institutional impediments to shorting may significantly increase costs on the short side and reduce the performance of real-world factor strategies.

In this section we adapt our Fama-MacBeth approach to evaluate the extent to which implementation gaps arise derive from shorting frictions. To do this, we consider two long-only variants of value, size, and momentum. Specifically, we first consider pure no-shorting strategies in which mutual funds borrow at the riskfree rate to invest in the long side of each factor. Our "long-only" factors are the excess returns on H, S, and U portfolios, all of which are accessible to short-sale constrained mutual funds, and we denote these long-only factors with a '+' superscript.

The typical mutual fund is highly exposed to the market—the mean and median correlations with the S&P 500 are 85% and 90%—and increasing exposure to H, S, or U may be financed by reducing a long position in other securities (e.g., the market) rather than by opening a short position. With this motivation, we also consider returns on "tilt" factors, defined as the difference between the long-factor portfolios and the market.[23] We denote the "tilt" factors with a '#' superscript. For both sets of factors, we do not modify MKT because the market factor is already in excess return form and accessible to long-only funds.

Table VI reports stock portfolio and mutual fund returns to the long-only Carhart factors. Focusing on the differences in premia earned, λ^{Δ}, relative to the baseline estimates, the long-only factor implementation costs are about 60% as large for HML^{+} and about 40% as large for UMD^{+}, but they are of comparable statistical significance.[24] As before, we find no evidence of significant implementation costs for market or long-only size factor exposures in value-weighted portfolios. Equal-weighted portfolio results are very similar, although we do find a significant SMB^{+} implementation gap because of the increased weight assigned to difficult-to-access microcaps

[23] Such tilt factors also have the advantage of closely tracking the traditional Carhart factors. For example, if H and L have comparable market capitalization for all dates, then the return to the tilt factor $HML^{\#}$ is $H-(H+L)/2$ or $HML/2$, and the accessible tilt factor is proportional to standard HML.

[24] Intriguingly, the scaling of long-only implementation costs relative to total implementation costs is in line with Israel and Moskowitz (2013)'s finding that roughly 60% of the value premium and 50% of the momentum premium are earned on the long side of the anomalies. We find that real-world trading costs are roughly proportional to premia earned in on-paper portfolios.

Table VI: Implementation Cost Estimates in Fama-MacBeth Regressions — Long-Only Factors

Table reports Fama-MacBeth estimates of the compensation for long-only factor exposure for positive and negative beta stock portfolios (second panel), domestic equity mutual funds (third panel), and their difference (top panel). Coefficients are the average cross-sectional slopes $\bar{\lambda}_k$ across monthly regressions of excess returns r_{it} on time-series betas $\hat{\beta}^+_{ik}$,

$$r_{it} = \sum_k \lambda^S_{kt} \hat{\beta}^+_{ik} 1_{i \in S} + \sum_k \lambda^{MF}_{kt} \hat{\beta}^+_{ik} 1_{i \in MF} + \epsilon_{it},\ t = 1, \ldots, T,$$

where k indexes the long-only versions of the Carhart (1997) factors—the excess returns on MKT, H, S, and U—and λ^Δ is defined as $\lambda^S - \lambda^{MF}$. Stock portfolio sets are described in Section III. All coefficients are annualized and reported in percent. Standard errors are Newey-West with three lags. t statistics are reported in parentheses.

(a) Value-Weighted Stock Portfolios

		1970 − 2016				**1993 − 2016**			
	N_S	MKT	HML^+	SMB^+	UMD^+	MKT	HML^+	SMB^+	UMD^+
λ^Δ	100	-0.61*	2.56***	0.52	3.09***	-0.32	2.10***	0.32	2.41***
t-stat		(-1.94)	(4.05)	(1.00)	(4.52)	(-0.92)	(3.85)	(0.56)	(2.98)
λ^Δ	269	-0.29	1.60***	0.02	2.85***	0.18	1.30***	-0.34	2.46***
t-stat		(-1.21)	(2.72)	(0.04)	(4.25)	(0.81)	(3.00)	(-0.61)	(2.85)
λ^S	100	6.22***	12.25***	9.19***	11.69***	7.32**	12.89***	10.84**	12.12***
t-stat		(2.59)	(4.33)	(2.85)	(4.11)	(2.24)	(3.19)	(2.54)	(3.22)
λ^S	269	6.54***	11.29***	8.68***	11.46***	7.82**	12.09***	10.17**	12.17***
t-stat		(2.73)	(3.95)	(2.68)	(4.02)	(2.41)	(2.95)	(2.38)	(3.24)
λ^{MF}	—	6.83***	9.69***	8.66***	8.60***	7.63**	10.80***	10.51**	9.71**
t-stat		(2.81)	(3.25)	(2.60)	(2.85)	(2.34)	(2.59)	(2.44)	(2.48)
T		564	564	564	564	282	282	282	282
$\bar{N}_{MF}$		1286	1286	1286	1286	2123	2123	2123	2123

$^*p < .10$, $^{**}\ p < .05$, $^{***}\ p < .01$

(b) Equal-Weighted Stock Portfolios

		1970 − 2016				**1993 − 2016**			
	N_S	MKT	HML^+	SMB^+	UMD^+	MKT	HML^+	SMB^+	UMD^+
λ^Δ	100	-0.83	3.82***	2.21***	3.64***	-0.25	3.11***	2.31***	3.27***
t-stat		(-1.66)	(6.06)	(3.37)	(5.46)	(-0.43)	(5.57)	(2.64)	(4.01)
λ^Δ	269	-0.26	3.40***	2.17***	4.25***	0.41	2.84***	2.38**	4.26***
t-stat		(-0.50)	(4.92)	(2.96)	(6.29)	(0.66)	(4.21)	(2.25)	(4.86)
λ^S	100	6.00**	13.50***	10.88***	12.24***	7.38**	13.91***	12.82***	12.98***
t-stat		(2.49)	(4.71)	(3.27)	(4.27)	(2.23)	(3.41)	(2.87)	(3.39)
λ^S	269	6.57***	13.09***	10.83***	12.86***	8.04**	13.64***	12.90***	13.98***
t-stat		(2.75)	(4.44)	(3.19)	(4.44)	(2.46)	(3.23)	(2.80)	(3.64)
λ^{MF}	—	6.83***	9.69***	8.66***	8.60***	7.63**	10.80***	10.51**	9.71**
t-stat		(2.81)	(3.25)	(2.60)	(2.85)	(2.34)	(2.59)	(2.44)	(2.48)
T		564	564	564	564	282	282	282	282
$\bar{N}_{MF}$		1286	1286	1286	1286	2123	2123	2123	2123

$^*p < .10$, $^{**}\ p < .05$, $^{***}\ p < .01$

with high average returns.

The second and third panels report the risk premia earned on the long side of each factor. On paper, long-only value and momentum premia are quite large relative to the equity premium, with gaps of 5%–6% depending on the choice of portfolio weighting and time period. These expected-return improvements earned by tilting away from the market portfolio are reflected much less strongly in mutual funds, however.

Table VII replaces long-only factors with tilt factors. Our conclusions are much the same as above: $HML^{\#}$ and $UMD^{\#}$ suffer large implementation costs in practice regardless of time period or portfolio weighting. Moreover the magnitude of these estimated costs is comparable to that of the long-only factors in Table VI: real-world underperformance is robust to assumptions on how funds implement the long side of anomalies. Focusing on the second and third panels delivers statistical assessments of the cross-column comparisons of Table VI. Mutual funds earn a marginally statistically significant premium on value tilts in the full sample and zero premium to factor tilts for all other factors and sample periods. By contrast, stocks earn robust premia to value and momentum tilts.

From both tables we conclude that the implementation costs of "long only" versions of standard factors are significant and comparable to short-side costs. Real-world shorting frictions hence explain as much as half of the high all-in implementation costs of value and momentum factors. The high cost of shorting restrictions may explain the growing popularity of levered mutual funds, e.g., "130/30" funds, for which these restrictions are less binding.

B. The Role of Investability Frictions

Implementation costs attenuate the returns to traded securities and motivate investors to depart from prescribed factor strategies. Frictions that reduce the set of investment opportunities are an important "shadow" implementation cost—analogous to the shadow price on a constraint on which stocks can be included in a portfolio—faced by real-world investors and missed by existing measures of costs. In this section we consider the role of security size in circumscribing mutual funds' investable universe. Security size is a natural candidate for explaining the performance gap between on-paper and real-world factor investing because (1) the highest returns to HML exposure are earned in the smallest stocks (Fama and French (2012), Israel and Moskowitz (2013)), and (2) low-market capitalization securities are too small to accommodate meaningful investment by large mutual funds.

The smallest stocks or "microcaps" present especially challenging environments for asset managers because of their particularly low carrying capacities and high transaction costs. Perhaps because of the challenges facing potential arbitrageurs in this space, the majority of academic anomalies only exist in these "dusty corners" of the stock market (Hou, Xue, and Zhang (2017)). To evaluate the effect of microcaps on our cost estimates, we exclude microcaps from our set of

Table VII: Implementation Cost Estimates in Fama-MacBeth Regressions — Tilt Factors

Table reports Fama-MacBeth estimates of the compensation for long-only factor exposure for positive and negative beta stock portfolios (second panel), domestic equity mutual funds (third panel), and their difference (top panel). Coefficients are the average cross-sectional slopes $\bar{\lambda}_k$ across monthly regressions of excess returns r_{it} on time-series betas $\hat{\beta}^{\#}_{ik}$,

$$r_{it} = \sum_k \lambda^S_{kt} \hat{\beta}^{\#}_{ik} 1_{i \in S} + \sum_k \lambda^{MF}_{kt} \hat{\beta}^{\#}_{ik} 1_{i \in MF} + \epsilon_{it}, \; t = 1, \ldots, T,$$

where k indexes "tilt" versions of the Carhart (1997) factors—the excess return on the market, and H, S, and U net of the market—and λ^{Δ} is defined as $\lambda^S - \lambda^{MF}$. Stock portfolio sets are described in Section III. All coefficients are annualized and reported in percent. Standard errors are Newey-West with three lags. t statistics are reported in parentheses.

(a) Value-Weighted Stock Portfolios

		1970 − 2016				**1993 − 2016**			
	N_S	MKT	$HML^{\#}$	$SMB^{\#}$	$UMD^{\#}$	MKT	$HML^{\#}$	$SMB^{\#}$	$UMD^{\#}$
λ^{Δ}	100	-0.61*	3.17***	1.13*	3.70***	-0.32	2.41***	0.64	2.72***
t-stat		(-1.94)	(4.34)	(1.77)	(5.08)	(-0.92)	(3.39)	(0.86)	(3.09)
λ^{Δ}	269	-0.29	1.89***	0.31	3.15***	0.18	1.11**	-0.52	2.27***
t-stat		(-1.21)	(3.06)	(0.58)	(4.81)	(0.81)	(2.36)	(-0.91)	(2.74)
λ^S	100	6.22***	6.03***	2.97*	5.47***	7.32**	5.58**	3.52	4.80***
t-stat		(2.59)	(4.19)	(1.94)	(4.57)	(2.24)	(2.49)	(1.64)	(2.72)
λ^S	269	6.54***	4.75***	2.15	4.92***	7.82**	4.28*	2.36	4.35**
t-stat		(2.73)	(3.28)	(1.42)	(4.18)	(2.41)	(1.85)	(1.11)	(2.54)
λ^{MF}	—	6.83***	2.86*	1.83	1.77	7.63*	3.16	2.88	2.08
t-stat		(2.81)	(1.92)	(1.17)	(1.47)	(2.34)	(1.36)	(1.34)	(1.24)
T		564	564	564	564	282	282	282	282
$\bar{N}_{MF}$		1286	1286	1286	1286	2123	2123	2123	2123

$^{*}p < .10$, $^{**}\ p < .05$, $^{***}\ p < .01$

(b) Equal-Weighted Stock Portfolios

		1970 − 2016				**1993 − 2016**			
	N_S	MKT	$HML^{\#}$	$SMB^{\#}$	$UMD^{\#}$	MKT	$HML^{\#}$	$SMB^{\#}$	$UMD^{\#}$
λ^{Δ}	100	-0.83	4.65***	3.04***	4.47***	-0.25	3.36***	2.56**	3.52***
t-stat		(-1.66)	(5.56)	(3.16)	(5.92)	(-0.43)	(4.31)	(2.18)	(3.81)
λ^{Δ}	269	-0.26	3.66***	2.43**	4.52***	0.41	2.43**	1.97	3.85***
t-stat		(-0.50)	(3.97)	(2.28)	(6.33)	(0.66)	(2.55)	(1.38)	(4.34)
λ^S	100	6.00**	7.50***	4.88***	6.24***	7.38**	6.53***	5.44**	5.60***
t-stat		(2.49)	(5.06)	(2.75)	(4.98)	(2.23)	(2.98)	(2.23)	(3.08)
λ^S	269	6.57***	6.52***	4.27**	6.29***	8.04**	5.60**	4.85*	5.93***
t-stat		(2.75)	(4.00)	(2.26)	(4.82)	(2.46)	(2.27)	(1.82)	(3.23)
λ^{MF}	—	6.83***	2.86*	1.83	1.77	7.63*	3.16	2.88	2.08
t-stat		(2.81)	(1.92)	(1.17)	(1.47)	(2.34)	(1.36)	(1.34)	(1.24)
T		564	564	564	564	282	282	282	282
$\bar{N}_{MF}$		1286	1286	1286	1286	2123	2123	2123	2123

$^{*}p < .10$, $^{**}\ p < .05$, $^{***}\ p < .01$

stock portfolios. We follow Fama and French (2008) and Hou, Xue, and Zhang (2017) in defining microcaps as stocks with market capitalization less than the 20th percentile of NYSE market capitalization, and we implement this filter by dropping the smallest-size portfolios from double-sorted size-value, size-beta, size-prior return, and size-Amihud portfolios. This exclusion eliminates a fifth of the portfolios but only 3% of market capitalization (Fama and French (2008)).

Table VIII reports Fama-MacBeth estimates of factor premia on this set of stock portfolios. We present only value-weighted results because we are interested in downweighting tiny stocks to reflect the investable universe. Our main finding is that microcaps indeed explain some of the measured performance attrition for value and momentum strategies, but not enough to close the measured implementation gap. As a useful placebo, the gap on replicating performance on the value-weighted market changes by at most a few basis points.

In the 1970–2016 sample, both value and momentum compensation are about 1% smaller in the stock portfolios in which microcaps are excluded. This difference persists for value in the more recent sample, echoing Fama and French (2012) and Israel and Moskowitz (2013), but it roughly halves for momentum. Nevertheless, the performance gap between non-microcap stock portfolios and mutual funds remains economically large and statistically robust. If mutual funds indeed cannot invest in microcap stocks, this narrowing of the investable universe explains about one third of the implementation gap for value and about one sixth of the implementation gap for momentum.

C. Tracking Error and the Performance of Factor Strategies

Mutual funds face a trade-off between following high-cost canonical factor strategies and deviating from those strategies to capture the bulk of factor premia at lower costs. Benchmark-based performance evaluation in particular pushes funds to mimic factor benchmarks despite the potentially lower Sharpe ratios of doing (Basak and Pavlova (2013)). In this section we split our sample into quintiles by Carhart four-factor R^2s to evaluate whether variation in tracking error is associated with factor strategy performance.

This split also serves a second function in combating bias in our implementation cost estimates that arises from misspecifying mutual fund strategies. Bias occurs if incidental factor exposures incurred by other activities are cross-sectionally correlated with mutual fund returns. For especially high R^2 values in the time-series regressions, the scope for omitted variable bias is small if coefficients are stable across specifications, as they are in our study (Oster (2017)).

To perform this split, we run the time-series regressions fund-by-fund as before using the Carhart (1997) model, and we sort funds into one of five equally-spaced bins at each date based on the R^2 of its time-series regression. Funds with high R^2 have returns nearly spanned by the academic strategies,[25] and these funds have low tracking error and little scope for omitted strategies that might

[25]The top R^2 quintile also includes many index funds (roughly 5% of observations in our sample). Because index funds seek to replicate factors at the lowest possible cost, we expect factor compensation estimates for this quintile to represent the best case achievable for passive or active mutual funds.

Table VIII: Implementation Cost Estimates in Fama-MacBeth Regressions — Microcaps Excluded

Table reports Fama-MacBeth estimates of the compensation for factor exposure in value-weighted stock portfolios in the baseline regressions (top panel) and regressions with liquidity principal components (bottom panel). Coefficients are the average cross-sectional slopes $\bar{\lambda}_k$ across monthly regressions of excess returns r_{it} on time-series betas $\hat{\beta}_{ik}$,

$$r_{it} = \sum_k \lambda_{kt}^S \hat{\beta}_{ik} 1_{i \in S} + \sum_k \lambda_{kt}^{MF} \hat{\beta}_{ik} 1_{i \in MF} + \epsilon_{it},\ t = 1, \ldots, T,$$

where k indexes the four Carhart (1997) factors and λ^Δ is defined as $\lambda^S - \lambda^{MF}$. First-stage regression estimates in the second panel include these factors, the first principal component of market liquidity proxies, and the first principal component of funding liquidity proxies. Liquidity proxies and stock portfolio sets are described in Section III, with the important distinction that all portfolios with the smallest market capitalization quintile are excluded in the $N_S = 80$ specifications. All coefficients are annualized and reported in percent. Standard errors are Newey-West with three lags. t statistics are reported in parentheses.

(a) Baseline Specification

		1970 − 2016				**1993 − 2016**			
	N_S	MKT	HML	SMB	UMD	MKT	HML	SMB	UMD
λ^Δ	80	-0.37	2.85***	0.70	6.14***	0.10	2.06**	-0.42	3.49**
t-stat		(-1.39)	(3.84)	(1.24)	(4.66)	(0.32)	(2.44)	(-0.61)	(2.12)
λ^S	80	6.61***	5.47***	1.71	7.68***	7.88***	4.37	1.78	5.23
t-stat		(2.74)	(3.03)	(1.07)	(3.31)	(2.40)	(1.57)	(0.80)	(1.41)
λ^S	100	6.60***	6.43***	1.27	8.72***	7.67**	5.43*	1.96	6.01
t-stat		(2.75)	(3.51)	(0.75)	(3.74)	(2.35)	(1.93)	(0.81)	(1.60)
λ^{MF}	–	6.98***	2.62	1.01	1.54	7.78**	2.31	2.20	1.73
t-stat		(2.86)	(1.51)	(0.59)	(0.63)	(2.38)	(0.83)	(0.92)	(0.45)

$^*p < .10$, $^{**}\ p < .05$, $^{***}\ p < .01$

(b) Liquidity PCs

		1970 − 2016				**1993 − 2016**			
	N_S	MKT	HML	SMB	UMD	MKT	HML	SMB	UMD
λ^Δ	80	-0.37	2.97***	0.55	6.42***	0.13	2.25***	-0.69	4.55***
t-stat		(-1.34)	(3.81)	(0.97)	(4.82)	(0.4)	(2.64)	(-0.96)	(2.65)
λ^S	80	6.63***	5.60***	1.45	7.70***	7.93**	4.34	1.53	5.31
t-stat		(2.76)	(3.10)	(0.91)	(3.31)	(2.43)	(1.56)	(0.68)	(1.43)
λ^S	100	6.55***	6.71***	1.26	8.77***	7.68**	5.38*	1.98	5.99
t-stat		(2.74)	(3.63)	(0.74)	(3.76)	(2.37)	(1.90)	(0.82)	(1.59)
λ^{MF}	–	6.99***	2.64	0.90	1.28	7.80**	2.09	2.22	0.76
t-stat		(2.87)	(1.51)	(0.53)	(0.52)	(2.41)	(0.74)	(0.92)	(0.20)
T		564	564	564	564	282	282	282	282
$\bar{N}_{MF}$		1286	1286	1286	1286	2123	2123	2123	2123

$^*p < .10$, $^{**}\ p < .05$, $^{***}\ p < .01$

complicate the interpretations of λ^{MF} and λ^{Δ}. Conversely funds with lower R^2 either (1) implement academic strategies with greater discretion and/or tracking error or (2) implement strategies that we cannot observe. We then construct cross-sectional mutual fund factor compensation estimates for each R^2 group as in Tables II–III.

Table IX presents results from the splits by explanatory power of the four-factor model on the full 1970–2016 sample. The decomposition by R^2 delivers two results related to how funds implement asset pricing factors. First, perhaps unsurprisingly, funds that track factors more closely are generally more efficient at earning factor premia. Performance differences across fund quintiles are statistically significant for MKT, SMB, and UMD factors. This result is reversed for the market factor, however. Funds with greater deviations from the academic factors typically achieve greater returns to market beta. Small deviations from the CRSP market go a long way toward improving returns: market exposure is compensated 41–81 basis points more in mutual funds that track the four factors less well. This finding reinforces the importance of using flexible approaches to measuring implementation costs that are robust to real-world departures from the academic factors.

More importantly for our study, relative to the average mutual fund, funds with the highest R^2s achieve economically similar performance on market and value factors and somewhat higher performance on size and momentum factors. Value premia are about 1% larger among the funds most closely mimicking academic factors, and compensation for value exposure is significantly different from zero at the 5% level for the highest-R^2 quintile. Likewise, returns to momentum exposure for this group are nearly triple those of the typical mutual fund, and they are statistically significant with strength depending on specification. Even so, the funds that most closely track the four academic factors continue to significantly underperform the on-paper factors. The best-performing R^2 segments for value see an implementation gap of 1%–2% relative to the stock portfolios, and the momentum implementation gap for these funds is roughly half the on-paper momentum premium.

VI. Cost Estimates Across Funds and Time

A. Implementation Costs Across Funds

With the exception of the breakdown by four-factor R^2s, our analysis thus far considers the implementation costs of factor strategies for an average mutual fund, with no attention paid to heterogeneous characteristics and costs. Variation in investors' trading technologies may drive a wedge between a typical asset manager and the marginal investor in an anomaly. By dividing asset managers into groups we can learn whether factors are broadly (in)accessible or whether they generate positive net-of-costs returns for a subset of managers. In this section, we briefly demonstrate the utility of our cross-sectional approach for examining segments of asset managers.

Motivated by extensive work relating fund size to gross-of-fees performance (e.g., Berk and

Table IX: Fama-MacBeth Slopes for Stocks and Mutual Funds — R^2 Quintile Splits

Table reports Fama-MacBeth estimates of the compensation for factor exposure for domestic equity mutual funds. Coefficients are the average cross-sectional slopes $\bar{\lambda}_k^g$ across monthly regressions of excess returns r_{it} on time-series betas $\hat{\beta}_{ik}$ for each group of mutual funds g,

$$r_{it} = \sum_k \lambda_{kt}^{MF,g} \hat{\beta}_{ik} + \epsilon_{it}, \; t = 1, \ldots, T, \; g = 1, \ldots, 5,$$

where k indexes the four Carhart (1997) factors. We partition mutual funds into five equal groups sorted by time-series regression R^2s from the Carhart model, where R^2 cutoffs are set at each date based on the sample of live funds. "5" indicates the highest R^2 funds, and "1" indicates the lowest R^2 funds. The first column reports the average R^2s across all fund-date observations within each R^2 group. Subsequent first-stage regression estimates include these factors only (first columns) and the first principal component of market and funding liquidity proxies (second columns). Liquidity proxies and stock portfolio sets are described in Section III. λ_5^{Δ} is the difference between compensation for factor exposure between the 269 stock portfolios and the highest R^2 fund group. All coefficients are annualized and reported in percent. Standard errors are Newey-West with three lags. t statistics are reported in parentheses. The final three rows report the p values of F tests of coefficients being jointly different from zero, of F tests of equality of coefficients, and of Patton and Timmermann (2010)'s test of non-monotonicity of coefficients.

		No Liquidity Proxies				**Liquidity PCs**			
	$\bar{R}^2$	MKT	HML	SMB	UMD	MKT	HML	SMB	UMD
λ_5^{MF}	94.2%	6.50***	3.60**	1.78	4.59*	6.52***	3.97**	1.74	5.36**
t-stat		(2.69)	(1.99)	(1.04)	(1.68)	(2.71)	(2.15)	(1.02)	(1.99)
λ_4^{MF}	89.9%	6.91***	2.93*	2.67	0.73	6.89***	3.04*	2.72	0.34
t-stat		(2.82)	(1.70)	(1.57)	(0.26)	(2.83)	(1.74)	(1.59)	(0.12)
λ_3^{MF}	86.0%	7.31***	3.00*	0.09	3.23	7.28***	2.89	0.11	2.30
t-stat		(2.96)	(1.68)	(0.05)	(1.20)	(2.95)	(1.59)	(0.06)	(0.86)
λ_2^{MF}	79.9%	7.29***	2.66	1.15	-0.81	7.27***	2.42	1.14	-1.45
t-stat		(2.98)	(1.48)	(0.64)	(-0.31)	(2.98)	(1.32)	(0.64)	(-0.55)
λ_1^{MF}	55.4%	7.00***	2.93	-0.98	2.08	7.14***	3.44*	-1.51	2.38
t-stat		(2.80)	(1.52)	(-0.49)	(0.72)	(2.85)	(1.75)	(-0.75)	(0.81)
λ^{MF}	81.1%	6.98***	2.62	1.01	1.54	6.99***	2.64	0.90	1.28
t-stat		(2.86)	(1.51)	(0.59)	(0.63)	(2.87)	(1.51)	(0.53)	(0.52)
λ^S	–	6.77***	5.20***	0.94	8.85***	6.77***	5.47***	0.89	8.84***
t-stat		(2.82)	(2.84)	(0.56)	(3.80)	(2.83)	(2.94)	(0.53)	(3.78)
λ_5^{Δ}	–	0.27	1.60**	-0.84	4.26***	0.25	1.49*	-0.85	3.48**
t-stat		(1.17)	(2.02)	(-1.45)	(2.80)	(1.08)	(1.85)	(-1.46)	(2.21)
$\lambda = 0$		0.00***	0.41	0.00***	0.02**	0.00***	0.22	0.00***	0.00***
$\lambda =$		0.00***	0.83	0.00***	0.01***	0.01**	0.47	0.00***	0.00***
$\Delta\lambda \lessgtr 0$		0.27	0.08*	0.79	0.68	0.13	0.37	0.85	0.78

$^*p < .10$, $^{**}\ p < .05$, $^{***}\ p < .01$

Green (2004), Pastor, Stambaugh, and Taylor (2015), and Berk and van Binsbergen (2015)), we split fund groups into groups based on lagged total net assets (TNA). We then run our second-stage cross-sectional regressions (Equation (2)) separately for each asset manager TNA group.[26] We set aside funds with less than \$10 million in assets because selection into this group implies that the fund has lost money (we retain observations only after funds reach \$10 million in assets to avoid incubation bias).

Table X presents results from these segmented regressions on the full 1970–2016 sample. As in Tables II–III, mutual funds generally achieve returns to market factor exposure comparable to those of on-paper stock portfolios. HML also earns positive compensation for most TNA groups, but returns to HML are not statistically different from zero in both specifications, with the possible exception of the mega-funds group. Point estimates for returns to SMB are positive for all fund size groups excluding micro funds, but SMB compensation estimates are not statistically distinguishable from zero or from each other.

Focusing on momentum, we estimate large differences in compensation across mutual fund size categories, with the smallest funds earning 5%–6% more per unit momentum beta than the largest funds. Notwithstanding the greater momentum-strategy performance of small funds, we nonetheless continue to reject the hypothesis that these funds perform as well as on-paper stock portfolios. We can also reject non-monotonicity of momentum compensation across size categories using the bootstrap test of Patton and Timmermann (2010): we find momentum strategy performance is significantly decreasing in fund size. This feature makes intuitive sense in that momentum is a high-turnover strategy, and larger funds suffer greater market impact costs in implementing momentum than smaller funds.

From this analysis we conclude that heterogeneity among asset managers is important when considering the net-of-costs returns to momentum. When we focus only on small mutual funds, net-of-costs compensation to momentum looks quite different from that of the average fund. Which momentum premium is of greater interest hinges on whether the researcher evaluates a broad set of firms, as in benchmarking applications, or marginal investors, as in discussions of market efficiency. Intriguingly we find that the largest mutual funds earn the most negative compensation for momentum exposure, suggesting that the firm examined in Frazzini, Israel, and Moskowitz (2015) is exceptional, or that non-mutual fund asset managers have different compensation schedules for factor exposure. More generally, our results reveal that the longstanding disagreement on the profitability of momentum strategies likely arises in part because market-wide and single-firm analyses, e.g., Lesmond, Schill, and Zhou (2004) and Frazzini, Israel, and Moskowitz (2015), respectively, focus on different sets of institutions with different factor implementation technologies.

[26] Groups are assigned separately for each date with cutoffs in terms of December 2016 dollars. The micro-fund group has $TNA_t < \$10M$ and comprises 5.2% of the data. The small-fund group has $\$10M < TNA_t < \$50M$ and comprises 22.8% of the data. The medium-fund group has $\$50M < TNA_t < \$250M$ and comprises 31.8% of the data. The large-fund group has $\$250M < TNA_t < \$1B$ and comprises 22.5% of the data. The mega-fund group has $TNA_t > \$1B$ and comprises 17.7% of the data.

Table X: Fama-MacBeth Slopes for Stocks and Mutual Funds — Size Splits

Table reports Fama-MacBeth estimates of the compensation for factor exposure for domestic equity mutual funds. Coefficients are the average cross-sectional slopes $\bar{\lambda}_k^g$ across monthly regressions of excess returns r_{it} on time-series betas $\hat{\beta}_{ik}$ for each group of mutual funds g,

$$r_{it} = \sum_k \lambda_{kt}^{MF,g} \hat{\beta}_{ik} + \epsilon_{it},\ t = 1, \ldots, T,\ g = 1, \ldots, 5,$$

where k indexes the four Carhart (1997) factors. We partition mutual funds into four groups based on one-month lagged total net assets (TNA), with TNA cutoffs specified in December 2016 USD. The micro-fund group has $TNA_t < \$10M$, the small-fund group has $\$10M < TNA_t < \$50M$, the medium-fund group has $\$50M < TNA_t < \$250M$, the large-fund group has $\$250M < TNA_t < \$1B$, and the mega-fund group has $TNA_t > \$1B$. First-stage regression estimates include these factors only (first columns) and the first principal component of market and funding liquidity proxies (second columns). Liquidity proxies and stock portfolio sets are described in Section III. λ_{small}^{Δ} is the difference between compensation for factor exposure between the 269 stock portfolios and the small-fund group. All coefficients are annualized and reported in percent. Standard errors are Newey-West with three lags. t statistics are reported in parentheses. The final three rows report the p values of F tests of coefficients being jointly different from zero, of F tests of equality of coefficients, and of Patton and Timmermann (2010)'s test of non-monotonicity of coefficients.

	No Liquidity Proxies				**Liquidity PCs**			
	MKT	HML	SMB	UMD	MKT	HML	SMB	UMD
λ_{mega}^{MF}	6.66***	3.11*	1.89	-2.53	6.67***	3.15*	1.94	-2.77
t-stat	(2.74)	(1.67)	(1.05)	(-0.75)	(2.75)	(1.66)	(1.08)	(-0.84)
λ_{large}^{MF}	6.85***	2.78	0.90	0.86	6.86***	2.83	0.91	0.04
t-stat	(2.78)	(1.54)	(0.52)	(0.31)	(2.79)	(1.54)	(0.53)	(0.02)
λ_{medium}^{MF}	7.02***	2.45	0.90	2.36	7.00***	2.45	0.96	1.76
t-stat	(2.87)	(1.41)	(0.52)	(0.92)	(2.86)	(1.37)	(0.55)	(0.68)
λ_{small}^{MF}	7.36***	2.94	1.20	3.40	7.30***	2.55	1.37	2.62
t-stat	(2.98)	(1.64)	(0.72)	(1.25)	(2.96)	(1.37)	(0.82)	(0.97)
λ_{micro}^{MF}	7.18***	2.60	-2.68	-0.24	7.18***	2.54	-3.29	-0.04
t-stat	(2.94)	(1.11)	(-1.32)	(-0.06)	(2.92)	(1.11)	(-1.59)	(-0.01)
λ_{small}^{Δ}	-0.59	2.26**	-0.26	5.45***	-0.53	2.92**	-0.48	6.22***
t-stat	(-1.59)	(2.22)	(-0.34)	(3.32)	(-1.4)	(2.58)	(-0.62)	(3.76)
λ^{MF}	6.98***	2.62	1.01	1.54	6.99***	2.64	0.90	1.28
t-stat	(2.86)	(1.51)	(0.59)	(0.63)	(2.87)	(1.51)	(0.53)	(0.52)
$\lambda = 0$	0.01***	0.46	0.56	0.11	0.01**	0.52	0.52	0.14
$\lambda =$	0.13	0.81	0.46	0.13	0.20	0.83	0.44	0.13
$\Delta\lambda \lessgtr 0$	0.01***	0.28	0.20	0.01***	0.01***	0.06*	0.28	0.01***

$^{*}p < .10$, ** $p < .05$, *** $p < .01$

B. Implementation Costs Over Time

The preceding analysis considers how implementation costs vary in the cross-section. In this section, we investigate determinants of time-series variation in implementation costs. Figure II plots the log return of the "on-paper" factor-mimicking portfolios minus the log return of the corresponding mutual-fund factor-mimicking portfolio. To do this we invoke the interpretation of Fama-MacBeth coefficients λ_{kt} as the date t return on a portfolio with a unit loading on factor k and zero loading on all other factors. Our series is the centered rolling difference in performance,

$$y_k(t) = \sum_{s=t-6}^{t+6} \log\left(1 + \lambda_{kt}^{S}\right) - \log\left(1 + \lambda_{kt}^{MF}\right) \approx \sum_{s=t-6}^{t+6} \lambda_{kt}^{\Delta}. \tag{9}$$

The four panels of Figure II depict factor implementation costs for each set of liquidity proxies using the 269 stock portfolios as the on-paper return benchmark. Although magnitudes vary slightly across specifications, the two slope series are highly similar for each factor. The implementation gap is clearly rank-ordered as UMD, HML, SMB, and MKT, with large and positive implementation gaps for UMD and HML, no implementation gap for SMB, and a small negative implementation gap for MKT. The difference series are also affected by macroeconomic events. All four implementation gaps fall before the end of the tech bubble of the late 1990s and rise during the subsequent crash and the Great Recession of 2007–2009. One interpretation of this feature is that factor returns are most accessible by investment managers when market liquidity is abundant and funding constraints are unlikely to be binding.

Perhaps the most intriguing feature of Figure II is the absence of a trend in strategy implementation costs. This feature contrasts with well-documented secular declines in bid-ask spreads and commissions since 1970 (e.g., Jones (2002) and Corwin and Schultz (2012)). An equilibrium perspective on the size of the asset management sector suggests why we instead obtain a stationary time series.[27] As trading technology improves and equity intermediation becomes more competitive, the cost of trading the first dollar of a factor strategy declines. Perceived sector-level alphas increase for factor investors, and aggregate inflows attract new entrants (as in Figure I) or contribute to the growth of existing fund managers (as in Berk and Green (2004)). These inflows increase the scale of factor investing, which in turn increases non-proportional transactions costs such as price impact. In equilibrium this process continues until factor alphas fall to zero for the marginal dollar. Consequently the *average* dollar invested in factor strategies may see no reduction in implementation costs despite improvements in trading technology.

This conjectured equilibrium adjustment mechanism hinges on non-proportional trading costs, but it generates a testable prediction that industry-level inflows increase implementation costs of factor strategies. We analyze this relationship between implementation costs, flows, and illiquidity

[27] Augmented Dickey-Fuller tests reject the null of a unit root in implementation costs at the 0.1% significance level in all series.

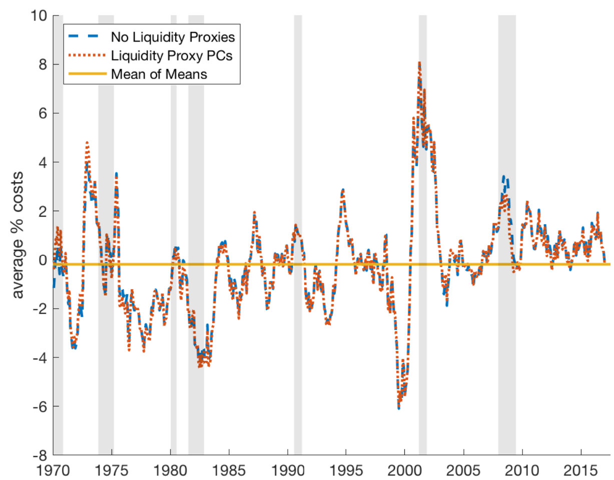
Figure II: Rolling Performance Difference Between Mutual Funds and Stocks ($\beta = 1$)

Figures plot the rolling difference between log Fama-MacBeth cross-sectional slopes for stock portfolios (S) and mutual funds (MF). Each series $y_k(t)$ equals the centered rolling difference

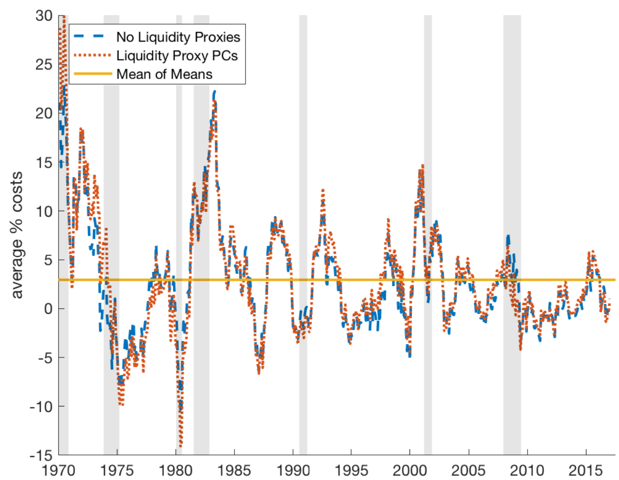
$$y_k(t) = \sum_{s=t-6}^{t+6} \log\left(1 + \lambda_{kt}^{S}\right) - \log\left(1 + \lambda_{kt}^{MF}\right),$$

where λ_{kt} are cross-sectional slopes from monthly regressions of excess returns r_{it} on time-series betas $\hat{\beta}_{ik}$. Each figure plots differences in slopes (1) when no liquidity proxies are included in the time-series regressions and (2) when the first principal component of market liquidity proxies and of funding liquidity proxies are included. Stock portfolio slopes are estimated using all 269 portfolios described in Section III. Solid lines depict averages of series means. NBER recessions are in gray.

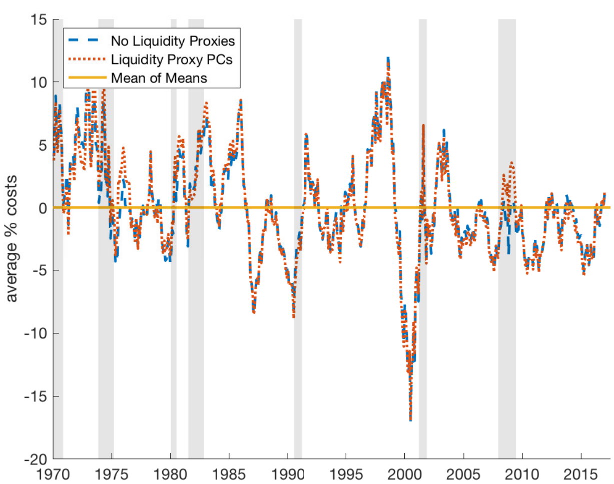
(a) MKT Costs



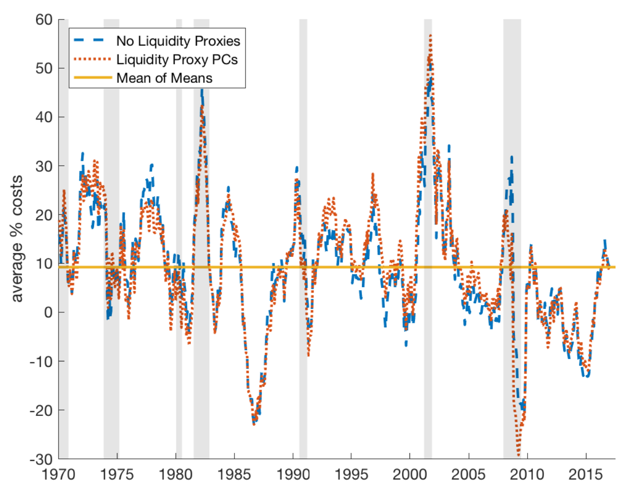
(b) HML Costs



(c) SMB Costs



(d) UMD Costs



more formally by examining relating the cost time series with liquidity and fund flow proxies. We start by constructing illiquidity proxies as the first principal components of market liquidity proxies and of funding liquidity proxies, as described in Section IV.C. We also construct flow variables to capture costs associated with fund inflows and outflows. Fund flows are the component of asset growth not explained by returns,

$$flow_{it} = \frac{TNA_{it}}{TNA_{i,t-1}} - (1 + R_{it}). \tag{10}$$

We summarize the distribution of flows with its first and second cross-sectional moments—the cross-sectional average fund flow ($MFLOW$) and the cross-sectional dispersion in fund flows ($DFLOW$). In addition to speaking to returns to scale, flow variables are a natural candidate for explaining trading costs because large flows into the mutual fund sector or reshuffling of assets among mutual funds generates liquidity demands. To enhance interpretability, we normalize all right-hand-side variables to have mean zero and standard deviation one.

Table XI reports results from a regression of λ_{kt}^{Δ} on the liquidity and fund flow proxies,

$$\lambda_{kt}^{\Delta} = \alpha + \beta_{MFLOW} MFLOW + \beta_{DFLOW} DFLOW + \beta_{ML} PC_{ML} + \beta_{FL} PC_{FL} + \epsilon_{kt}. \tag{11}$$

We report only value-weighted results for the 269 stock portfolios because relations between costs and liquidity proxies are quite similar for value-weighted and equal-weighted stock portfolios and for 100 and 269 stock portfolios.

We draw four lessons from Table XI. First, the *time-invariant* component of implementation costs from Equation (4) is large and positive for these factors, as evidenced by the constant terms for HML and UMD. Second, focusing on flows, average inflows are weakly associated with higher implementation costs for value and momentum factors, and cross-sectional dispersion in flows is weakly associated with lower implementation costs for these factors. We find no flow-cost relations for market and size factors, as is expected because these costs are small in magnitude to begin with. We interpret these relations as suggestive evidence that (1) inflows are expensive from a transactions-cost standpoint for funds trading value and momentum strategies, thereby contributing to diseconomies of scale and stationary average implementation costs, and (2) reallocation of funds within the mutual fund sector may increase liquidity trading (in a Kyle (1985) sense), thereby reducing average transactions costs for value and momentum traders. Third, focusing on illiquidity principal components, market illiquidity increases implementation costs, and particularly so for value and momentum strategies. Intuitively trading becomes more expensive when market liquidity is low. Fourth, funding illiquidity *decreases* implementation costs (again most strongly for HML and UMD). We conjecture that mutual funds are insulated from funding liquidity shocks that affect highly levered institutional asset managers like hedge funds (Sadka (2010) and Boyson, Stahel, and Stulz (2010), among others, discuss hedge funds' particular vulnerability to funding

Table XI: Liquidity, Flows, and the Implementation Gap

Table reports estimates of a regression of annualized implementation gaps λ^{Δ}_{kt} on the first principal component of funding liquidity proxies (FL) and market liquidity proxies (ML), as well as the cross-sectional average and standard deviation of fund inflows ($MFLOW$ and $DFLOW$, respectively),

$$\lambda^{\Delta}_{kt} = \alpha + \beta_{MFLOW} MFLOW + \beta_{DFLOW} DFLOW + \beta_{ML} PC_{ML} + \beta_{FL} PC_{FL} + \epsilon_{kt},$$

where k indexes the four Carhart (1997) factors. Implementation costs are estimated with principal components of market and funding liquidity proxies in the time-series regressions for mutual fund and stock portfolio betas. Liquidity proxies and stock portfolio sets are described in Section III. Illiquidity principal components have unit standard deviation and are constructed to be negatively correlated with the market return and positively correlated with the VIX/VXO. Fund flows are ${}^{TNA_{it}}/{}_{TNA_{i,t-1}} - (1 + R_{it})$. Standard errors are Newey-West with three lags. t statistics are reported in parentheses.

(a) Baseline Specification

	MKT			HML			SMB			UMD		
β_{MFLOW}	0.07		-0.02	1.91**		1.30	-0.02		-0.19	0.65		-0.03
t-stat	(0.27)		(-0.06)	(2.21)		(1.47)	(-0.04)		(-0.34)	(0.50)		(-0.03)
β_{DFLOW}	-0.05		-0.10	-2.16*		-2.45**	0.13		0.09	-1.81		-2.09
t-stat	(-0.13)		(-0.27)	(-1.87)		(-2.15)	(0.23)		(0.17)	(-0.92)		(-1.08)
β_{ML}		0.53*	0.55*		2.80***	3.03***		0.44	0.44		2.75	3.03
t-stat		(1.65)	(1.68)		(2.81)	(3.31)		(0.59)	(0.59)		(1.53)	(1.64)
β_{FL}		-0.46	-0.49		-3.08***	-3.08***		-0.68	-0.73		-2.92	-0.73
t-stat		(-1.34)	(-1.35)		(-3.28)	(-3.12)		(-0.88)	(-0.88)		(-1.57)	(-0.88)
α	-0.21	-0.21	-0.21	2.59***	2.59***	2.59***	-0.07	-0.07	-0.07	7.30***	7.30***	7.30***
t-stat	(-0.88)	(-0.88)	(-0.88)	(3.85)	(3.87)	(3.92)	(-0.14)	(-0.14)	(-0.14)	(5.59)	(5.57)	(5.65)

$^{*}p < .10$, $^{**}\ p < .05$, $^{***}\ p < .01$

(b) Liquidity PCs

	MKT			HML			SMB			UMD		
β_{MFLOW}	0.10		-0.02	2.38**		1.55	-0.14		-0.25	1.87		0.00
t-stat	(0.39)		(-0.09)	(2.50)		(1.64)	(-0.29)		(-0.46)	(1.42)		(0.00)
β_{DFLOW}	-0.02		-0.07	-1.95		-2.28*	0.01		-0.07	-1.50		-1.70
t-stat	(-0.05)		(-0.19)	(-1.57)		(-1.90)	(0.02)		(-0.12)	(-0.76)		(-0.91)
β_{ML}		0.57*	0.58*		3.42***	3.61***		0.76	0.79		2.38	2.61
t-stat		(1.82)	(1.83)		(3.38)	(3.88)		(0.90)	(0.93)		(1.41)	(1.51)
β_{FL}		-0.61*	-0.63*		-4.12***	-4.01***		-0.54	-0.64		-6.69***	-0.64
t-stat		(-1.79)	(-1.77)		(-4.28)	(-4.05)		(-0.69)	(-0.76)		(-3.77)	(-0.76)
α	-0.22	-0.22	-0.22	2.83***	2.83***	2.83***	-0.02	-0.02	-0.02	7.55***	7.55***	7.55***
t-stat	(-0.91)	(-0.92)	(-0.92)	(3.92)	(4.03)	(4.08)	(-0.03)	(-0.03)	(-0.03)	(5.73)	(5.92)	(5.98)

$^{*}p < .10$, $^{**}\ p < .05$, $^{***}\ p < .01$

liquidity shocks), and hence mutual funds can acquire the ingredients of factor strategies from distressed asset managers at a discount during times of strained funding liquidity.

VII. Conclusion

Existing methods for assessing the implementation costs of financial market anomalies use proprietary trading data for single firms or market-wide trading data combined with parametric transactions cost models. We propose an extension of Fama-MacBeth regression to estimate implementation costs using only returns data from stocks and mutual funds. Doing so frees us from the requirement of specifying factor trading strategies and transaction costs models that may be incomplete or misspecified. Moreover, the ready availability of returns data for a large universe of investment managers enables detailed investigation of factor implementation costs in the cross-section and over time.

We demonstrate that mutual funds are generally poorly compensated for exposure to some common risk factors. Our estimates based on Fama-MacBeth regressions imply that implementation costs erode almost the entirety of the return to value and momentum strategies for typical mutual funds, but have little effect on market and size factor strategies. Taken together, these results paint a sobering picture of the real-world returns to the most important financial market anomalies. These costs derive in part from institutional constraints often ignored in studies of academic anomalies, such as shorting and investability constraints, but even these frictions do not fully explain high observed costs. We find suggestive evidence that unprofitable deviations from standard academic strategies and greater market impact associated with fund growth contribute to the remainder.

Sample splits reveal considerable heterogeneity in implementation costs among funds. Using a very different estimation method and data set, our results agree with Lesmond, Schill, and Zhou (2004)'s analysis that momentum profits in particular may be out of reach for a typical asset manager. However, we find smaller funds and funds that better track academic factors tend to perform significantly better in earning factor premia than larger funds and funds with greater tracking error. In this respect markets may be efficient from the perspective of an average mutual fund, even if some segments of the mutual fund space see a very different picture of risk and return net-of-costs. Analyses using proprietary data from single funds cannot reveal such heterogeneity.

The nonparametric, market-based method for estimating all-in implementation costs proposed in this paper can be viewed as an independent check on prior work because it differs in both estimation strategy and data employed. The assumptions underlying our approach are few and transparent, and our stark findings on realizable factor premia obtain under a wide range of alternative specifications. While we do not anticipate resolving a decades-old dispute on whether momentum is accessible to typical investors, our approach forces a conversation about the palatability of assumptions on representativeness, price impact, and the like made in existing work.

References

Abadie, Alberto and Guido W. Imbens. 2006. "Large Sample Properties of Matching Estimators for Average Treatment Effects." *Econometrica* 74 (1):235–267.

———. 2011. "Bias-Corrected Matching Estimators for Average Treatment Effects." *Journal of Business & Economic Statistics* 29 (1):1–11.

Amihud, Yakov. 2002. "Illiquidity and Stock Returns: Cross-Section and Time-Series Effects." *Journal of Financial Markets* 5 (1):31 – 56.

Arnott, Robert D., Vitali Kalesnik, and Lillian J. Wu. 2017. "The Incredible Shrinking Factor Return." Working paper.

Asness, Clifford S., Tobias J. Moskowitz, and Lasse Heje Pedersen. 2013. "Value and Momentum Everywhere." *Journal of Finance* 68 (3):929–985.

Banz, Rolf W. 1981. "The Relationship Between Return and Market Value of Common Stocks." *Journal of Financial Economics* 9 (1):3 – 18.

Barber, Brad M., Xing Huang, and Terrance Odean. 2016. "Which Factors Matter to Investors? Evidence from Mutual Fund Flows." *Review of Financial Studies* 29 (10):2600–2642.

Barroso, Pedro and Pedro Santa-Clara. 2015. "Momentum Has Its Moments." *Journal of Financial Economics* 116 (1):111 – 120.

Basak, Suleyman and Anna Pavlova. 2013. "Asset Prices and Institutional Investors." *American Economic Review* 103 (5):1728–58. URL http://www.aeaweb.org/articles?id=10.1257/aer.103.5.1728.

Basu, S. 1977. "Investment Performance of Common Stocks in Relation to Their Price-Earnings Ratios: A Test of the Efficient Market Hypothesis." *Journal of Finance* 32 (3):663–682.

Berk, Jonathan B. 2000. "Sorting Out Sorts." *Journal of Finance* 55 (1):407–427.

Berk, Jonathan B. and Richard C. Green. 2004. "Mutual Fund Flows and Performance in Rational Markets." *Journal of Political Economy* 112 (6):1269–1295.

Berk, Jonathan B. and Jules H. van Binsbergen. 2015. "Measuring Skill in the Mutual Fund Industry." *Journal of Financial Economics* 118 (1):1 – 20.

———. 2016. "Assessing Asset Pricing Models Using Revealed Preference." *Journal of Financial Economics* 119 (1):1 – 23.

Boyson, Nicole M., Christof W. Stahel, and René M. Stulz. 2010. "Hedge Fund Contagion and Liquidity Shocks." *Journal of Finance* 65 (5):1789–1816.

Breen, William J., Laurie Simon Hodrick, and Robert A. Korajczyk. 2002. "Predicting Equity Liquidity." *Management Science* 48 (4):470–483.

Brunnermeier, Markus K. 2009. "Deciphering the Liquidity and Credit Crunch 2007–2008." *Journal of Economic Perspectives* 23 (1):77–100.

Bühlmann, Peter and S. van de Geer. 2011. *Statistics for High-Dimensional Data: Methods, Theory, and Applications*. Springer Series in Statistics.

Carhart, Mark M. 1997. "On Persistence in Mutual Fund Performance." *Journal of Finance* 52 (1):57–82.

Chen, Zhiwu, Werner Stanzl, and Masahiro Watanabe. 2002. "Price Impact Costs and the Limit of Arbitrage." Working paper.

Chetverikov, Denis, Zhipeng Liao, and Victor Chernozhukov. 2017. "On Cross-Validated Lasso." Working paper.

Cochrane, John H. 2011. "Presidential Address: Discount Rates." *Journal of Finance* 66 (4):1047–1108.

Corwin, Shane A. and P. Schultz. 2012. "A Simple Way to Estimate Bid-Ask Spreads from Daily High and Low Prices." *Journal of Finance* 67 (2):719–760.

Daniel, Kent, Mark Grinblatt, Sheridan Titman, and Russ Wermers. 1997. "Measuring Mutual Fund Performance with Characteristic-Based Benchmarks." *Journal of Finance* 52 (3):1035–1058.

Daniel, Kent and Sheridan Titman. 1997. "Evidence on the Characteristics of Cross Sectional Variation in Stock Returns." *Journal of Finance* 52 (1):1–33.

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.

Elton, Edwin J., Martin J. Gruber, and Christopher R. Blake. 2001. "A First Look at the Accuracy of the CRSP Mutual Fund Database and a Comparison of the CRSP and Morningstar Mutual Fund Databases." *Journal of Finance* 56 (6):2415–2430.

Engle, Robert, Robert Ferstenberg, and Jeffrey Russell. 2012. "Measuring and Modeling Execution Cost and Risk." *Journal of Portfolio Management* 38 (2):14–28.

Fama, Eugene F. 1970. "Efficient Capital Markets: A Review of Theory and Empirical Work." *Journal of Finance* 25 (2):383–417.

Fama, Eugene F. and K. R. French. 2010. "Luck Versus Skill in the Cross-Section of Mutual Fund Returns." *Journal of Finance* 65 (5):1915–1947.

Fama, Eugene F. and Kenneth R. French. 1992. "The Cross-Section of Expected Stock Returns." *Journal of Finance* 47 (2):427–465.

———. 2008. "Dissecting Anomalies." *Journal of Finance* 63 (4):1653–1678.

———. 2012. "Size, Value, and Momentum in International Stock Returns." *Journal of Financial Economics* 105 (3):457 – 472.

Fama, Eugene F. and James D. MacBeth. 1973. "Risk, Return, and Equilibrium: Empirical Tests." *Journal of Political Economy* 81 (3):607–636.

Frazzini, Andrea, Ronen Israel, and Tobias J. Moskowitz. 2015. "Trading Costs of Asset Pricing Anomalies." Working paper.

Frazzini, Andrea and Lasse Heje Pedersen. 2014. "Betting Against Beta." *Journal of Financial Economics* 111 (1):1 – 25.

Glosten, Lawrence R and Lawrence E Harris. 1988. "Estimating the Components of the Bid/Ask Spread." *Journal of Financial Economics* 21 (1):123 – 142.

Grinblatt, Mark, Sheridan Titman, and Russ Wermers. 1995. "Momentum Investment Strategies, Portfolio Performance, and Herding: A Study of Mutual Fund Behavior." *American Economic Review* 85 (5):1088–1105.

Grundy, Bruce D. and J. Spencer Martin. 2001. "Understanding the Nature of the Risks and the Source of the Rewards to Momentum Investing." *Review of Financial Studies* 14 (1):29–78.

Harvey, Campbell R. 2017. "Presidential Address: The Scientific Outlook in Financial Economics." *Journal of Finance* 72 (4):1399–1440.

Harvey, Campbell R., Yan Liu, and Heqing Zhu. 2016. "... and the Cross-Section of Expected Returns." *Review of Financial Studies* 29 (1):5–68.

Hasbrouck, Joel. 2009. "Trading Costs and Returns for U.S. Equities: Estimating Effective Costs from Daily Data." *Journal of Finance* 64 (3):1445–1477.

Hastie, T., R. Tibshirani, and M. Wainwright. 2015. *Statistical Learning with Sparsity: The Lasso and Generalizations*. CRC Press.

He, Zhiguo, Bryan Kelly, and Asaf Manela. 2017. "Intermediary Asset Pricing: New Evidence from Many Asset Classes." *Journal of Financial Economics* .

Hou, Kewei, Chen Xue, and Lu Zhang. 2017. "Replicating Anomalies." Working paper.

Imbens, Guido W. and Donald B. Rubin. 2015. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction.* Cambridge University Press, 1st ed. ed.

Israel, Ronen and Tobias J. Moskowitz. 2013. "The Role of Shorting, Firm Size, and Time on Market anomalies." *Journal of Financial Economics* 108 (2):275 – 301.

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.

———. 2001. "Profitability of Momentum Strategies: An Evaluation of Alternative Explanations." *Journal of Finance* 56 (2):699–720.

Jones, Charles M. 2002. "A Century of Stock Market Liquidity and Trading Costs." Working paper.

Keim, Donald B. 2003. "The Cost of Trend Chasing and The Illusion of Momentum Profits." Working paper.

Keim, Donald B. and Ananth Madhavan. 1997. "Transactions Costs and Investment Style: An Inter-Exchange Analysis of Institutional Equity Trades." *Journal of Financial Economics* 46 (3):265 – 292.

Korajczyk, Robert A. and Ronnie Sadka. 2004. "Are Momentum Profits Robust to Trading Costs?" *Journal of Finance* 59 (3):1039–1082.

Kyle, Albert S. 1985. "Continuous Auctions and Insider Trading." *Econometrica* 53 (6):1315–1335.

Lesmond, David A., Joseph P. Ogden, and Charles A. Trzcinka. 1999. "A New Estimate of Transaction Costs." *Review of Financial Studies* 12 (5):1113–1141.

Lesmond, David A., Michael J. Schill, and Chunsheng Zhou. 2004. "The Illusory Nature of Momentum Profits." *Journal of Financial Economics* 71 (2):349 – 380.

Lettau, Martin, Matteo Maggiori, and Michael Weber. 2014. "Conditional Risk Premia in Currency Markets and Other Asset Classes." *Journal of Financial Economics* 114 (2):197 – 225.

Lewellen, Jonathan, Stefan Nagel, and Jay Shanken. 2010. "A Skeptical Appraisal of Asset Pricing Tests." *Journal of Financial Economics* 96 (2):175 – 194.

Nagel, Stefan. 2012. "Evaporating Liquidity." *Review of Financial Studies* 25 (7):2005–2039.

Newey, Whitney K. and Kenneth D. West. 1987. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* 55 (3):703–708.

Novy-Marx, Robert and Mihail Velikov. 2016. "A Taxonomy of Anomalies and Their Trading Costs." *Review of Financial Studies* 29 (1):104–147.

Oster, Emily. 2017. "Unobservable Selection and Coefficient Stability: Theory and Evidence." *Journal of Business & Economic Statistics* Forthcoming.

Pastor, Lubos and Robert F. Stambaugh. 2003. "Liquidity Risk and Expected Stock Returns." *Journal of Political Economy* 111 (3):642–685.

———. 2012. "On the Size of the Active Management Industry." *Journal of Political Economy* 120 (4):740–781.

Pastor, Lubos, Robert F. Stambaugh, and Lucian A. Taylor. 2015. "Scale and Skill in Active Management." *Journal of Financial Economics* 116 (1):23 – 45.

Patton, Andrew J. and Allan Timmermann. 2010. "Monotonicity in Asset Returns: New Tests with Applications to the Term Structure, the CAPM, and Portfolio Sorts." *Journal of Financial Economics* 98 (3):605 – 625.

Perold, Andre F. 1988. "The Implementation Shortfall: Paper Versus Reality." *Journal of Portfolio Management* 14 (3):4–9.

Rosenbaum, Paul R. and Donald B. Rubin. 1984. "Reducing Bias in Observational Studies Using Subclassification on the Propensity Score." *Journal of the American Statistical Association* 79 (387):516–524.

Sadka, Ronnie. 2010. "Liquidity Risk and the Cross-Section of Hedge-Fund Returns." *Journal of Financial Economics* 98 (1):54 – 71.

Tibshirani, Robert. 1994. "Regression Shrinkage and Selection Via the Lasso." *Journal of the Royal Statistical Society, Series B* 58:267–288.

Wermers, Russ. 2000. "Mutual Fund Performance: An Empirical Decomposition into Stock-Picking Talent, Style, Transactions Costs, and Expenses." *Journal of Finance* 55 (4):1655–1695.

Zou, Hui. 2006. "The Adaptive Lasso and Its Oracle Properties." *Journal of the American Statistical Association* 101 (476):1418–1429.

A. Mutual Fund Filters

We clean the CRSP mutual fund database at the individual fund and fund group levels. We first clean at the lowest level of aggregation to deal with missing and erroneous data, and then we filter our sample based on fund group-level information.

Cleaning Procedures at the Fund Level We first flag fund-dates with reporting frequencies less than monthly in the monthly returns file (6,526,393 observations). As discussed by Elton, Gruber, and Blake (2001) and Fama and French (2010), about 15% of funds before 1983 report returns only annually, and we mark as missing the fund returns for which neither adjacent month has a non-zero and non-missing return. These annual returns comprise 1.71% of fund-month observations.

Next we construct current and lagged total net asset (TNA) values for value-weighting fund returns within and across fund groups. Nearly a tenth of total net asset (TNA) values are undefined, and we interpolate TNA values to avoid discarding such a large fraction of the data. Before interpolating, we flag as missing invalid TNA values that arise because of recording errors or bottom coding. As noted in the CRSP mutual fund database documentation, entries of $100,000 denote TNAs of less than or equal to this value. Although not documented, entries of $1,000 seem to serve a similar role. We eliminate bottom-coded TNAs by setting to missing values less than or equal to $100,000 USD. Likewise, we set to missing TNA values exceeding $1 trillion USD, as no single fund has ever reached this value. Imposing these filters, 14.9% of TNA observations are flagged as missing.

We interpolate TNAs in three steps. First, we compute a "predicted" TNA by multiplying the last available TNA value by cumulative returns since that date. This predicted TNA value misses inflows and outflows from the fund. Second, when available, we reconcile predicted TNAs and the next filled TNA observation. The ratio of true TNA to predicted TNA (minus one) is a discrepancy associated with fund inflows or outflows. We assume flows are constant between known TNA values, and we multiply predicted TNA by $(1 + \textit{discrepancy})^{s/\Delta t}$, where s is the number of months since the last known TNA value and Δt is the number of months between TNA values. We assume a discrepancy of zero if there is no next known TNA. In the third step we run the first and second steps backward to use return data to fill in TNAs before the first reported TNA value. Given the interpolated values, we again set as missing any TNA values smaller than $100,000 or greater than $1 trillion. The filling and cleaning procedures reduce the number of missing TNA values to 2.8% of the data.

Share classes differ from one another in their fee structures, and we account for this variation before aggregating across share classes within a fund. We convert net returns to gross returns by adding to net returns the annual expense ratio divided by 12, following Fama and French (2010). The fund summary file has missing or non-positive expense ratios for 16.9% of observations, however, and we take several steps to fill in the missing data. First, as before, we fill missing expense ratios with the nearest observation with a non-missing value within each CRSP fund number group. This operation reduces the number of missing expense ratios to 8.4% of the summary data. We then merge the monthly return data with the summary data by fund number and calendar quarter. This merge assigns expense ratios to 76.2% of fund-month observations. For unmerged observations, we merge again on fund number and year, where we take the average expense ratio within the fund number-year in the fund summary file. This operation boosts the number of fund-month observations with an expense ratio to 88.5% of the data or 5,774,820 observations. We then drop the 89 observations with expense ratios exceeding 50% as these are almost certainly data errors.

We next filter out extreme return observations resulting from data errors. For example, we do not wish to include the recorded return of 533% on the Deutsche Equity 500 Index Fund in September 1997. Berk and van Binsbergen (2015) and Pastor, Stambaugh, and Taylor (2015) address these errors in part using external Bloomberg and Morningstar databases. We take a simpler approach to eliminate errors. We drop the 23 observations with reported returns exceeding one (i.e., 100%) in absolute value. This approach is inspired by the shape of the tail of extreme returns in the data depicted in Figure A.I: the frequency of extreme returns decays roughly exponentially until $|r| = 100\%$, with a smattering of randomly spaced returns beyond this value. These observations appear to come from a different distribution, and for this reason, we classify them as likely errors.

Because our analysis concerns mutual funds, we filter out exchange-traded funds (ETFs), exchange-traded notes (ETNs), and variable annuity underlying (VAU) funds. To do this, we discard any observations for which et_flag indicates an ETF or ETN or vau_fund indicates a VAU at any time in a fund's life. These exclusions total 9.1% of observations.

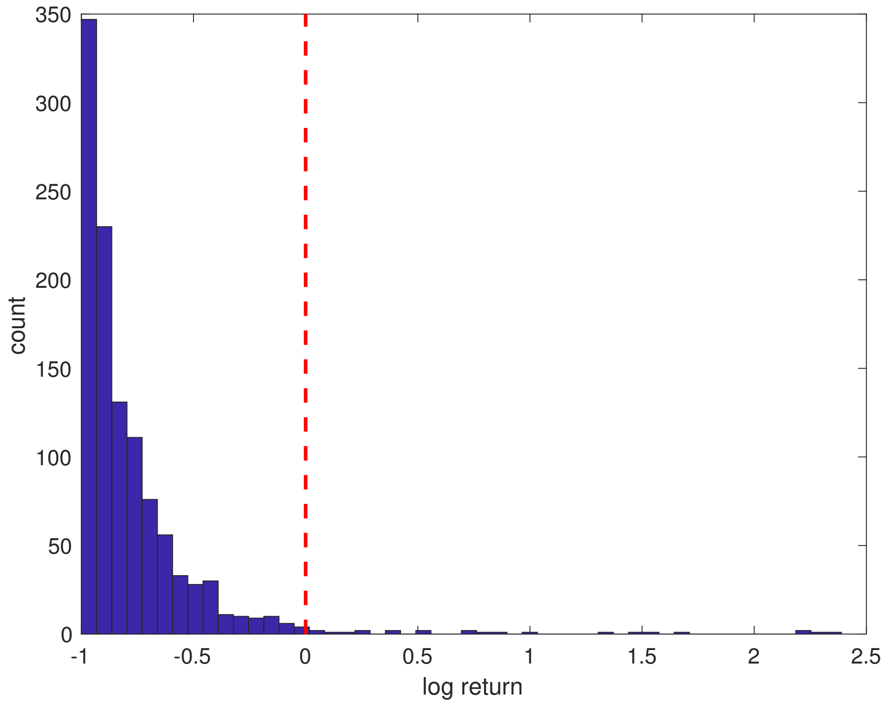
Aggregation into Fund Groups Having accounted for the salient variation across share classes, we next identify share classes of the same fund. As a preliminary step, we fill missing fund names using the nearest observation with a non-missing fund name within each CRSP fund number group. Of the 1,859,702 observations in the fund summary file, we assign fund names for 19,460 observations and remove 242 observations without recoverable names. We then repeat this procedure for missing fund tickers. This matching step assigns 116,238 of the 274,875 observations with missing tickers. By contrast with observations missing names, we retain observations with missing tickers.

We then follow almost verbatim the fund class grouping procedure of Berk and van Binsbergen (2015) and Pastor, Stambaugh, and Taylor (2015). Their procedure consists of two steps:

1. We identify share classes following three mutually exclusive rules. First, if the CRSP fund name contains a semicolon and the phrase after the last semicolon does not contain a forward slash, we retain the fund name prior to the last semicolon as the fund group name. Second, if the CRSP fund name contains a forward slash, and the entire phrase after the last forward slash does not contain a space or a semicolon, we use the word prior to the last forward slash as the fund group name. Third, if neither rule applies, we assume that the CRSP fund name

Figure A.I: Distribution of Log Fund Returns

Figure plots the distribution of the log of absolute monthly mutual fund returns. We truncate the plot to -1 on the left to maintain resolution on the extreme returns on the right. The dashed line represents of cutoff at $|r| = 100\%$.



does not include a share class.

We make a minor adjustment to their methodology before applying these rules. Although handling semicolons is straightforward, forward slashes—the other class-name delimiter used in CRSP—require more care. For example, fund names include "Franklin/Templeton" and "M/M" (money market), so "/" does not serve only as a delimiter, and the absence of a space does not guarantee that the subsequent string is a share class. So as a preliminary step, we replace forward slashes in T/F, T/E, M/M, L/S, Small/Mid, Long/Short, S/T, and L/T with backslashes in fund names.

2. We define equivalent funds as those sharing an adjusted name or a ticker symbol. To do so, we iteratively build equivalence classes of funds with equivalent names and/or ticker symbols. Because equivalence is transitive, a pair of funds that shares a name, and another fund that shares a ticker with the second fund, are all considered to be of the same group.

This mapping reduces the 61,734 surviving unique fund IDs in the CRSP monthly returns file to 23,613 unique fund groups. Of the 6,522,095 observations in the monthly return file, only 4,298 of these are not assigned a fund group, and these observations are dropped.

Cleaning Procedures at the Fund Group Level We construct fund group returns and total net assets by taking a weighted average of returns across component fund IDs. The return weights are one-month lagged TNAs. We retain observations for which the lagged TNA is undefined but the fund group only has one fund ID, that is, the one fund ID has an effective weight of 100%. Likewise, fund group TNAs are the sum of current TNA values across component fund IDs. Aggregating funds across share classes delivers 2,244,101 monthly fund-group observations.

As Fama and French (2010) note, "incubation bias arises because funds typically open to the public—and their pre-release returns are included in mutual fund databases—only if the returns turn out to be attractive." We follow their approach to countering incubation bias by keeping observations only after a fund group achieves a TNA of at least $10 million (in December 2016 dollars).[28] We retain data from funds that later drop below this threshold to avoid introducing a selection bias. Dropping fund groups that never achieve a $10 million TNA eliminates 2.3% of fund group-month observations. Dropping observations from potential incubation periods before the $10 million threshold is achieved eliminates another 3.9% of the sample.

Next, we filter fund groups based on fund name and objective. We first exclude all funds with names containing ETF, ETN, exchange-traded fund, exchange traded fund, exchange-traded note, exchange traded note, iShares, and PowerShares (not case sensitive) as a redundant filter on top of the CRSP-based ETF/ETN filter. These exclusions eliminate 3,006 observations. We then exclude any funds with names that have clear international or non-equity connotations: international, intl, bond, emerging, frontier, rate, fixed income, commodity, oil, gold, metal, world, global, China, Europe, Japan, real estate, absolute return, government, exchange, euro, India, Israel, treasury, Australia, Asia, pacific, money, cash, yield, U.K., UK, kingdom, municipal, Ireland, LIBOR, govt, obligation, money, cash, yield, mm, m/m, diversified (but not diversified equity), and short term (not case sensitive). This filter complements our requirement that a fund have a domestic equity "ED" CRSP objective code.[29] These filters reduce the number of valid funds from 12,691 to 4,282, and the corresponding number of non-missing return observations decreases to 740,899 for the entire December 1961 to December 2016 CRSP mutual fund database.

Lastly, we restrict the set of funds to those with at least two years of monthly data in our 1970–2016 sample period. This filter reduces our sample to 4,267 mutual funds with 724,995 non-missing return observations. Summary statistics for this sample are reported in Table I.

B. Bias of Symmetric Fama-MacBeth Regressions with General h_{it}

Section IV.C implements an asymmetric Fama-MacBeth regression in which the first stage includes liquidity proxies, and the second stage does not. If instead we were to also include the

[28]Our inflation index is the Consumer Price Index for All Urban Consumers (CPIAUCSL) series provided by the Federal Reserve Bank of St. Louis' FRED database.

[29]The CRSP objective code unifies Wiesenberger objective codes for 1962–1993 data, Strategic Insight objective codes for 1993–1998 data, and Lipper objective codes for 1998–2016 data.

loadings on the liquidity proxies in Equation (2), the second-stage regression becomes

$$r_{it} = \sum_k \lambda^S_{kt}\hat{\beta}_{ik}1_{i\in S} + \sum_k \lambda^{MF}_{kt}\hat{\beta}_{ik}1_{i\in MF} + \sum_l \lambda^S_{lt}\hat{\gamma}_{il}1_{i\in S} + \sum_l \lambda^{MF}_{lt}\hat{\gamma}_{il}1_{i\in MF} + \epsilon_{it}.\ t = 1,\dots,T. \quad (12)$$

From the conjectured return process of Equation (5), $\hat{\lambda}^S_t = \lambda^S_t$, $\hat{\lambda}^{\Delta}_{kt} = \bar{\eta} + \frac{cov((\eta_i - \bar{\eta})\beta_i, \beta_i)}{var(\beta_i)}$, and $\hat{\lambda}^{\Delta}_{lt} = \eta_{lt}$. The problem with this approach is that the $\hat{\lambda}^{\Delta}_{lt}$ terms absorb the time-varying part of η_{it}, so we can no longer cleanly attribute time-varying costs to each return factor. Moreover the logic of mutual funds scaling down strategies in the face of high costs applies to η_{it} rather than to η_i.

To resolve the first issue we need to decompose η_{st} into factor-specific parts. The sum of all time-varying costs is

$$TVC_{it} \equiv \sum_l \eta_{lt}\gamma_{il} = \sum_l \eta_{lt}\left(\sum_k \gamma_{ikl}\beta_{ik}\right). \quad (13)$$

Regressing total time-varying costs on β_is decomposes costs into *factor-specific* time-varying parts,

$$TVC_{it} = \sum_t \sum_k \eta_{kt}\beta_{ik}1_t + \epsilon_{it}. \quad (14)$$

This regression can be interpreted as projecting time-varying liquidity costs onto the factor-exposure space. However, this rotation is imperfect because of cross-sectional variation in γ_is. To see why dispersion in γ_i is problematic, consider a single coefficient estimate in a one-return factor case of Equation (14),

$$\hat{\eta}_t = \frac{cov\left(\sum_l \eta_{lt}\gamma_{il}, \beta_i\right)}{cov\left(\beta_i\right)} = \sum_l \eta_{lt}\bar{\gamma}_l + \sum_s \frac{cov\left(\beta_i\eta_{lt}\left(\gamma_{il} - \bar{\gamma}_l\right), \beta_i\right)}{cov\left(\beta_i\right)}. \quad (15)$$

The first term represents the average exposure to liquidity factors multiplied by the factors' time-t realizations. This is the term of interest, but instead we identify this term plus a cross-sectional bias term.

Focusing on the bias for each l, we might expect higher-than-average cost-factor sensitivities $\gamma_{il} > \bar{\gamma}_l$ to be associated with lower betas if firms are risk averse. Although we would expect betas to be negatively associated with *total costs* per unit of risk exposure η_{it}, it is not clear what relation the *time-varying component* alone should have with betas. Because of this ambiguous sign and the additional complexity of this approach, it is preferable not to include the liquidity exposures in the cross-sectional regression step.

C. Spanning Variation in η Using Many Liquidity Proxies

Including more covariates increases the likelihood that we span variation in implementation costs, η_{it}, by including all salient liquidity proxies. At the same time, including additional highly correlated cost proxies may overfit the first-stage regression and deliver nonsensical cross-sectional slopes in the second stage.

Sparse regression techniques offer a solution to this challenge. We supplement the standard first-stage regression with a Lasso or l_1-penalized regression (Tibshirani (1994)). We augment the least-squares minimization problem in the time-series regressions with additional terms to penalize liquidity coefficients,

$$\min_{\beta,\tilde{\gamma}} \frac{1}{T_i} \sum_t \left(r_{it} - \sum_k f_{kt}\beta_{ik} - \sum_l \tilde{\eta}_{lt}\tilde{\gamma}_{il} \right)^2 + \kappa \left(\sum_k \omega_k \left|\beta_{ik}\right| + \sum_l \omega_l \left|\tilde{\gamma}_{il}\right| \right), \tag{16}$$

where κ represents a penalty term for coefficients different from zero, and ω_k and ω_l represent additional relative penalties explained below. The problem reduces to least squares when $\kappa = 0$; otherwise, liquidity coefficients are compressed toward zero. Note that we do not require a penalization in the cross-sectional step because the second-stage regression omits liquidity proxies. As before, we normalize all liquidity proxies to give them similar scales and an equal chance of entering the Lasso regression.[30]

Lasso simultaneously prevents overfitting in the time-series regressions by shrinking coefficients and selects covariates by zeroing out coefficients that would otherwise be close to zero. Both features facilitate the use of many liquidity proxies even when a mutual fund is relatively short-lived. Moreover, we no longer need to choose which measure(s) best approximate the costs faced by each fund, and indeed, different liquidity measures may be more salient for different mutual funds. First-stage penalization also knocks out spurious strategy loadings for funds that take on risk exposures unintentionally—a small non-zero loading taken en route to implementing a different strategy will be zeroed out.

The original Lasso implementation sets $\omega_k = \omega_l = 1$ for all k and l. Unfortunately Lasso is not guaranteed to deliver consistent estimates of β and γ, and it does not have the "oracle property" by which the variable selection step identifies the correct model and estimates converge at the optimal rate. By contrast, Zou (2006)'s adaptive Lasso has these desirable features, which enables us to construct confidence intervals for cross-sectional slopes as though the first-stage regression were OLS. Adaptive Lasso differs from Lasso in placing higher penalties on parameters with little explanatory power by setting $\omega = \left|\hat{\beta}\right|^{-\gamma}$. Our penalization weights use OLS $\hat{\beta}$s (as in Zou (2006)) and a penalty exponent of $\gamma = 1$.

The obvious concern when using Lasso is the selection of the penalization parameter κ. Follow-

[30] We interpolate missing elements of the VXO/TED series using their matrix-completed values $\phi'_{VXO}g_{ML}$ and $\phi'_{TED}g_{FL}$ from the PCA-ALS procedure described in Footnote 17.

ing standard practice (e.g., Bühlmann and van de Geer (2011), Hastie, Tibshirani, and Wainwright (2015)), we use k-fold cross-validation to select κ. Cross-validation works as follows. First, select a candidate value of κ_m and partition the sample into k equal "folds"; in our case, we choose the MATLAB default of $k = 10$. Next, for each fold, estimate the model on the set difference of the full sample and the partition. Then calculate the mean-squared error (MSE) of the estimated model on the fold that was set aside. This procedure provides k pseudo-out-of-sample (POOS) MSEs as a function of κ_m. Finally, repeat this procedure for a range of κ_m, and select κ as the value κ_m that maximizes the average POOS MSE. Intuitively this process tames overfitting by selecting the model with the best out-of-sample predictive properties.[31]

Table A.I presents results using the adaptive Lasso first stage described by Equation (16). Most importantly, the coefficients on λ^{Δ} are of similar size and statistical significance as they are in the preceding two tables. Using the adaptive Lasso results in one key change from Table III, however: the point estimate for UMD compensation for mutual funds becomes negligible in the full sample and negative in the recent sample. This feature is consistent with mutual funds earning compensation for momentum exposure only to the extent that momentum also embeds liquidity risk. By including a rich set of liquidity and liquidity risk proxies rather than two principal components, we allow this source of compensation to be spanned in the first stage, thereby effectively kicking out UMD as a compensated factor for mutual funds.

D. Matched Pairs Estimates of Implementation Costs

Our cross-sectional approach compares the return compensation for an incremental unit of risk exposure taken in on-paper portfolios versus in mutual funds. Such an approach does not address whether mutual funds achieve more favorable risk-reward trade-offs for investments in stocks with high book-to-market ratios, small size, or high prior returns. To answer this question, we consider the building blocks for many tradeable return factors in academia—long-short portfolios implied by characteristic sorts—and conduct a matched pairs analysis of characteristic-sorted stocks and matched mutual funds.

Our analysis is similar in spirit to Daniel, Grinblatt, Titman, and Wermers (1997), who examine the origins of mutual fund performance by comparing mutual fund returns against those of characteristic-matched portfolios of stocks. Rather than using holdings data to build and match with benchmark portfolios, we use a formal matched pairs design to directly compare mutual funds and "high-characteristic" stocks with similar risk attributes. This approach has the advantage of

[31]Remarkably, Chetverikov, Liao, and Chernozhukov (2017) demonstrate that time-series betas estimated using the cross-validated Lasso converge to the true betas at rate $\sqrt{n}$, up to a negligible log term. Because the convergence rate is comparable to that of OLS, using (adaptive) Lasso in the first stage does not exacerbate the errors-in-variables problem endemic to Fama-MacBeth regression. We therefore follow standard practice in taking betas as "known" inputs into the Fama-MacBeth cross-sectional regressions and adjusting standard errors for heteroskedasticity and serial correlation by Newey-West.

Table A.I: Implementation Cost Estimates in Fama-MacBeth Regressions — Liquidity Lasso

Table reports Fama-MacBeth estimates of the compensation for factor exposure for stock portfolios (second panel), domestic equity mutual funds (third panel), and their difference (top panel). Coefficients are the average cross-sectional slopes $\bar{\lambda}_k$ across monthly regressions of excess returns r_{it} on time-series betas $\hat{\beta}_{ik}$,

$$r_{it} = \sum_k \lambda_{kt}^S \hat{\beta}_{ik} 1_{i \in S} + \sum_k \lambda_{kt}^{MF} \hat{\beta}_{ik} 1_{i \in MF} + \epsilon_{it},\ t = 1, \ldots, T,$$

where k indexes the four Carhart (1997) factors and λ^Δ is defined as $\lambda^S - \lambda^{MF}$. First-stage regression estimates include these factors and all market and funding liquidity proxies in an adaptive Lasso regression with portfolio-specific penalty parameters chosen by 10-fold cross validation. Liquidity proxies and stock portfolio sets are described in Section III. All coefficients are annualized and reported in percent. Standard errors are Newey-West with three lags. t statistics are in parentheses.

(a) Value-Weighted Stock Portfolios

		1970 − 2016				**1993 − 2016**			
	N_S	MKT	HML	SMB	UMD	MKT	HML	SMB	UMD
λ^Δ	100	-0.22	4.97***	0.09	8.71***	-0.02	3.85***	-0.36	6.67***
t-stat		(-0.71)	(6.16)	(0.14)	(6.14)	(-0.06)	(4.17)	(-0.42)	(3.48)
λ^Δ	269	-0.06	3.71***	-0.30	8.57***	0.33	2.93***	-1.13	7.18***
t-stat		(-0.24)	(4.88)	(-0.54)	(6.14)	(1.32)	(4.10)	(-1.56)	(3.74)
λ^S	100	6.70***	6.96***	1.11	8.61***	7.76**	5.34*	2.05	5.88
t-stat		(2.80)	(3.64)	(0.64)	(3.67)	(2.39)	(1.80)	(0.83)	(1.56)
λ^S	269	6.86***	5.70***	0.72	8.47***	8.10**	4.43	1.27	6.38*
t-stat		(2.86)	(2.99)	(0.42)	(3.60)	(2.50)	(1.43)	(0.52)	(1.70)
λ^{MF}	—	6.92**	1.99	1.02	-0.10	7.78**	1.50	2.41	-0.79
t-stat		(2.83)	(1.01)	(0.58)	(-0.04)	(2.39)	(0.48)	(0.97)	(-0.19)
T		564	564	564	564	282	282	282	282
$\bar{N}_{MF}$		1286	1286	1286	1286	2123	2123	2123	2123

$^*p < .10$, $^{**}\ p < .05$, $^{***}\ p < .01$

(b) Equal-Weighted Stock Portfolios

		1970 − 2016				**1993 − 2016**			
	N_S	MKT	HML	SMB	UMD	MKT	HML	SMB	UMD
λ^Δ	100	-0.27	5.15***	2.64***	8.26***	0.31	2.68**	2.22	6.05***
t-stat		(-0.58)	(5.92)	(2.65)	(5.68)	(0.51)	(2.26)	(1.61)	(2.99)
λ^Δ	269	0.26	4.91***	2.50**	9.86***	1.07	2.03*	1.99	7.69***
t-stat		(0.5)	(4.92)	(2.26)	(6.35)	(1.60)	(1.72)	(1.34)	(3.48)
λ^S	100	6.65***	7.14***	3.66*	8.16***	8.09**	4.17	4.62	5.26
t-stat		(2.77)	(3.69)	(1.85)	(3.46)	(2.49)	(1.43)	(1.65)	(1.38)
λ^S	269	7.18***	6.90***	3.51*	9.76***	8.85***	3.52	4.40	6.89*
t-stat		(3.01)	(3.25)	(1.70)	(4.03)	(2.75)	(1.10)	(1.52)	(1.76)
λ^{MF}	—	6.92***	1.99	1.02	-0.10	7.78**	1.50	2.41	-0.79
t-stat		(2.83)	(1.01)	(0.58)	(-0.04)	(2.39)	(0.48)	(0.97)	(-0.19)
T		564	564	564	564	282	282	282	282
$\bar{N}_{MF}$		1286	1286	1286	1286	2123	2123	2123	2123

$^*p < .10$, $^{**}\ p < .05$, $^{***}\ p < .01$

exploiting a wider range of stock-level variation to more precisely match the properties of mutual funds, and it builds on recent advances in matched-pairs theory to adjust for differences between matched entities.

A. Matched Pairs Methodology

We begin by constructing characteristics for each stock and sorting stocks into quintile portfolios. Our characteristics are 60-month rolling market beta (requiring at least 24 observations), book-to-market ratio,[32] market capitalization (with scale reversed to place small stocks in the top quintile), and prior return over the previous year, skipping the latest month (the "2-12" return). We follow the methodology of Ken French's website in constructing these characteristics, and we use the provided breakpoints based on NYSE quintiles where available. In the case of rolling market betas we construct our own quintile breakpoints. For the first three characteristics, we assign portfolios at the end of June and retain assignments for July through the end of the following June. Momentum is a higher-frequency anomaly, and we sort on prior returns and retain assignments for the next month only. We then estimate two sets of betas on monthly return data for all common stocks in the CRSP universe and all U.S. domestic equity mutual funds: univariate betas with respect to a single factor f_k and multivariate betas with respect to all four Carhart factors.

For each stock in quintile q for factor k in month t, we find the three closest mutual funds active in that month. We assess proximity using the Mahalanobis distance on betas with covariances estimated using the full sample,[33] where β_k determines distance in our univariate analysis, and all four βs determine distance in our multivariate analysis. By matching betas rather than characteristics, our approach remains rooted in factor-based explanations for differences in average returns: the matched-pairs analysis answers the question of whether mutual funds can replicate high-characteristic stock performance by taking similar risk exposures.

Implicitly we select stocks as our matched pairs "treatment group" because we want to mimic on-paper factor portfolios as best as possible using mutual funds. Each stock is matched to three mutual funds rather than to one to improve precision of the estimated average return for mutual funds with the same risk characteristics. To ensure high match quality, we impose a maximum distance or caliper of 0.25 standard deviations for univariate matches (following Rosenbaum and Rubin (1984)'s rule-of-thumb in the context of propensity score matching) and $0.25 \times \sqrt{4} = 0.5$ standard deviations for four-factor matching. Stocks with fewer than three matched mutual funds within these radii are dropped.[34]

[32] We source book-to-market ratios from the WRDS Financial Ratios Suite.

[33] The Mahalanobis distance is $\sqrt{(x-y)' \Sigma^{-1} (x-y)}$ for two vectors x and y and covariance matrix Σ, and it reduces to the Euclidean distance when $\Sigma = \mathbf{I}$. Σ^{-1} accounts for differences in standard deviations and nonzero correlations among attributes.

[34] The stocks that do not match are those with the most extreme characteristics, e.g., microcaps, but they also have particularly high returns on average. For this reason our estimates using matched stocks likely overstate the ability of mutual funds to replicate the performance of stocks in the highest-characteristic quintiles.

Matching of multivariate betas controls for possible variation in several risk factors simultaneously at the cost of worsening match quality along any particular dimension. Because betas for stocks and matched mutual funds do not perfectly coincide in either case, we employ a bias-adjusted matched pairs estimator for each month's implementation gap. To a first approximation, this matched-pairs estimate is the average difference between next-month returns for stocks and mutual funds. We follow Abadie and Imbens (2006, 2011) to bias adjust this estimate by shifting mutual fund returns by the compensation for a "local" difference in betas. This adjustment factor is the difference in matched betas multiplied by the slope of cross-sectional regressions of returns on betas. In this way, our modern matched-pairs approach weds older differencing techniques with cross-sectional regression models. Finally, armed with monthly performance differences, we take the time-series average value as our full-sample implementation gap estimate for quintile q and factor k. We also consider value-weighted performance differences using the lagged market capitalization of the matched stocks.

We do not directly compare the performance of long-short strategies for stocks and mutual funds. We cannot short mutual funds, and underperformance on both ends of a long-short strategy, for example, because of transactions costs, would be incorrectly obscured by differencing. Instead, we compare the performance of stocks in high-characteristic portfolios and matched mutual funds. These differences in high-characteristic quintile returns represent a lower bound on the underperformance of mutual fund implementations of factor investing. To see why, consider the difference in factor returns for stocks and mutual funds, Δ^{S-MF},

$$\begin{aligned}\Delta^{S-MF} &\equiv \left(\mu^{S}_{(5)} - \mu^{S}_{(1)}\right) - \left(\mu^{MF}_{(5)} - \mu^{MF}_{(1)}\right) \\ &\geq \left(\mu^{S}_{(5)} - \mu^{S}_{(1)}\right) - \left(\mu^{MF}_{(5)} - \mu^{S}_{(1)}\right) = \mu^{S}_{(5)} - \mu^{MF}_{(5)} \equiv \Delta^{S-MF}_{(5)},\end{aligned} \tag{17}$$

where Δ^{S-MF} is the implementation gap and $\mu_{(x)}$ denotes the average returns for quintile-x stocks (S) and mutual funds (MF). The inequality in Equation (17) holds if mutual funds are weakly less able to implement the short side of strategies than paper shorting returns would indicate. We expect underperformance on selling the low-beta quintiles because shorting entails relatively high transaction costs. Short-side underperformance is especially plausible if we find that mutual funds also underperform on the long side. In addition, some firms implement positive-cost versions of anomalies such as long-only momentum, in which only the extreme "buy" portfolio is traded.

B. Results

Table A.II reports the results of our matched pairs analysis. The LMS value in the upper-left corner indicates that mutual funds outperform stocks with the same market beta exposure by 2.66% per year when stocks are in the 80–100th percentile of the distribution of rolling market betas. We designate "LMS" as long high-market beta stocks and short low-market beta stocks to distinguish

Table A.II: Returns of Matched Stocks and Matched Mutual Funds

Tables report bias-adjusted differences in performance between stocks and matched mutual funds. The first row of each panel denotes the difference in returns between stocks and mutual funds in quintile five ($\Delta_{(5)}^{S-MF}$) of the distribution of stock characteristics. Differences are estimated using matched pairs on the four Carhart (1997) factors with bias adjustment by linear regression in the matching variable(s), where we designate "LMS" as long high-market beta stocks and short low-market beta stocks to distinguish long-short market beta portfolios from the excess return on the market. Differences are equal- or value-weighted within each month and averaged across months. $\mu_{(5)-(1)}^{S}$ rows denote the difference in equal- or value-weighted performance between stocks in quintiles 5 and 1 of the distribution of characteristics. All coefficients are annualized and reported in percent, and standard errors are Newey-West with three lags. The top panel matches only on β_k, whereas the bottom panel matches on all four factors. "VW" rows value weight returns by market capitalization, and "EW" rows equal weight returns. t statistics are reported in parentheses.

(a) Sorting and Matching on Univariate Beta

		1970 − 2016				**1993 − 2016**			
		LMS	HML	SMB	UMD	LMS	HML	SMB	UMD
VW	$\Delta_{(5)}^{S-MF}$	-2.66*	3.05**	1.71	3.83***	-0.16	0.25	1.97	3.41*
	t-stat	(-1.95)	(2.30)	(1.20)	(3.15)	(-0.09)	(0.15)	(0.90)	(1.95)
	$\mu_{(5)-(1)}^{S}$	-0.63	4.77**	1.61	8.75***	2.32	1.51	2.34	4.47
	t-stat	(-0.18)	(2.32)	(0.69)	(3.00)	(0.44)	(0.52)	(0.74)	(1.00)
EW	$\Delta_{(5)}^{S-MF}$	-0.32	8.78***	3.39**	7.94***	2.77	8.31***	3.93	6.76***
	t-stat	(-0.15)	(4.47)	(2.14)	(5.42)	(0.93)	(3.37)	(1.62)	(3.33)
	$\mu_{(5)-(1)}^{S}$	-0.70	11.32***	3.82	6.17**	2.26	10.29***	4.00	1.91
	t-stat	(-0.23)	(5.82)	(1.55)	(2.32)	(0.47)	(3.67)	(1.25)	(0.43)

(b) Sorting and Matching on Multivariate Beta

		1970 − 2016				**1993 − 2016**			
		LMS	HML	SMB	UMD	LMS	HML	SMB	UMD
VW	$\Delta_{(5)}^{S-MF}$	-1.67	2.28*	6.21***	2.29*	3.37**	3.41***	4.11***	3.25**
	t-stat	(-1.07)	(1.87)	(5.36)	(1.91)	(2.02)	(2.96)	(4.13)	(2.53)
	$\mu_{(5)-(1)}^{S}$	-0.63	4.77**	1.61	8.75***	2.32	1.51	2.34	4.47
	t-stat	(-0.18)	(2.32)	(0.69)	(3.00)	(0.44)	(0.52)	(0.74)	(1.00)
EW	$\Delta_{(5)}^{S-MF}$	3.79***	6.81***	7.91***	7.44***	5.76***	7.56***	5.70***	6.12***
	t-stat	(2.71)	(5.66)	(6.17)	(7.90)	(3.70)	(6.43)	(4.82)	(5.80)
	$\mu_{(5)-(1)}^{S}$	-0.70	11.32***	3.82	6.17**	2.26	10.29***	4.00	1.91
	t-stat	(-0.23)	(5.82)	(1.55)	(2.32)	(0.47)	(3.67)	(1.25)	(0.43)

long-short market beta portfolio returns from the excess return on the market. This outperformance is larger than the on-paper return to an value-weighted long-short strategy based on market beta quintiles, and it is roughly 40% of the annual equity premium over this period (6.35% per year, not tabulated). Moving across the upper-left panel we see that the implementation gap is positive for all other factors considered. Costs are particularly high for momentum strategies, as prior literature suggests, and they are similarly high for value strategies. Mutual funds underperform on the long side for both value and momentum by 3%–4%, and these values are statistically robust.

The top panel reflects performance differences between stocks and mutual funds matched on betas from a one-factor model of returns. This matching is akin to a portfolio sort in which a single characteristic is used. Other return-relevant variables are not held fixed as we vary one characteristic, so it may be that stocks and matched mutual funds vary substantially on other dimensions. For example, mutual funds trading continuation strategies like momentum are less likely to trade contrarian strategies like value ($\rho_{\beta_{HML},\beta_{UMD}} = -15.7\%$ for mutual funds). The bottom panel addresses this concern by reporting return differences in high-characteristic portfolios between stocks and mutual funds matched on the four Carhart (1997) factors. In this analysis we estimate multifactor betas using four-factor time-series regressions as in Equation (1).

Matching on multifactor betas reduces (increases) our estimated implementation costs for the full (recent) sample, which suggests that controlling for differences in multiple risk exposures is important to make the stock and mutual fund samples comparable. In the full sample we see that the implementation gap is large and statistically significant for SMB only, and it is economically large but marginally statistically significant for momentum and value. In the recent sample, the implementation gaps are so large that they swamp or severely attenuate factor returns for all non-market factors considered (tabulated in $\mu^{S}_{(5)-(1)}$ rows). *No non-market factors* earn returns after real-world costs for the 1993–2016 period in which the four academic factors are known and the mutual fund universe is far more developed (see Figure I). This finding mirrors our main result in Section IV, and it accords with the evidence of Berk and van Binsbergen (2016) and Barber, Huang, and Odean (2016) that investors perceive only the market factor to be risky and should (eventually) eliminate other risk premia through fund flows.

The bottom half of each panel presents equal-weighted results. Differences in performance widen dramatically when small, harder-to-access stocks are upweighted. Focusing first on the univariate matches, mutual fund underperformance on value, size, and momentum strategies doubles relative to the corresponding value-weighted results. Turning to the multivariate matches also strengthens our finding that mutual funds underperform matched stocks for the non-market factors. Implementation gaps are again larger than in the value-weighted results, but the magnitudes are nonetheless economically large for the three main anomalies in both: the 2.3%–3.4% implementation gap for value increases to 6.8%–7.6% against a time-series average return of 4.8% for value-weighted HML; 4.1%–6.2% increases to 5.7%–7.9% for size against a time-series average return of 1.6% for SMB;

and 2.3%–3.3% increases to 6.1%–7.4% against a time-series average return of 8.8% for UMD. In short, we replicate the high implementation costs of these factors, and such performance attrition is a stark departure from the muted effects of trading costs often considered in the academic literature.

Taken together, these matched-pair results agree qualitatively with the cross-sectional results for three of the four factors (MKT/LMS, HML, and UMD), but they disagree for size. This disagreement is likely attributable to the fact that SMB beta is not associated with cross-sectional differences in average returns—and the cross-sectional approach thus reveals no difference in compensation to SMB exposure—whereas the small-size characteristic is associated with higher average returns. Consequently we observe high returns on small-stock portfolios in the matched pairs approach, and mutual funds clearly cannot capture these returns well in practice.

E. Matched Pairs Estimates of Implementation Costs: Additional Details

A. Bias Adjustment for Imperfect Matches

Characteristics are not perfectly matched between stocks and mutual funds, and match characteristics may differ systematically between stocks and matched mutual funds. We follow Imbens and Rubin (2015)'s guidance to bias correct our matching estimator using a linear regression of outcomes on mutual fund ("control-group") attributes.[35] For each date t, we bias-adjust mutual fund returns using a factor model for returns in which the estimated betas serve as risk exposures,

$$r_{it} = \alpha_i + \sum_k \delta_{kt}\hat{\beta}_{ik} + \epsilon_{it},\ i \in MF,\ t = 1, \ldots, T. \tag{18}$$

Despite its matched-pairs origin, Equation (18) is our usual Fama-MacBeth cross-sectional regression for the set of domestic equity mutual funds.

Using Equation (18), we shift our estimate of mutual fund returns by the difference in betas between matched stocks and mutual funds multiplied by the time t return to a unit of beta exposure,

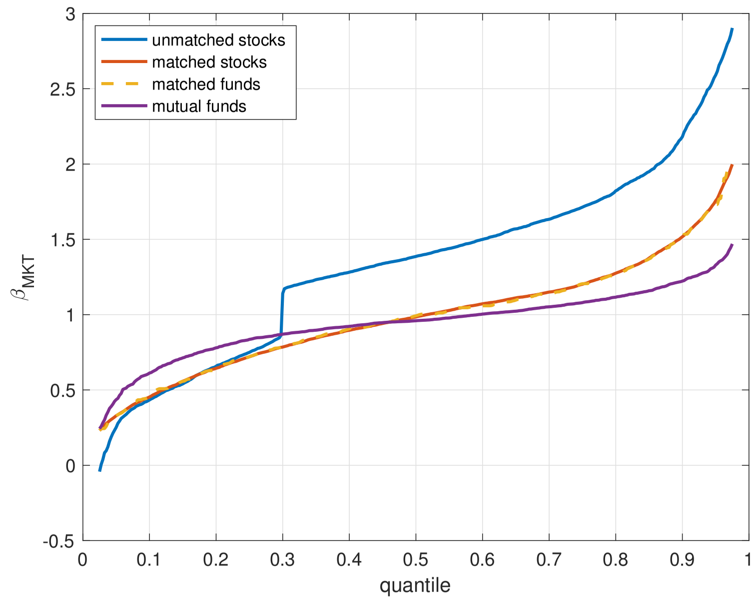
$$\frac{1}{3}\sum_{j\in\Omega(i)}\sum_k \delta_{kt}\left(\beta_{ik}^{S} - \beta_{jk}^{MF}\right), \tag{19}$$

where $\Omega(i)$ denotes the three-element set of mutual funds matched to stock i. In effect, bias-correction combines unadjusted matched pairs with the factor model approach of Section IV. By contrast with the cross-sectional analysis of Section IV, however, these adjustments are "local" because differences in betas between stocks and matched mutual funds are small by construction.

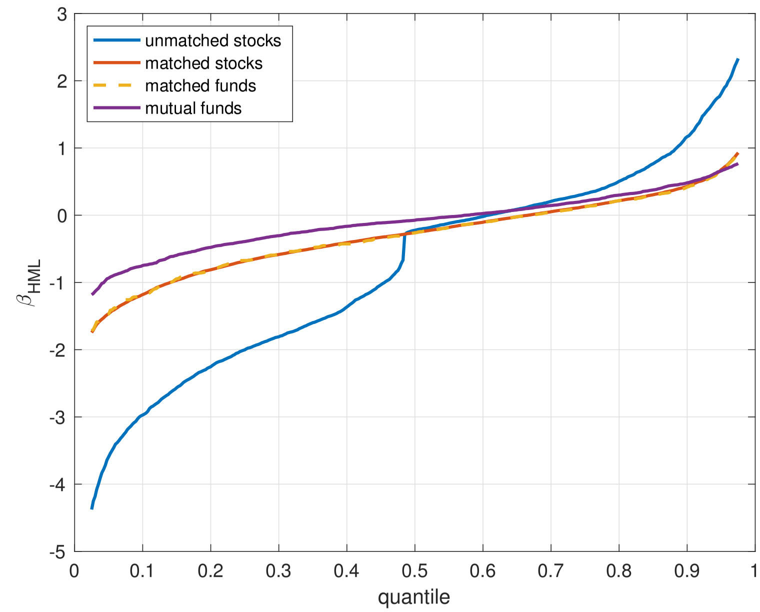
Figure A.II: Comparison of Samples on Matching Variable — Univariate β Matching

Figure plots the distribution of the matching variable for unmatched stocks, matched stocks, matched mutual funds, and all mutual funds. Matches are constructed monthly. Stocks are considered "matched" at date t if and only if they have at least three mutual funds within a Mahalanobis distance of 0.25σ of the matching variable during month t. Covariances for the Mahalanobis metric are calculated across all stocks and mutual funds and all dates.

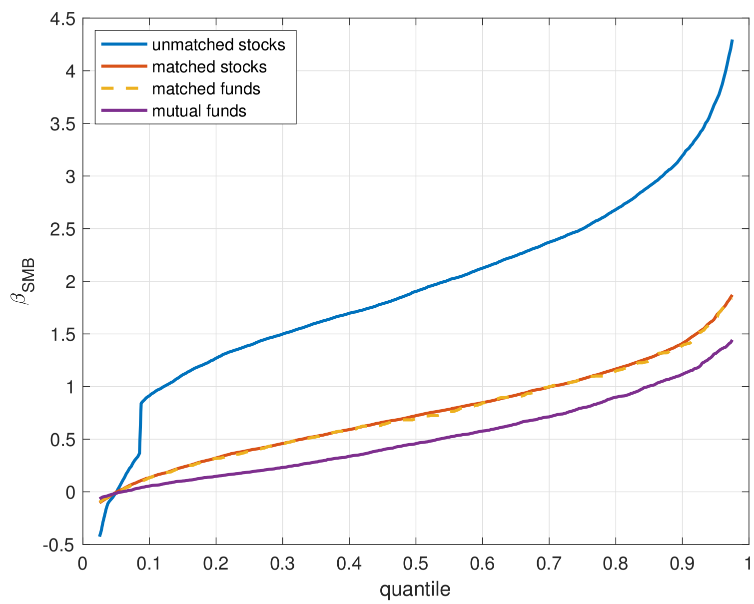
(a) Match on β_{MKT}



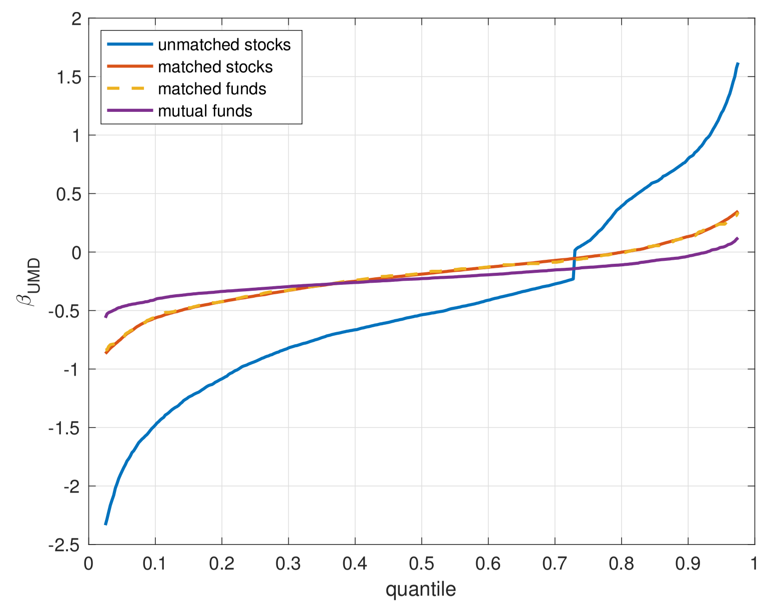
(b) Match on β_{HML}



(c) Match on β_{SMB}



(d) Match on β_{UMD}



B. Evaluation of Match Quality

Figure A.II plots the distribution of each matching variable (β) for unmatched stocks, matched stocks, matched mutual funds, and all mutual funds. We see immediately that stocks have more variable factor exposures than mutual funds, so the most extreme stocks on either side of the beta distributions cannot be matched to mutual funds. Of stocks that are matched, their beta distributions line up well with those of mutual funds: there are no systematic biases at any point in the distribution of stock betas, as evidenced by the absence of over- and undershooting of the red line by the yellow-dashed line. Matching in the tails of the stock beta distributions is achieved by oversampling relatively extreme mutual funds. This feature manifests as a counterclockwise rotation of the purple quantile plot for all mutual funds to achieve the yellow-dashed quantile plot of matched mutual funds.

Table A.III quantifies match quality depicted in Figure A.II. All quintiles and factors have highly similar means and standard deviations between matched stocks (S_M) and matched mutual funds (S_U). All factors are matched well in all quintiles, and overall match percentages are high (62%–76%). Match rates deteriorate in the most extreme quintiles, and particularly so in quintile five of market beta (the largest betas) and market capitalization (the smallest stocks). In these quintiles the proportion of matched to unmatched stocks becomes less favorable, and stocks match more frequently to the same extreme mutual funds. For example, the typical matched mutual fund in the smallest stock group is used more than 12 times: 989,081 stock-months are matched to three of 246,168 unique mutual fund-months.

The drawback to matching on a single variable is that other factors determining returns may differ considerably between stocks and matched mutual funds. Figure A.III confirms this issue by plotting bivariate distributions of four-factor betas when matches are constructed based only on β_{MKT}. Perfect matching between stocks and mutual funds would appear visually as complete coverage of the green regions by the red region. Instead we see green clouds surrounding the red region, indicating that matched mutual funds do not cover the same range of non-market betas as matched mutual funds. Focusing on the third column of the first row as an example, we see that matched stocks tend to have significantly larger β_{SMB} than matched mutual funds, so the existence of a size premium would create a positive wedge between the returns on mutual funds and stocks.

Table A.IV quantifies these visual disparities. Taking the same (1,3) coordinate, we see that the typical matched-stock size betas are consistently 0.4–0.6 larger for stocks than for mutual funds when entities are matched exclusively on market beta. Such differences are rife throughout the table. An apples-to-apples comparison of stocks and mutual funds thus requires multivariate matching if the true model of average returns has factors other than the market.

Figure A.IV reports match quality when matching uses all four factor betas and a wider caliper

[35] Bias-correction is optional in univariate matches, but it considered to be best practice. It is required to correct for a "large-sample bias" for multivariate matches (Abadie and Imbens (2011)).

Table A.III: Comparison of Samples on Matching Variable — Univariate β Matching

Table presents the distribution of the matching variable for unmatched stocks (S_U), matched stocks (S_M), and matched mutual funds (MF_M) for each quintile (first column) and sorting variable (top row). Matches are constructed monthly. Stocks are considered "matched" at date t if and only if they have at least three mutual funds within a Mahalanobis distance of 0.25σ of the matching variable during month t. Covariances for the Mahalanobis metric are calculated across all stocks and mutual funds and all dates. μ is the average value of betas within a bin for each sorting variable, σ is the standard deviation of these betas, and N is the count of bin elements. The final row tabulates the fraction of stocks successfully matched to mutual funds. All summary statistics are constructed across all dates.

Beta		MKT			HML			SMB			UMD		
		μ	σ	N	μ	σ	N	μ	σ	N	μ	σ	N
Sort variable		Rolling β_{MKT}			B/M Ratio			Market Capitalization			Prior 2-12 Return		
Q1	S_U	0.48	0.54	117623	-1.33	1.89	211377	0.42	0.82	34662	-0.47	1.18	290081
	S_M	0.64	0.34	328482	-0.48	0.69	496027	0.42	0.43	147673	-0.22	0.33	478300
	MF_M	0.64	0.34	159820	-0.47	0.67	218975	0.42	0.43	153337	-0.22	0.32	200744
Q2	S_U	0.94	0.52	68325	-0.93	1.69	102010	1.24	0.95	45201	-0.46	0.87	121630
	S_M	0.85	0.30	384677	-0.37	0.64	322581	0.61	0.45	166031	-0.21	0.27	323205
	MF_M	0.85	0.30	200223	-0.37	0.63	199108	0.61	0.44	164850	-0.21	0.27	196867
Q3	S_U	1.28	0.46	74585	-0.54	1.57	82297	1.66	0.90	79294	-0.44	0.80	99315
	S_M	1.01	0.29	378894	-0.26	0.63	298930	0.74	0.46	189659	-0.20	0.25	291625
	MF_M	1.01	0.29	207395	-0.27	0.62	199268	0.73	0.46	165391	-0.20	0.25	194710
Q4	S_U	1.49	0.41	112962	-0.27	1.50	84935	1.92	0.90	144050	-0.43	0.81	101922
	S_M	1.17	0.33	338522	-0.19	0.63	308536	0.85	0.48	264145	-0.20	0.26	294305
	MF_M	1.17	0.33	170017	-0.19	0.62	208690	0.84	0.47	169842	-0.20	0.25	193558
Q5	S_U	1.88	0.62	212070	-0.18	1.66	101835	2.10	1.22	769150	-0.40	0.97	184801
	S_M	1.42	0.44	236074	-0.18	0.64	403905	0.81	0.50	989081	-0.21	0.29	406564
	MF_M	1.42	0.44	102654	-0.18	0.63	223381	0.79	0.50	246168	-0.21	0.28	197637
% Matched		74.0%			75.6%			61.5%			69.0%		

Figure A.III: Comparison of Samples on All Variables — Univariate β Matching on β_{MKT}

Figure plots the distribution of factor betas for unmatched stocks, matched stocks, and matched mutual funds. Matches are constructed monthly using the single match variable β_{MKT}, and plots depict all bivariate distributions of Carhart (1997) four-factor betas. Stocks are considered "matched" at date t if and only if they have at least three mutual funds within a Mahalanobis distance of 0.25σ of the matching variable during month t. Covariances for the Mahalanobis metric are calculated across all stocks and mutual funds and all dates. To enhance visual clarity we clip the distribution of betas at the 2.5 and 97.5 percentiles and plot every 10th data point for unmatched and matched stocks. We plot every 30th data point for matched mutual funds because each matched stock has three associated mutual funds in its approximating set. Diagonal elements plot univariate histograms on a single beta rather than bivariate distributions.

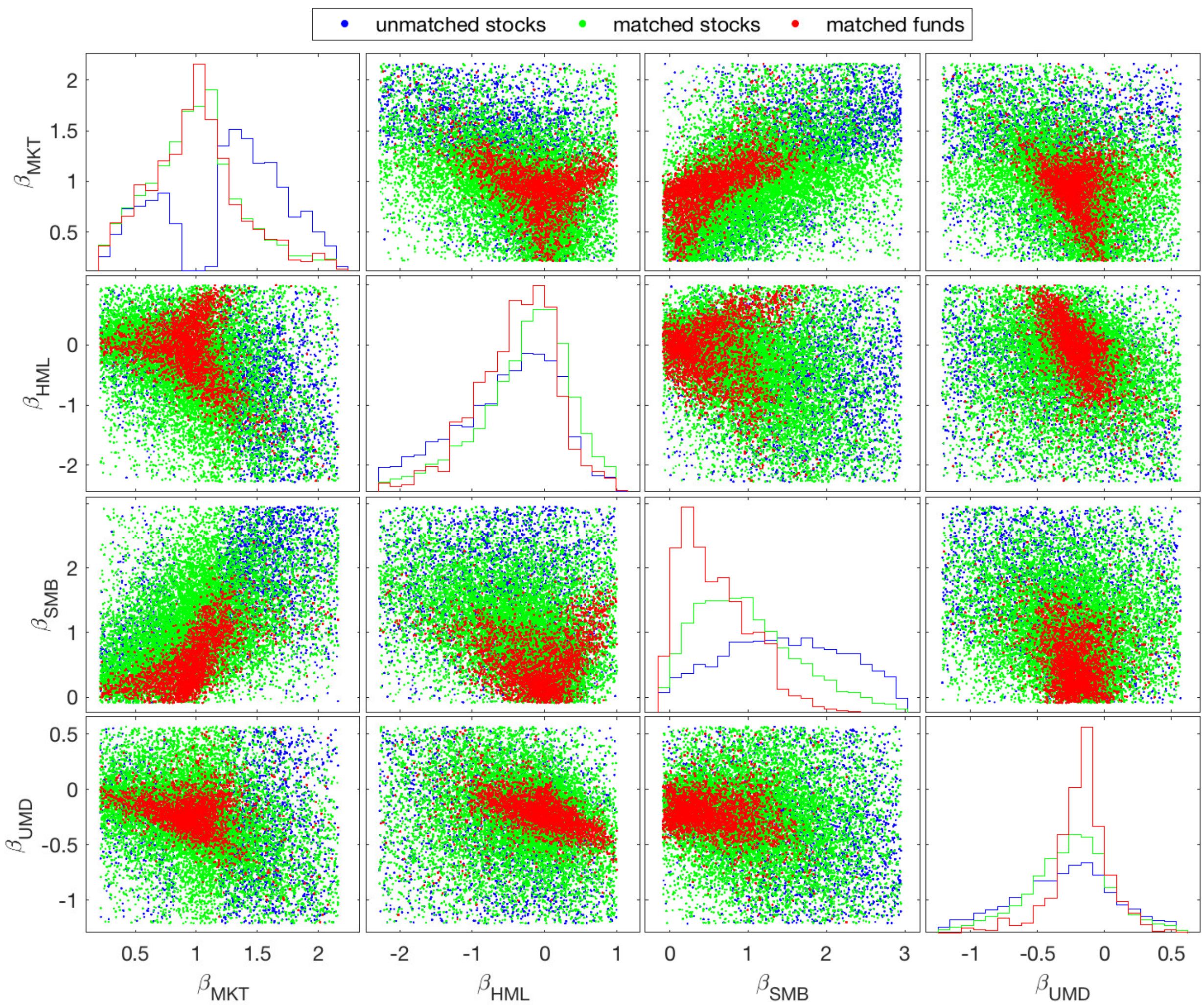


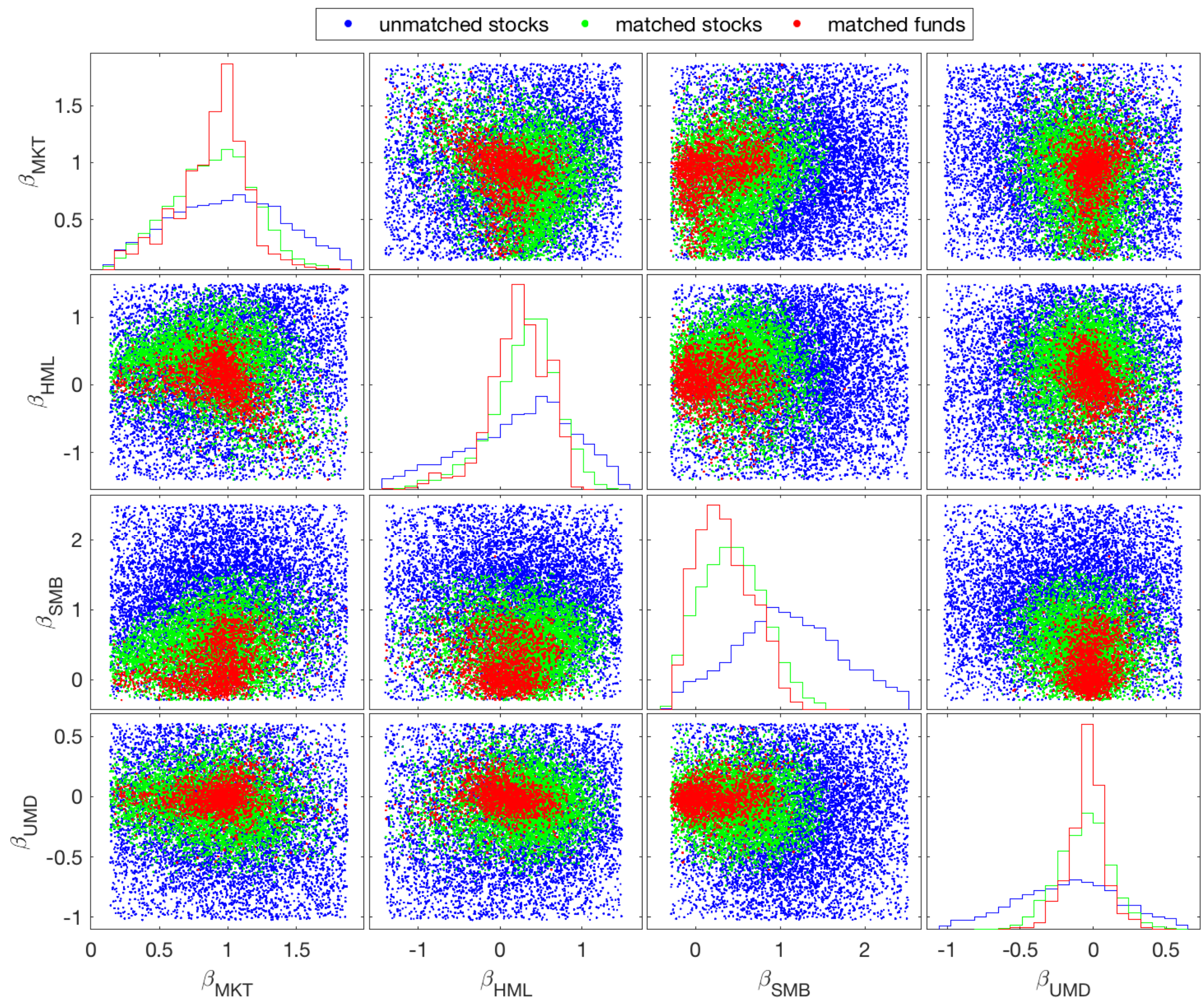
Table A.IV: Comparison of Samples on Matching Variable — Univariate β Matching on β_{MKT}

Table presents the distribution of factor betas for unmatched stocks (S_U), matched stocks (S_M), and matched mutual funds (MF_M) for each quintile (first column) of the rolling market beta distribution. Matches are constructed monthly on lagged market beta. Stocks are considered "matched" at date t if and only if they have at least three mutual funds within a Mahalanobis distance of 0.25σ of the matching variable during month t. Covariances for the Mahalanobis metric are calculated across all stocks and mutual funds and all dates. μ is the average value of betas within a bin for each factor, σ is the standard deviation of these betas, and N is the count of bin elements. The final row tabulates the fraction of stocks successfully matched to mutual funds. All summary statistics are constructed across all dates.

		MKT			HML			SMB			UMD		
		μ	σ	N	μ	σ	N	μ	σ	N	μ	σ	N
Sort variable		Rolling β_{MKT}			Rolling β_{MKT}			Rolling β_{MKT}			Rolling β_{MKT}		
Q1	S_U	0.48	0.54	117623	-0.07	0.80	117623	0.77	0.88	117623	-0.15	0.55	117623
	S_M	0.64	0.34	328482	-0.07	0.70	328482	0.75	0.80	328482	-0.17	0.44	328482
	MF_M	0.64	0.34	159820	-0.16	0.38	159820	0.29	0.33	159820	-0.12	0.16	159820
Q2	S_U	0.94	0.52	68325	-0.31	0.83	68325	1.16	0.93	68325	-0.28	0.56	68325
	S_M	0.85	0.30	384677	-0.13	0.65	384677	0.86	0.70	384677	-0.22	0.38	384677
	MF_M	0.85	0.30	200223	-0.27	0.45	200223	0.42	0.37	200223	-0.15	0.18	200223
Q3	S_U	1.28	0.46	74585	-0.56	0.87	74585	1.57	0.88	74585	-0.36	0.61	74585
	S_M	1.01	0.29	378894	-0.25	0.69	378894	1.00	0.71	378894	-0.27	0.40	378894
	MF_M	1.01	0.29	207395	-0.38	0.54	207395	0.58	0.42	207395	-0.17	0.22	207395
Q4	S_U	1.49	0.41	112962	-0.76	0.95	112962	1.83	0.85	112962	-0.39	0.65	112962
	S_M	1.17	0.33	338522	-0.42	0.83	338522	1.18	0.78	338522	-0.30	0.47	338522
	MF_M	1.17	0.33	170017	-0.53	0.65	170017	0.75	0.45	170017	-0.20	0.27	170017
Q5	S_U	1.88	0.62	212070	-1.18	1.44	212070	2.15	1.13	212070	-0.44	0.93	212070
	S_M	1.42	0.44	236074	-0.65	1.14	236074	1.48	0.92	236074	-0.40	0.60	236074
	MF_M	1.42	0.44	102654	-0.71	0.83	102654	0.95	0.49	102654	-0.29	0.35	102654
% Matched			74.0%			75.6%			61.5%			69.0%	

Figure A.IV: Comparison of Samples on All Variables — Multivariate β Matching

Figure plots the distribution of factor betas for unmatched stocks, matched stocks, and matched mutual funds. Matches are constructed monthly using all Carhart (1997) four-factor betas, and plots depict all bivariate distributions of these betas. Stocks are considered "matched" at date t if and only if they have at least three mutual funds within a Mahalanobis distance of 0.5σ of the matching variables during month t. Covariances for the Mahalanobis metric are calculated across all stocks and mutual funds and all dates. To enhance visual clarity we clip the distribution of betas at the 2.5 and 97.5 percentiles and plot every 10th data point for unmatched and matched stocks. We plot every 30th data point for matched mutual funds because each matched stock has three associated mutual funds in its approximating set. Diagonal elements plot univariate histograms on a single beta rather than bivariate distributions.



of 0.5σ. The figure clarifies the trade-off between high multivariate match quality and sample coverage. The blue region of unmatched stocks is quite small in the univariate matches, but it visually dominates here. Likewise the matching along any single dimension is not quite as good as in the univariate matches (e.g., Figure A.II), as the red and green regions of matched mutual funds and matched stocks do not perfectly coincide. However, these regions are much more similar than in the previous figure, and the differences between matched stock and matched mutual fund betas are small enough to be tamed by our local bias-adjustment methodology.

Table A.V quantifies this trade-off. About three quarters of the sample is matched (the size of the blue region in Figure A.IV overstates the sparse-matching problem because the red and green regions are more densely populated). The distributions of matched stocks and matched mutual funds are mostly comparable, but they differ in the tails as more extreme stock betas are matched with less extreme mutual fund betas within our generous caliper. The table confirms the necessity of bias adjustment for this high-dimensional match.

Table A.V: Comparison of Samples on All Variables — Multivariate β Matching

Table presents the distribution of factor betas for unmatched stocks (S_U), matched stocks (S_M), and matched mutual funds (MF_M) for each quintile (first column) of the corresponding characteristic. Matches are constructed monthly on lagged betas for all four Carhart (1997) factors. Stocks are considered "matched" at date t if and only if they have at least three mutual funds within a Mahalanobis distance of 0.5σ of matching variables during month t. Covariances for the Mahalanobis metric are calculated across all stocks and mutual funds and all dates. μ is the average value of betas within a bin for each factor, σ is the standard deviation of these betas, and N is the count of bin elements. The final row tabulates the fraction of stocks successfully matched to mutual funds. All summary statistics are constructed across all dates.

		MKT			HML			SMB			UMD		
		μ	σ	N	μ	σ	N	μ	σ	N	μ	σ	N
Sort variable		Rolling β_{MKT}			B/M Ratio			Market Capitalization			Prior 2-12 Return		
Q1	S_U	0.52	0.79	233575	-0.05	0.69	463476	0.11	0.40	48861	-0.23	0.16	563395
	S_M	0.60	0.78	212530	0.10	0.41	243928	-0.01	0.18	133474	-0.07	0.06	204986
	MF_M	0.65	0.86	89922	0.07	0.30	135833	0.06	0.18	104771	-0.04	0.02	104320
Q2	S_U	0.82	1.08	210159	0.08	0.71	241485	0.58	0.94	75238	-0.18	0.10	245100
	S_M	0.81	0.98	242843	0.25	0.53	183106	0.23	0.45	135994	-0.06	0.06	199735
	MF_M	0.83	0.97	120591	0.19	0.44	113808	0.22	0.37	103756	-0.04	0.02	113036
Q3	S_U	0.98	1.23	228389	0.24	0.80	205448	0.90	1.25	127198	-0.16	0.10	195728
	S_M	0.96	1.11	225090	0.34	0.61	175779	0.40	0.63	141755	-0.05	0.06	195212
	MF_M	0.95	1.06	118484	0.26	0.50	103057	0.33	0.52	91137	-0.04	0.02	115284
Q4	S_U	1.12	1.38	271331	0.33	0.87	207796	1.10	1.49	238202	-0.16	0.11	198351
	S_M	1.07	1.22	180153	0.38	0.62	185675	0.55	0.78	169993	-0.05	0.06	197876
	MF_M	1.03	1.12	92282	0.29	0.52	99544	0.44	0.66	87228	-0.04	0.02	115689
Q5	S_U	1.41	1.69	361787	0.39	0.98	288691	1.28	1.76	1254622	-0.18	0.14	367664
	S_M	1.18	1.35	86357	0.38	0.62	217049	0.53	0.78	503609	-0.06	0.07	223701
	MF_M	1.11	1.20	50835	0.29	0.52	104153	0.39	0.60	124666	-0.04	0.02	112299
% Matched		82.5%			76.4%			65.3%			74.4%		