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

Andrew J. Patton and Brian M. Weller*

Duke University

December 10, 2017

Abstract

Is there a gap between the profitability of a trading strategy "on paper" and that which can be achieved in practice? We answer this question by developing two new techniques to measure the real-world implementation costs of financial market anomalies. The first method extends Fama-MacBeth regressions to compare the on-paper returns to factor exposures with those achieved by mutual funds. The second method estimates average return differences between stocks and mutual funds matched on risk characteristics. Unlike existing approaches, these techniques deliver estimates of implementation costs without estimating parametric microstructure models from trading data or explicitly specifying 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

*Email: andrew.patton@duke.edu, brian.weller@duke.edu. We thank Alex Chinco, Campbell Harvey, David Hsieh, Robert Korajczyk, Juhani Linnainmaa, Toby Moskowitz, and seminar participants at the Yale Junior Finance Conference, Duke Fuqua, and John Cochrane's 60th Birthday Conference for valuable comments and discussion.

I. Introduction

Empirical asset pricing overflows with explanations for differences in average returns across securities. The proliferation of predictors distracts from bona fide market anomalies from which we might draw lessons about risks, preferences, and beliefs. Increasing computing power accelerates the rate of factor discovery and the urgency of separating empirical regularities from journal-friendly fictions. 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 are real, statistical accidents, or intentional deceptions.

We give on-paper trading strategies the benefit of the doubt and instead question whether these are actionable 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 transactions 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 the transactions 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 are almost by definition not representative of asset managers as a whole, such analyses provide an informative lower bound on the transactions costs of factor strategies. The second approach uses market-wide trading data such as NYSE Trade and Quote (TAQ) to estimate price impact functions for individual securities. These papers then accumulate simulated costs of trades 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 camp either establish a lower bound on trading costs by dispensing with non-proportional costs of trade or estimate parametric price impact models and extrapolate trading costs beyond small portfolio sizes.

We introduce two new methodologies for measuring real-world implementation costs for factor trading strategies. Our approaches differ from existing methods in that they do not utilize specialized trading data or parameteric transaction cost models, nor do they require the user to take a stand on the exact form of factor trading strategies. This latter feature is particularly important because high-turnover academic factors like momentum may have industry variants with significantly lower implementation costs. Together these features facilitate the estimation of the all-in implementation costs for several academic factors and a representative population of asset managers.

Our first approach is an extension of the familiar Fama and MacBeth (1973) procedure. Fama-MacBeth regressions estimate factor loadings β_{ik} with time-series regressions for each test asset i and each factor k, and then estimate the compensation per unit of factor exposure λ_{kt} using cross-sectional regressions at each date t. Standard test assets are based on stock portfolios, and the resulting estimates of of factor exposure compensation, denoted λ_{kt}^{S}, represent the "on-paper" profitability of a given factor strategy. We augment the set of test assets to include all the set of all 7,331 U.S. domestic mutual funds, and we allow the compensation earned by mutual funds for the same exposure to a given factor λ_{kt}^{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 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.

The Fama-MacBeth approach described above compares *slopes* or incremental compensation for risk for stocks and mutual funds. Our second approach directly compares *levels* of compensation for stocks and mutual funds with similar risk characteristics. To make this comparison, we sort stocks into quintile portfolios based on characteristics, as in portfolio sorts or "high minus low" factor construction. Then, for each stock in quintile q at date t, we construct a matched sample of mutual funds consisting of the three nearest mutual funds as assessed by Mahalanobis distance on factor betas.[3] The difference in returns between stocks and matched mutual funds Δ_{kt}^{q} is our matched pairs estimate of implementation costs associated with real-world trading in beta-quintile q of factor k at date t. Because implementation costs are likely to be significantly higher on the short side of factor strategies, the difference in long-only returns Δ_{kt}^{5} delivers a lower bound on the implementation costs of a 5–1 long-short strategy for factor k. This analysis has the ancillary benefit of controlling for differences in the distribution of betas between stock portfolios and mutual funds, which may be important if high compensation for factor exposure is earned only in some segments of this distribution.

Our empirical analysis focuses on the implementation costs of mutual funds for the market (MKT),[4] 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 they serve as the basis for hundreds of billions of dollars in quantitatively managed assets. We study mutual funds as our set of asset managers because they collectively manage more than $16 trillion of capital in the United States,[5] and the mutual fund industry

[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] Betas for stocks and matched mutual funds do not perfectly coincide. We follow Abadie and Imbens (2006, 2011) to shift mutual fund returns by the compensation for a "local" difference in betas. Because this return-adjustment takes the form of a cross-sectional regression of returns on betas, our approach marries matched pairs techniques with traditional factor models.

[4] We use the CRSP value-weighted market portfolio net of the one-month Treasury bill rate as our zero-cost market factor proxy.

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

has been better populated for a longer period of time than alternative asset managers such as hedge funds. Our approaches are readily extended to other factors and other market participants, however.

Our analysis delivers three new empirical facts on the implementation costs of anomalies for typical mutual funds. First, momentum strategies suffer extreme underperformance in practice: our estimates of annual implementation costs are in the range of 2.2%–8.3%, which eliminates most profits accruing to momentum during the 1970–2016 period. We conclude—as Lesmond, Schill, and Zhou (2004) do—that transactions costs make momentum strategies inaccessible to typical asset managers. Second, mutual fund implementation costs sharply reduce returns to the value factor, although our approaches differ on the extent of underperformance; the corresponding average annual differences between on-paper and mutual fund performance for value are 2.4%–5.3%.[6] By contrast, mutual funds implement market and size factor exposure relatively well compared to the academic benchmark, although they still underperform in capturing outsize returns to the small-size characteristic. Third, average implementation costs are stationary despite secular declines in bid-ask spreads and commissions since 1970. Industry inflows increase transactions costs, which neutralize reductions in costs to the first dollar traded in factor strategies. This channel provides a non-proportional costs of trading rationale for Pastor and Stambaugh (2012)'s industry-level diseconomies of scale in asset management.

As a second empirical contribution, we focus on cross-sectional heterogeneity in implementation costs and demonstrate the importance of benchmarking performance and considering market efficiency as a function of investors' trading technology. 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 market segments is 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.

While our new approaches deliver simple, nonparametric, estimates of the implementation costs for factor trading strategies, they do face some limitations. Firstly, as mentioned above, both approaches deliver lower bounds on implementation costs. In our empirical analysis these bounds do not greatly limit the 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

[6] These results are robust to the exclusion of microcap stocks. Microcap stocks explain at most 17%–27% of the measured implementation gap.

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 implementation problem with a factor, but they cannot deliver a clean bill of health.

Secondly, our techniques rely on real-world asset managers to reveal implementation costs through realized returns to their chosen factor exposures. We require a subset of asset managers to invest in the factor of interest over an extended period of time. This requirement is not likely to be satisfied when studying a factor that is new to the academic literature and fund managers have not had an opportunity to trade on this factor.[7]

Finally, like much of the literature on performance evaluation, our methods are 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. In our application we verify that omitted mutual fund strategies do not drive our high implementation cost estimates by replicating 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 95%). For these funds, the scope for omitted strategies is too small to explain the observed real-world performance gaps. In general settings, however, our methods speak only to implementation costs of the projection of the returns to traded strategies on the included factors.

Confronted with hundreds of cross-sectional return predictors, recent papers have proposed techniques to help focus attention on the most robust anomalies.[8] Notwithstanding the limitations of our approach mentioned above, we recommend our methodologies as complements to these suggestions for three reasons. First, our methodologies provide an easy test of the real-world applicability of a conjectured factor. If mutual funds are not compensated for factor exposure, a factor is less likely to be real or implementable. We anticipate that our implementability test generalizes Hou, Xue, and Zhang (2017)'s suggestion to exclude microcap stocks in that strategies relying on the smallest stocks would see large real-world performance attrition relative to paper portfolios. Second, our approaches provide orthogonal information to existing asset pricing tests. While it might be possible to reconfigure empirical choices to elevate a t-statistic from 2 to 3, we view it as less likely that an entirely new hurdle can be cleared for spurious factors. Third, the computational burden of our technique is low, and mutual fund performance data is readily available to empirical researchers. With our techniques the barriers to entry for cross-sectional

[7]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).

[8]Harvey, Liu, and Zhu (2016) advocate raising statistical significance thresholds. Harvey (2017) endorses mixing standard thresholds with Bayesian priors on the plausibility of a factor. Giglio and Xiu (2017) suggest "cleaning" factors of noise using variation in test asset returns.

asset pricing work are not appreciably raised.

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.[9] 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 transactions 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 transactions costs of return factors and anomalies.

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. 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), who analyze the transactions costs of a variety of investment styles for \$83 billion of trades.[10] 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, 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 transactions 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

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

[10] 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 transactions costs meaningfully erode mutual fund returns.

estimated price impact functions to factor trading strategies.[11] 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 trades in "disproportionately high cost securities" rather than the typical-transactions cost securities Jegadeesh and Titman (1993) use for approximating the costs of trading momentum. Using 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) estimate effective and quoted spreads and non-proportional trading costs functions of Glosten and Harris (1988) and Breen, Hodrick, and Korajczyk (2002) using TAQ data. 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. While much smaller than Frazzini, Israel, and Moskowitz (2015)'s estimates, these carrying capacities are also measured based on older data for which transactions costs are significantly higher (Lou and Sadka (2016)). Novy-Marx and Velikov (2016) follows a similar approach, but they focus only on proportional costs to establish a lower bound. The authors 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. Although the paper focuses on performance evaluation, Novy-Marx and Velikov (2016) find a momentum strategy carrying capacity of $5 billion (as in Korajczyk and Sadka (2004)), and they find size and value carrying capacities of $170 billion and $50 billion respectively (which are comparable to Frazzini, Israel, and Moskowitz (2015)'s estimates).

Our work complements the two existing approaches to measuring implementation costs with new cross-sectional techniques that combine 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, Og-

[11] 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.

den, and Trzcinka (1999)'s zero-trading day measure fails). Like papers that estimate transaction cost functions using market data, our methodology captures representative practitioners of factor investing rather than single special investment managers. In contrast with both approaches, our methodologies facilitate 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. This latter feature allows us to examine asset manager size strata separately in Section VI. In so doing we reconcile the contrasting findings of Lesmond, Schill, and Zhou (2004) and Frazzini, Israel, and Moskowitz (2015) by finding no net-of-costs return to momentum for *typical* mutual funds, but *positive* net-of-costs returns to the marginal mutual fund investor.

In concurrent work, Arnott, Kalesnik, and Wu (2017) propose a method similar to our Fama-MacBeth two-stage regressions. They 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, in our Fama-MacBeth approach we compare cross-sectional slopes of mutual fund and stock portfolio returns with respect to factor exposures, whereas Arnott, Kalesnik, and Wu (2017) compare mutual fund return slopes and on-paper time-series return realizations. The latter comparison sheds little light on implementation costs because realized factor slopes and factor returns may have very different means, as they do for the market factor. Second, our procedures address the omitted variable bias resulting from comovement in factor realizations and liquidity costs. In particular, the mismatch between first- and second-stage Fama-MacBeth regressions and the adaptive Lasso approach to accommodating heterogeneity in firm-relevant liquidity proxies are new to the implementation cost conversation. Third, we develop an entirely new approach to measuring implementation costs based on matching stocks and mutual funds with similar attributes to mimic the performance of characteristic-based portfolios in practice, and this approach reveals important differences between the size factor and characteristic. Finally, we slice the cross section of mutual funds to reconcile previous work on implementation costs, and we connect our stationary implementation costs empirically to recent work on industry-level diseconomies of scale in institutional investing.

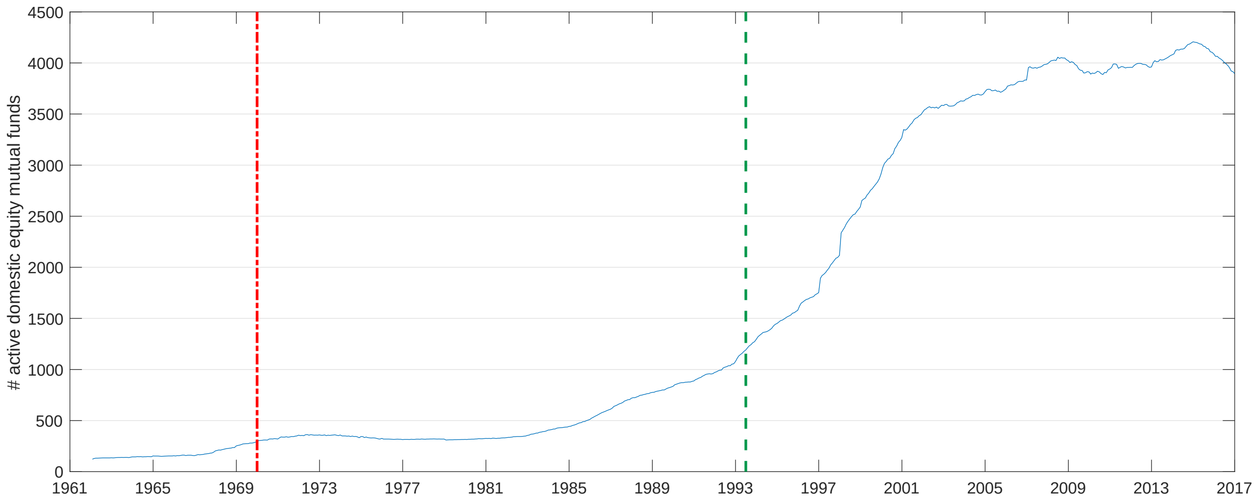
III. Data

Our data primarily come from the CRSP mutual fund and stock databases. Our mutual fund sample consists of 7,320 United States domestic equity mutual funds with at least 24 non-missing monthly gross returns during the period January 1970 to December 2016. Our stock sample, also monthly to match the mutual fund sample, consists of all CRSP stocks (share codes 10 or 11) for the corresponding period and subject to the same non-missing requirement, which gives a total of 22,121 unique PERMNOs.

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),

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.



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. CRSP provides returns net of management and 12b fees, and we convert these into gross returns by adding expense ratios divided by 12, following Fama and French (2010).

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.[12] 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 298 in January 1970 to 1,140 in July 1993 to 3,896 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 latter 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 1,071,818 fund-month

[12]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 7,331 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	1901	146.2	129.79	436.02	1554.00
Std. Dev.	1555	108.34	318.16	1337.10	6855.50
25%	357	64	3.81	30.47	49.80
50%	1202	119	22.78	88.00	201.40
75%	3724	200	89.60	290.26	861.70

Unit	Mean Return % / Month	Return Vol. % / Month	Sharpe Ratio Annualized	$\bar{\rho}_{MF}$ %	$\rho_{S\&P500}$ %
Mean	0.48	4.95	0.42	76.83	86.64
Std. Dev.	0.62	1.86	0.43	16.28	18.09
25%	0.33	3.98	0.24	74.79	84.28
50%	0.59	4.71	0.45	80.44	91.13
75%	0.80	5.61	0.61	84.49	95.25

observations and the 1993–2016 sample consists of 929.446 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 test portfolios. Because our factor set includes value (HML), size (SMB), and momentum (UMD), our baseline analysis uses 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 loadings to identify risk premia in the cross section. 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 also include 49 industry portfolios, 25 size-operating profitability portfolios, 25 size-investment portfolios, 10 beta-sorted portfolios, 10 market capitalization-sorted portfolios, 10 book equity to market equity 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 by monthly averages of non-missing daily absolute returns over dollar volume, and stocks are assigned for the following year using deciles the end of June, by parallel with the timing convention of the other portfolio data. The 25 size-illiquidity portfolios sort on both lagged market capitalization and Amihud illiquidity quintile. 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).[13] 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

We measure the returns to factor investing using two approaches. 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 the next section we evaluate the gross-of-fee returns to investing in factor portfolios of stocks and mutual funds with similar risk attributes.

A. Baseline Estimates

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

[13] 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.

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 is standard except for the choice of test assets. As discussed in the preceding section, our sorting characteristics pertain to each factor in f_{kt}, and we have $N_S = 100$ and $N_S = 269$ stock portfolios depending on regression specification. In addition to stock portfolios, we also include as test assets $N_{MF} = 7,331$ mutual funds, of which several thousand are active in the typical period. 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 $N_S + N_{MF}$ time-series regressions are

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

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,

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

λ_{kt}^S is the realized price of risk for factor k and date t based on stock returns, and λ_{kt}^{MF} is the corresponding estimate based on mutual fund returns.[14] The difference $\hat{\lambda}_{kt}^{\Delta} \equiv \hat{\lambda}_{kt}^S - \hat{\lambda}_{kt}^{MF}$ is our estimate of the implementation costs for strategy k: it is the difference between the on-paper profitability 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.[15] Our point estimates are the average of the monthly differences in factor compensation $\bar{\lambda}_k^{\Delta}$, and we construct standard errors for this difference using Newey and West (1987) with 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 as-

[14] Because the indicators partition the set of observations, Regression (2) is equivalent to two separate cross-sectional regressions run on stocks and mutual funds.

[15] Julian Robertson's Tiger Management, one of the world's largest hedge funds in the late 1990s, collapsed precisely because it could not continue to service outflows associated with its value investment strategy. Shleifer and Vishny (1997) formalize the role of agency frictions in perpetuating anomalies.

suming 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. However, in using long estimation windows to estimate βs, we nonetheless take on potential measurement error arising from time-variation in stock portfolio or mutual fund risk exposures.

Following Lettau, Maggiori, and Weber (2014) and others, we omit the constant term 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, $\hat{\beta}_{ik}$ has a large and positive risk price regardless of whether a constant is included. None of the other factor risk premia are meaningfully affected.

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 0.26% per year greater for risk exposures taken in practice via mutual funds than on paper in (100 equal-weighted) stock portfolios. This difference changes sign to favor on-paper portfolios by 0.34% per year when assessed against the full set of 269 portfolios because the average compensation for market beta is higher in the larger sample. Neither effect is statistically or economically significant, and the absence of a performance gap is robust to using value-weighted portfolios (bottom panel) rather than equal-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 return to each factor 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 equal-weighted paper portfolios in this period. Compensation for size factor (SMB) exposure has a small positive point estimate for mutual funds, but these values are not reliably different from zero or from on-paper compensation for size.

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 reflects the netting out of common variation in factor realizations within each cross section. 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

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) Equal-Weighted Stock Portfolios

		1970 – 2016				**1993 – 2016**			
	N_S	MKT	HML	SMB	UMD	MKT	HML	SMB	UMD
λ^Δ	100	-0.26	5.01***	1.74*	7.34***	-0.12	4.38***	1.92	4.91***
t-stat		(-0.54)	(5.65)	(1.77)	(5.31)	(-0.17)	(4.45)	(1.39)	(3.18)
λ^Δ	269	0.34	3.41***	1.88*	8.66***	0.87	2.51**	2.03	6.94***
t-stat		(0.67)	(3.55)	(1.76)	(6.29)	(1.33)	(2.40)	(1.36)	(4.00)
λ^S	100	6.72***	7.76***	2.99	9.48***	7.70**	6.57**	4.12	6.18
t-stat		(2.77)	(4.25)	(1.53)	(3.95)	(2.32)	(2.42)	(1.48)	(1.63)
λ^S	269	7.31***	6.17***	3.13	10.80***	8.69***	4.69	4.23	8.21**
t-stat		(3.04)	(3.15)	(1.52)	(4.41)	(2.65)	(1.61)	(1.45)	(2.10)
λ^{MF}	—	6.98***	2.75	1.25	2.14	7.81**	2.19	2.21	1.27
t-stat		(2.87)	(1.61)	(0.71)	(0.86)	(2.40)	(0.79)	(0.92)	(0.34)
T		564	564	564	564	282	282	282	282
$\bar{N}_{MF}$		1901	1901	1901	1901	3299	3299	3299	3299

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

(b) Value-Weighted Stock Portfolios

		1970 – 2016				**1993 – 2016**			
	N_S	MKT	HML	SMB	UMD	MKT	HML	SMB	UMD
λ^Δ	100	-0.35	4.31***	-0.32	7.08***	-0.38	4.00***	-0.31	4.69***
t-stat		(-1.12)	(5.22)	(-0.45)	(5.29)	(-0.95)	(4.68)	(-0.33)	(3.13)
λ^Δ	269	-0.20	2.71***	-0.49	7.00***	0.18	2.42***	-1.03	5.40***
t-stat		(-0.8)	(3.86)	(-0.83)	(5.37)	(0.77)	(3.72)	(-1.32)	(3.37)
λ^S	100	6.62***	7.06***	0.94	9.23***	7.43**	6.18**	1.89	5.96
t-stat		(2.75)	(3.81)	(0.55)	(3.90)	(2.27)	(2.20)	(0.77)	(1.58)
λ^S	269	6.78***	5.46***	0.77	9.14***	7.99**	4.61	1.18	6.67*
t-stat		(2.83)	(2.98)	(0.45)	(3.91)	(2.46)	(1.62)	(0.49)	(1.80)
λ^{MF}	—	6.98***	2.75	1.25	2.14	7.81**	2.19	2.21	1.27
t-stat		(2.87)	(1.61)	(0.71)	(0.86)	(2.40)	(0.79)	(0.92)	(0.34)
T		564	564	564	564	282	282	282	282
$\bar{N}_{MF}$		1901	1901	1901	1901	3299	3299	3299	3299

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

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, but these returns are now universally statistically indistinguishable from zero (in part because the sample length is cut in half). For stock portfolios, the compensation for HML and UMD (SMB) exposures also decrease (increase) 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 insignificant). In sum, 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 implementing market exposure and size and the low real-world efficacy of implementing value and momentum.

B. Time- and Mutual-Fund Varying Per-Unit Cost Estimates

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. The η_{it} term in turn has components

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

The first component is the fixed, firm-specific cost of trading a factor. The second component is the 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 liquidity demand shock that thwarts or facilitates firm i's trading strategy for factor k.

Funds trade and earn returns

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

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 through investment managers. In the constant-cost setting of Section IV.A, we achieve this ideal: η_{it} simplifies to η, and Fama-MacBeth regressions with a K-factor model recovers consistent estimates of $f - h$ as in Equation (2).

By contrast, in this general setting we face two key challenges that complicate the straightforward comparison of f_t and h_{it}. First, trading costs vary over time, and these costs may covary with factor realizations. For example, several measures of funding and market liquidity deteriorate significantly during the 2007–2008 Financial Crisis, and the aggregate market return likewise is consistently negative at that time. 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 or is particularly skilled at trading a given factor cost-effectively is more likely to select a larger factor exposure, all else equal. For this reason we would expect mutual fund-specific trading gains η_i to be increasing in β_i, and the cross-sectional slopes of returns with respect to β_i are biased upward ($\hat{\lambda}_t^{MF} > \lambda_t^{MF}$).

We now address these two sources of bias. 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 discussed in Section III. Throughout we use liquidity levels rather than innovations because high illiquidity rather than increases in illiquidity relative to the previous month likely contribute to high factor implementation costs.[16] We then run Fama-MacBeth regressions as before, but we extend the factor model to include liquidity proxies in the time-series regressions,

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

where $\tilde{\eta}_{lt}$ are the liquidity factor proxies at time t. To avoid overfitting in the first stage by including too many correlated liquidity proxies, we start with two: 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). We normalize all liquidity variables to have unit standard deviation before taking principal components because liquidity proxies vary widely in their scales.[17]

[16] By contrast, if we sought to explain returns, *innovations* to liquidity expectations would be the correct variable to use (as emphasized by Pastor and Stambaugh (2003) and He, Kelly, and Manela (2017), among others).

[17] The CBOE VXO and the TED spread series start in January 1986. Our principal components procedure accom-

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. 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

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

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 are 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.[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 reports results using 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) compensation for UMD exposure decreases from 2.14 to 1.98 in the long sample and from 1.27 to 0.27 in the recent sample, suggesting that liquidity factor exposure at least partly explains mutual funds' compensation for momentum. This finding extends one result of Asness, Moskowitz, and Pedersen (2013) to the universe of mutual funds. 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.

Ideally we would use all liquidity variables rather than their principle components because we want to time-varying determinants of η_{it} to lie in the span of the liquidity-augmented factor model. Including more covariates increases the likelihood that we span 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 Equation (2).

Sparse regression techniques offer a solution to this challenge. We supplement the standard

modates 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. By construction illiquidity principal components have unit standard deviation, and we assign these components an illiquidity interpretation by normalizing them to be positively correlated with the VIX/VXO.

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

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) Equal-Weighted Stock Portfolios

		1970 – 2016				**1993 – 2016**			
	N_S	MKT	HML	SMB	UMD	MKT	HML	SMB	UMD
λ^Δ	100	-0.43	5.27***	2.06**	7.51***	-0.06	4.11***	1.97	5.57***
t-stat		(-0.90)	(5.81)	(2.10)	(5.34)	(-0.08)	(3.97)	(1.46)	(3.27)
λ^Δ	269	0.15	3.85***	2.29**	8.73***	0.78	2.56**	2.20	6.99***
t-stat		(0.30)	(3.86)	(2.14)	(6.13)	(1.17)	(2.37)	(1.54)	(3.72)
λ^S	100	6.55***	8.06***	3.22*	9.49***	7.77**	5.97**	4.22	5.84
t-stat		(2.71)	(4.34)	(1.67)	(3.97)	(2.35)	(2.20)	(1.54)	(1.54)
λ^S	269	7.13***	6.64***	3.45*	10.71***	8.61***	4.42	4.46	7.26*
t-stat		(2.99)	(3.31)	(1.69)	(4.38)	(2.65)	(1.50)	(1.57)	(1.85)
λ^{MF}	—	6.98***	2.79	1.16	1.98	7.83**	1.86	2.26	0.27
t-stat		(2.88)	(1.62)	(0.66)	(0.80)	(2.41)	(0.67)	(0.93)	(0.07)
T		564	564	564	564	282	282	282	282
$\bar{N}_{MF}$		1901	1901	1901	1901	3299	3299	3299	3299

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

(b) Value-Weighted Stock Portfolios

		1970 – 2016				**1993 – 2016**			
	N_S	MKT	HML	SMB	UMD	MKT	HML	SMB	UMD
λ^Δ	100	-0.43	4.70***	-0.24	7.30***	-0.38	4.35***	-0.34	5.69***
t-stat		(-1.36)	(5.55)	(-0.33)	(5.40)	(-0.94)	(5.00)	(-0.36)	(3.48)
λ^Δ	269	-0.21	2.99***	-0.45	7.15***	0.18	2.77***	-1.01	6.24***
t-stat		(-0.83)	(4.03)	(-0.76)	(5.46)	(0.78)	(4.09)	(-1.30)	(3.73)
λ^S	100	6.56***	7.49***	0.92	9.28***	7.45**	6.21**	1.91	5.96
t-stat		(2.73)	(3.99)	(0.54)	(3.93)	(2.29)	(2.19)	(0.78)	(1.58)
λ^S	269	6.78***	5.78***	0.71	9.13***	8.01**	4.63	1.24	6.52*
t-stat		(2.83)	(3.11)	(0.42)	(3.90)	(2.48)	(1.60)	(0.51)	(1.76)
λ^{MF}	—	6.98***	2.79	1.16	1.98	7.83**	1.86	2.26	0.27
t-stat		(2.88)	(1.62)	(0.66)	(0.80)	(2.41)	(0.67)	(0.93)	(0.07)
T		564	564	564	564	282	282	282	282
$\bar{N}_{MF}$		1901	1901	1901	1901	3299	3299	3299	3299

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

first-stage regression with a Lasso or l_1-penalized regression (Tibshirani (1994)). The least-squares minimization problem of Equation (6) adds an additional term penalizing the liquidity coefficients,

$$\min_{\beta,\tilde{\gamma}} \frac{1}{T_i} \sum_t \left(r_{it} - \sum_k f_{kt}\beta_{ik} - \sum_l \tilde{\eta}_{lt}\tilde{\gamma}_{il} \right)^2 + \kappa \left(\sum_k \omega_k \left|\beta_{ik}\right| + \sum_l \omega_l \left|\tilde{\gamma}_{il}\right| \right), \tag{8}$$

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

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 are OLS $\hat{\beta}$s (as in Zou (2006)), and our penalty exponent is $\gamma = 1$.

The obvious concern when using Lasso or adaptive Lasso is the selection of the penalization parameter κ. Following 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 of the estimated model on the fold that was set aside. This procedure provides k pseudo-out-of-sample (POOS) R^2s 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 R^2. Intuitively this process tames overfitting by selecting the model with the best out-of-sample predictive properties.[19]

Table IV presents results using the adaptive Lasso first stage described by Equation (8). 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 the prior table, 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 priced factor on the set of mutual fund test assets. Because λ_{UMD}^{MF} is an upper bound on mutual fund compensation for UMD exposure, Table (IV) implies that momentum strategies are not implementable for the typical mutual fund.

V. Matched Pairs Estimates of Implementation Costs

The cross-sectional approach of the previous section 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 the question of whether *large* investments by mutual funds achieve more favorable risk-reward trade-offs than marginal factor investments. For example, mutual funds may excel at taking on moderate risk exposures to a factor, but their performance may deteriorate for extreme exposures for which taking on additional leverage is needed (Frazzini and Pedersen (2014)). 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.[20]

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-

[19] 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 adjust standard errors for heteroskedasticity and serial correlation.

[20] 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 approach to directly compare mutual funds and stocks with similar multifactor betas.

Table IV: 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 $\hat{\beta}_{ik}$,

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

where k indexes the four Carhart (1997) factors and λ^{Δ} is defined as $\lambda^S - \lambda^{MF}$. First-stage regression estimates include these factors and all market and funding liquidity proxies in an adaptive Lasso regression with portfolio-specific penalty parameters κ_i 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 reported in parentheses.

(a) Equal-Weighted Stock Portfolios

		1970 – 2016				**1993 – 2016**			
	N_S	MKT	HML	SMB	UMD	MKT	HML	SMB	UMD
λ^{Δ}	100	-0.04	5.64***	1.90*	8.78***	0.25	3.84***	1.96	6.89***
t-stat		(-0.09)	(6.24)	(1.88)	(5.72)	(0.34)	(3.25)	(1.39)	(3.28)
λ^{Δ}	269	0.40	4.93***	2.19**	9.95***	1.05	2.35*	2.14	8.14***
t-stat		(0.79)	(4.84)	(2.02)	(6.44)	(1.55)	(1.95)	(1.46)	(3.82)
λ^S	100	6.88***	7.95***	3.11	9.64***	8.07**	5.41*	4.25	6.44*
t-stat		(2.85)	(4.12)	(1.58)	(3.97)	(2.46)	(1.89)	(1.51)	(1.65)
λ^S	269	7.33***	7.24***	3.41*	10.81***	8.88***	3.91	4.43	7.69*
t-stat		(3.06)	(3.41)	(1.66)	(4.41)	(2.74)	(1.24)	(1.54)	(1.95)
λ^{MF}	—	6.93***	2.30	1.22	0.86	7.83***	1.57	2.29	-0.45
t-stat		(2.85)	(1.18)	(0.67)	(0.33)	(2.41)	(0.51)	(0.92)	(-0.11)
T		564	564	564	564	282	282	282	282
$\bar{N}_{MF}$		1901	1901	1901	1901	3299	3299	3299	3299

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

(b) Value-Weighted Stock Portfolios

		1970 – 2016				**1993 – 2016**			
	N_S	MKT	HML	SMB	UMD	MKT	HML	SMB	UMD
λ^{Δ}	100	-0.10	5.29***	-0.62	8.25***	-0.28	4.64***	-0.34	6.12***
t-stat		(-0.3)	(6.38)	(-0.87)	(5.72)	(-0.68)	(4.89)	(-0.34)	(3.29)
λ^{Δ}	269	-0.03	3.68***	-0.75	7.86***	0.19	3.21***	-1.06	6.68***
t-stat		(-0.09)	(4.87)	(-1.25)	(5.77)	(0.75)	(4.32)	(-1.30)	(3.67)
λ^S	100	6.82***	7.59***	0.59	9.11***	7.54**	6.21**	1.95	5.67
t-stat		(2.83)	(3.94)	(0.34)	(3.83)	(2.32)	(2.10)	(0.78)	(1.48)
λ^S	269	6.90***	5.98***	0.47	8.72***	8.02**	4.77	1.23	6.23*
t-stat		(2.88)	(3.14)	(0.27)	(3.71)	(2.48)	(1.55)	(0.50)	(1.65)
λ^{MF}	—	6.93***	2.30	1.22	0.86	7.83**	1.57	2.29	-0.45
t-stat		(2.85)	(1.18)	(0.67)	(0.33)	(2.41)	(0.51)	(0.92)	(-0.11)
T		564	564	564	564	282	282	282	282
$\bar{N}_{MF}$		1901	1901	1901	1901	3299	3299	3299	3299

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

market ratio,[21] 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,[22] where β_k determines distance in our univariate analysis, all four βs determine distance in our multivariate analysis. 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 one to improve precision of the 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 $0.25 \times 4 = 1$ standard deviation for four-factor matching. Stocks with fewer than three matched mutual funds within these radii are dropped.

We follow Abadie and Imbens (2006, 2011) to establish a bias-adjusted matched pairs estimator for each month's implementation gap by calculating the average difference between next-month returns for stocks and mutual funds. Armed with monthly implementation gap estimates, we take the average value as our full-sample estimate. We also consider differences in value-weighted returns within each month using the lagged market capitalization of the matched stocks. Appendix C provides a detailed description of the matched pairs methodology and evaluation of match quality.

Our analysis includes univariate and multivariate matches because each mimics a different standard asset pricing approach and because each method entails making a particular set of trade-offs. Intuitively, matching on univariate betas answers the question of whether mutual funds are comparably good as stocks at isolating a particular characteristic; for example, do mutual funds with high (low) "value" betas achieve the same excess returns as high (low) value beta stocks? This analysis parallels a comparison between Q5–Q1 returns from two standard portfolio sorts, where one set of test assets (mutual funds) is transformed to have as similar risk attributes as possible to the

[21] We source our book-to-market ratios from the WRDS Financial Ratios Suite. We make small modifications to the provided code to extend the ratios to all stocks through the end of 2016.

[22] The Mahalanobis distance is $\sqrt{(x-y)' \, \Sigma^{-1} \, (x-y)}$ for two vectors x and y and covariance matrix Σ. When Σ is diagonal, it normalizes each dimension to have a unit standard deviation, and we adopt this terminology in the main text. It reduces to the Euclidean distance when $\Sigma = \mathbf{I}$.

other set of test assets (stocks). In favor of univariate matching is its high match rate and accurate matches along the dimension of interest, and this approach dominates when one matching feature is more important than others in determining stock and mutual fund returns. Matching on all betas answers the same question when excess returns are assessed with respect to a richer, nonparametric model that controls for possible variation in other risk factors. This multivariate-beta matching approach parallels multifactor tests such as Fama-MacBeth cross-sectional regressions in which we assess the returns to a single factor holding all other factors fixed. In favor of multivariate matching is its ability to simultaneously control for several determinants of returns to conduct a true "all else equal" analysis for stocks and mutual funds. This approach—which we favor in our application—trades off match quality in the dimension of interest with an elimination of systematic biases in the other dimensions. We discuss trade-offs of each approach in more detail in Appendix C.B.

For each matched pairs analysis, 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,

$$r_S - r_{MF} = \left(r_S^{long} - r_S^{short}\right) - \left(r_{MF}^{long} - r_{MF}^{short}\right) \geq \left(r_S^{long} - r_S^{short}\right) - \left(r_{MF}^{long} - r_S^{short}\right) = r_S^{long} - r_{MF}^{long}. \tag{9}$$

The inequality in Equation (9) 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.

Note that 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, as such costs are additive. Instead, we settle for comparing performance on the long portfolios to establish a lower bound on the implementation costs of asset pricing factors.

B. Results

Table V reports the results of our matched pairs analysis. The Δ^{LMS} value in the upper-left corner indicates that mutual funds underperform stocks with the same market beta exposure by 2.5% per year when stocks are in the 60–80th 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 long-short market beta portfolios from the equity premium $R_m - R_f$. This implementation gap

Table V: Returns of Matched Stocks and Matched Mutual Funds

Tables report differences in performance between stocks and matched mutual funds. The first two columns of each panel denote the difference in returns between stocks and mutual funds in quintiles four ("4") and five ("5") 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 equity premium $R_m - R_f$. Differences are equal- or value-weighted within each month and averaged across months. $5-1$ columns 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 first panel matches only on β_k, whereas the second panel matches on all four factors. Columns 1–6 equal weight returns, and columns 7–12 value weight returns by lagged market capitalization. t statistics are reported in parentheses.

(a) Sorting and Matching on Univariate Beta

	Equal Weighted						Value Weighted					
	1970 – 2016			**1993 – 2016**			**1970 – 2016**			**1993 – 2016**		
	4	5	$5-1$	4	5	$5-1$	4	5	$5-1$	4	5	$5-1$
Δ^{LMS}	2.50*	1.01	-0.70	4.41**	2.84	2.26	-1.01	-2.53**	-0.63	0.24	-0.61	2.32
t-stat	(1.65)	(0.49)	(-0.23)	(2.01)	(0.92)	(0.47)	(-1.00)	(-2.00)	(-0.18)	(0.16)	(-0.33)	(0.44)
Δ^{HML}	5.10***	9.14***	11.32***	4.54**	8.57***	10.29***	1.39	3.74***	4.77**	-0.71	0.55	1.51
t-stat	(3.69)	(4.69)	(5.82)	(2.57)	(3.43)	(3.67)	(1.60)	(2.89)	(2.32)	(-0.73)	(0.33)	(0.52)
Δ^{SMB}	3.07***	3.63**	3.82	2.88	3.61	4.00	3.06***	1.62	1.61	2.78	1.52	2.34
t-stat	(2.76)	(2.35)	(1.55)	(1.62)	(1.49)	(1.25)	(2.79)	(1.22)	(0.69)	(1.58)	(0.71)	(0.74)
Δ^{UMD}	5.96***	8.12***	6.17**	6.40***	7.06***	1.91	2.24**	3.88***	8.75***	2.18	3.43*	4.47
t-stat	(5.74)	(5.61)	(2.32)	(4.27)	(3.44)	(0.43)	(2.16)	(3.23)	(3.00)	(1.44)	(1.91)	(1.00)

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

(b) Sorting and Matching on Multivariate Betas

	Equal Weighted						Value Weighted					
	1970 – 2016			**1993 – 2016**			**1970 – 2016**			**1993 – 2016**		
	4	5	$5-1$	4	5	$5-1$	4	5	$5-1$	4	5	$5-1$
Δ^{LMS}	2.86***	3.55***	-0.70	4.44***	5.69***	2.26	-0.72	-0.40	-0.63	1.29	3.03*	2.32
t-stat	(4.57)	(3.51)	(-0.23)	(5.21)	(3.62)	(0.47)	(-1.04)	(-0.37)	(-0.18)	(1.55)	(1.92)	(0.44)
Δ^{HML}	3.99***	7.58***	11.32***	4.18***	7.69***	10.29***	0.07	3.30***	4.77	-0.05	3.19***	1.51
t-stat	(6.36)	(8.38)	(5.82)	(5.23)	(6.56)	(3.67)	(0.11)	(3.93)	(2.32)	(-0.06)	(2.90)	(0.52)
Δ^{SMB}	3.54***	5.86***	3.82	4.07***	5.88***	4.00	3.39***	4.03***	1.61	3.92***	4.28***	2.34
t-stat	(5.22)	(6.83)	(1.55)	(4.60)	(4.95)	(1.25)	(5.04)	(5.49)	(0.69)	(4.50)	(4.08)	(0.74)
Δ^{UMD}	5.05***	6.68***	6.17**	5.51***	6.21***	1.91	1.84***	2.19**	8.75***	1.89**	2.89**	4.47
t-stat	(9.30)	(8.76)	(2.32)	(7.60)	(5.98)	(0.43)	(2.60)	(2.22)	(3.00)	(2.23)	(2.25)	(1.00)

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

is larger than the on-paper return to an equal-weighted long-short strategy based on market beta quintiles, and roughly 40% of the annual equity premium over this period (6.35% per year, not tabulated). Moving to the next column on the right, mutual funds continue to underperform stocks by a much-smaller 1.0% per year, and this value is no longer statistically significant. Moving down the upper-left panel we see that the implementation gap is positive and statistically significant for all non-market factors considered. Costs are particularly high for momentum strategies, as prior literature suggests, and they are similarly high for value strategies.

The smaller point estimates in quintile four relative to quintile five is a common feature throughout the panels, and it reflects the balance of two competing forces. On one side, if per-unit costs are fixed across firms (as in Section IV.A), higher betas translate into higher total implementation gaps $\beta(h-f)$. On the other side, mutual funds select into the highest β group, so high-β group membership likely reflects lower per-unit underperformance $h-f$. For these reasons the product of $\beta(h-f)$ could increase or decrease as we move to more extreme β quintiles, and empirically, the scale effect tends to dominate the selection effect.

The upper-left 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 ($\rho_{\beta_{HML},\beta_{UMD}} = -22.7\%$), whereas momentum and value betas are positively correlated among stocks ($\rho_{\beta_{HML},\beta_{UMD}} = +16.1\%$). The bottom-left 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).

Controlling for the three non-sorting variables has a large effect on the statistical reliability of the market and size factor implementation costs, which suggests that funds with high-beta strategies likely differ from high-beta stocks on other dimensions. In the equal-weighted specification we see that the implementation gap is large and statistically significant for quintiles four and five for all anomalies. The implementation gaps are so large, in fact, that they swamp or severely attenuate factor returns for all four factors considered, regardless of whether we include all stocks or focus only on the investable set of matched stocks. In short, at least from an equal-weighted perspective, *no factors* earn returns after real-world costs during the 1970–2016 period. This finding also holds in the right panel 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).

Columns 7–12 of the top panel present value-weighted results. The contrast between equal- and value-weighted results testifies to the importance of size in dictating the ability of mutual funds to mimic return factors; differences in performance attenuate dramatically when small, harder-toaccess stocks are downweighted. Focusing first on the univariate matches, mutual fund underperformance on value, size, and momentum strategies falls by roughly half relative to the corresponding equal-weighted results, and mutual fund underperformance on the long-short market beta strategy reverts to zero or even to a significantly negative value. As we find in Section IV, mutual funds capture returns to market beta quite well.

Turning to the multivariate matches in the value-weighted specification restores our finding that mutual funds underperform matched stocks for the non-market factors. Implementation gaps are again smaller than in the equal-weighted results, but the magnitudes are nonetheless economically large for the three main anomalies: 0.1%–3.3% for value against a time-series average return of 4.8% for value-weighted HML; 3.4%–4.0% for size against a time-series average return of 1.6% for SMB; and 1.8%–2.2% against a time-series average return of 8.8% for UMD. These differences carry over to the more recent 1993–2016 interval, as well. In sum, 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.

VI. Cost Estimates Over Time and Across Funds

A. Comparison with Previous Implementation Cost Estimates

Our analysis thus far considers the implementation costs of factor strategies for representative mutual funds, 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, and 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 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), analogous to the R^2 splits in the previous section. We then run our second-stage cross-sectional regressions (2) separately for each asset manager TNA group.[23] We set aside funds with less than \$10 million in assets because selection into this group implies that the fund has lost money (recall that we retain observations only after funds reach \$10 million in assets to avoid incubation bias).[24]

Table VI presents results from these segmented regressions on the full 1970–2016 sample. As in Tables II–IV, 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, and returns to HML are collectively different from zero in all specifications. Micro funds drive our finding of low overall mutual fund compensation for value. Point estimates for returns to SMB are positive for all fund size groups, but SMB compensation estimates are not statistically distinguishable from zero or from each other.

Focusing on momentum, differences in compensation across mutual fund size categories are statistically significant and economically large, with the smallest funds earning 9%–10% 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)—momentum strategy performance is decreasing in fund size.[25] 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. We conclude that variation across mutual funds is important when considering the net-of-costs returns to the momentum factor.

Table VII compares these real-world factor return estimates with estimates from selected influential 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 and is considered by others in the literature.

Papers that consider a wider range of trading costs reach mixed conclusions on the implementability of momentum. Korajczyk and Sadka (2004) suggest that momentum profits exist only at small scales (the table reflects only proportional costs, and by their reckoning, non-proportional

[23]Groups are assigned separately for each date with cutoffs based on December 2016 USD. The micro-fund group has $TNA_t < \$10M$ and comprises 5.2% of the data. This group is selected in that it consists of past losers whose assets had at one time exceeded $\$10M$. 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.

[24]In principle we could complement this analysis with a matched pairs approach, but dividing the set of possible mutual funds into several groups significantly reduces match quality in the extreme-quintile portfolios.

[25]Intriguingly the Patton and Timmermann (2010) non-monotonicity test is much more sensitive to differences in λ^{MF} across size groups than the standard F-test for differences.

Table VI: Fama-MacBeth Slopes for Stocks and Mutual Funds — Size 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 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), the first principal component of market and funding liquidity proxies (second columns), and all liquidity proxies in adaptive Lasso regression (third 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 tests of coefficients being jointly different from zero, of equality of coefficients, and of non-monotonicity of coefficients (using the test of Patton and Timmermann (2010)), where we exclude the micro-fund group because funds must be past losers to select into that group.

	No Liquidity Proxies				**Liquidity PCs**				**Liquidity Lasso**			
	MKT	HML	SMB	UMD	MKT	HML	SMB	UMD	MKT	HML	SMB	UMD
λ_{mega}^{MF}	6.65***	3.05*	2.33	-5.39	6.66***	2.98	2.31	-5.68*	6.62***	3.26	2.31	-5.92*
t-stat	(2.75)	(1.65)	(1.27)	(-1.58)	(2.76)	(1.58)	(1.26)	(-1.70)	(2.75)	(1.56)	(1.23)	(-1.66)
λ_{large}^{MF}	6.83***	2.98*	1.10	1.63	6.83***	2.91	1.19	0.53	6.85***	2.16	1.04	-1.55
t-stat	(2.79)	(1.65)	(0.61)	(0.54)	(2.80)	(1.57)	(0.66)	(0.18)	(2.80)	(1.03)	(0.56)	(-0.52)
λ_{medium}^{MF}	6.96***	2.89*	1.39	2.84	6.93***	3.10*	1.46	2.75	6.95***	2.42	1.28	1.11
t-stat	(2.86)	(1.68)	(0.78)	(1.09)	(2.84)	(1.75)	(0.82)	(1.05)	(2.84)	(1.23)	(0.71)	(0.4)
λ_{small}^{MF}	7.37***	3.13*	1.21	4.26*	7.31***	2.82	1.40	3.47	7.19***	3.15	1.76	4.47
t-stat	(2.99)	(1.78)	(0.71)	(1.65)	(2.97)	(1.57)	(0.82)	(1.34)	(2.92)	(1.52)	(1.00)	(1.55)
λ_{micro}^{MF}	6.99***	1.91	-2.10	1.75	7.03***	1.69	-2.87	1.82	6.26**	0.95	-0.83	-2.41
t-stat	(2.87)	(0.85)	(-1.07)	(0.45)	(2.87)	(0.78)	(-1.43)	(0.48)	(2.50)	(0.33)	(-0.39)	(-0.60)
λ_{small}^{Δ}	-0.59*	2.33***	-0.44	4.88***	-0.53	2.95***	-0.69	5.66***	-0.29	2.84***	-1.29*	4.25***
t-stat	(-1.68)	(2.59)	(-0.61)	(3.35)	(-1.50)	(2.97)	(-0.96)	(3.76)	(-0.80)	(2.72)	(-1.76)	(2.55)
λ^{MF}	6.98***	2.75	1.25	2.14	6.98***	2.79	1.16	1.98	6.93***	2.30	1.22	0.86
t-stat	(2.87)	(1.61)	(0.71)	(0.86)	(2.88)	(1.62)	(0.66)	(0.80)	(2.85)	(1.18)	(0.67)	(0.33)
$\lambda = 0$	0.00***	0.01***	0.50	0.15	0.00***	0.01**	0.46	0.18	0.00***	0.07*	0.47	0.16
$\lambda =$	1.00	1.00	0.96	0.09*	1.00	1.00	0.97	0.10*	1.00	0.97	0.96	0.09*
$\Delta\lambda \lessgtr 0$	0.01***	0.08*	0.26	0.01***	0.01**	0.11	0.24	0.00***	0.02**	0.48	0.43	0.00***

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

Table VII: 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 VI. 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}	2.79	1.16	1.98
1970–2016	t-stat	(1.62)	(0.66)	(0.80)
	λ_{small}^{MF}	2.84	1.34	3.54
	t-stat	(1.58)	(0.78)	(1.37)
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)

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

costs quickly overwhelm strategy returns), and Lesmond, Schill, and Zhou (2004) argue that high transactions costs preclude profitable momentum strategies altogether. Because these studies estimate transactions costs functions using all TAQ transactions, their average implementation cost estimates are smoothed over size quintiles and over trades unrelated to momentum strategies. As a consequence these authors find a result like the λ^{MF} of Tables II–IV, whereby momentum has an economically unimportant premium for real-world asset managers.

When we focus only on small asset managers, the picture looks quite different. As noted in Table VI, small funds earn nearly double the average-fund compensation to momentum, and 9%–10% more per year than the largest funds per unit of momentum exposure. Combining these results we find that both sets of authors are correct; longstanding disagreement on the profitability of momentum strategies arises 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 investors. 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 an outlier, or that non-mutual fund asset managers have different compensation schedules for factor exposure.

B. Implementation Costs Over Time

Figure II adds a time-series dimension to the average risk price differences of Tables II–IV.[26] For each panel we plot the log return of the "on-paper" investments in each factor with a unit risk loading minus the corresponding log return on mutual fund investments. 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,

$$y_k(t) = \sum_{s=t-6}^{t+6} \log\left(1 + \lambda_{kt}^{S}\right) - \log\left(1 + \lambda_{kt}^{MF}\right) \approx \sum_{s=t-6}^{t+6} \lambda_{kt}^{\Delta}. \tag{10}$$

This quantity has an equivalent interpretation as the relative cost associated with real-world rather than on-paper investment in factor strategies.

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 three slope series are remarkably similar for each factor. The implementation gap is clearly rank-ordered as UMD, HML, MKT, and SMB, with large and positive implementation gaps for UMD and HML, no implementation gap for MKT, and a small negative implementation gap for SMB. The difference series are also affected by macroeconomic

[26]In the interest of brevity we suppress the corresponding plots for the matched pairs analysis.

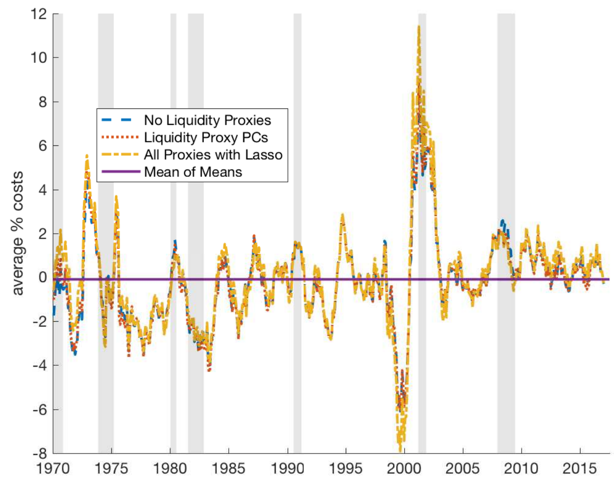
Figure II: Rolling Performance Difference Between Mutual Funds and Stocks ($\beta = 1$)

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

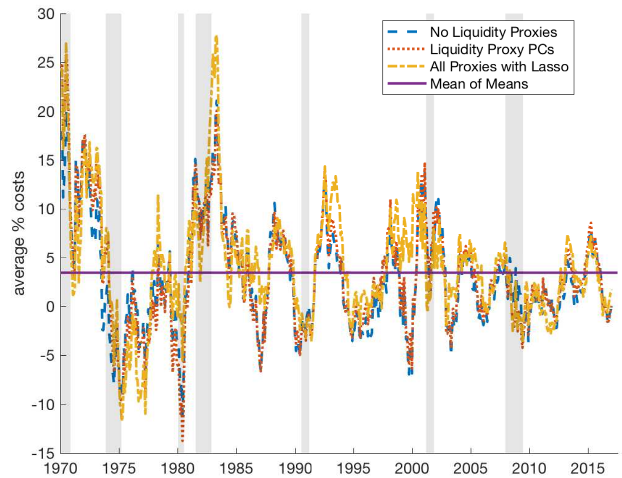
$$y_k(t) = \sum_{s=t-6}^{t+6} \log\left(1 + \lambda_{kt}^{S}\right) - \log\left(1 + \lambda_{kt}^{MF}\right),$$

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 for which the time-series regression includes no liquidity proxies, the first principal component for market liquidity proxies and funding liquidity proxies, and all liquidity proxies with adaptive Lasso penalization (κ_i chosen by 10-fold cross validation). 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.

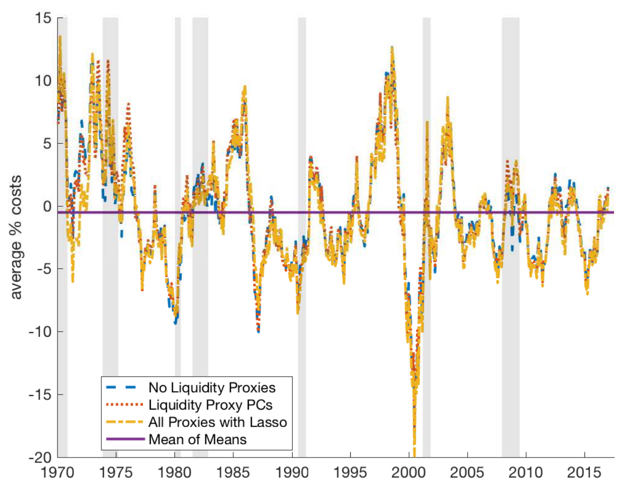
(a) MKT Costs



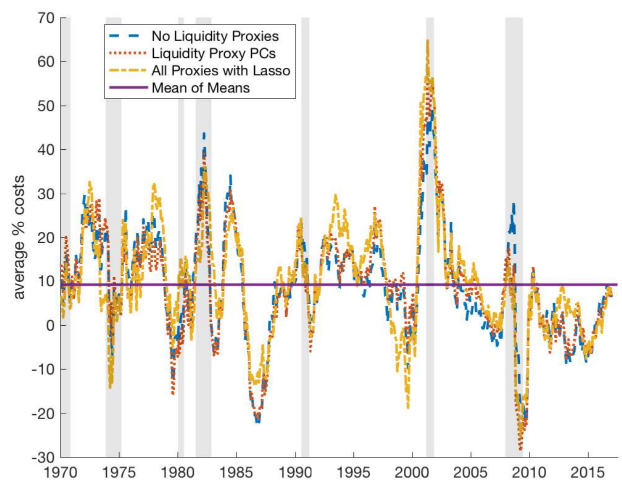
(b) HML Costs



(c) SMB Costs



(d) UMD Costs



events. All four implementation gaps fall before the end of the tech bubble of the late 1990s and rise during the subsequent crash and/or 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 reveals 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.

The conjectured equilibrium adjustment mechanism hinges on non-proportional trading costs—omitted by studies such as Novy-Marx and Velikov (2016)—and it rationalizes industry-level decreasing returns to scale suggested by Pastor and Stambaugh (2012) and documented by Pastor, Stambaugh, and Taylor (2015). It also generates a testable prediction that industry-level inflows increase implementation costs of factor strategies.

We analyze this relationship between implementation costs, flows, and illiquidity 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.B. 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,

$$flow_{it} = \frac{TNA_{it}}{TNA_{i,t-1}} - (1 + R_{it}) \,. \tag{11}$$

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 reflecting returns-to-scale, flow variables are a natural candidate for explaining trading costs because large flows 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.

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

Table VIII: 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) Implementation Costs — Baseline Specification

	MKT			HML			SMB			UMD		
β_{MFLOW}	0.02		-0.12	1.97**		0.94	0.56		0.36	2.66**		1.66
t-stat	(0.07)		(-0.45)	(2.43)		(1.08)	(0.98)		(0.55)	(2.06)		(1.23)
β_{DFLOW}	-0.01		-0.05	-2.22**		-2.48**	-0.08		-0.17	-2.76		-2.99
t-stat	(-0.01)		(-0.12)	(-2.07)		(-2.35)	(-0.13)		(-0.27)	(-1.43)		(-1.56)
β_{ML}		0.42	0.44		2.53**	2.76***		0.98	0.96		2.31	2.52
t-stat		(1.27)	(1.31)		(2.40)	(2.67)		(1.14)	(1.08)		(1.13)	(1.19)
β_{FL}		-0.42	-0.48		-3.35***	-3.48***		-0.90	-0.78		-3.42*	-0.78
t-stat		(-1.18)	(-1.26)		(-3.47)	(-3.21)		(-1.04)	(-0.80)		(-1.93)	(-0.80)
α	-0.20	-0.20	-0.20	2.71***	2.71***	2.71***	-0.49	-0.49	-0.49	7.00***	7.00***	7.00***
t-stat	(-0.80)	(-0.80)	(-0.80)	(3.92)	(3.90)	(3.95)	(-0.83)	(-0.83)	(-0.83)	(5.46)	(5.42)	(5.52)
R^2	0.00	0.01	0.01	0.02	0.04	0.06	0.00	0.00	0.01	0.01	0.01	0.02

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

(b) Implementation Costs — Liquidity PCs

	MKT			HML			SMB			UMD		
β_{MFLOW}	0.05		-0.13	2.45***		1.13	0.51		0.35	3.80***		1.29
t-stat	(0.20)		(-0.50)	(2.78)		(1.22)	(0.91)		(0.54)	(2.88)		(0.97)
β_{DFLOW}	0.02		-0.03	-1.95*		-2.24**	-0.23		-0.35	-2.30		-2.54
t-stat	(0.05)		(-0.07)	(-1.70)		(-2.01)	(-0.38)		(-0.57)	(-1.19)		(-1.36)
β_{ML}		0.47	0.49		3.00***	3.17***		1.31	1.32		2.10	2.29
t-stat		(1.44)	(1.47)		(2.95)	(3.25)		(1.40)	(1.35)		(1.09)	(1.16)
β_{FL}		-0.55	-0.62		-4.37***	-4.37***		-0.84	-0.77		-7.36***	-0.77
t-stat		(-1.58)	(-1.63)		(-4.50)	(-4.09)		(-0.96)	(-0.78)		(-4.44)	(-0.78)
α	-0.21	-0.21	-0.21	2.99***	2.99***	2.99***	-0.45	-0.45	-0.45	7.15***	7.15***	7.15***
t-stat	(-0.83)	(-0.83)	(-0.83)	(4.10)	(4.16)	(4.21)	(-0.76)	(-0.77)	(-0.77)	(5.56)	(5.73)	(5.82)
R^2	0.00	0.01	0.01	0.02	0.06	0.07	0.00	0.01	0.01	0.01	0.05	0.06

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

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

$$\lambda^{\Delta}_{kt} = \alpha + \beta_{MFLOW} MFLOW + \beta_{DFLOW} DFLOW + \beta_{ML} PC_{ML} + \beta_{FL} PC_{FL} + \epsilon_{kt}. \tag{12}$$

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. Likewise to be succinct we report only implementation costs in the baseline specification and with liquidity proxy principal components in Fama-MacBeth time-series regression step.

We draw four lessons from Table VIII.[28] First, the constant terms are large and positive for HML and UMD, confirming that the *time-invariant* component of implementation costs from Equation (4) is large and positive for these factors. Second, focusing on flows, average inflows are 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 or unreliable 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 other-funds' liquidity trading (in a Kyle (1985) sense), thereby reducing 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 conversely *decreases* implementation costs (again most strongly for HML and UMD). We conjecture that mutual funds are insulated from funding liquidity shocks that affect more 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 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. Robustness Tests

A. Do Omitted Strategies Explain High Measured Costs?

Omitted mutual fund strategies may bias our implementation cost estimates. Bias occurs if incidental factor exposures incurred by other activities are cross-sectionally correlated with mutual fund returns. However, the potential for meaningful contamination of cross-sectional slopes de-

[28] Low R^2s throughout are attributable in large part to the noisiness of differences in monthly factor realizations for stocks and mutual funds.

creases with the power of the Carhart (1997) model for explaining time-series variation in 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)).

We split our sample into R^2 quintiles to bound potential bias arising from omitted fund strategies. 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 one or more academic strategies, and these funds have little scope for omitted strategies that might 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–IV.

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 omitted strategies. First, funds with the highest R^2s achieve similar performance on the market factor to the typical mutual fund, and somewhat higher performance on all non-market factors. These differences are not statistically distinguishable, however, nor is performance statistically distinguishable across any set of segmented funds for any factor and first-stage specification. Second, even the funds that most closely track the four academic factors continue to significantly underperform the on-paper factors. Value premia are more than 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 R^2 quintiles four and five. Even so, the best-performing R^2 segments for value still see an implementation gap of 1%–2% relative to the stock portfolios. Returns to momentum exposure for the highest R^2 group are nearly double those of the typical mutual fund, but they are not statistically significant in any specification, and the implementation gap for these funds $\lambda_5^{\Delta} \equiv \lambda^S - \lambda_5^{MF}$ is different from zero at the 1% level. Taken together these findings suggest that our results are not explained by an omitted strategies bias.

Table IX also adds color to how funds implement asset pricing factors. For the market factor, 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 reducing implementation costs for the aggregate market: the market factor is compensated 20–88 basis points more in mutual funds that track the four factors less well. This finding reinforces the importance of using omnibus approaches to measuring implementation costs that are robust to real-world departures from the academic factors. Finally, for the non-market factors, strategy deviations from the four factors decrease factor compensation, sometimes substantially. Funds that take on factor risk while pursuing non-factor strategies are rarely compensated well for that risk: destroying returns relative to an index strategy is common.

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), the first principal component of market and funding liquidity proxies (second columns), and all liquidity proxies in adaptive Lasso regression (third 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 tests of coefficients being jointly different from zero, of equality of coefficients, and of non-monotonicity of coefficients (using the test of Patton and Timmermann (2010)).

		No Liquidity Proxies				Liquidity PCs				Liquidity Lasso			
	$\bar{R}^2$	MKT	HML	SMB	UMD	MKT	HML	SMB	UMD	MKT	HML	SMB	UMD
λ_5^{MF}	94.6%	6.55***	3.78**	2.13	3.88	6.55***	4.00**	2.03	4.20	6.55***	3.76*	1.82	3.89
t-stat		(2.72)	(2.11)	(1.24)	(1.43)	(2.73)	(2.21)	(1.18)	(1.57)	(2.73)	(1.90)	(1.05)	(1.34)
λ_4^{MF}	90.7%	6.75***	4.04**	2.77*	2.52	6.71***	4.29**	2.83*	2.54	6.70***	4.47**	2.79*	2.98
t-stat		(2.77)	(2.19)	(1.65)	(0.92)	(2.76)	(2.28)	(1.69)	(0.92)	(2.76)	(2.17)	(1.65)	(1.03)
λ_3^{MF}	87.1%	7.43***	2.48	-0.21	3.22	7.41***	2.48	-0.13	2.68	7.39***	2.27	-0.25	2.30
t-stat		(3.01)	(1.43)	(-0.12)	(1.23)	(3.00)	(1.40)	(-0.08)	(1.04)	(2.99)	(1.12)	(-0.15)	(0.81)
λ_2^{MF}	82.1%	7.20***	2.51	1.87	0.46	7.16***	2.39	1.86	-0.08	7.14***	2.08	1.79	-0.64
t-stat		(2.94)	(1.47)	(1.08)	(0.18)	(2.93)	(1.36)	(1.07)	(-0.03)	(2.91)	(1.02)	(1.01)	(-0.23)
λ_1^{MF}	60.0%	7.09***	3.10	-0.59	2.40	7.25***	3.36*	-1.13	2.58	7.16***	1.57	-0.82	-0.75
t-stat		(2.87)	(1.63)	(-0.27)	(0.81)	(2.94)	(1.76)	(-0.51)	(0.88)	(2.86)	(0.71)	(-0.37)	(-0.25)
λ_5^{Δ}	–	0.23	1.68**	-1.36**	5.26***	0.23	1.77**	-1.32**	4.92***	0.35	2.23***	-1.35**	4.83***
		(0.99)	(2.07)	(-2.28)	(3.49)	(0.96)	(2.13)	(-2.20)	(3.13)	(1.37)	(2.65)	(-2.21)	(2.90)
λ^{MF}	82.9%	6.98***	2.75	1.25	2.14	6.98***	2.79	1.16	1.98	6.93***	2.30	1.22	0.86
t-stat		(2.87)	(1.61)	(0.71)	(0.86)	(2.88)	(1.62)	(0.66)	(0.80)	(2.85)	(1.18)	(0.67)	(0.33)
$\lambda = 0$		0.00***	0.00***	0.38	0.38	0.00***	0.00***	0.35	0.37	0.00***	0.02**	0.44	0.56
$\lambda =$		1.00	0.94	0.57	0.92	1.00	0.90	0.47	0.85	1.00	0.79	0.55	0.67
$\Delta\lambda \lessgtr 0$		0.51	0.21	0.98	0.40	0.51	0.38	0.98	0.60	0.57	0.57	0.98	0.01***

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

B. Do Microcap Stocks Explain Superior On-Paper Performance?

Implementation frictions 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 the capital of 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 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 X reports Fama-MacBeth estimates of factor premia on this set of stock portfolios. We present only value-weighted results because we are interested in shaping our stock portfolios to reflect the investible 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 sans-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 investible universe explains (at most) 23%–27% of the implementation gap for value and 17%–18% of the implementation gap for momentum.

Table X: 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), regressions with liquidity principal components (middle panel), and Lasso regressions with all liquidity proxies. 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. First-stage regression estimates in the third panel include the Carhart (1997) factors and all market and funding liquidity proxies in an adaptive Lasso regression with portfolio-specific penalty parameters κ_i chosen by 10-fold cross validation. 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.

			1970 – 2016				**1993 – 2016**		
	N_S	MKT	HML	SMB	UMD	MKT	HML	SMB	UMD
$\lambda^{\Delta}_{baseline}$	80	-0.40	3.32***	0.33	5.85***	-0.15	3.06***	-0.59	4.26***
t-stat		(-1.38)	(4.17)	(0.51)	(4.38)	(-0.38)	(3.47)	(-0.74)	(2.67)
$\lambda^{S}_{baseline}$	80	6.58***	6.07***	1.58	7.99***	7.67**	5.25*	1.62	5.54
t-stat		(2.72)	(3.32)	(0.99)	(3.41)	(2.33)	(1.89)	(0.72)	(1.47)
$\lambda^{S}_{baseline}$	100	6.62***	7.06***	0.94	9.23***	7.43**	6.18**	1.89	5.96
t-stat		(2.75)	(3.81)	(0.55)	(3.90)	(2.27)	(2.20)	(0.77)	(1.58)
λ^{Δ}_{PCs}	80	-0.41	3.48***	0.23	5.97***	-0.11	3.34***	-0.82	5.26***
t-stat		(-1.38)	(4.28)	(0.36)	(4.42)	(-0.27)	(3.80)	(-1.00)	(3.11)
λ^{S}_{PCs}	80	6.58***	6.27***	1.39	7.95***	7.72**	5.20*	1.44	5.53
t-stat		(2.73)	(3.42)	(0.87)	(3.39)	(2.36)	(1.87)	(0.63)	(1.47)
λ^{S}_{PCs}	100	6.56***	7.49***	0.92	9.28***	7.45**	6.21**	1.91	5.96
t-stat		(2.73)	(3.99)	(0.54)	(3.93)	(2.29)	(2.19)	(0.78)	(1.58)
λ^{Δ}_{Lasso}	80	-0.10	3.85***	-0.10	6.82***	-0.03	3.53***	-0.83	5.60***
t-stat		(-0.30)	(4.77)	(-0.16)	(4.89)	(-0.07)	(3.62)	(-0.98)	(2.93)
λ^{S}_{Lasso}	80	6.83***	6.15***	1.11	7.68***	7.80**	5.10*	1.46	5.15
t-stat		(2.83)	(3.31)	(0.69)	(3.27)	(2.39)	(1.77)	(0.64)	(1.36)
λ^{S}_{Lasso}	100	6.82***	7.59***	0.59	9.11***	7.54**	6.21**	1.95	5.67
t-stat		(2.83)	(3.94)	(0.34)	(3.83)	(2.32)	(2.10)	(0.78)	(1.48)
T		564	564	564	564	282	282	282	282
$\bar{N}_{MF}$		1901	1901	1901	1901	3299	3299	3299	3299

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

VIII. 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 two new methods (an extension of the Fama-MacBeth approach, and a matched pairs approach) 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 number of investment managers enables the examination of factor implementation costs for a large investment management universe.

Both of our proposed approaches demonstrate that mutual funds are generally poorly compensated for exposure to 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. Our estimates based on matched pairs suggest comparable performance attrition for value and momentum strategies, and they differ in revealing high costs to investing in small-stock characteristic portfolios.

Taken together, these results paint a sobering picture of the real-world implementability of the most important financial market anomalies. We agree with Lesmond, Schill, and Zhou (2004)'s analysis that momentum profits in particular may be out of reach for the typical asset manager. In this respect markets may be efficient from the perspective of representative mutual funds, even if outlier funds see a very different picture of risk and return net-of-costs (e.g., Frazzini, Israel, and Moskowitz (2015)'s asset manager).

Our approaches are readily extended to a wide range of candidate return factors provided that these are reflected in mutual fund exposures. Common anomalies like size, value, and momentum clearly meet this requirement. In future research we will apply these tools more broadly to investigate whether other residents of the factor zoo truly can survive in the wild.

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." *The 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.

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

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

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.

Boyson, Nicole M., Christof W. Stahel, and René M. Stulz. 2010. "Hedge Fund Contagion and Liquidity Shocks." *The 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." *The 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." *The 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." *The 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." *The 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." *The Journal of Finance* 52 (3):1035–1058.

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." *The Journal of Finance* 56 (6):2415–2430.

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

Fama, Eugene F. 1970. "Efficient Capital Markets: A Review of Theory and Empirical Work." *The 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." *The 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." *The Journal of Finance* 63 (4):1653–1678.

———. 2012. "Size, Value, and Momentum in International Stock Returns." *Journal of Financial Economics* 105 (3):457 – 472. URL http://www.sciencedirect.com/science/article/pii/S0304405X12000931.

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.

Giglio, Stefano and Dacheng Xiu. 2017. "Inference on Risk Premia in the Presence of Omitted Factors." Working paper.

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." *The 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." *The Review of Financial Studies* 14 (1):29–78.

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

Harvey, Campbell R., Yan Liu, and Heqing Zhu. 2016. "... and the Cross-Section of Expected Returns." *The 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." *The 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." *The Journal of Finance* 48 (1):65–91.

———. 2001. "Profitability of Momentum Strategies: An Evaluation of Alternative Explanations." *The 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. 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?" *The 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." *The 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.

Lou, Xiaoxia and Ronnie Sadka. 2016. *Portfolio Construction, Measurement, and Efficiency: Essays in Honor of Jack Treynor*, chap. Invisible Costs and Profitability. Springer, 135–153.

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." *The 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.” *The 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.

Shleifer, Andrei and Robert W. Vishny. 1997. “The Limits of Arbitrage.” *The Journal of Finance* 52 (1):35–55.

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.” *The 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 As a first 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 flag fund-dates with reporting frequencies less than monthly. 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 observation has a nonzero value. These annual reporters 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 set to missing 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 missing or 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, where 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 $(1 + discrepancy)^{s/\Delta t}$, 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. Lagging the TNA value by one month increases the share of missing observations to 3.6% of 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 then 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.

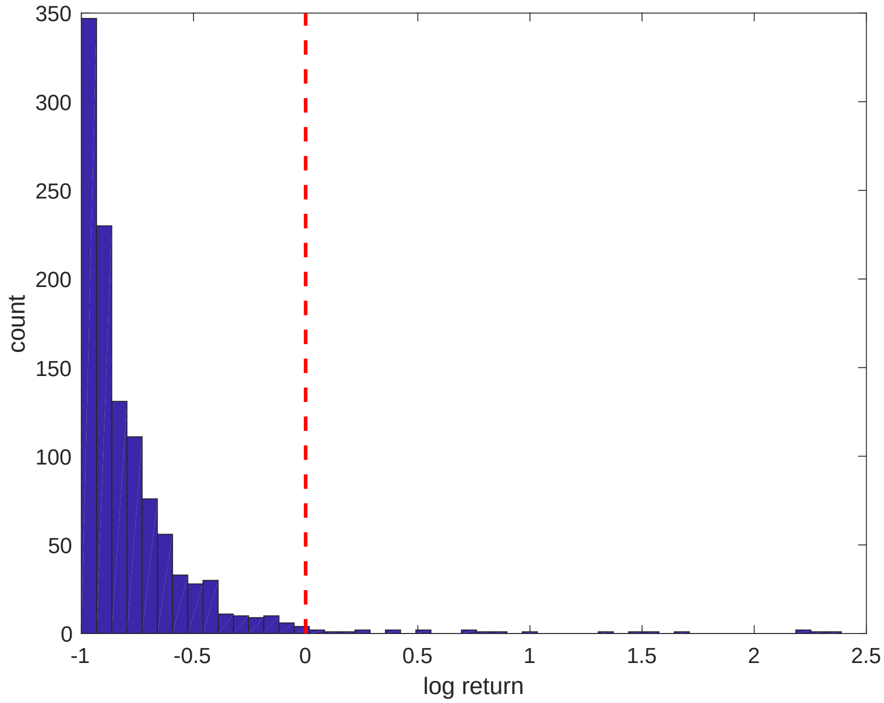
Because our analysis concerns mutual funds, we filter out exchange-traded funds (ETFs), exchange-traded notes (ETNs), and variable annuity underlying (VAU) funds. To do this, we discard any observations for which et_flag indicates an ETF or ETN or vau_fund indicates a VAU at any time in a fund's life.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

Aggregation into Fund Groups Having accounted for the salient variation across share classes, we identify share class names of the same fund by extracting prefixes of CRSP fund names from the fund summary file. Suffixes represent different share classes of the same fund, e.g., "investor" and "admiral" shares of a Vanguard fund. We assign all funds with the same prefix to a unique fund group.

We take the following steps to obtain fund name prefixes. First, we cut off all fund names after a semicolon. Names after the semicolon are exclusively class names. We add to this dictionary of possible class names the Arabic and Roman numerals 1–10, letters A–Z, and a set of known share class markers, inv, inst, investor, institutional, trust, corp, advisor, admin, part, restricted, retail, shares, adv, common, series, prim, and primary. Although handling semicolons is straightforward,

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 $|r| = 100\%$.



forward slashes—the other class-name delimiter used in CRSP—require more care.[29] For example, fund names include "Franklin/Templeton" and "M/M" (money market), so "/" does not serve only as a delimiter. As a preliminary step, we replace forward slashes in T/F, T/E, M/M, L/S, Long/Short, S/T, and L/T with backslashes in fund names. Then we mark all fund names with a "/" for which (1) we have not already eliminated share classes under the semicolon rule and (2) the post-slash string is in the set of available suffixes. These are the set of funds with class information, and we clip the post-slash string for these elements. Finally we take the unique prefixes as our set of fund names sans class information. This mapping reduces the 61,734 surviving unique fund IDs in the CRSP monthly returns file to 33,538 unique fund groups. Of the 6,522,095 observations in the

[29]We deviate from Berk and van Binsbergen (2015)'s handling of fund subclasses because they do not properly handle forward slashes. From their page 9:

> In most cases, the only difference among subclasses is the amount of expenses charged to investors, so simply including them as separate funds would artificially increase the statistical significance of any identified effect. For funds that appear in the CRSP database, identifying subclasses is a relatively easy process. **CRSP provides a separator in the fund name in the form of either a colon (":") or a slash ("/"). Information after the separator denotes a subclass.**

We attribute the difference between "colon" and "semicolon" to a minor database tweak or typographical error. The claim about forward slashes is incorrect because forward slashes often serve as abbreviations or concatenations rather than as delimiters, so we must devise new rules for separating out fund subclasses.

monthly return file, only 4,298 of these are not assigned a fund group, and these observations are dropped.

We use fund names rather than CUSIP as our unique identifier to capture funds that have identical or near-identical investments but different CUSIPs or CRSP portfolio numbers. Fund names also have fewer missing values than either CUSIP (11.1% missing) or portfolio number (27.8% missing) identifiers. This matching on name prefix also mimics the behavior of CRSP's Class Group code designed for this purpose, but which only becomes available in August 1998. For the observations for which class groups are defined, matching on name prefix gives 44,218 unique codes, whereas matching on both name prefix and class group gives 45,484 unique pairs. The overlap between methods is quite good despite the simplicity of our prefix approach.

Cleaning Procedures at the Fund Group Level We construct fund group returns and total net assets by taking a weighted average of the returns and a sum of TNAs 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%. Fund group TNAs are the sum of current TNA values for all fund IDs for which the TNA is defined. Aggregating funds across share classes delivers 3,693,705 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).[30] 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 5.5% of fund group-month observations. Dropping observations from potential incubation periods before the $10 million threshold is achieved eliminates another 7.8% 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 10,823 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

[30] 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.

"ED" CRSP objective code.[31] These filters reduce the number of valid funds from 22,342 to 7,346, and the corresponding number of non-missing return observations decreases to 1,088,834 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 7,331 mutual funds with 1,071,818 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}

If instead we were to include the loadings on the liquidity proxies in Equation (2), the second-stage regression becomes

$$r_{it} = \sum_k \lambda_{kt}\hat{\beta}_{ik}1_{i\in S} + \sum_k \lambda_{kt}^{MF}\hat{\beta}_{ik}1_{i\in MF} + \sum_l \lambda_{lt}\hat{\gamma}_{il}1_{i\in S} + \sum_l \lambda_{lt}^{MF}\hat{\gamma}_{il}1_{i\in MF} + \epsilon_{it}.\ t = 1,\ldots,T. \quad (13)$$

From the conjectured return process of Equation (5), $\hat{\lambda}_t^S = \lambda_t^S$, $\hat{\lambda}_{kt}^{\Delta} = \bar{\eta} + \frac{cov((\eta_i - \bar{\eta})\beta_i, \beta_i)}{var(\beta_i)}$, and $\hat{\lambda}_{lt}^{\Delta} = \eta_{lt}$. The problem with this approach is that the $\hat{\lambda}_{lt}^{\Delta}$ term drops 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 to factor investing using mutual funds is

$$TVC_{it} \equiv \sum_l \eta_{lt}\gamma_{il} = \sum_l \eta_{lt}\left(\sum_k \gamma_{ikl}\beta_{ik}\right). \quad (14)$$

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

$$TVC_{it} = \sum_t \sum_k \eta_{kt}\beta_{ik}1_t + \epsilon_{it}. \quad (15)$$

This regression can be interpreted as rotating the transaction cost space onto the return factor 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 (15),

$$\hat{\eta}_t = \frac{cov\left(\sum_l \eta_{lt}\gamma_{il}, \beta_i\right)}{cov\left(\beta_i\right)} = \sum_l \eta_{lt}\bar{\gamma}_l + \sum_s \frac{cov\left(\beta_i\eta_{lt}\left(\gamma_{il} - \bar{\gamma}_l\right), \beta_i\right)}{cov\left(\beta_i\right)}. \quad (16)$$

The first term represents the average exposure to liquidity or performance factors multiplied by

[31] 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.

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. Match Quality Adjustments and Reporting for Matched Pairs

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.[32] 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,

$$r_{it} = \alpha_i + \sum_k \delta_{kt}\hat{\beta}_{ik} + \epsilon_{it},\ i \in MF,\ t = 1, \ldots, T. \tag{17}$$

Despite its matched pairs origin, Equation (17) is our usual Fama-MacBeth cross-sectional regression on the set of mutual funds.

Using Equation (17), 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,

$$\frac{1}{3}\sum_{j\in\Omega(i)}\sum_k \delta_{kt}\left(\beta_{ik}^{S} - \beta_{jk}^{MF}\right), \tag{18}$$

where $\Omega(i)$ denotes the three-element set of mutual funds matched to stock i. In effect, bias-correction marries the matched pairs approach 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.

B. Evaluation of Match Quality

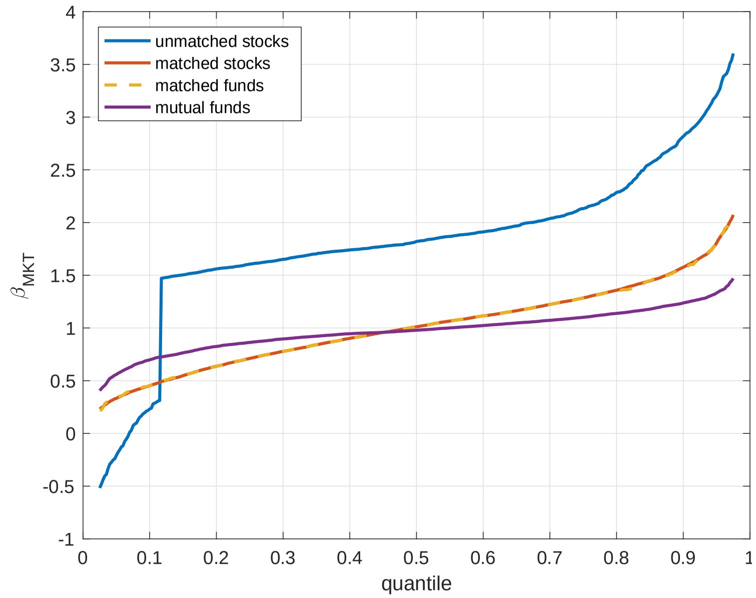
Figure A.II plots the distribution of each matching variable (β) for unmatched stocks, matched stocks, matched mutual funds, and all mutual funds. We see immediately that stocks have more variable factor exposures than mutual funds, so the most extreme stocks on either side of the

[32]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)).

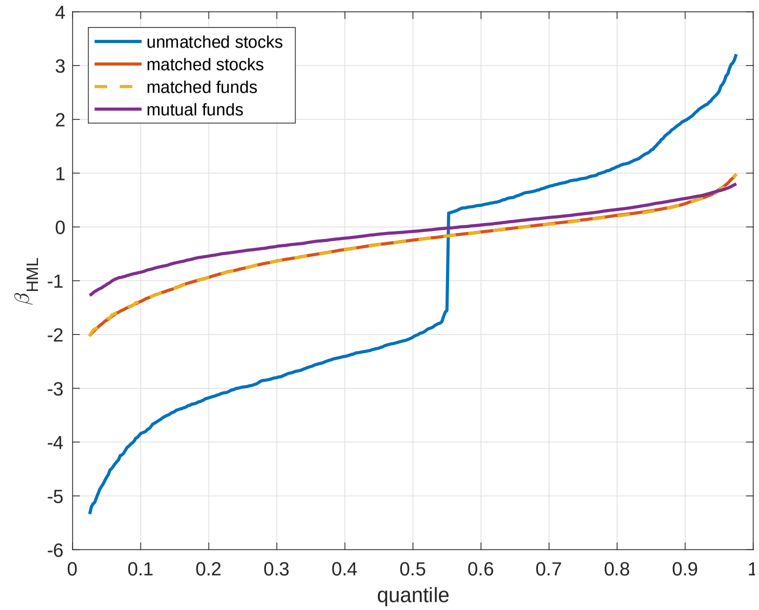
Figure A.II: Comparison of Samples on Matching Variable — Univariate β Matching

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.

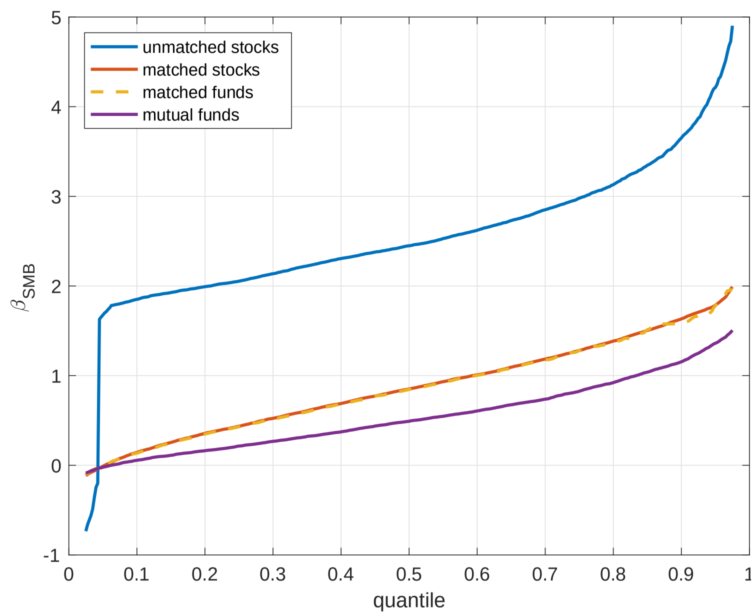
(a) Match on β_{MKT}



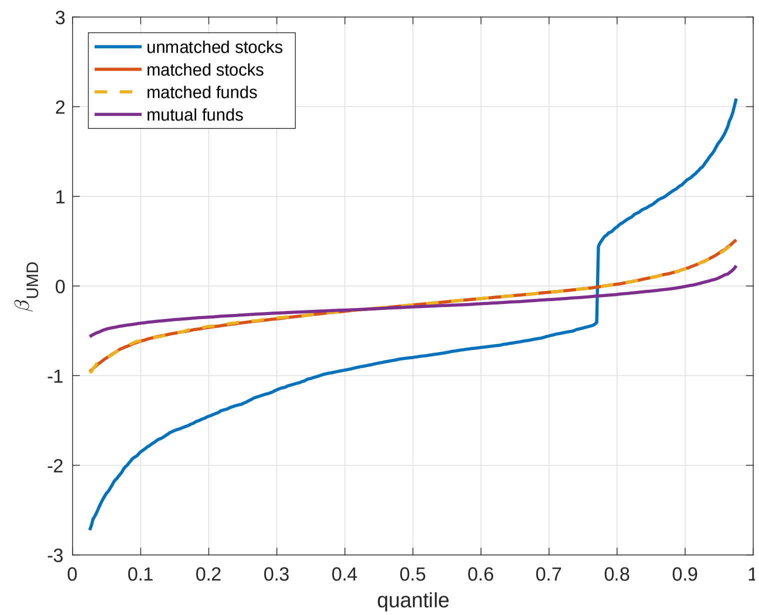
(b) Match on β_{HML}



(c) Match on β_{SMB}



(d) Match on β_{UMD}



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.I 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 quintiles one through four, and overall match percentages are high (79%–92%). Match quality deteriorates slightly in the most extreme quintiles, and particularly so in quintile five of market capitalization (the smallest stocks). Imperfect matching in the extreme quintiles also manifests in the imbalance between the number of matched stocks and matched mutual funds. For example, the typical matched mutual fund in the smallest stock group is used more than five times: 1,295,528 stock-months are matched to three of 771,365 unique mutual fund-months.

The drawback to matching on a single variable is that other factors determining returns may differ wildly 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.II 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 and Table A.III report match quality when matching uses all four factor betas and a wider caliper of 1σ. 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 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.I: 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	0.07	0.99	19440	-1.68	2.49	100034	0.80	1.46	5202	-0.63	1.44	169003
	S_M	0.62	0.35	426665	-0.58	0.79	607370	0.41	0.47	177133	-0.23	0.39	599378
	MF_M	0.62	0.34	360578	-0.58	0.78	567050	0.41	0.47	307892	-0.22	0.39	522791
Q2	S_U	1.42	1.11	5488	-1.42	2.43	36035	2.21	1.05	10690	-0.66	1.12	57071
	S_M	0.85	0.32	447514	-0.42	0.72	388556	0.67	0.51	200542	-0.22	0.32	387764
	MF_M	0.85	0.32	488984	-0.42	0.71	485974	0.67	0.50	343640	-0.22	0.32	480910
Q3	S_U	1.71	0.67	10562	-0.86	2.35	26623	2.46	0.81	28384	-0.66	1.05	44085
	S_M	1.04	0.31	442917	-0.28	0.69	354604	0.84	0.52	240569	-0.21	0.30	346855
	MF_M	1.04	0.31	533610	-0.29	0.69	483005	0.83	0.52	360147	-0.21	0.30	464071
Q4	S_U	1.80	0.49	30101	-0.41	2.31	27339	2.54	0.84	69305	-0.64	1.05	45686
	S_M	1.21	0.34	421383	-0.19	0.68	366132	0.96	0.52	338890	-0.21	0.30	350541
	MF_M	1.21	0.34	438132	-0.20	0.67	508044	0.95	0.52	410747	-0.21	0.30	464824
Q5	S_U	2.14	0.65	108281	-0.08	2.40	39019	2.57	1.32	462703	-0.56	1.21	97951
	S_M	1.48	0.45	339863	-0.19	0.69	466721	0.95	0.56	1295528	-0.21	0.34	493414
	MF_M	1.48	0.45	255089	-0.19	0.69	591750	0.93	0.56	771365	-0.21	0.34	508307
% Matched			92.3%			90.4%			79.3%			83.9%	

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 25th data point for unmatched and matched stocks. We plot every 75th 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.

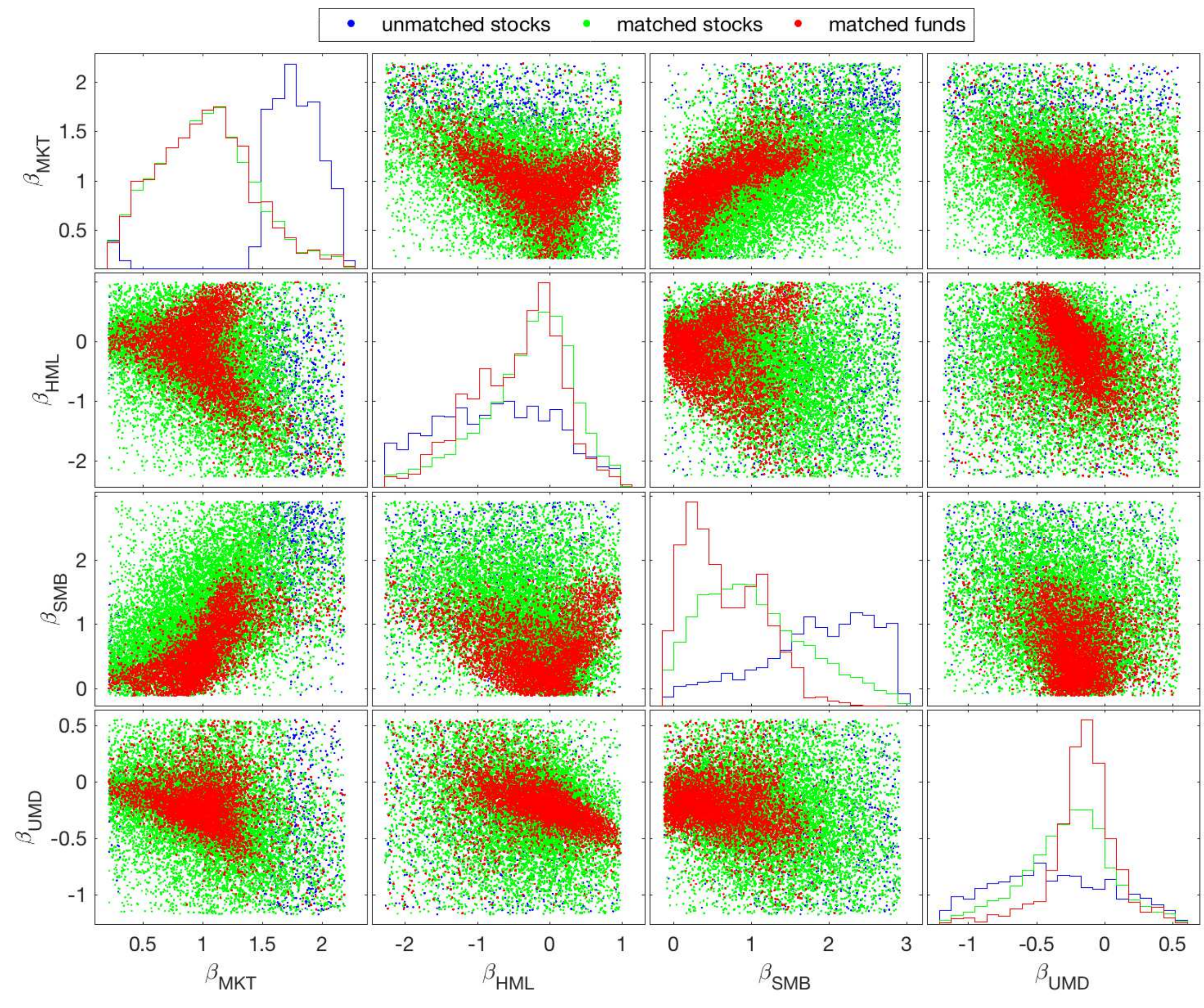


Table A.II: Comparison of Samples on Matching Variable — Univariate β Matching on β_{MKT}

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	0.07	0.99	19440	0.08	1.50	19440	0.96	1.38	19440	-0.22	1.04	19440
	S_M	0.62	0.35	426665	-0.08	0.67	426665	0.75	0.78	426665	-0.16	0.43	426665
	MF_M	0.62	0.34	360578	-0.17	0.40	360578	0.28	0.35	360578	-0.09	0.18	360578
Q2	S_U	1.42	1.11	5488	-0.93	1.83	5488	1.99	1.48	5488	-0.46	1.23	5488
	S_M	0.85	0.32	447514	-0.15	0.65	447514	0.89	0.72	447514	-0.23	0.39	447514
	MF_M	0.85	0.32	488984	-0.29	0.50	488984	0.43	0.41	488984	-0.13	0.20	488984
Q3	S_U	1.71	0.67	10562	-0.97	1.32	10562	2.16	0.97	10562	-0.53	0.94	10562
	S_M	1.04	0.31	442917	-0.28	0.71	442917	1.07	0.74	442917	-0.28	0.43	442917
	MF_M	1.04	0.31	533610	-0.45	0.61	533610	0.62	0.46	533610	-0.15	0.24	533610
Q4	S_U	1.80	0.49	30101	-1.00	1.21	30101	2.21	0.89	30101	-0.51	0.84	30101
	S_M	1.21	0.34	421383	-0.47	0.83	421383	1.28	0.81	421383	-0.31	0.49	421383
	MF_M	1.21	0.34	438132	-0.64	0.71	438132	0.82	0.47	438132	-0.17	0.29	438132
Q5	S_U	2.14	0.65	108281	-1.36	1.61	108281	2.38	1.13	108281	-0.55	1.06	108281
	S_M	1.48	0.45	339863	-0.75	1.18	339863	1.61	1.00	339863	-0.38	0.65	339863
	MF_M	1.48	0.45	255089	-0.89	0.88	255089	1.04	0.50	255089	-0.25	0.38	255089
% Matched			92.3%			90.4%			79.3%			83.9%	

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 1σ 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 25th data point for unmatched and matched stocks. We plot every 75th 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.

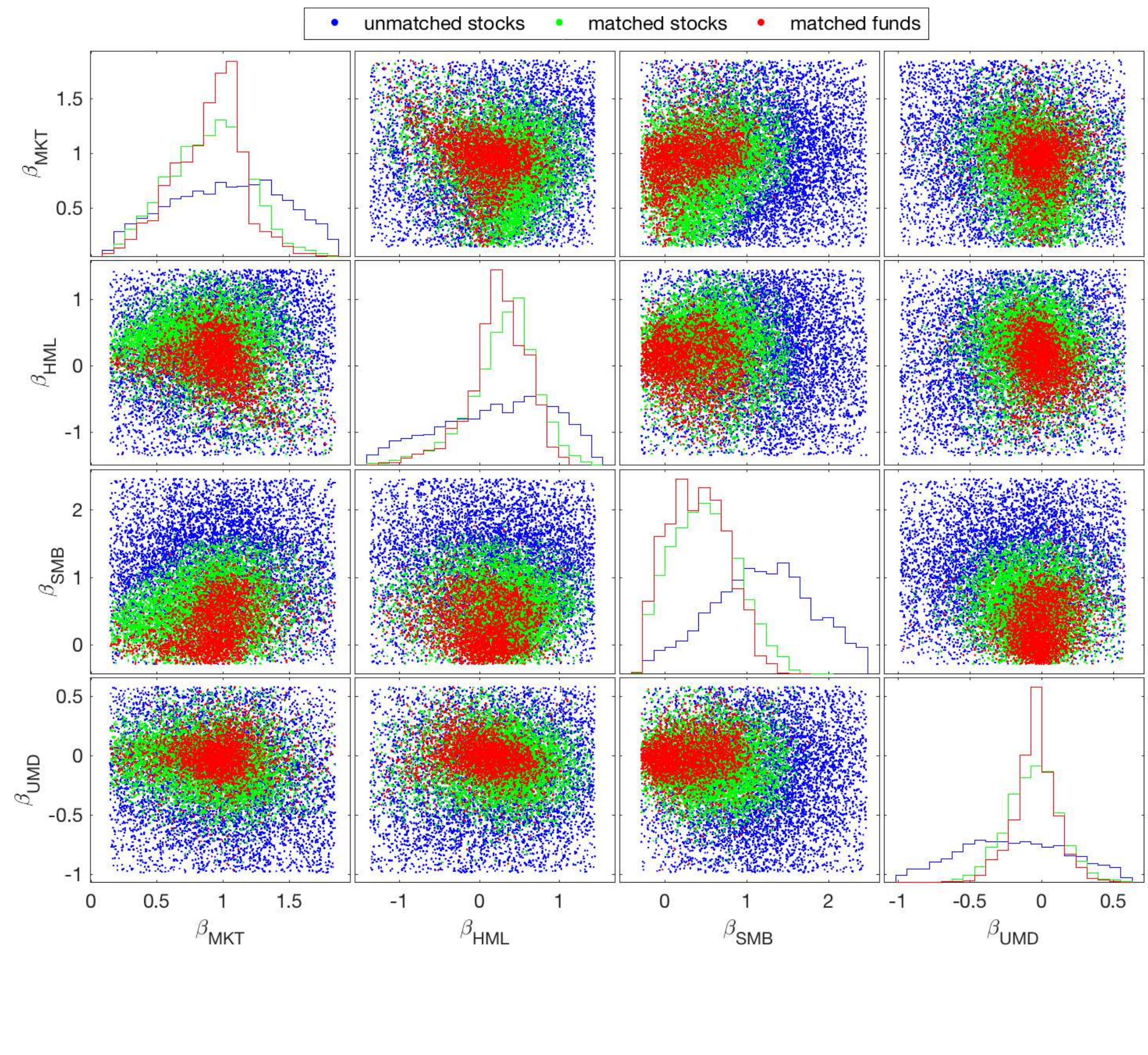


Table A.III: Comparison of Samples on All Variables — Multivariate β Matching

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 1σ 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	0.49	0.59	168304	-0.08	1.55	405758	0.23	0.66	29281	-0.25	0.92	496405
	S_M	0.60	0.27	277801	0.10	0.47	301646	-0.02	0.34	153054	-0.07	0.21	271976
	MF_M	0.64	0.26	171608	0.07	0.40	292111	0.04	0.27	198820	-0.05	0.17	217905
Q2	S_U	0.81	0.46	151681	0.04	1.24	196712	0.71	0.73	51037	-0.21	0.68	190596
	S_M	0.81	0.24	301321	0.25	0.43	227879	0.24	0.35	160195	-0.06	0.18	254239
	MF_M	0.83	0.22	245554	0.20	0.38	235285	0.23	0.30	207829	-0.05	0.14	230926
Q3	S_U	0.97	0.46	175911	0.21	1.14	161542	1.03	0.79	90954	-0.19	0.65	145094
	S_M	0.97	0.23	277568	0.35	0.41	219685	0.43	0.37	177999	-0.06	0.18	245846
	MF_M	0.95	0.19	255107	0.28	0.36	207244	0.37	0.33	190252	-0.04	0.14	233756
Q4	S_U	1.13	0.48	224445	0.31	1.12	162780	1.22	0.86	181929	-0.19	0.65	148511
	S_M	1.08	0.23	227039	0.39	0.39	230691	0.60	0.36	226266	-0.05	0.18	247716
	MF_M	1.04	0.19	200255	0.31	0.35	197346	0.50	0.32	187119	-0.04	0.14	235388
Q5	S_U	1.43	0.59	328783	0.38	1.17	244166	1.36	1.23	1086091	-0.20	0.74	304168
	S_M	1.19	0.28	119361	0.39	0.39	261574	0.60	0.38	672140	-0.06	0.20	287197
	MF_M	1.14	0.25	113097	0.31	0.34	200136	0.48	0.34	275598	-0.04	0.16	238987
% Matched		53.4%			51.0%			48.4%			50.1%		

Table A.III quantifies this trade-off. Just over half 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.