Model Selection with Transaction Costs *

Andrew Detzel[†] Robert Novy-Marx[‡] Mihail Velikov[§]

June 4, 2021

Abstract

Failing to account for transaction costs materially impacts inferences drawn when evaluating asset pricing models, biasing tests in favor of those employing high cost factors. Ignoring transaction costs, the Hou, Xue, and Zhang (2015) q-factor model and the Barillas and Shanken (2018) six-factor model models have high maximum squared Sharpe ratios and small alphas across 120 anomalies. They do not, however, come close to spanning the *achievable* mean-variance efficient frontier. Accounting for transaction costs, the Fama and French (2015, 2018) five-factor model has a significantly higher squared Sharpe ratio than either of these alternative models, while variations employing cash profitability perform better still. More generally, these results highlight the importance of incorporating real-world concerns into financial research.

JEL classification: G11, G12, G14
Keywords: Factor Models, Trading Costs, Mispricing

*Novy-Marx provides consulting services to Dimensional Fund Advisors, an investment firm headquartered in Austin, Texas with strong ties to the academic community. The thoughts and opinions expressed in this article are those of the authors alone, and no other person or institution has any control over its content. For helpful comments and suggestions, we thank Hank Bessembinder, Andrew Chen (Discussant), Victor DeMiguel (Discussant), Wayne Ferson, Andrei Gonçalves (Discussant), Michael Halling (Discussant), Avi Kamara, Raymond Kan, Lukas Kremens, Dong Lou, Claudia Moise, Alessio Saretto, Stephan Siegel, Yang Song, Fabio Trojani (Discussant), Raman Uppal, Chen Xue, and Irina Zviadadze along with conference and seminar participants at the Adam Smith Asset Pricing Workshop, Northern Finance Association, Midwest Finance Association, Federal Reserve Bank of Richmond, and University of Washington.

[†]Daniels College of Business, University of Denver, 2101 S University Blvd., Denver, CO 80122. Email: andrew.detzel@du.edu.

[‡]Simon Business School and NBER, University of Rochester, 500 Joseph C. Wilson Blvd., Box 270100, Rochester, NY 14627. Email: robert.novy-marx@simon.rochester.edu.

[§]Smeal College of Business, Pennsylvania State University, 366 Business Building, University Park, PA 16802. Email: velikov@psu.edu.

1. Introduction

The finance literature documents hundreds of cross-sectional anomalies in stock returns (see, for example, Harvey, Liu, and Zhu, 2016). This setting led researchers to propose numerous asset pricing models, which in turn generated multiple studies dedicated to choosing from among these models using progressively more sophisticated statistical methodology.[1]

A model's ability to price assets ultimately reflects how close its factors come to spanning the efficient frontier. If a test asset generates abnormal returns relative to a model, then that asset improves the investment opportunity set. If some combination of the model's factors are ex post mean-variance efficient, then no other asset can be used to improve performance, and the model prices everything. As a result, factor models are now typically evaluated based on how close their factors come to spanning the ex post efficient frontier, most commonly using the maximum squared Sharpe ratio criterion of Barillas and Shanken (2017).

Unfortunately, this methodology produces misleading results, at least as typically applied, regarding how close a model comes to spanning the efficient frontier. This fact is best illustrated with a simple example. A single factor based on one month industry-relative reversals, constructed using only stocks with below median volatility as in Novy-Marx and Velikov (2016), has a squared Sharpe ratio that is more than nine times the maximum squared Sharpe ratio of the Fama and French (1993) three-factor model, and almost four times that of the Fama and French (2015) five-factor model (4.81 versus 0.53 and 1.24, respectively). However, no one seriously thinks that the single-factor model based on low volatility industry-relative reversals is highly promising, despite its apparent superior performance to all other candidate models

[1] See., for example, Fama and French (1993); Hou, Xue, and Zhang (2015); Barillas and Shanken (2018); Fama and French (1993, 2015); Stambaugh and Yuan (2017); Fama and French (2018); Hou, Mo, Xue, and Zhang (2019); Kozak, Nagel, and Santosh (2019); Barillas, Kan, Robotti, and Shanken (2019); Daniel, Hirshleifer, and Sun (2019); Feng, Giglio, and Xiu (2020); Bryzgalova, Huang, and Julliard (2020).

under the criterion of ex post efficiency.[2]

The basic problem is how prior studies define the investment opportunity set, without regard to implementation costs. The theoretical underpinning of linear factor models, Ross' (1976) Arbitrage Pricing Theory, is based on the idea that investment opportunities that generate abnormal returns attract arbitrage capital until they are eliminated. These opportunities only attract arbitrage capital in practice, however, if they generate abnormal returns in the *real world.* While the low volatility industry-relative reversal has enormous gross alpha relative to standard factor models, it is also extremely expensive to trade. As a result, the strategy does not actually represent an attractive trading opportunity, and its supposed "alpha" does not indicate that it expands the achievable investment frontier.

This paper starts from the premise that arbitrage capital can only be expected to eliminate true abnormal trading opportunities, i.e., those that can be exploited profitably in practice. This is also the view taken by Fama (1991) when arguing that an "[economically] sensible version of the efficiency hypothesis says that prices reflect information to the point where the marginal benefits of acting on information (the profits to be made) do not exceed the marginal costs." Starting from this premise lowers the bar on what we should expect from an asset pricing model, because even the right model may allow for strategies with large gross alphas. There is no reason to expect even the correct model to price strategies that cannot be traded profitably; the best models are those that parsimoniously span the achievable efficient frontier, fully describing the true investment opportunity set as simply as possible.

This paper reevaluates the performance of asset pricing models from this perspective, providing the first study on model selection that accounts for transaction costs.

[2] This strategy's high gross Sharpe ratio also makes it attractive to machine learning algorithms designed to select factors. Kozak, Nagel, and Santosh (2019) estimate an SDF using 50 anomalies that puts more weight on this strategy than any other. The second- and third- largest weights are assigned to factors based on "industry momentum reversal" and "[ignoring vol] industry-momentum reversal," related factors that are also expensive to trade.

Accounting for transaction costs materially impacts our results, reversing several major conclusions drawn by the prior literature that ignores transaction costs. Ignoring transaction costs biases prior studies in favor of models with factors constructed to maximize gross returns, even when this increases trading costs more than gross performance. For example, the mean-variance efficient portfolios of the factors in the Hou et al. (2015) four-factor model and the Fama and French (2018) five-factor model take similar positions in market, size, profitability, and investment factors. The size, investment, and profitability factors in the Hou et al. (2015) model are re-balanced monthly instead of annually, however, and are constructed using a more complicated three-way sorting procedure that increases the weight of small stocks. Both of these differences contribute to gross factor performance, but increase trading costs even more. As a result, while these differences improve gross performance, they diminish the performance an investor would actually realize. So while the Hou et al. (2015) model has a maximum squared Sharpe ratio that is almost 60% higher than the Fama and French (2018) model ignoring transaction costs, the maximum squared Sharpe ratio that the model's factors could have actually delivered in practice was almost 40% lower.

The Fama and French (2018) models also outperform the alternative candidate models pricing a broad cross-section of 120 anomalies net-of-costs. Accounting for costs improves the apparent ability of all models to price this cross-section, because many of the anomalies with the most impressive gross returns are also expensive to trade, and thus have less anomalous net returns. The impact of diminished anomaly performance is partly offset by the lower net returns earned by the explanatory factors, but this effect is more acute for factors with high turnover, resulting in improved relative performance for models employing low cost factors like those of Fama and French (2018).

We also find, consistent with Fama and French (2018), that replacing the accruals-based operating profitability factor with one based on cash-based profitability significantly improves model performance. The net-of-costs maximum squared Sharpe ratios for variations of the Fama and French (2018) five- and six-factor models that employ a cash-based profitability factor are more than a third higher than those on any of the other models we consider. These variations also span the other models, in the sense that an investor already trading these models' factors cannot significantly expand their opportunity set by additionally trading another model's factors. The cash-based variations on the Fama and French model also generally perform better than other models pricing our set of 120 anomalies accounting for transaction costs.

Overall, the evidence in this paper suggests caution when interpreting results from model selection studies. Arbitrage capital only flows to opportunities that investors can actually exploit, so even the right factor model should, at best, explain returns that compensate for bearing risk, and not returns that merely reflect implementation frictions, no matter how elaborate the statistical techniques used to evaluate model performance. Once this insight is incorporated the model selection landscape fundamentally changes, because models that appear to perform strongly ignoring costs often do so only because of high turnover that would be expensive in practice.

More generally, our results highlight serious issues associated with ignoring real world concerns when doing financial research. Finance is a practical field. Implementation has a first order impact on the performance realized by investors. As argued by Harvey (2017), strong incentives to find positive results, which are far more likely to get published, may tempt researchers to design experiments that are more likely to "succeed." One of the easiest ways to do this is to ignore important frictions. For example, most anomalies look stronger among the smallest stocks. Exploiting these anomalies using schemes that overweight these hard-to-trade firms can dramatically improve a strategy's "paper" performance, but generally does nothing to improve the investor's actual realized performance (see, for example, Novy-Marx and Velikov, 2020). Inferences drawn from tests employing these strategies are not reliable, and cannot be generalized into real economic insights. Finance research is most useful when it incorporates real world concerns to the greatest extent possible.

2. Model comparison using abnormal returns and maximum squared Sharpe ratios

Factor models are often judged by how well they price test assets. Admitting only small abnormal returns relative to the model is the hallmark of success, indicating that the factors come close to spanning the mean-variance efficient frontier. Conversely, if a test asset has a large abnormal return relative to the model, then an investor trading a model's factors can significantly expand her investment opportunity set by additionally trading the test asset.

Comparing factor models by how well they price test assets is problematic, however, for several reasons. First, the procedure generally lacks a formal statistical criteria for model selection. Given even a moderately challenging set of test assets, formal statistical tests generally reject all models. Models that are rejected less emphatically are consequently deemed "better" than models that are rejected more emphatically, even when tests do not actually directly compare models. The formal statistical tests used to test each individual model also perversely reward models that explain *less* test asset return variation. Greater residual variation reduces the precision with which the test assets' factor model alphas are estimated, and thus their significance, which makes rejection less likely (or at least less emphatic).

Comparing factor models using test assets is also sensitive to which anomalies are used to test the models, often yielding contradictory answers when employing different sets of test assets. Using "anomalies" to test asset pricing models also always raises prima facie selection bias concerns, particularly when comparing alternative candidate models to the canonical Fama and French (1993) three-factor model and its variations. Anomalies are usually defined by their high abnormal returns relative to the Fama and French (1993) model, biasing tests employing anomalies as test assets against this model.

2.1. The Barillas and Shanken (2017) maximum SR^2 test

In response to these issues, Barillas and Shanken (2017) introduce a summary statistic of model quality, the maximum squared Sharpe ratio (SR^2). This metric quantifies how close the span of a factor model is to the ex post mean variance efficient frontier of *all* assets. Higher SR^2 indicates smaller alphas relative to the model, in a precise sense that is strictly true for the appropriate combination of all test asset alphas.

Gibbons, Ross, and Shanken (1989) quantify the gains (ignoring transaction costs) from adding test assets to a set of factors using the increase in the maximum squared Sharpe ratio. They show that given a set of tradable factors with excess returns f, then adding test assets with excess returns Π yields an increase in the maximum squared Sharpe ratio of

$$SR^2(\Pi, f) - SR^2(f) \quad = \quad \alpha' \Sigma^{-1} \alpha,$$

where $\mathrm{SR}^2(X)$ denotes the maximum possible Sharpe ratio attainable from X, α denotes the vector of intercepts from regressions of the test assets' excess returns on the factor returns, and Σ is the covariance matrix of residuals from these regressions.

While this metric of investment frontier expansion can be sensitive to the choice of test assets (e.g., Lewellen, Nagel, and Shanken, 2010; Fama and French, 2018), Barillas and Shanken (2017) note that if Π is the entire universe of excess returns, then $\mathrm{SR}^2(\Pi, f) = SR^2(\Pi)$ for any set of factor returns f, and the model with the lowest mispricing is thus precisely the one with the highest maximum squared Sharpe ratio. They consequently use the maximum squared Sharpe ratio as a model selection criterion, and this metric has seen wide adoption in the literature.[3] We use the criterion here, but unlike earlier studies we do so accounting for transaction costs.

2.2. Issues with tests based on alpha and maximum SR^2

Figure 1 depicts the before- and after-costs maximum squared Sharpe ratios of the Fama and French (1993) three-factor model (FF3), and a low-volatility industry-relative-reversal factor (LV-IRR) constructed following Novy-Marx and Velikov (2016).[4] FF3 has a squared Sharpe ratio, before costs, of 0.52, while the gross squared Sharpe ratio of LV-IRR is an astounding 4.81. That is, in spite of the fact that hundreds of papers price assets with FF3 while none do so with LV-IRR, the single-factor model consisting of LV-IRR clearly dominates FF3 by the maximum squared-Sharpe ratio criterion—at least ignoring costs. The reader is therefore forced to either accept the implausible conclusion that a low volatility industry relative reversal factor by itself represents a better asset pricing model than the FF3, or to conclude that the current practice of comparing models based on gross squared Sharpe ratios is deeply flawed.

Figure 1 also provides a simple resolution for this tension, however, by showing that FF3 has a significantly higher *achievable* squared Sharpe ratio than LV-IRR. After accounting for transaction costs, the FF3 squared Sharpe ratio is 0.44, four times as high as the 0.11 observed on LV-IRR. Overall, Figure 1 shows that ignoring costs clearly biases comparisons based on squared Sharpe ratios in favor of models with high before-costs performance, even if this performance is unachievable by investors,

[3] See, for example, Barillas and Shanken (2018), Fama and French (2018), Ferson, Siegel, and Wang (2019), and Barillas, Kan, Robotti, and Shanken (2019).

[4] Detailed methodology of all the procedures employed in this paper are provided in the next section.

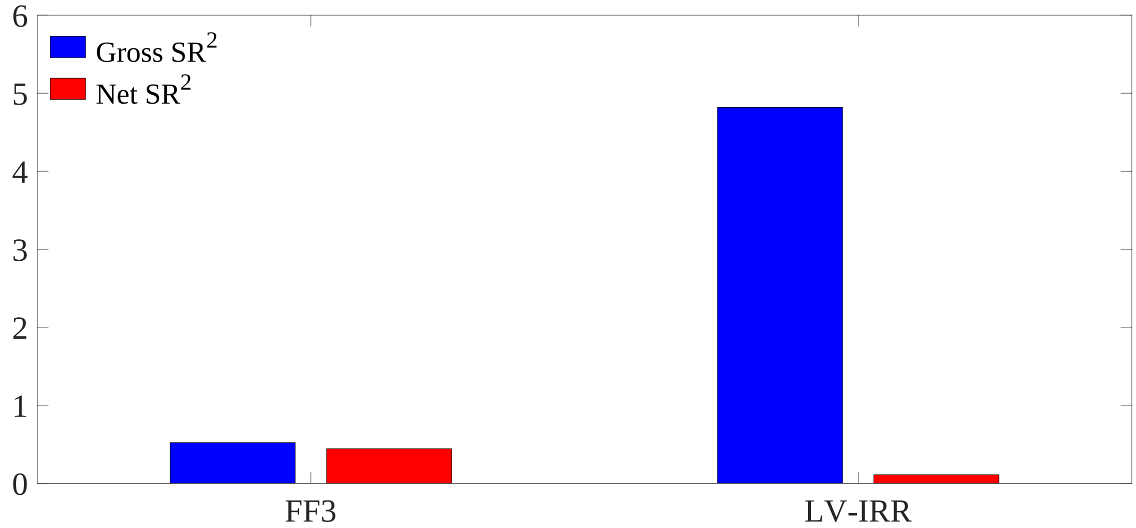


Figure 1: Before-costs (gross) and after-costs (net) maximum squared Sharpe ratios of the Fama-French three-factor model (FF3) and the low-volatility industry-relative-reversal factor (LV-IRR). The blue bars, 'Gross SR^2', use factor returns that ignore transaction costs. The red bars, 'Net SR^2', use factor returns that account for transaction costs. The sample period is January 1972 to December 2017.

and thus does not represent a meaningful asset pricing benchmark.[5]

3. Candidate factor models

The candidate models used in this paper come from Barillas and Shanken (2018) and Fama and French (2018). Barillas and Shanken (2018) compare the Fama and French (2015) five-factor model, sometimes augmented with a momentum factor, to the four-factor "*q*-theory" model of Hou et al. (2015). They also construct their own preferred model, using the strongest combination of factors from the other models. We analyze these four models and, following Fama and French (2018), additionally

[5] Li, DeMiguel, and Martín-Utrera (2020) expand on the motivation we use in this paper, focusing on how incorporating price impacts alters inferences regarding model selection. They show that when trading impacts prices, then the maximum squared Sharpe ratio criterion for model selection is problematic, because even for a single model the metric differs across investors based on the intensity with which they trade. Kan, Wang, and Zheng (2019) argue that the metric is problematic even absent transaction frictions, because investors cannot achieve the ex-post tangency portfolio in real time, so inferences change when comparing models based on out-of-sample Sharpe ratios instead of those observed in-sample.

Table 1: Factor models employed in tests

For each of the asset pricing models we consider in this paper, this table lists the non-market factors used by the model, denoted by a 'x' in the column below the model name, and three properties of the factors' construction: the primary characteristics used to form the factor, 'Primary sorting characteristic', the frequency that each factor is updated, 'Rebalance frequency', and the sorting method used to form each factor, 'Portfolio construction'. With the exception of SMB, factors with a portfolio construction of 2x3 are based on the six value-weighted portfolios resulting from the intersections of independent sorts of stocks into two groups based on size and three groups based on the primary sorting characteristic. With the exception of ME, factors with portfolio construction of 2x3x3 are based on the 18 value-weighted portfolios obtained from the intersections of independent sorts of stocks into two size groups, three groups based on a secondary characteristic besides size, and three groups based on the primary sorting characteristic. Breakpoints for size are based on the median market capitalization of NYSE stocks at the time of rebalancing and breakpoints for all other characteristics are based on the 30^{th} and 70^{th} percentiles of NYSE stocks. In all models, the factor returns are obtained as the equal-weighted average of the returns on the portfolios with high (or low) values of the primary sorting characteristic minus the equal-weighted average of the portfolios with low (or high) values. SMB returns are given by the simple average of the returns on all portfolios with low size minus the average of the returns on all portfolios with large size in three independent 2x3 sorts of stocks on size and each of the following characteristics: book-to-market ratio, growth in book assets, and operating profitability. ME returns are given by the simple average of the returns on all portfolios with low size minus those with large size in 2x3x3 sorts on size, growth in book assets, and return on equity All models employ a market factor, MKT, which is the return on the CRSP value-weighted index in excess of the return on the 30-day Treasury-bill. FF5 and FF6 denote the Fama and French (2015, 2018) five- and six-factor models, respectively. HXZ4 dentoes the Hou et al. (2015) four-factor q-model. BS6 denotes the Barillas and Shanken (2018) six-factor model. FF5C and FF6C denote versions of the FF5 and FF6, respectively, that use cash-based operating profitability instead of accruals operating profitability.

				Models that employ the factor					
Factor	Sorting signal	Rebalancing	Construction	FF5	FF6	HXZ4	BS6	$FF5_C$	$FF6_C$
SMB	Market capitalization	annual	2x3	x	x		x	x	x
HML	Book-to-market	annual	2x3	x	x			x	x
RMW	Accruals operating profitability	annual	2x3	x	x				
CMA	Growth in book assets	annual	2x3	x	x			x	x
MOM	Prior year's stock performance, excluding most recent month month	monthly	2x3		x		x		x
ME	Market capitalization	monthly	2x3x3			x			
IA	Growth in book assets	monthly	2x3x3			x	x		
ROE	Quarterly returns-on-equity	monthly	2x3x3			x	x		
HML(m)	Book-to-market	monthly	2x3				x		
RMW_C	Cash operating profitability	annual	2x3					x	x

Table 2: Factor summary statistics

For each asset pricing factor we consider, this table presents average monthly returns and t-statistics, both gross and net of transaction costs, along with average monthly turnover, TO, and transaction costs, TC. MKT, SMB, HML, RMW, CMA denote the Fama and French (2015) market, size, value, profitability, and investment factors, respectively. MOM denotes the Fama and French (2018) momentum factor. RMW_{C} denotes the Fama and French (2018) cash profitability factor, which is constructed similarly to RMW but based on cash based, not accruals based, operating profitability. ME, ROE, and IA denote the Hou et al. (2015) size, profitability, and investment factors, respectively. HML(m) denotes the monthly updated value factor of Asness and Frazzini (2013). The units for average returns, TO, and TC are % per month. The sample period is January 1972 to December 2017.

	Average returns					
	Gross	t-statistic	Net	t-statistic	TO	TC
MKT	0.56	2.95	0.56	2.95		
SMB	0.20	1.52	0.17	1.31	2.85	0.03
HML	0.36	2.89	0.31	2.45	5.10	0.05
RMW	0.28	2.79	0.22	2.23	5.27	0.05
CMA	0.31	3.75	0.22	2.61	9.76	0.09
MOM	0.66	3.56	0.19	1.03	51.40	0.47
ME	0.27	2.07	0.12	0.89	18.62	0.15
ROE	0.55	5.00	0.22	2.01	36.36	0.33
IA	0.38	4.82	0.16	2.03	24.59	0.22
HML(m)	0.34	2.21	0.14	0.90	19.37	0.20
RMW_{C}	0.36	4.52	0.29	3.58	7.36	0.07

consider versions of the Fama and French five- and six-factor factor models modified by replacing the standard profitability factor (RMW) with a variation based on cash profitability (RMW_{C}). Table 1 provides a simple summary of the factors employed in all six models. The table includes, for each factor, the stock-level characteristic used for portfolio construction (e.g., size, book-to-mark ratio), the basic construction methodology, and the rebalancing frequency. A more detailed description of the factors is provided in Appendix A.1.

Table 2 provides summary statistics for the factors. It reports each factor's average monthly returns (with t-statistics), both ignoring transaction costs ("gross") and net-of transaction costs ("net"), along with average monthly turnover (TO) and average monthly trading costs (TC). All the factors except for SMB have statistically significant average gross returns, ranging from 0.28% to 0.66% per month (RMW and MOM, respectively).[6] The turnover and transaction costs tend to be relatively high for factors that rebalance monthly, which also tend to be those with higher average gross returns. The turnover of the factors rebalanced monthly—MOM, ROE, IA, and HML(m)—ranges from 22.3% to 52.7% per month. This is an order of magnitude higher than the the 3.7% to 10.7% average monthly turnover observed on the annually rebalanced factors, SMB, HML, RMW, and CMA. As a result, the net-of-costs average returns of the former are generally insignificant after costs. This extends prior evidence from Lesmond et al. (2004) that momentum, while seemingly highly profitable before costs, is hardly profitable after costs.

4. Main results

4.1. Maximum squared Sharpe ratios

Figure 2 graphically depicts a key result of the paper, that the ranking of model performance ignoring transaction costs differs dramatically from the ranking accounting for transaction costs. The Hou-Xue-Zhang four-factor model and the Barrillas-Shanken six-factor model are both constructed to get closer to the ex post mean-variance frontier than their predecessors. However, they are also constructed without any concern for transaction costs. As a result, both of these models have higher gross squared Sharpe ratios than the Fama and French models, despite actually moving an investor farther from the achievable investment opportunity set.

Moving across the dark (blue) bars, Figure 2 shows an upward sloping pattern.

[6] Transaction costs are calculated as in Novy-Marx and Velikov (2016), using the stock-level effective spread measure of Hasbrouck (2009). This measure captures the substantial cross-sectional heterogeneity in transaction costs and is broadly available over our whole sample. A more detailed description of the methodology is provided in Appendix A.2.

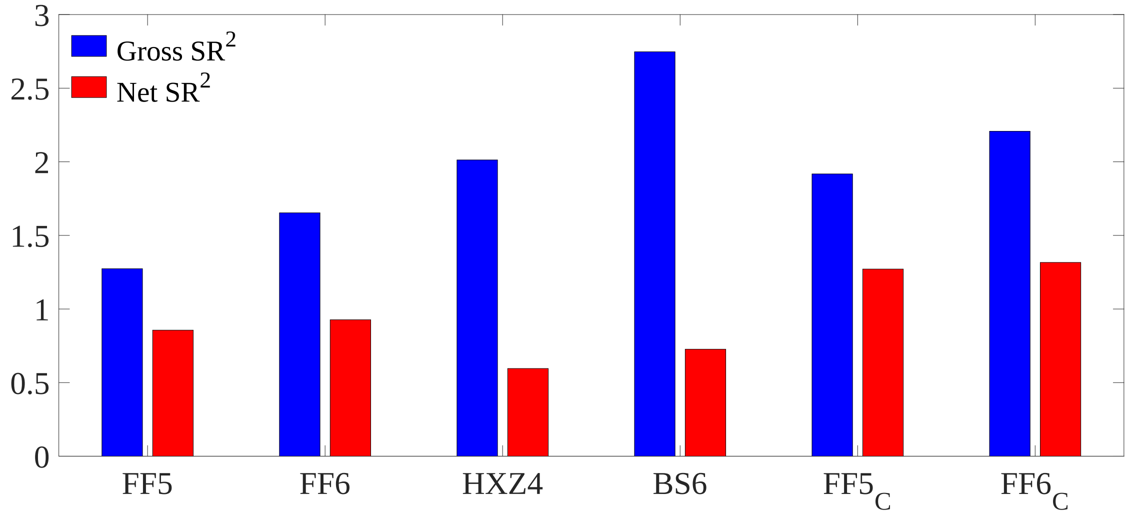


Figure 2: Maximum squared Sharpe ratios of factor models. This figure presents maximum squared Sharpe ratios from the factors in the models listed on the x-axis. The blue columns, 'Gross SR^2', use factor returns that ignore transaction costs. The red columns, 'Net SR^2', use factor returns that account for transaction costs. The sample period is January 1972 to December 2017.

The FF5 model's gross SR^2 is 1.24. Adding a momentum factor (FF6) increases this to 1.62. The HXZ4 model's gross SR^2 is 2.01, and the BS6 model's is 2.75. These squared Sharpe ratios are illusory, however, because they do not actually represent what an investor could have achieved. The achievable investment opportunity set is what is relevant for directing arbitrage capital.

The pattern for the net squared Sharpe ratios, which directly reflect the best risk-reward trade-off that an investor could have achieved in practice, looks very different. While the HXZ4 and BS6 models have higher gross squared Sharpe ratios than the Fama and French models, these models' factors are more expensive to trade, so these models' spans are actually farther from the achievable efficient frontier. The Fama and French five- and six-factor models' net SR^2 of 0.81 and 0.89 exceed the HXZ4 and BS6 models' 0.60 and 0.74, respectively.

The performance of the variations of the Fama and French model that include the cash profitability factor are more impressive still. While the gross SR^2 of the $FF5_C$ is 1.94, slightly lower than that of the HXZ4 model, its net SR^2 is 1.29, more than twice that of the HXZ4 model. Adding the momentum factor only marginally increases the net SR^2 of the $FF6_C$, to 1.33. The $FF6_C$ model dominates all other models in the sense that adding all of the other models' factors does not expand the investment opportunity set. Untabulated results show that the net squared Sharpe ratio of the 11-factor model that includes all the factors from all four models is the same 1.33 as that of the $FF6_C$ model alone.

Table 3 reports portfolio weights in ex post mean-variance efficient portfolios constructed using factors from the six candidate asset pricing models, i.e., the holdings of the portfolios that yield the squared Sharpe ratios shown in Figure 2. Panel A gives the weights an investor would have liked to have held in each factor, provided they could have traded completely without cost. It shows that the HXZ4 model improves on the gross squared Sharpe ratio of the FF5 model essentially by swapping out the Fama and French size, profitability, and investment factors for their own versions of these factors, which rebalance more frequently and are constructed using the more complicated three-way sorting procedure that further over-weights smaller stocks. Similarly, the BS6 model improves on the gross squared Sharpe ratio of the FF6 model essentially by swapping out the Fama and French profitability factor for HXZ4 profitability factor, and shifting the weight on the Fama and French investment factor to a value factor that rebalances monthly, largely in response to stock price performance, yielding a large negative correlation with momentum. The Fama and French model variations that use cash profitability improve on the standard versions essentially by moving all the of weight on RMW, and a third of the weight on CMA, to RMW_C, which has a higher Sharpe ratio than the accrual-based profitability factor.

Panel B reports results accounting for transaction costs. While the portfolio hold-

Table 3: Ex post mean-variance efficient portfolios

For each of the asset pricing models specified by the row headings, this table presents the weights of each factor in the portfolio consisting of the model's factors that maximizes the ex-post squared Sharpe ratio, SR^2. Panel A uses factor returns that ignore transaction costs. Panel B uses factor returns that account for transaction costs. The sample period spans January 1972 through December 2017.

Panel A: Results ignoring transaction costs

	Optimal factor weight (% of holdings in the ex post MVE portfolio)											
	MKT	SMB	HML	RMW	CMA	MOM	ME	ROE	IA	HML(m)	RMW_C	SR^2
FF5	18	10	-1	29	42							1.24
FF6	18	8	7	23	32	12						1.62
HXZ4	15						13	30	42			2.01
BS6	13	10				14		29	9	25		2.75
$FF5_C$	16	13	-1		26						44	1.94
$FF6_C$	16	11	4		22	8					39	2.22

Panel B: Results accounting for transaction costs

	Optimal factor weight (% of holdings in the ex post MVE portfolio)											
	MKT	SMB	HML	RMW	CMA	MOM	ME	ROE	IA	HML(m)	RMW_C	SR^2
FF5	21	10	4	30	35							0.83
FF6	20	9	8	26	29	7						0.90
HXZ4	22						9	27	42			0.60
BS6	17	12				8		28	17	18		0.74
$FF5_C$	18	13	2		20						46	1.29
$FF6_C$	18	13	4		18	4					43	1.33

ings are similar, the net factor returns are lower by the trading costs they incur, and the squared Sharpe ratios that could actually have been realized are consequently reduced by 33-73%. This reduction is greater for models employing high turnover factors. As a result, the FF5 and FF6 models have higher net squared Sharpe ratios than the HXZ4 and BS6 models (0.83 and 0.90 versus 0.60 and 0.74), despite having gross squared Sharpe ratios that are significantly lower. The net squared Sharpe ratios of the Fama and French five-factor model variation that employs cash profitability is much higher, 1.29. For this model, including momentum only yields marginal performance improvements, increasing the net squared Sharpe ratio to 1.33.[7]

4.2. Model comparison using frontier expansion

Barillas and Shanken (2017) show that, under the maximum squared Sharpe ratio criterion, one asset pricing model is superior to another if the first model's factors price the second model's factors but the converse is false. Adding the superior model's factors to the inferior model expands the investment opportunity set, and thus the SR^2, but adding the inferior model's factors to the superior model, relative to which they have no alpha, does not improve investment opportunities.

Transaction costs complicate quantifying frontier expansion. In the presence of trading frictions, the common notion of alpha can not be interpreted as a measure of abnormal investment opportunities, because a significant alpha does not necessarily indicate the existence of a trading strategy an investor could actually exploit. We can measure the extent to which adding one model's ($M1$'s) factors to those of another model ($M0$) expands the frontier, however, using a multi-factor version of the generalized alpha measure of Novy-Marx and Velikov (2016). Specifically, we use the alpha from a regression of the excess returns of the ex post mean-variance efficient

[7] HML also adds little to the Fama and French model variation based on cash profitability, because CMA and HML are positively correlated and thus play similar roles. The four factor model employing MKT, SMB, CMA, and $\mathrm{RMW_C}$ has a squared Sharpe ratio of 1.28, statistically indistinguishable from the 1.29 on the five factor model that additionally includes HML.

portfolio constructed from the union of the two models' factors ($MVE_{M1\cup M0}$) on the returns of the efficient portfolio constructed using the factors from the base model alone (MVE_{M0}):

$$MVE_{M1\cup M0,t} = \alpha + \beta MVE_{M0,t} + \epsilon_t. \quad (1)$$

The α's magnitude cannot be directly interpreted in general, but its statistical significance would indicate that the factors in $M1$ significantly improve the achievable investment frontier beyond that available from trading just the factors in $M0$.[8]

Panel A of Table 4 presents results of these spanning regressions for each pair of models, both ignoring and accounting for transaction costs. To quantify economic significance, Panel B also reports the percent increase in the maximum squared Sharpe ratio that results from adding the additional factors to the base model for each pair, which reflects the utility gain that could be realized by a mean-variance investor. The left panel, which presents results ignoring transaction costs, shows that on a pre-cost basis most of the models expand the investment frontier when added to others, though the HXZ4 model spans the standard Fama and French five- and six-factor models, and the BS6 model spans the HXZ4 model. The right panel, which presents results accounting for transaction costs, shows very different results. Here the standard Fama and French models, FF5 and FF6, significantly expand the achievable investment opportunity space relative to the HXZ4 of BS6 models alone, while these models' factors do not expand the achievable investment opportunity space relative to the FF6 model, and only marginally improve that of the FF5 model. Perhaps most strikingly, the bottom two rows show that none of the models significantly expand the achievable investment opportunity space for an investor already trading the variation of the Fama and French five-factor model that

[8] Novy-Marx and Velikov (2016) show that if $M1$ is a single test asset and there are no transaction costs, then the alpha from this regression, scaled by the weight of the test asset in the mean-variance efficient portfolio of the factor model and the test asset, agrees exactly with the common notion of alpha.

Table 4: Spanning regressions

Panel A reports intercepts from regressions of the form $MVE_{M0\cup M1,t} = \alpha + \beta \cdot MVE_{M0,t} + \epsilon_t$, where MVE_{M0} denotes the excess returns to the ex-post mean-variance-efficient portfolio of the factors in models $M0$, and $MVE_{M0\cup M1}$ denotes the same for the model made by augmenting $M0$ with the factors from model $M1$. The t-statistics, given in square brackets, are robust to heteroskedasticity. Column headings specify the $M1$, while rows specify the $M0$. Panel B quantifies the economic significance of these intercepts, using the fractional increase in the maximum squared Sharpe ratio resulting from augmenting the column model with the factors from the row model, $\%\Delta SR^2(M0, M1) = SR^2(M0, M1)/SR^2(M0) - 1$. Results are given both ignoring transaction costs (Gross), and accounting for these costs (Net). The sample spans January 1972 through December 2017.

	Suplamentary model ($M1$)											
	Gross						Net					
Base model ($M0$)	FF5	FF6	HXZ4	BS6	$FF5_C$	$FF6_C$	FF5	FF6	HXZ4	BS6	$FF5_C$	$FF6_C$
Panel A: Alphas from regressing $MVE_{M0\cup M1}$ onto MVE_{M0}												
FF5		1.10	1.62	2.14	1.31	1.67		0.28	0.36	0.40	1.04	1.12
		[3.46]	[5.93]	[9.58]	[5.93]	[6.75]		[1.59]	[2.03]	[2.07]	[4.55]	[4.84]
FF6	0.00		0.99	1.75	1.11	1.11	0.00		0.14	0.14	0.94	0.94
	[0.00]		[4.53]	[7.64]	[5.08]	[5.08]	[0.00]		[1.28]	[1.28]	[4.23]	[4.23]
HXZ4	0.12	0.18		1.18	0.75	0.88	0.93	0.97		0.48	1.78	1.84
	[1.50]	[1.73]		[5.11]	[3.63]	[4.15]	[4.23]	[4.28]		[2.50]	[5.51]	[5.68]
BS6	0.60	0.60	0.04		0.81	0.81	0.61	0.61	0.00		1.55	1.55
	[3.90]	[3.90]	[0.80]		[4.43]	[4.43]	[2.90]	[2.90]	[0.00]		[4.83]	[4.83]
$FF5_C$	0.25	0.64	0.90	1.58		0.59	0.00	0.12	0.04	0.12		0.12
	[2.55]	[3.45]	[4.25]	[8.41]		[2.74]	[0.00]	[1.16]	[0.74]	[1.16]		[1.16]
$FF6_C$	0.22	0.22	0.48	1.31	0.00		0.00	0.00	0.00	0.00	0.00	
	[2.32]	[2.32]	[2.95]	[6.51]	[0.00]		[0.00]	[0.00]	[0.06]	[0.06]	[0.00]	
Panel B: $\%\Delta SR^2(M0, M1)$												
FF5		31.0	67.9	169.9	70.2	92.2		8.8	11.6	13.2	55.7	60.9
FF6	0.0		29.7	105.9	46.7	46.7	0.0		4.1	4.1	48.0	48.0
HXZ4	3.2	4.5		37.9	19.0	23.1	53.8	56.0		23.3	116.7	121.7
BS6	21.3	21.3	0.9		30.6	30.6	26.4	26.4	0.0		79.8	79.8
$FF5_C$	8.5	22.5	23.4	85.3		14.6	0.0	3.4	1.0	3.4		3.4
$FF6_C$	6.9	6.9	11.4	61.6	0.0		0.0	0.0	0.0	0.0	0.0	

employs cash profitability. That is, this model spans the others, and even adding a momentum factor does not significantly improve the model's performance.

4.3. Model comparison using bootstrap simulations

While the maximum squared Sharpe ratio is a useful summary statistic for model comparison, it does not accompany a natural distribution theory to identify whether two models have statistically different maximum squared Sharpe ratios. It is informative about which models' span was closer to the ex post mean variance efficient frontier of all assets, but provides little information regarding how likely it is that a "good" model will beat a "bad" model going forward. We can answer questions of these sorts, however, using bootstraps. These simulations allow us to get a sense of the full distribution of potential maximum squared Sharpe ratios for each model, and to make estimates of the probability of any one model beating any other model, or of beating all the other models, over some sample going forward.

Bootstraps have the further advantage that they can be used to generate both in-sample (IS) and out-of-sample (OS) results. The IS results are comparable to the full in-sample results reported in Figure 2 and Table 3 in the sense that, like the full in-sample results, they are biased upward (see, for example, DeMiguel et al., 2009). For these, in each bootstrap run we use the same simulated return series both to select the optimal weights on any model's factors, and to calculate this portfolio's returns. This biases the SR^2 upward, because the MVE weights are overfit, yielding the highest possible in-sample Sharpe ratio, despite being only noisy estimates of the true optimal weights. The IS maximum squared Sharpe ratios are consequently higher than what an investor can expect going forward.

Unlike the IS results, the OS results are not upward biased, and thus more accurately reflect what investors can hope to achieve in practice. For these, in each bootstrap run we estimate optimal portfolio weights as in the IS runs, but calculate portfolio performance using an alternative set of simulated factor returns. The performance of these portfolios, which are largely free from the look-ahead bias that plagues the IS portfolios, is significantly less impressive, but more representative of true potential performance.[9]

The actual bootstrap procedure follows Fama and French (2018). To be certain that the OS results do not use any in-sample information, we need to ensure that the returns from months used to calculate portfolio performance are distinct from the those used to estimate the optimal portfolio weights. At the start of each bootstrap run we consequently partition the whole sample into IS months and OS months. To do this, one of each consecutive pair of months is randomly assigned to the IS sample, with its partner month assigned to the OS sample. The IS factor returns are created by sampling, with replacement, from the IS months, with the OS returns constructed using the corresponding partner months. For each run we compute maximum squared Sharpe ratios, SR^2, both IS and OS, for each model, and compare these to each other. When two models have the same Sharpe ratio in a given run, we break the tie in favor of the more parsimonious model. For example, when MOM is not used in the optimal in-sample portfolio for FF6 in a given simulation run, the FF6 Sharpe ratio must equal that of the FF5, and in such cases we declare the latter model, which employed one fewer potential factors, the "winner." Reported results are based off 100,000 runs of the longest practicable sample size, 23 years (276 months), which is half the length of the full data sample, January 1972 through December 2017.

Table 5 presents the formal statistical comparison of the differences between the net-of-costs Sharpe ratios of each model. Panel A presents in-sample (IS) results. The IS SR^2 are biased upwards, so their level should be interpreted with caution.[10]

[9] The OS results are themselves not completely free from all potential biases, since the factors under consideration were themselves identified with the benefit of hindsight in studies that use the full sample data.

[10] These results are even more biased than the true in-sample results given in Table 3, because

With that caveat, the average IS net SR^2 is higher for the FF5 and FF6 models than for the HXZ4 and BS6 models. Moreover, the FF6 model outperforms the HXZ4 in over 95% of the samples, and the BS6 model in over 79% of the samples.

Panel B presents the out-of-sample (OS) results. The average OS net-of-costs SR^2 of the FF5 and FF6 models are closer, but still higher compared to those of the HXZ4 and BS6 models. The net-of-costs OS SR^2 of the FF5 and FF6 are greater than those of the HXZ4 and BS6, respectively, in three-quarters of the bootstrap samples. Overall, the evidence in Figure 2 and Tables 3 and 5 shows that the FF5 and FF6 models tend to dominate the HXZ4 and BS6 on a net basis, because of the latter models' high transaction costs. This finding essentially reverses the comparisons of Hou et al. (2019) and Barillas and Shanken (2018).

The Fama and French model variations that employ cash-based profitability have stronger performance still, with average maximum squared Sharpe ratios exceeding one, more than 50% higher than their counterparts employing the accruals-based profitability factor. In fact, in 92.5% of the runs the $FF5_C$ or $FF6_C$ model is the best performing model. The performance of these two models is also surprisingly similar. The $FF5_C$ has a slightly higher average OS squared Sharpe ratio than the six-factor version that also includes momentum, and with ties breaking in favor of the more parsimonious model, the $FF5_C$ outperforms the $FF6_C$ more often than it underperforms it. In 28.6% of runs, the MVE portfolio of the $FF6_C$ factors excludes momentum. In the remaining runs, 41.3% of the time it improves OS performance, but three-quarters as often, 30.1% of the time, trading momentum diminishes OS performance. That is, almost half the time an investor in these factors takes a position in momentum, they end up regretting it ex post.

the sample is shorter, yielding more acute overfitting of the optimal weights. Appendix Table B.1 shows IS bootstrap results for a sample length of 552 months, the same length as our true in-sample results, and reports similar maximum squared Sharpe ratios.

Table 5: Bootstrapped net-of-costs maximum squared Sharpe ratios

For each model we consider, the first column presents average net-of-costs maximum squared Sharpe ratios, SR^2, from 100,000 in-sample (IS) or out-of-sample (OS) simulation runs. IS and OS simulations split the 552 sample months of our sample period, January 1972 through December 2017, into 276 adjacent pairs: months (1, 2), (3, 4), ... (551, 552). A simulation run draws a random sample with replacement of 276 pairs. The IS simulation run chooses a month randomly from each pair in the run. We calculate IS SR^2 on that sample of months and then apply the corresponding portfolio weights in the unused months of the simulation pairs to produce the corresponding OS estimate of the Sharpe ratio for the IS tangency portfolio. The six columns labeled by model names, 'FF5' to 'FF6C', present the percentage of bootstrap simulations in which the squared Sharpe ratio of the model defined by the row heading is greater than that of the model defined by the column heading. The last column, 'Best', presents the percentage of bootstrap simulation runs in which the model specified by the row heading has the highest squared Sharpe ratio among all models in the run. Panel A presents IS results and Panel B presents OS results. The sample period is January 1972 through December 2017.

Panel A: In-sample bootstrap results

		Probability (%) that the row model performs better than the column model						
	Mean-SR^2	FF5	FF6	HXZ4	BS6	$FF5_C$	$FF6_C$	Best
FF5	1.23		21.0	86.1	58.4	2.4	1.5	0.6
FF6	1.41	79.0		95.5	79.1	19.4	3.1	2.1
HXZ4	0.89	13.9	4.5		0.7	2.8	0.9	0.0
BS6	1.18	41.6	20.9	99.3		13.0	4.4	4.2
$FF5_C$	1.68	97.5	80.6	97.2	87.0		28.6	26.8
$FF6_C$	1.83	98.5	96.8	99.1	95.6	71.4		66.3

Panel B: Out-of-sample bootstrap results

		Probability (%) that the row model performs better than the column model						
	Mean-SR^2	FF5	FF6	HXZ4	BS6	$FF5_C$	$FF6_C$	Best
FF5	0.65		47.9	78.6	73.5	2.8	7.7	1.2
FF6	0.66	52.1		80.5	80.1	7.6	3.1	1.6
HXZ4	0.45	21.4	19.5		35.8	4.6	4.8	2.5
BS6	0.49	26.5	19.9	64.2		5.4	3.6	2.2
$FF5_C$	1.03	97.2	92.4	95.4	94.6		58.7	54.9
$FF6_C$	1.01	92.3	96.8	95.2	96.4	41.3		37.6

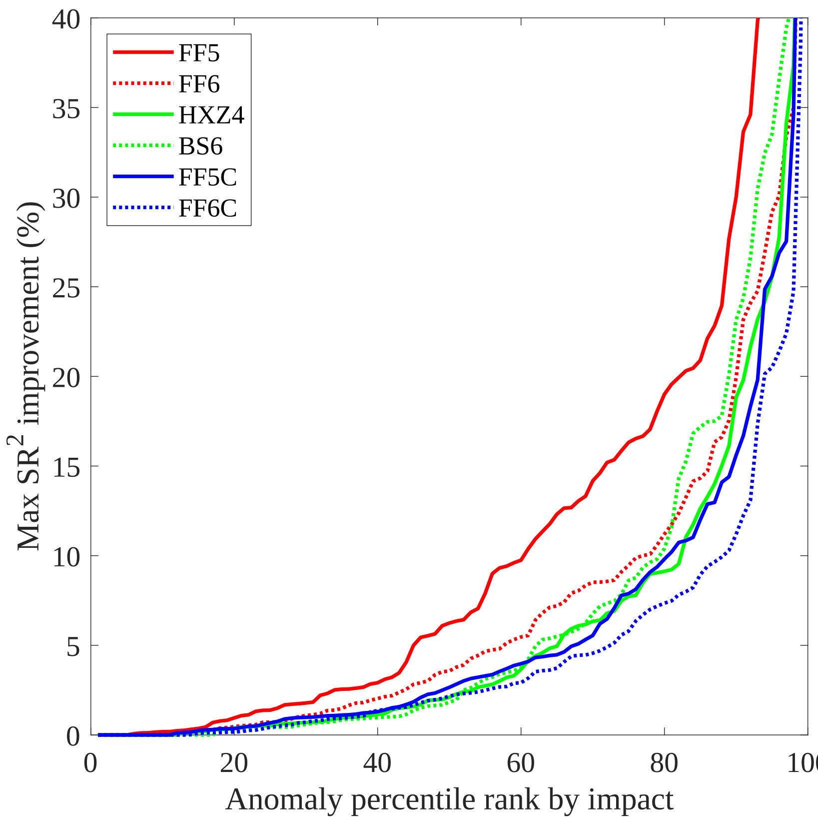
5. Model performance explaining anomalies

This section takes an alternative throwback approach to factor model comparison. Instead of studying the ability of the factor models to price each other's factors, this alternative approach compares models' ability to price various "test assets." For these test assets we use 120 anomaly long/short portfolios from Chen and Velikov (2019).[11] These are constructed from the usual data sources; more than half of the predictors are constructed using Compustat data, while almost a third use only market pricing data available from CRSP. Most of the remainder use analyst forecasts, though several focus on institutional ownership data, trading volume, or specialized data. For further details, see Appendix C.

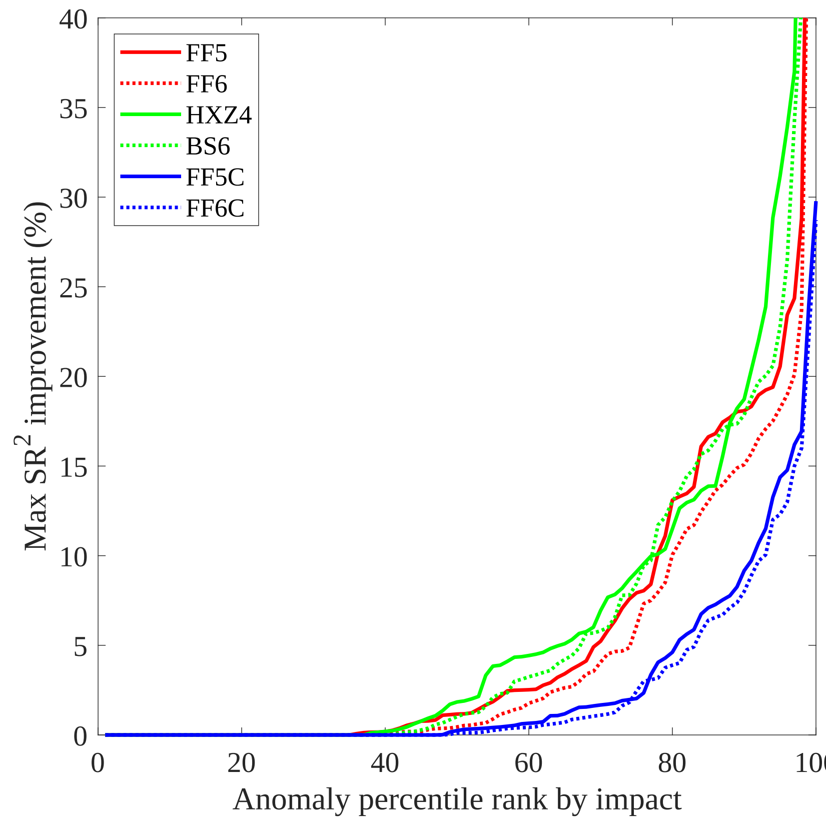
For this approach we still rely on squared Sharpe ratio as a comparison criterion, since the common notion of alpha does not measure abnormal returns in the presence of transaction costs. Specifically, for each model M taken from the six candidates, and for each test asset A taken from the 120 anomalies, we compute the maximum squared Sharpe ratios attainable from the model's factors alone, $\text{SR}^2(M)$, and from the model's factors augmented with the anomaly, $\text{SR}^2(M, A)$. We then compute the extent to which adding the anomaly expands the model's ex-post mean-variance frontier, $\%\Delta\text{SR}^2(M, A) = \text{SR}^2(M, A)/\text{SR}^2(M) - 1$. We do this using both gross- and net-of-costs returns. In each case, we then rank the anomalies for each model based on this measure.

Figure 3 plots the extent to which the anomalies improve each model's squared Sharpe ratio. Anomalies that improve a model's maximum SR^2 the least are shown on the left, while those that yield the greatest improvements are shown to the right. If one model's line is below another, then across the distribution, the anomalies

[11] These anomalies are distilled from a set of 156 cross-sectional return predictors employed by Chen and Zimmerman (2018), taken from 115 publications in accounting, economics, and finance journals. Chen and Velikov (2019) remove 34 predictors that do not reliably produce return spreads in decile sorts, and 2 predictors they deem are due to risk, to arrive at the final set of 120 anomalies.



Panel A: Ignoring costs



Panel B: Net of costs

Figure 3: Frontier expansion (squared Sharpe ratio improvement) from adding anomalies to asset-pricing models. For each asset-pricing model we consider, M, and each of the 120 anomalies, A, described in Section 5, we compute the maximum ex-post squared Sharpe ratio attainable from the model's factors, $\mathrm{SR}^2(M)$, along with the maximum squared Sharpe ratio attainable from the model's factors and the anomaly, $\mathrm{SR}^2(M, A)$. We then compute the squared Sharpe ratio improvement, $\%\Delta\mathrm{SR}^2(M, A) = \mathrm{SR}^2(M, A)/\mathrm{SR}^2(M) - 1$. This figure depicts plots of the percentiles from the distribution of the 120 $\%\Delta\mathrm{SR}^2(M, A)$ for each model specified by the plot legend. Panel A (B) ignores (accounts for) transaction costs. For ease of visualization, the y-axes are truncated at 40%.

expand the frontier of the former model less than they do the second, indicating that it better prices anomalies.

Panel A, which shows results ignoring transaction costs, is consistent with the work of Hou et al. (2015). It shows that the anomalies generally expand the frontier for the Fama and French five-factor model more than they do for the HXZ4 or BS6 models, and that these later models perform similarly to the Fama and French six-factor model.

The right panel, which shows results net of transaction costs, looks very different. First, in each case nearly half the anomalies contribute nothing to frontier expansion. That is, only half of anomalies seem at all anomalous after accounting for costs. In the right half of the distribution, where the anomalies do expand the investment frontier, the lines for the Fama and French five- and six-factor models tend to be below those for the HXZ4 and BS6 models, suggesting that these later models, presented as improvements on the standard models, actually perform worse at explaining achievable investment performance. This is particularly remarkable, given that our test assets are selected in large part on the basis of their anomalous performance relative to the Fama and French models, and consequently biased against these models. Perhaps most strikingly, the lines corresponding to the Fama and French model variations that employ cash profitability dominate the other models, admitting much smaller frontier expansions across the whole distribution.

6. Robustness

The HXZ4 and BS6 models do not outperform the standard Fama-French models on a net basis. They were built to perform well on a gross basis, without regards for transaction costs, and are consequently expensive to trade, impairing their performance in practice. They should perform better, however, if designed in a manner that reduces the cost of trading their factors. In this section, we consider the robustness of the inferences drawn above to techniques traders could use to mitigate trading costs.

6.1. *Models using cost-mitigated factors*

We first consider the performance of models that employ factors constructed with "banding," the single most effective cost mitigation technique investigated by Novy-Marx and Velikov (2016). The banded strategies employ a buy/hold spread at each break point. This introduces trading hysteresis, dramatically reducing turnover. Specifically, at each break point we introduce a 20% spread. For a tertile break point, instead of trading at each rebalance point to hold stocks with sorting characteristic more extreme than 30% of NYSE firms, we buy stocks that have entered the top 20%, and continue to hold any stocks that we already held provided they remain in the top 40%. For the median breakpoint, we buy stocks entering the top 40% and continue to hold current positions until they leave the top 60%.

Table 6 gives summary statistics for factors constructed using the banding technique, including average monthly returns, turnover, and transaction costs. These factors have similar average gross performance, but roughly a third less turnover, and commensurately lower transaction costs, than their un-mitigated counterparts.

Table 6: Summary statistics of cost-mitigated factors

This table presents average monthly returns and t-statistics, both gross and net of transaction costs, along with average monthly turnover, TO, and transaction costs, TC, of versions of the asset-pricing factors defined in Table 2 that employ the 20% banding cost-mitigation strategy described in Section 6.1. The units for average returns, TO, and TC are % per month. The sample period is January 1972–December 2017.

	Average monthly excess returns					
	Gross	t-statistic	Net	t-statistic	TO	TC
MKT	0.56	2.95	0.56	2.95		
SMB	0.16	1.22	0.14	1.07	1.88	0.02
HML	0.35	2.78	0.32	2.50	3.09	0.03
RMW	0.22	2.37	0.18	1.94	3.46	0.04
CMA	0.31	3.74	0.23	2.77	7.68	0.07
MOM	0.74	3.91	0.46	2.46	28.1	0.28
ME	0.24	1.91	0.11	0.89	14.5	0.13
ROE	0.58	4.88	0.33	2.80	25.3	0.24
IA	0.47	5.50	0.29	3.36	19.0	0.18
HML(m)	0.30	2.06	0.23	1.54	6.56	0.08
RMW_C	0.32	3.62	0.26	2.98	5.38	0.05

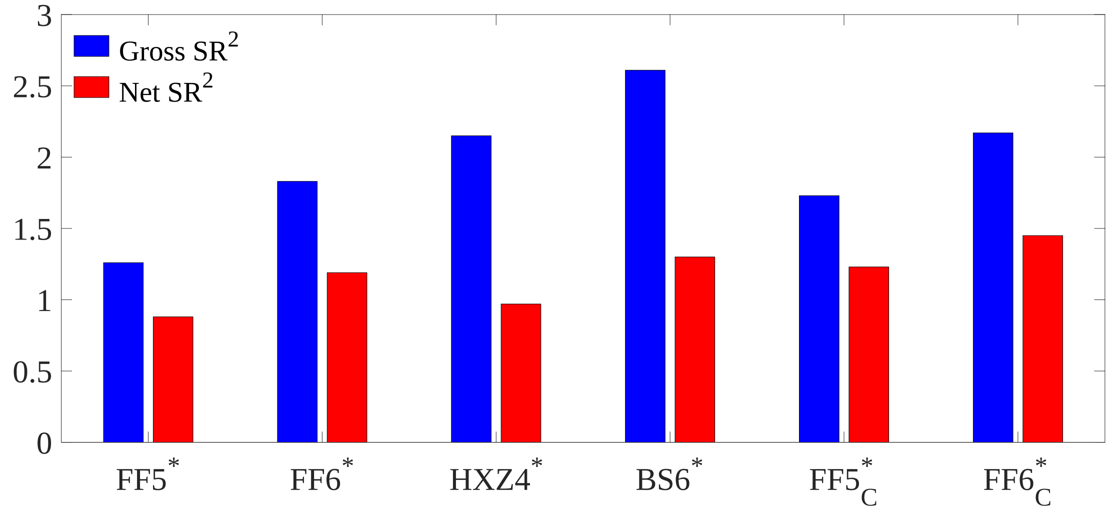


Figure 4: Maximum squared Sharpe ratios of cost-mitigated factor models. This figure presents maximum squared Sharpe ratios from factors in the models listed on the x-axis. The blue columns, 'Gross SR^2', use factor returns that ignore transaction costs. The red columns, 'Net SR^2', use factor returns that account for transaction costs. The factors are all versions that use the 20% banding cost-mitigation strategy described in Section 6.1. The sample period is January 1972 to December 2017.

Figure 4, which replicates Figure 2 for the models using the cost mitigated factors, shows the maximum squared Sharpe ratios, both accounting for and ignoring transaction costs.[12] Overall, the results are similar to those presented earlier, with a few important differences. All but one ($FF5_C$) of the models perform better on a net basis, because they are more transactionally efficient. The improvements are more pronounced for the models that employ higher turnover factors, because transactional efficiency yields greater benefits when costs are higher. As a result, there is less distinction between model performance, though the $FF6_C$ is still the best performing model. It outperforms its five-factor counterpart more here because the momentum factor, which has the highest turnover and highest trading costs of all the factors, benefits the most from efficient construction. It consequently adds more

[12] Appendix Table D.3 reports the weights in the ex-post MVE portfolios for each models' factors in Figure 4. Model spanning tests analogous to those provided in Table 4, which show similar results, can be found in Appendix Table D.1.

to models designed explicitly to account for and mitigate trading frictions.

Panel A of Table 7 presents the full-sample maximum squared Sharpe ratios net of costs from Figure 4 and their improvement relative to those from Table 3 that do not use cost mitigation. Consistent with Figure 4, cost mitigation improves the net-of-costs performance of most models, especially the HXZ4 and BS6, which experience more than 60% increases in squared Sharpe ratios from cost mitigation. Panels B and C of Table 7 shows bootstrap results analagous to those provided in Table 5, both in-sample (IS) and out of sample (OS), for the models that employ the cost mitigated factors. It shows that, while the $FF6_C$ is still the best performing model more than 55% of the time in-sample, and a little over 40% out of sample, the BS6 model is a stronger contender, performing best in-sample about a third of the time and out of sample a little over a quarter of the time. This partly reflects the greater contribution of momentum to model performance when factors are designed to be transactionally efficient. Investors that chose to trade momentum because of its in-sample performance are significantly less likely to regret trading it ex post. While trading the momentum factor designed without regards for reducing transaction costs diminishes overall out-of-sample portfolio performance almost half the time in Table 5, trading the banded momentum factor does so barely more than a quarter of the time in Table 7.

Table 7: Bootstrapped net-of-costs maximum squared Sharpe ratios using cost-mitigated factors

For each model we consider, Panel A presents full-sample maximum squared Sharpe ratios, SR^2. The row, '% Improvement', contains the percentage change in squared Sharpe ratios from Panel B of Table 3 to those in Panel A. The first column of Panels B and C presents average net-of-costs maximum squared Sharpe ratios, SR^2, from 100,000 in-sample (IS) or out-of-sample (OS) simulation runs. IS and OS simulations split the 552 months of our sample period, January 1972 to December 2017, into 276 adjacent pairs: months (1, 2), (3, 4), ... (551, 552). A simulation run draws a random sample with replacement of 276 pairs. The IS simulation run chooses a month randomly from each pair in the run. We calculate IS SR^2 on that sample of months and then apply the corresponding portfolio weights in the unused months of the simulation pairs to produce the corresponding OS estimate of the Sharpe ratio for the IS tangency portfolio. The six columns labeled with model names, 'FF5' to 'FF6C', present the percentage of bootstrap simulations in which the squared Sharpe ratio of the model defined by the row heading is greater than or equal to that of the model defined by the column heading. The last column of Panels B and C, 'Max', presents the percentage of bootstrap simulation runs in which the model specified by the row heading has the highest squared Sharpe ratio of all models in the run. Panel B presents IS results and Panel C presents OS results. All models use versions of the factors that employ the 20% banding strategy described in Section 6.1. Panel A presents IS results and Panel B presents OS results. The sample period is January 1972 to December 2017.

Panel A: Full-sample maximum-squared Sharpe ratios								
		FF5	FF6	HXZ4	BS6	$FF5_C$	$FF6_C$	
SR^2		0.88	1.19	0.97	1.30	1.23	1.45	
% Improvement		6%	32%	62%	76%	-4%	9%	
Panel B: In-sample bootstrap results								
		Probability (%) that the row model performs better than the column model						
	Mean-SR^2	FF5	FF6	HXZ4	BS6	$FF5_C$	$FF6_C$	Max
FF5	1.27		4.6	48.6	11.8	4.1	1.5	0.2
FF6	1.70	95.4		85.0	34.3	48.3	8.4	3.4
HXZ4	1.30	51.4	15.0		0.1	21.4	5.4	0.0
BS6	1.84	88.2	65.7	99.9		60.9	35.6	34.7
$FF5_C$	1.68	95.9	51.7	78.6	39.1		9.5	6.5
$FF6_C$	2.02	98.5	91.6	94.6	64.4	90.5		55.3
Panel C: Out-of-sample bootstrap results								
		Probability (%) that the row model performs better than the column model						
	Mean-SR^2	FF5	FF6	HXZ4	BS6	$FF5_C$	$FF6_C$	Max
FF5	0.66		18.6	33.4	17.6	4.1	6.7	0.6
FF6	0.86	81.4		58.5	32.7	34.5	8.3	2.6
HXZ4	0.80	66.6	41.5		21.9	31.4	20.0	8.1
BS6	0.99	82.4	67.3	78.1		50.4	35.1	26.8
$FF5_C$	0.99	95.9	65.5	68.6	49.6		28.9	20.7
$FF6_C$	1.12	93.3	91.6	80.0	64.9	71.1		41.2

6.2. Trading diversification

The previous tests, like most of the prior literature, account for transaction costs under the implicit assumption that investors trade factors, or factor-based funds, as opposed to individual stocks (e.g., Berk and van Binsbergen, 2015; Novy-Marx and Velikov, 2016). Under this approach, the costs of investing in a given factor are equal to the costs of trading individual stocks according to that factor's portfolio weights. However, DeMiguel et al. (2020) show that when investors trade individual stocks based on multiple characteristics at the same time, buys based on one characteristic can offset sells based on another characteristic. As a result, trading individual stocks based on the portfolio weights of multiple factors can be less expensive than the sum of the costs associated with investing in each factor in isolation. In this section, we examine the impact of this effect, which DeMiguel et al. (2020) refer to as "trading diversification," on our main results.

Figure 5 shows maximum squared Sharpe ratios similar to those in Figures 2 and 4, but under the assumption that investors realize the benefits of trading diversification, i.e., after netting offsetting trades across the individual factors. Said differently, investors weight individual stocks by summing the relevant weights from the factors in each of the candidate models. Similar to the results using cost-mitigation strategies, after trading diversification benefits, the squared Sharpe ratios of the HXZ4 and BS6 become closer to those of the FF5 and FF6 models, respectively. This finding is consistent with the intuition that trading diversification should provide a larger cost reduction for factors that trade more frequently. The $FF5_{C}$ and $FF6_{C}$ continue to dominate the other models with squared Sharpe ratios that are 40% larger than their next-best competitors.

Table 8 presents model performance results net-of-costs, similar to those in Tables 3 and 5, but under the assumption that investors realize the benefits of trading diversification, i.e., after netting offsetting trades across the individual factors. Panel

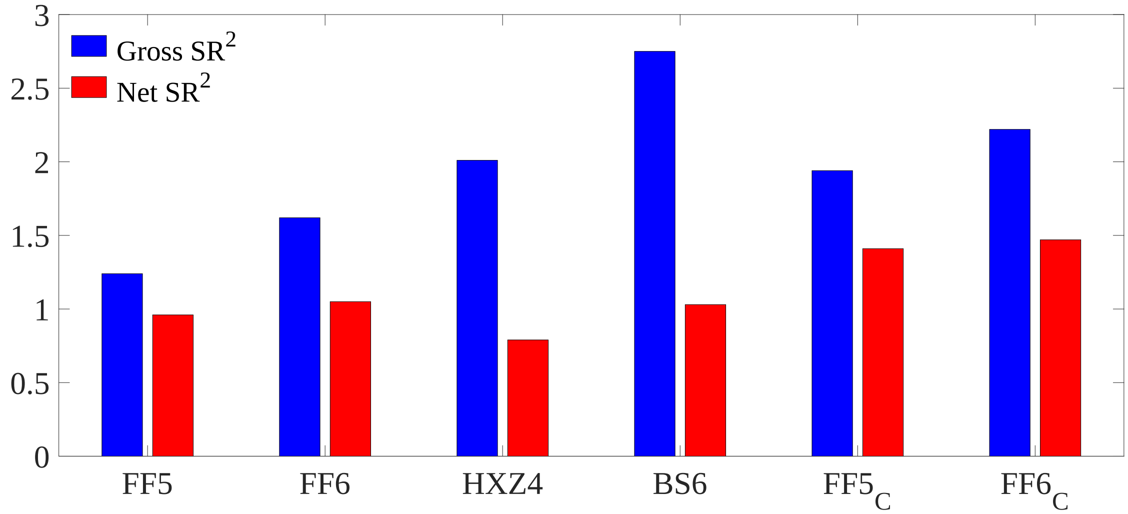


Figure 5: Maximum squared Sharpe ratios of factor models incorporating trading diversification. This figure presents maximum squared Sharpe ratios from the factors in the models listed on the x-axis. The blue columns, 'Gross SR^2', use factor returns that ignore transaction costs. The red columns, 'Net SR^2', use factor returns that account for transaction costs and incorporate the trading diversification effect described in Section 6.2. The sample period is January 1972 to December 2017.

A of Table 8 tabulates the full-sample maximum squared Sharpe ratios net of costs depicted in Figure 5 and their improvement relative to those from Table 3 that ignore trading diversification. Consistent with Figure 5, cost mitigation improves the net-of-costs performance of all models, especially the HXZ4 and BS6, which experience increases in squared Sharpe ratios from trading diversification that exceed 30%.

Panels B and C of Table 8 present the in-sample and out-of-sample bootstrap comparisons of Sharpe ratios across models. These panels show that $FF5_C$ and $FF6_C$ have higher Sharpe ratios than the other models in the vast majority of bootstrap simulations. While the $FF6_C$ can not have a lower Sharpe ratio than the nested $FF5_C$ in-sample, its out-of-sample squared Sharpe ratio is effectively the same on average (1.13 vs. 1.14), and lower in over half of bootstrap runs, than that of the $FF5_C$. Similarly, the FF6 does not significantly outperform the FF5 in out-of-sample bootstraps, suggesting that investors would find it difficult to improve the mean-variance frontier

Table 8: Full-sample and bootstrapped net-of-costs maximum squared Sharpe ratios accounting for trading diversification

For each model we consider, Panel A presents full-sample maximum squared Sharpe ratios, SR^2. The row, '% Improvement', contains the percentage change in squared Sharpe ratios from Panel B of Table 3 to those in Panel A. The first column of Panels B and C presents average net-of-costs maximum squared Sharpe ratios, SR^2, from 100,000 in-sample (IS) or out-of-sample (OS) simulation runs. IS and OS simulations split the 552 months of our sample period, January 1972 to December 2017, into 276 adjacent pairs: months (1, 2), (3, 4), ... (551, 552). A simulation run draws a random sample with replacement of 276 pairs. The IS simulation run chooses a month randomly from each pair in the run. We calculate IS SR^2 on that sample of months and then apply the corresponding portfolio weights in the unused months of the simulation pairs to produce the corresponding OS estimate of the Sharpe ratio for the IS tangency portfolio. The six columns in Panels B and C labeled with model names, 'FF5' to 'FF6C', present the percentage of bootstrap simulations in which the squared Sharpe ratio of the model defined by the row heading is greater than or equal to that of the model defined by the column heading. The last column of Panels B and C, 'Max', presents the percentage of bootstrap simulation runs in which the model specified by the row heading has the highest squared Sharpe ratio of all models in the run. Panel B presents IS results and Panel C presents OS results. The returns on all models' factors are corrected for transaction costs accounting for the trading diversification benefit described in Section 6.2.

Panel A: Full-sample maximum-squared Sharpe ratios

	FF5	FF6	HXZ4	BS6	$FF5_C$	$FF6_C$
SR^2	0.96	1.05	0.79	1.03	1.41	1.47
% Improvement	16%	17%	32%	39%	10%	11%

Panel B: In-sample bootstrap results

		Probability (%) that the row model performs better than the column model						
	Mean-SR^2	FF5	FF6	HXZ4	BS6	$FF5_C$	$FF6_C$	Max
FF5	1.34		17.7	77.4	39.3	2.4	1.4	0.4
FF6	1.53	82.3		91.1	58.5	18.9	3.1	1.7
HXZ4	1.08	22.6	8.9		0.8	5.2	1.7	0.1
BS6	1.46	60.7	41.5	99.2		23.0	10.7	10.3
$FF5_C$	1.82	97.6	81.1	94.8	77.0		24.0	21.7
$FF6_C$	1.99	98.6	96.9	98.3	89.3	76.0		65.8

Panel C: Out-of-sample bootstrap results

		Probability (%) that the row model performs better than the column model						
	Mean-SR^2	FF5	FF6	HXZ4	BS6	$FF5_C$	$FF6_C$	Max
FF5	0.74	0.0	42.8	67.1	48.7	2.7	7.4	0.9
FF6	0.76	57.2	0.0	70.9	52.9	8.1	3.2	1.1
HXZ4	0.61	32.9	29.1	0.0	22.7	8.4	8.0	3.3
BS6	0.75	51.3	47.1	77.3	0.0	15.3	11.8	8.6
$FF5_C$	1.14	97.3	91.9	91.6	84.7	0.0	52.4	46.9
$FF6_C$	1.13	92.6	96.8	92.0	88.2	47.6	0.0	39.1

with the momentum factor after costs out-of-sample. Panel C further shows that the HXZ4 tends to have a lower out-of-sample Sharpe ratio than the other models, but no statistically significant winner emerges from the FF5, FF6, and BS6. Overall, Table 8 shows that trading diversification benefits can close the performance gaps between some models, but does not significantly alter our main inferences.

7. Conclusion

In a world with implementation costs, even the "right" risk-based asset-pricing model should not completely explain the cross-section of before-cost expected returns, but only the portion that arises as compensation for risk. We investigate which asset pricing models explain the cross-section of returns taking transaction costs into account. Our results show that correcting for costs fundamentally alters the outcome of model selection exercises. Prior model-selection studies tend to pick factors that update at the monthly frequency, because they account for the benefits of frequent updating while ignoring the costs incurred doing so. This gives these models an unrealizable and illusory advantage over models with factors that update only annually. After costs are considered, models with less frequent re-balancing tend to outperform.

More generally, these results highlight serious problems associated with ignoring real world concerns when researching financial markets. Accounting for frictions can completely reverse conclusions obtained ignoring constraints. Performance investors realize is often far inferior to that promised by active managers, who support their claims using backtests that fail to adequately account for slippage. Implementation issues are not just an annoyance; they have first order impacts on outcomes. Compounding these issues are strong incentives to find positive results, in both academia and industry. Negative results rarely get published; mandates are won by promising spectacular returns. These incentives can bias researchers, even unintentionally, toward experimental designs that are more likely to yield "positive" results. Ignoring implementation issues often yields stronger results. Many high turnover strategies look attractive ignoring trading costs, but are impossible to profitably trade. Inferences drawn from test ignoring frictions cannot be generalized into any real economic insights. To be useful, financial research must incorporate real world concerns to the greatest extent possible.

A. Data and methods

This section describes our data sources and methods. Based on data availability, our sample is January 1972 through December 2017.

We compare the net-of-costs performance of six empirical factor models: the Fama and French (2015) five-factor model, the Fama and French (2018) six-factor model, the Hou, Xue, and Zhang (2015) four-factor model, and the Barillas and Shanken (2018) six-factor model, and versions of the two Fama and French models that employ a cash profitability factor in place of the one based on accruals used in the baseline models. To calculate the transaction costs associated with trading each factor, we follow Novy-Marx and Velikov (2016) and use the effective bid-ask spread estimator of Hasbrouck (2009). The anomalies we use for as test assets in Section 5 are from Chen and Velikov (2019). Finally, we employ the Novy-Marx and Velikov (2016) generalized alpha to evaluate the pricing performance of the models we study.

A.1. Factor models

We compare the performance before and after applying our trading cost measure to six empirical asset-pricing models that are used in recent model-selection studies.

- FF5: The five-factor model of Fama and French (2015) has factors MKT, SMB, HML, RMW, and CMA. MKT denotes the return on the CRSP value-weighted stock market index in excess of the risk-free rate. SMB is a size factor that is long stocks with small market capitalization and short stocks with large market capitalization, HML is a value factor that is long stocks with high book-to-market ratios and short stocks with low book-to-market ratios. RMW is a profitability factor that is long stocks with high (robust) operating profitability and short stocks with low (weak) operating profitability. CMA is an investment factor that is long stocks with low (conservative) investment and short stocks with high (aggressive) investment.

- FF6: The six-factor model of Fama and French (2018) constructed as FF5 augmented by MOM, the momentum factor of Carhart (1997). The momentum factor is re-balanced monthly and is long (short) stocks with the highest (lowest) returns over the prior 12 months excluding the prior month.

- HXZ4: The four-factor q-model of Hou et al. (2015) has factors MKT, ME, IA, and ROE. ME is a size factor that is long stocks with low market capitalization and short stocks with high market capitalization, IA is an investment factor that is long stocks with low investment and short stocks with high investment, and ROE is a factor that is long stocks with high profitability (from the most recent quarterly earnings) and short stocks with low profitability.

- BS6: The six-factor model of Barillas and Shanken (2018). Barillas and Shanken (2018) use a Bayesian technique to compare the factors of FF6, HXZ4, and HML(m), which is the monthly-updated value factor of Asness and Frazzini (2013). They conclude that the dominant model includes six factors: MKT, SMB, IA, ROE, MOM, and HML(m).

- FF5_C: The five-factor model of Fama and French (2015) where the RMW factor is replaced by a profitability factor RMW_C that is constructed similarly, but replaces operating profitability with cash operating profitability following Ball et al. (2016).

- FF6_C: The six-factor model of Fama and French (2015) where the RMW factor is replaced by a profitability factor RMW_C that is constructed similarly, but replaces operating profitability with cash operating profitability following Ball et al. (2016).

The details of the construction of the factors used in the four factor models above plays a crucial role in our analysis. All the characteristic-based factors in FF5 (SMB, HML, RMW, and CMA) are constructed by rebalancing the underlying portfolios once a year, at the end of each June. By contrast, the MOM factor, as well as the characteristic-based factors in HXZ4 (ME, IA, ROE) and the value factor in BS6 (HML(m)), are rebalanced monthly.

We obtain before-costs returns on each Fama-French factor except RMW_C from the Kenneth French's website, along with the returns to the small-cap portfolios formed on size and each of the other relevant characteristics necessary to make HML_S, MOM_S, RMW_S, and CMA_S. We replicate the before-cost returns on the RMW_C factor using CRSP and COMPUSTAT data following Fama and French (2018). We obtain before-cost returns to the HXZ4 factors directly from Chen Xue, and the HML(m) factor returns from AQR's website. We also replicate all these factors, and find correlations with the externally obtained versions ranging from 0.97 to 1.00.

Below we argue that the most important difference in performance stems from the construction of the respective profitability factors in FF5 and HXZ4—RMW and ROE. While RMW rebalances portfolios annually by sorting stocks on a measure of operating profitability derived from the COMPUSTAT annual files, the ROE factors rebalances portfolios monthly by sorting stocks on a measure of return-on-book-equity derived from the COMPUSTAT quarterly files. The monthly updating using quarterly data requires much higher turnover, thereby inducing significantly higher trading costs.

Another major difference between the Fama-French models (FF5 and FF6) and the HXZ4 model comes from the sorting methodology used to build the underlying portfolios used for factor construction. While the Fama-French models use a 2x3 sort on size and one of the other factor characteristics, the HXZ4 model uses a 2x3x3 sort on size, investment, and profitability.

A.2. Transaction costs

To measure trading costs, we follow Novy-Marx and Velikov (2016) and use the effective bid-ask spread estimator from Hasbrouck (2009).[13] The costs are estimated using a Bayesian-Gibbs sampler on a generalized Roll (1984) model of stock price dynamics:

$$V_t = V_{t-1} + \epsilon_t, \tag{2}$$
$$P_t = V_t + cQ_t,$$

where V_t is the "efficient value" of a (log) stock price, P_t is the trade price, $Q_t = +1$ (-1) if the trade was a buy (sell), ϵ_t is a random public shock to the efficient value, and c is the effective one-way transaction cost. It follows from Eq. (2) that:

$$\Delta P_t = c\Delta Q_t + \epsilon_t, \tag{3}$$

Hasbrouck estimates c via Bayesian methods applied to an augmented daily-frequency version of Eq. (3)[14]:

$$\Delta P_t = c\Delta Q_t + \beta r_{m,t} + \epsilon_t, \tag{4}$$

where $r_{m,t}$ denotes the market return on day t.

Hasbrouck (2009) documents that the effective spreads estimated using this procedure achieve a 96.5% correlation with the effective spread estimated using the Trade and Quote (TAQ) intraday data. While this measure does not account for the price impact of large trades, its estimation does not rely on intraday or proprietary data, and is available for the full sample of publicly traded companies.[15] The

[13] This procedure is also used by a growing literature. See, e.g., Detzel and Strauss (2018); Detzel et al. (2019), Novy-Marx and Velikov (2019); Chen and Velikov (2019); Barroso and Detzel (2020)

[14] The Bayesian methods are necessary because Q_t is unobservable.

[15] For additional details on the trading cost estimation, see Hasbrouck (2009) and/or Novy-Marx and Velikov (2016). SAS code to estimate the effective bid-ask spreads is available on Joel

asset-pricing factors implicitly assume market orders, because they transact at exactly the moment of closing. Thus, our net-of-costs factors can be interpreted as the net-of-costs return on the marginal dollar's worth of investment in the literal factors. This is arguably the relevant correction for traders using the benchmark factors as a proxy for the opportunity cost of investing in a given test asset

The Hasbrouck procedure yields a number of missing observations, and these need to be filled in to compute trading-strategy costs. To do so, in each month t we assign to each stock i missing an estimate of effective spread that of the stock j with the closest match in terms of market cap and idiosyncratic volatility, which are the main observable correlates of transactions costs.[16] Specifically, each month we rank all stocks' market values and idiosyncratic volatilities, referring to the ranks as $rankME_i$, and $rankIVOL_i$, respectively. Then, we assign to stock i the estimated spread of the stock j with the smallest value of

$$\sqrt{(rankME_i - rankME_j)^2 + (rankIVOL_i - rankIVOL_j)^2}.$$

For each asset-pricing factor, f, we first replicate the returns to the gross factors as a long/short dollar neutral portfolio. For example, the Fama and French (1993) HML factor is constructed out of the six value-weight portfolios formed on size and

Hasbrouck's website at http://pages.stern.nyu.edu/ jhasbrou/.

[16] See Novy-Marx and Velikov (2016) for more details. Idiosyncratic volatility is defined as the standard deviation of residuals from a CAPM based on the prior 90-days of a given stock's returns.

book-to-market. Its return in month $t+1$ is given by:

$$HML_{t+1} = \frac{1}{2}\left(\underbrace{\sum_{i=1}^{N_{SV,t}} w_{it}^{SV} \times r_{it+1}}_{R_{\text{SMALL,VALUE}}} + \underbrace{\sum_{i=1}^{N_{BV,t}} w_{it}^{BV} \times r_{it+1}}_{R_{\text{BIG,VALUE}}}\right) - \frac{1}{2}\left(\underbrace{\sum_{i=1}^{N_{SG,t}} w_{it}^{SG} \times r_{it+1}}_{R_{\text{SMALL,GROWTH}}} + \underbrace{\sum_{i=1}^{N_{BG,t}} w_{it}^{BG} \times r_{it+1}}_{R_{\text{BIG,GROWTH}}}\right) \tag{5}$$

where r_{it+1} is the return of stock i at time $t+1$, w_{it} is the weight of stock i within one of the four base portfolios, N_t is the number of stocks in each of the four base portfolios, and SV, BV, SG, and BG stands for small-value, big-value, small-growth, and big-growth, respectively. We can express the return in month $t+1$ as a long/short portfolio as:

$$HML_{t+1} = \underbrace{\sum_{i=1}^{N_{SV,t}} \frac{w_{it}^{SV}}{2} \times r_{it+1} + \sum_{i=1}^{N_{BV,t}} \frac{w_{it}^{BV}}{2} \times r_{it+1}}_{\text{Long portfolio}} - \left(\underbrace{\sum_{i=1}^{N_{SG,t}} \frac{w_{it}^{SG}}{2} \times r_{it+1} + \sum_{i=1}^{N_{BG,t}} \frac{w_{it}^{BG}}{2} \times r_{it+1}}_{\text{Short portfolio}}\right) \tag{6}$$

For each asset-pricing factor, f, we compute turnover (TO) and transaction costs (TC) following DeMiguel et al. (2009) as:

$$TO_{long,t} = \frac{1}{2}\sum_{i=1}^{N_t} |w_{i,t} - \tilde{w}_{i,t-1}|, \text{ and} \tag{7}$$

$$TC_{long,t} = (1 + f_{long,t})\sum_{i=1}^{N_t} |w_{i,t} - \tilde{w}_{i,t-1}| \cdot c_{i,t}, \tag{8}$$

where $c_{i,t}$ (r_{it}) is the one-way transaction cost (return) of stock i at time t, $N_t =$

number of stocks at time t, $w_{i,t}$ is the weight of stock i in the leg at time t after rebalancing, $\tilde{w}_{i,t-1} = \frac{w_{i,t-1}(1+r_{it})}{\sum_{j=1}^{N_t} w_{j,t-1}(1+r_{jt})}$ is the weight of stock i in the leg at time t before rebalancing, and $f_{long,t}$ denotes the return on the long leg of f. The transaction costs for the corresponding short leg ($TC_{Short,t}$) are defined similarly. The net-of-costs return on f (f_t^{net}) are:

$$f_t^{net} = f_t^{gross} - TC_{Long,t} - TC_{Short,t}, \tag{9}$$

where f_t^{gross} denotes the return ignoring costs. We estimate the w_{it} of the portfolios by replicating the factors following the respective studies (Specifically, Asness and Frazzini, 2013; Fama and French, 2015, 2018; Hou et al., 2015). In doing so, we obtain necessary accounting data from COMPUSTAT and stock-market data from CRSP.

B. Full-sample simulation results

Table B.1: Bootstrapped net-of-costs maximum squared Sharpe ratios: Full-sample simulations

For each model we consider, the first column presents average net-of-costs maximum squared Sharpe ratios, SR^2, from 100,000 simulation runs. Each run draws, with replacement, 552 random months from our sample period, January 1972 through December 2017. The six columns labeled by model names, 'FF5' to 'FF6C', present the percentage of bootstrap simulations in which the squared Sharpe ratio of the model defined by the row heading is greater than or equal to that of the model defined by the column heading. The last column, 'Max', presents the percentage of bootstrap simulation runs in which the model specified by the row heading has the highest squared Sharpe ratio among all models in the run.

	SR^2	FF5	FF6	HXZ4	BS6	$FF5_C$	$FF6_C$	Max
FF5	0.96		5.4	93.4	64.0	0.0	0.0	0.0
FF6	1.07	100.0		99.0	86.7	3.3	0.0	0.0
HXZ4	0.70	6.6	1.0		0.0	0.1	0.0	0.0
BS6	0.90	36.0	13.3	100.0		1.5	0.2	0.2
$FF5_C$	1.43	100.0	96.7	99.9	98.5		12.0	11.9
$FF6_C$	1.51	100.0	100.0	100.0	99.8	100.0		99.8

C. Anomalies data and detailed results

C.1. Anomalies used as test assets

To compare models based on how well they price other strategies we need a set of test assets. To that end, we use the 120 anomaly signals from Chen and Velikov (2019).[17] In contrast to Chen and Velikov (2019), which examines equal-weighted quintile strategies, we focus on strategies that sort stocks into deciles based on NYSE breakpoints and value-weight stocks within the long and short portfolios. This procedure more reliably produces significant net-of-costs returns. The anomalies are listed in Table C.1.

The anomalies are constructed from the usual data sources. More than half of the predictors focus on Compustat data, and about 30% use purely price data. Most of

[17] These 120 anomalies come from the Chen and Zimmerman (2018) set of 156 cross-sectional return predictors from 115 publications in accounting, economics, and finance journals. Chen and Velikov (2019) remove 34 predictors that do not reliably produce decile sorts and 2 predictors they deem are due to risk to arrive at the final set of 120 anomalies. To the best of our knowledge, this is the most comprehensive set of net-of-cost anomaly returns.

the remainder use analyst forecasts, though several focus on institutional ownership data, trading volume, or specialized data. For further details, please see Chen and Zimmerman (2018).

Table C.1: Anomaly definitions

Acronym	Description	Freq	Author(s)	Publication
AccrAbn	Abnormal Accruals	A	Xie	2001 AR
AccrOper	Percent Operating Accruals	A	Hafzalla et al	2011 AR
AccrPct	Percent Total Accruals	A	Hafzalla et al	2011 AR
Accruals	Accruals	A	Sloan	1996 AR
AdExpGr	Growth in advertising expenses	A	Lou	2014 RFS
AnnounRet	Earnings announcement return	Q	Chan et al	1996 JF
AssetCGr	Change in current operating assets	A	Richardson et al	2005 JAE
InvestAG	Asset Growth	A	Cooper et al	2008 JF
ATurn	Asset Turnover	A	Soliman	2008 AR
BEgrowth	Sustainable Growth	A	Lockwood Prombutr	2010 JFR
BetaSquared	CAPM beta squred	M	Fama MacBeth	1973 JPE
BidAskSpread	Bid-ask spread	M	Amihud Mendelsohn	1986 JFE
BM	Book to market	A	Fama French	1992 JF
BMent	Enterprise component of BM	A	Penman et al	2007 JAR
BMlev	Leverage component of BM	A	Penman et al	2007 JAR
CAPXgr	Change in capex (two years)	A	Anderson Garcia-Feijoo	2006 JF
Cash	Cash to assets	Q	Palazzo	2012 JFE
CF2Price	Cash flow to market	A	Lakonishok et al	1994 JF
CFOper2Price	Operating Cash flows to price	A	Desai et al	2004 AR
DebtFinC	Composite debt issuance	A	Lyandres Sun Zhang	2008 RFS
DeferRev	Deferred Revenue	A	Prakash Sinha	2012 CAR
DepGr	Change in depreciation to gross PPE	A	Holthausen Larcker	1992 JAE
EarnCons	Earnings Consistency	Q	Alwathainani	2009 BAR
EarnSupBig	Earnings surprise of big firms	M	Hou	2007 RFS
EarnSurp	Earnings Surprise	Q	Foster et al	1984 AR
EffFrontier	Efficient frontier index	A	Nguyen Swanson	2009 JFQA
EntMult	Enterprise Multiple	A	Loughran Wellman	2011 JFQA
EP	Earnings-to-Price Ratio	A	Basu	1977 JF
EPforecast	Earnings Forecast	M	Elgers Lo Pfeiffer	2001 AR
EPSDisp	EPS Forecast Dispersion	M	Diether et al	2002 JF
EPSForeLT	Long-term EPS forecast	M	La Porta	1996 JF
EPSrevise	Earnings forecast revisions	M	Chan et al	1996 JF
Eq2AGr	Change in equity to assets	A	Richardson et al	2005 JAE
ExcludExp	Excluded Expenses	M	Doyle et al	2003 RAS
ExtFinNet	Net external financing	A	Bradshaw et al	2006 JAE
FailurePr	Failure probability	Q	Campbell et al	2008 JF
FinLiabGr	Change in financial liabilities	A	Richardson et al	2005 JAE
GIndex	Governance Index	A	Gompers et al	2003 QJE
GM2SaleGr	Gross Margin growth over sales growth	A	Abarbanell Bushee	1998 AR
Herf	Industry concentration (Herfindahl)	A	Hou Robinson	2006 JF

Table C.1: Continued

Acronym	Description	Freq	Author(s)	Publication
High52	52 week high	M	George Hwang	2004 JF
IdioVol	Idiosyncratic risk	M	Ang et al	2006 JF
Illiquid	Amihud's illiquidity	M	Amihud	2002 JFM
IndMom	Industry Momentum	M	Grinblatt Moskowitz	1999 JFE
IndRetBig	Industry return of big firms	M	Hou	2007 RFS
InstOwnSI	Inst own among high short interest	Q	Asquith Pathak Ritter	2005 JFE
IntanBM	Intangible return using BM	A	Daniel Titman	2006 JF
IntanCFP	Intangible return using CFtoP	A	Daniel Titman	2006 JF
IntanEP	Intangible return using EP	A	Daniel Titman	2006 JF
IntanSP	Intangible return using Sale2P	A	Daniel Titman	2006 JF
InvestGr	Change in capital inv (ind adj)	A	Abarbanell Bushee	1998 AR
Invntory	Inventory Growth	A	Thomas Zhang	2002 RAS
InvToRev	Investment to revenue	A	Titman et al	2004 JFQA
KZ	Kaplan Zingales index	A	Lamont et al	2001 RFS
LaborGr	Employment growth	A	Bazdresch Belo Lin	2014 JPE
Leverage	Market leverage	A	Bhandari	1988 JFE
LiabCGr	Change in current operating liabilities	A	Richardson et al	2005 JAE
LTAssetGr	Change in Noncurrent Operating Assets	A	Soliman	2008 AR
LTNOAgr	Growth in Long term net operating assets	A	Fairfield et al	2003 AR
MaxRet	Maximum return over month	M	Bali et al	2010 JF
Mom12m	Momentum (12 month)	M	Jegadeesh Titman	1993 JF
Mom12to7	Intermediate Momentum	M	Novy-Marx	2012 JFE
Mom1813	Momentum-Reversal	M	De Bondt Thaler	1985 JF
Mom1m	Short term reversal	M	Jegedeesh	1989 JF
Mom36m	Long-run reversal	A	De Bondt Thaler	1985 JF
Mom6Jnk	Junk Stock Momentum	M	Avramov et al	2007 JF
Mom6m	Momentum (6 month)	M	Jegadeesh Titman	1993 JF
MomVol	Momentum and Volume	M	Lee Swaminathan	2000 JF
MomYoung	Firm Age - Momentum	M	Zhang	2004 JF
NDebtFin	Net debt financing	A	Bradshaw et al	2006 JAE
NDebtPrice	Net debt to price	A	Penman et al	2007 JAR
NEqFin	Net equity financing	A	Bradshaw et al	2006 JAE
NOA	Net Operating Assets	A	Hirshleifer et al	2004 JAE
NPayYield	Net Payout Yield	A	Boudoukh et al	2007 JF
NWCgr	Change in Net Working Capital	A	Soliman	2008 AR
OperLeverage	Operating Leverage	A	Novy-Marx	2010 ROF
OptVol	Option Volume to Stock Volume	M	Johnson So	2012 JFE
OptVolGr	Option Volume relative to recent average	M	Johnson So	2012 JFE
OrderBacklog	Order backlog	A	Rajgopal et al	2003 RAS
OrgCap	Organizational Capital	A	Eisfeldt Papanikolaou	2013 JF
OScore	O Score	A	Dichev	1998 JFE
PayYield	Payout Yield	A	Boudoukh et al	2007 JF
PensionFunding	Pension Funding Status	A	Franzoni Marin	2006 JF

Table C.1: Continued

Acronym	Description	Freq	Author(s)	Publication
PMGrowth	Change in Profit Margin	A	Soliman	2008 AR
Price	Price	M	Blume Husic	1972 JF
PriceDelay	Price delay	M	Hou Moskowitz	2005 RFS
ProfCash	Cash-based operating profitability	A	Ball et al	2016 JFE
ProfGross	gross profits / total assets	A	Novy-Marx	2013 JFE
ProfitMargin	Profit Margin	A	Soliman	2008 AR
ProfOper	operating profits / book equity	A	Fama French	2006 JFE
RDirtSurp	Real dirty surplus	A	Landsman et al	2011 AR
RealEstate	Real estate holdings	A	Tuzel	2010 RFS
RetConglomerate	Conglomerate return	M	Cohen Lou	2012 JFE
Rev2Price	Sales-to-price	A	Barbee et al	1996 FAJ
RevG2InvG	Sales growth over inventory growth	A	Abarbanell Bushee	1998 AR
RevG2OHG	Sales growth over overhead growth	A	Abarbanell Bushee	1998 AR
RevGrowth	Revenue Growth Rank	A	Lakonishok et al	1994 JF
RevSurprise	Revenue Surprise	Q	Jegadeesh Livnat	2006 JFE
RoA	earnings / assets	Q	Balakrishnan et al	2010 JAE
RoE	net income / book equity	A	Haugen Baker	1996 JFE
Seasonality	Return Seasonality	M	Heston Sadka	2008 JFE
ShareIs1	Share issuance (5 year)	A	Daniel Titman	2006 JF
ShareIs5	Share issuance (1 year)	A	Pontiff Woodgate	2008 JF
VolumeShare	Share Volume	Q	Datar Naik Radcliffe	1998 JFM
ShortInterest	Short Interest	Q	Dechow et al	2001 JFE
Size	Size	A	Banz	1981 JFE
OSmirkNTM	Volatility smirk near the money	M	Xing Zhang Zhao	2010 JFQA
OSmirkCP	Put volatility minus call volatility	M	Yan	2011 JFE
Tangibility	Tangibility	A	Hahn Lee	2009 JF
Tax2E	Taxable income to income	A	Lev Nissim	2004 AR
TaxGr	Change in Taxes	Q	Thomas Zhang	2011 JAR
ATurnGr	Change in Asset Turnover	A	Soliman	2008 AR
TurnovVol	Share turnover volatility	M	Chordia et al	2001 JFE
CF2Pvar	Cash-flow to price variance	A	Haugen Baker	1996 JFE
Volume2Mkt	Volume to market equity	M	Haugen Baker	1996 JFE
VolumeDol	Past trading volume	M	Brennan et al	1998 JFE
VolumeSD	Volume Variance	M	Chordia et al	2001 JFE
VolumeTrend	Volume Trend	M	Haugen Baker	1996 JFE
ZeroTrade	Days with zero trades	M	Liu	2006 JFE
ZScore	Altman Z-Score	A	Dichev	1998 JFE

C.2. Detailed anomaly squared Sharpe ratio improvement results

Table C.2 provides summary statistics for the data depicted in Figure 3. It reports the mean, standard deviation, and select percentiles from the distribution of $\%\Delta\text{SR}(M, A)$ across the 120 anomalies for each model, and tells a similar story. While ignoring costs the HXZ4 and BS6 perform somewhat better than the standard Fama and French models pricing anomalies, this advantage disappears after accounting for the cost of trading, and the Fama and French variations that employ cash profitability explain far more of anomaly performance an investor could actually capture than any of the other models.

Table C.2: Anomaly frontier expansion summary statistics.

For each model, M, specified by the row heading, and each of the 120 anomalies, A, described in Section 5, we compute the ex-post maximum squared Sharpe ratio attainable from the model's factors, $\text{SR}^2(M)$, and that attainable from the model's factors and the anomaly, $\text{SR}^2(M, A)$. We then compute the squared Sharpe ratio improvement, $\%\Delta\text{SR}(M, A) = \text{SR}^2(M, A)/\text{SR}^2(M) - 1$. For each model we consider, this table reports the mean, standard deviation, 'Std', and select percentiles from the distribution of $\%\Delta\text{SR}(M, A)$ across the 120 anomalies. Panels A and B, respectively, present results using before-costs and net-of-costs returns.

Panel A: Gross

			Percentiles						
Model	Mean	Std	25	50	75	90	95	99	100
FF5	11.9	16.8	1.4	6.2	16.3	30.0	43.1	83.3	123.6
FF6	7.4	10.9	0.7	3.6	9.5	19.9	29.2	50.9	83.3
HXZ4	6.3	10.1	0.5	2.1	7.7	18.8	25.5	51.9	68.3
BS6	7.1	11.0	0.4	1.8	8.6	23.2	33.5	46.3	55.1
FF5_C	6.2	10.6	0.7	2.7	7.9	15.6	25.6	55.6	86.2
FF6_C	4.8	8.1	0.4	2.2	5.8	11.2	20.5	38.5	67.7

Panel B: Net

			Percentiles						
Model	Mean	Std	25	50	75	90	95	99	100
FF5	5.7	9.8	0.0	1.2	7.9	18.1	20.6	54.1	64.7
FF6	4.7	8.4	0.0	0.4	6.1	15.1	18.2	49.0	54.9
HXZ4	6.8	11.4	0.0	1.8	9.1	18.7	31.2	57.0	59.4
BS6	6.2	10.7	0.0	1.0	8.4	17.8	22.8	51.9	74.7
FF5_C	2.6	5.0	0.0	0.2	2.0	9.2	14.4	24.0	29.8
FF6_C	2.3	4.6	0.0	0.1	2.4	8.0	12.3	22.1	28.7

Table C.3: Gross squared Sharpe ratio improvement: Detailed results

Acronym	FF5	FF6	HXZ4	BS6	$FF5_C$	$FF6_C$
AbnormalAccruals	32.02	22.14	20.03	17.68	14.08	10.96
Accruals	22.56	12.42	13.26	5.76	7.78	4.92
AnnouncementReturn	44.07	30.13	25.50	23.07	33.72	25.99
AssetGrowth	0.12	0.12	0.00	0.29	0.29	0.00
AssetTurnover	13.10	8.51	6.10	17.51	4.35	3.27
BetaSquared	0.18	0.72	0.90	3.56	0.48	2.31
BidAskSpread	1.81	6.77	4.79	5.35	0.01	7.29
BM	9.22	0.71	0.05	7.00	3.03	0.09
BPEBM	0.12	0.22	0.00	0.03	1.00	1.03
Cash	39.59	26.26	13.30	23.23	26.28	20.48
CBOperProf	59.11	37.06	28.66	37.42	18.50	13.15
CF	1.84	4.69	9.26	0.38	0.77	2.39
cfp	1.74	2.80	0.73	1.01	0.02	0.10
ChAssetTurnover	9.40	3.02	0.88	0.09	3.27	1.03
ChEQ	0.34	0.06	0.31	0.00	0.27	0.00
ChInv	17.80	9.59	7.75	3.84	10.97	7.14
ChInvIA	1.69	0.44	0.99	0.05	0.03	0.02
ChNCOA	12.68	4.25	1.11	0.30	4.44	1.47
ChNWC	12.65	8.99	5.93	5.47	2.65	2.34
ChPM	2.50	1.20	0.19	0.06	0.22	0.07
ChTax	18.72	3.91	0.07	0.72	9.71	2.55
CompositeDebtIssuance	6.18	5.27	4.95	3.42	4.68	4.45
DelCOA	6.05	2.45	3.21	7.35	3.20	1.54
DelCOL	42.10	24.60	23.62	32.43	14.87	9.92
DelDRC	0.10	0.11	0.79	0.22	0.07	0.06
DelEqu	0.09	0.06	0.00	0.03	0.00	0.00
DelFINL	5.54	4.63	1.74	1.98	1.56	1.60
DolVol	2.61	0.05	0.67	1.65	1.43	0.01
EarningsConsistency	27.55	17.72	15.00	9.63	16.84	12.58
EarningsSurprise	1.76	0.08	0.44	0.44	0.01	0.92
EarnSupBig	5.42	1.69	0.48	0.02	2.32	0.80
EBM	0.08	0.81	0.51	0.00	1.24	2.19
EntMult	3.85	6.50	7.80	0.02	1.12	2.68
EP	2.84	1.06	4.25	0.61	1.73	0.81
ExclExp	19.68	11.27	6.86	5.61	10.78	6.71
FailureProbability	16.68	2.94	1.06	5.31	3.56	0.04
fgr5yrLag	9.73	13.16	9.10	26.81	5.73	8.06
FirmAgeMom	66.01	28.41	23.13	16.75	42.42	24.59
ForecastDispersion	16.48	7.15	0.29	0.95	6.40	2.90
FR	0.38	0.56	1.97	1.09	0.07	0.01
Frontier	0.79	0.72	3.22	0.00	0.13	0.73
G	16.16	11.13	4.56	8.77	12.90	9.47

Table C.3: Continued

Acronym	FF5	FF6	HXZ4	BS6	$FF5_C$	$FF6_C$
GP	23.55	14.34	12.00	30.23	5.30	3.59
GrAdExp	3.22	1.11	1.05	1.27	2.66	1.29
grcapx	0.15	2.20	1.95	6.12	0.01	0.93
GrGMToGrSales	21.09	11.94	10.97	15.09	7.00	4.47
GrLTNOA	14.50	10.38	19.22	16.97	4.05	3.56
GrSaleToGrInv	2.55	0.40	0.20	0.02	1.04	0.16
GrSaleToGrOverhead	0.72	0.27	0.00	0.01	0.35	0.14
Herf	19.28	14.30	8.57	18.82	6.48	5.81
High52	3.16	10.04	0.99	6.52	0.97	6.99
hire	5.45	7.47	6.36	7.56	3.14	4.72
IdioRisk	15.72	4.35	1.51	0.16	3.68	0.67
Illiquidity	2.98	0.30	2.42	0.42	4.50	1.52
IndMom	1.10	7.95	0.30	2.56	0.44	4.65
IndRetBig	123.63	83.29	68.28	55.11	86.22	67.70
IntanBM	1.53	0.71	2.71	0.06	0.02	2.43
IntanCFP	0.18	1.92	5.88	0.67	0.00	1.73
IntanEP	6.31	0.49	0.00	1.74	1.23	0.00
IntanSP	13.84	1.14	0.03	1.01	4.36	0.10
IntMom	51.82	17.47	9.15	5.01	24.86	9.67
Investment	13.04	3.87	2.85	2.60	8.70	3.59
IOShortInterest	20.70	16.52	15.13	17.38	12.84	11.42
KZ	6.94	9.99	7.72	12.02	2.14	4.21
Leverage	20.31	9.33	6.41	32.52	5.06	2.29
MaxRet	9.32	3.34	1.56	0.05	2.28	0.69
MeanRankRevGrowth	1.39	0.23	0.00	0.06	0.98	0.27
Mom12m	34.63	3.57	2.64	1.62	16.34	1.83
Mom18m13m	2.25	5.42	9.57	3.48	3.90	6.59
Mom1m	0.77	5.57	2.32	1.03	0.99	4.41
Mom36m	9.55	7.10	2.77	3.20	3.32	2.88
Mom6m	19.97	0.02	1.42	0.24	10.83	0.18
Mom6mJunk	9.76	0.00	0.14	0.09	5.36	0.10
MomSeas	27.97	14.87	14.08	9.41	12.98	7.99
MomVol	3.42	7.24	0.72	3.23	1.13	5.09
NetDebtFinance	4.58	3.51	0.44	0.53	1.68	1.42
NetDebtPrice	15.26	8.26	6.83	14.56	11.16	7.41
NetEquityFinance	0.01	0.06	0.00	0.41	0.49	0.32
NetPayoutYield	1.31	2.36	3.02	0.09	0.01	0.19
NOA	14.66	8.53	3.82	0.69	14.20	10.00

Table C.3: Continued

Acronym	FF5	FF6	HXZ4	BS6	$FF5_C$	$FF6_C$
OperProf	0.46	0.12	0.00	0.23	1.96	1.96
OPLeverage	8.48	4.81	4.42	10.24	0.94	0.50
OptionVolume1	0.19	2.10	3.50	0.90	0.30	1.99
OptionVolume2	20.35	29.93	36.49	34.46	10.34	17.06
OrderBacklog	0.88	1.79	1.61	2.83	0.64	1.30
OrgCap	11.98	8.57	6.09	17.50	4.43	3.76
OScore	15.31	16.50	11.62	10.55	7.79	9.26
PayoutYield	0.10	0.06	0.62	1.46	0.70	0.54
pchdepr	6.43	4.77	6.31	5.52	3.39	2.93
PctAcc	1.71	0.54	0.31	0.55	0.03	0.14
PctTotAcc	0.06	0.88	0.51	0.43	0.03	0.56
PM	7.30	5.51	1.18	0.84	1.07	0.98
Price	6.37	1.39	1.49	0.91	1.52	1.92
PriceDelay	1.38	1.79	1.99	1.91	1.08	1.43
Profitability	12.65	2.14	0.69	0.35	0.90	0.06
RDS	2.57	1.96	0.64	1.53	1.09	0.94
realestate	45.80	34.75	34.72	42.53	27.57	22.46
retConglomerate	2.30	0.52	0.47	0.44	1.31	0.41
REV6	5.59	0.07	0.68	0.09	1.17	0.12
RevenueSurprise	0.26	0.00	1.19	0.95	0.14	0.00
RoE	0.24	0.23	0.06	0.01	0.35	0.26
sfe	34.34	23.60	21.84	39.65	24.89	20.00
ShareIss1Y	16.55	8.02	9.03	3.03	9.91	5.77
ShareIss5Y	1.02	0.26	0.52	0.71	0.37	0.09
ShareVol	1.13	1.01	0.99	3.66	2.91	2.61
ShortInterest	10.96	8.59	8.97	5.90	4.12	3.69
Size	0.30	0.38	0.63	0.03	0.12	0.04
Skew1	42.05	33.76	44.84	40.82	27.15	21.77
SmileSlope	22.86	24.15	25.59	24.83	19.25	20.52
SP	10.64	1.36	0.00	8.76	8.19	2.14
stdturn	2.84	3.56	1.95	4.38	3.86	4.46
tang	11.32	8.52	7.37	8.47	9.24	7.85
Tax	2.64	1.39	2.42	0.98	0.02	0.00
VarCF	2.56	1.57	0.12	0.84	1.28	0.88
VolMkt	1.37	5.11	2.21	7.46	2.43	5.49
VolSD	11.68	10.01	4.94	0.90	7.97	7.51
VolumeTrend	16.96	14.09	18.42	9.63	9.06	8.61
XFIN	0.03	0.05	0.15	0.08	0.56	0.44
zerotrade	0.00	0.04	0.05	1.64	0.35	0.13
ZScore	6.83	2.85	0.10	7.25	4.97	2.68

Table C.4: Net squared Sharpe ratio improvement: Detailed results

Acronym	FF5	FF6	HXZ4	BS6	$FF5_C$	$FF6_C$
AbnormalAccruals	19.66	17.59	18.09	18.24	7.56	7.08
Accruals	17.67	13.84	18.95	15.64	5.79	4.84
AnnouncementReturn	19.23	15.96	12.99	11.85	14.88	13.28
AssetGrowth	0.17	0.00	7.57	3.38	0.35	0.11
AssetTurnover	13.48	11.56	10.55	17.45	4.42	4.04
BetaSquared	0.15	0.00	0.07	0.17	0.00	0.00
BidAskSpread	0.00	0.00	0.00	0.00	0.00	0.00
BM	2.80	0.38	1.00	0.00	0.32	0.00
BPEBM	0.00	0.00	0.00	0.00	0.00	0.00
Cash	19.29	16.60	3.89	9.07	12.85	11.89
CBOperProf	64.71	54.87	55.98	74.74	21.54	19.32
CF	1.43	3.01	22.14	9.74	0.49	1.14
cfp	1.16	1.83	7.85	3.49	0.00	0.00
ChAssetTurnover	5.95	3.40	3.79	2.00	1.66	0.92
ChEQ	0.84	0.19	8.68	6.03	0.65	0.26
ChInv	18.40	14.42	27.96	20.16	11.36	9.76
ChInvIA	1.10	0.55	6.31	3.20	0.00	0.00
ChNCOA	7.46	4.30	5.88	4.16	1.91	1.04
ChNWC	3.72	3.42	1.92	3.06	0.09	0.12
ChPM	0.00	0.00	0.00	0.00	0.00	0.00
ChTax	3.26	0.50	0.00	0.00	1.07	0.13
CompositeDebtIssuance	1.86	2.05	0.78	0.91	1.56	1.71
DelCOA	0.13	0.01	0.00	0.00	0.00	0.00
DelCOL	26.52	21.19	13.08	20.03	7.75	6.53
DelDRC	0.00	0.00	0.12	0.00	0.00	0.00
DelEqu	0.00	0.00	0.00	0.00	0.00	0.00
DelFINL	2.56	2.67	7.73	6.63	0.34	0.42
DolVol	0.00	0.00	0.00	0.00	0.00	0.00
EarningsConsistency	16.55	13.98	20.98	14.73	9.23	8.32
EarningsSurprise	0.00	0.00	0.00	0.00	0.00	0.00
EarnSupBig	0.00	0.00	0.00	0.00	0.00	0.00
EBM	0.00	0.00	0.00	0.00	0.00	0.09
EntMult	5.55	8.09	34.97	19.06	1.95	3.07
EP	2.25	1.43	17.33	4.50	1.26	0.92
ExclExp	0.00	0.00	0.00	0.00	0.00	0.00
FailureProbability	7.86	4.88	0.43	5.66	0.66	0.15
fgr5yrLag	1.17	1.27	0.00	0.00	0.52	0.57
FirmAgeMom	15.90	7.35	8.68	7.79	10.79	7.25
ForecastDispersion	3.20	1.70	0.14	1.25	0.74	0.40
FR	0.15	0.21	2.02	1.23	0.00	0.00
Frontier	0.00	0.00	4.35	0.10	0.00	0.00
G	17.51	14.82	1.66	2.99	14.24	12.31

Table C.4: Continued

Acronym	FF5	FF6	HXZ4	BS6	$FF5_C$	$FF6_C$
GP	24.28	20.22	12.44	27.61	5.52	4.83
GrAdExp	0.00	0.00	0.00	0.00	0.01	0.00
grcapx	0.00	0.00	1.32	0.00	0.00	0.00
GrGMToGrSales	13.61	10.65	9.52	13.26	3.46	2.82
GrLTNOA	8.00	7.27	13.87	15.72	1.60	1.62
GrSaleToGrInv	0.77	0.19	0.76	0.14	0.19	0.03
GrSaleToGrOverhead	0.21	0.10	1.08	0.65	0.06	0.03
Herf	19.02	17.48	8.13	16.58	6.62	6.57
High52	0.39	0.00	0.00	0.00	0.02	0.00
hire	0.00	0.00	0.00	0.00	0.00	0.00
IdioRisk	0.00	0.00	0.00	0.00	0.00	0.00
Illiquidity	0.10	0.00	12.74	0.23	1.09	0.36
IndMom	0.00	0.00	0.00	0.00	0.00	0.00
IndRetBig	0.00	0.00	0.00	0.00	0.00	0.00
IntanBM	0.08	0.00	1.83	0.00	0.00	0.00
IntanCFP	0.00	0.01	9.55	2.34	0.00	0.05
IntanEP	1.59	0.15	0.00	0.00	0.00	0.00
IntanSP	8.07	2.59	0.00	0.82	1.78	0.31
IntMom	6.42	0.62	0.02	0.00	1.55	0.01
Investment	8.44	4.66	5.01	3.29	5.63	3.81
IOShortInterest	12.97	12.27	10.29	12.83	7.90	7.68
KZ	7.02	9.45	15.74	16.02	2.02	3.11
Leverage	18.13	13.16	1.85	14.52	4.23	3.08
MaxRet	0.00	0.00	0.00	0.00	0.00	0.00
MeanRankRevGrowth	2.50	1.41	4.52	2.35	1.73	1.18
Mom12m	16.70	7.52	5.75	11.70	7.15	3.73
Mom18m13m	0.00	0.00	0.00	0.00	0.00	0.00
Mom1m	0.00	0.00	0.00	0.00	0.00	0.00
Mom36m	2.88	2.70	0.17	1.43	0.37	0.39
Mom6m	0.58	0.00	0.00	0.00	0.16	0.00
Mom6mJunk	0.00	0.00	0.00	0.00	0.00	0.00
MomSeas	0.00	0.00	0.00	0.00	0.00	0.00
MomVol	0.00	0.00	0.00	0.00	0.00	0.00
NetDebtFinance	2.52	2.38	5.17	3.88	0.61	0.60
NetDebtPrice	10.30	7.95	5.69	9.62	7.33	6.27
NetEquityFinance	0.28	0.33	5.38	3.56	0.00	0.00
NetPayoutYield	3.91	4.86	30.94	21.60	0.43	0.68
NOA	18.04	15.11	23.44	17.31	16.42	15.19

Table C.4: Continued

Acronym	FF5	FF6	HXZ4	BS6	$FF5_C$	$FF6_C$
OperProf	1.12	0.75	4.39	4.26	0.00	0.00
OPLeverage	13.23	10.77	13.89	17.12	2.47	2.07
OptionVolume1	0.00	0.00	0.04	0.00	0.00	0.00
OptionVolume2	0.00	0.00	0.00	0.00	0.00	0.00
OrderBacklog	0.02	0.20	0.00	0.03	0.02	0.12
OrgCap	16.86	15.05	13.88	19.77	7.04	6.72
OScore	10.13	11.64	10.02	17.30	4.81	5.57
PayoutYield	0.06	0.05	4.80	0.60	0.69	0.64
pchdepr	1.23	1.20	2.04	2.19	0.45	0.48
PctAcc	0.78	0.39	0.55	1.13	0.00	0.00
PctTotAcc	0.00	0.00	0.00	0.00	0.00	0.00
PM	4.90	4.61	3.12	5.73	0.40	0.42
Price	1.18	0.00	0.00	0.00	0.03	0.00
PriceDelay	0.00	0.00	0.00	0.00	0.00	0.00
Profitability	2.52	0.37	0.00	0.40	0.00	0.00
RDS	0.07	0.06	0.00	0.00	0.00	0.00
realestate	49.61	46.47	37.23	42.10	29.78	28.72
retConglomerate	0.00	0.00	0.00	0.00	0.00	0.00
REV6	0.00	0.00	0.00	0.00	0.00	0.00
RevenueSurprise	0.00	0.00	0.00	0.00	0.00	0.00
RoE	0.01	0.02	0.34	1.18	0.00	0.00
sfe	21.44	19.04	13.51	23.92	16.34	15.63
ShareIss1Y	24.36	18.87	59.35	40.54	14.52	12.30
ShareIss5Y	3.57	2.48	18.50	7.81	1.69	1.27
ShareVol	0.00	0.00	0.00	0.00	0.00	0.00
ShortInterest	4.04	3.82	4.13	4.89	1.07	1.07
Size	0.00	0.00	4.95	0.00	0.00	0.00
Skew1	0.00	0.00	0.00	0.00	0.00	0.00
SmileSlope	0.00	0.00	0.00	0.00	0.00	0.00
SP	2.48	0.36	4.62	0.00	2.05	0.73
stdturn	0.00	0.00	0.00	0.04	0.03	0.11
tang	4.90	4.66	4.47	5.66	4.03	3.97
Tax	2.49	1.93	10.07	7.79	0.01	0.00
VarCF	0.78	0.64	0.24	0.07	0.30	0.24
VolMkt	0.00	0.00	0.00	0.00	0.00	0.00
VolSD	13.32	12.87	31.40	13.74	9.10	9.14
VolumeTrend	18.03	17.03	55.77	34.99	9.94	9.85
XFIN	0.53	0.54	4.34	5.88	0.00	0.00
zerotrade	0.00	0.00	3.89	0.33	0.00	0.00
ZScore	1.85	1.03	0.23	0.00	1.42	0.98

D. Extra cost-mitigation results

Table D.1: Spanning regressions using cost-mitigated factor models

Panel A of this table presents intercepts, with heteroskedasticity-robust t-statistics in brackets below, from regressions of the form: $MVE_{M0\cup M1,t} = \alpha + \beta \cdot MVE_{M0,t} + \epsilon_t$, where MVE_{M0} and $MVE_{M0\cup M1}$ denote the excess return on the ex-post mean-variance efficient portfolio consisting of the factors in models $M0$ and the union of $M0$ and $M1$, respectively. To measure the economic significance of the intercepts, Panel B presents squared Sharpe ratio improvements, $\%\Delta\text{SR}^2(M0, M1) = \text{SR}^2(M0, M1)/\text{SR}^2(M0) - 1$. The model ($M1$) is specified by the column heading, and the ($M0$) by the row heading. All models use the cost-mitigated versions of the factors defined in Section 6.1. The sample period is January 1972 to December 2017.

	Gross						Net					
	FF5	FF6	HXZ4	BS6	FF5$_\text{C}$	FF6$_\text{C}$	FF5	FF6	HXZ4	BS6	FF5$_\text{C}$	FF6$_\text{C}$
Panel A: Alphas												
FF5		1.38	2.08	1.87	0.96	1.71		0.92	1.01	1.30	0.80	1.35
		[4.27]	[6.66]	[9.11]	[4.24]	[5.74]		[3.25]	[3.93]	[4.62]	[3.68]	[4.74]
FF6	0.00		0.98	1.34	0.65	0.65	0.00		0.29	0.50	0.57	0.57
	[0.00]		[4.59]	[6.74]	[3.25]	[3.25]	[0.00]		[2.10]	[3.00]	[3.04]	[3.04]
HXZ4	0.26	0.35		0.76	0.56	0.72	0.64	0.80		0.95	1.05	1.26
	[2.34]	[2.51]		[4.37]	[3.15]	[3.52]	[3.47]	[3.75]		[3.72]	[4.20]	[4.63]
BS6	0.62	0.62	0.10		0.77	0.77	0.23	0.23	0.00		0.63	0.63
	[4.32]	[4.32]	[1.59]		[5.23]	[5.23]	[2.15]	[2.15]	[0.00]		[3.34]	[3.34]
FF5$_\text{C}$	0.07	0.83	1.39	1.50		0.91	0.00	0.58	0.44	0.78		0.58
	[1.19]	[3.40]	[5.31]	[8.65]		[3.38]	[0.00]	[2.57]	[2.70]	[3.85]		[2.57]
FF6$_\text{C}$	0.03	0.03	0.68	1.13	0.00		0.00	0.00	0.09	0.28	0.00	
	[0.68]	[0.68]	[3.71]	[6.26]	[0.00]		[0.00]	[0.00]	[1.21]	[2.41]	[0.00]	
Panel B: $\%\Delta SR^2(M0, M1)$												
FF5		45.9	81.3	160.9	40.9	73.8		34.3	38.6	59.7	39.5	64.7
FF6	0.0		26.7	78.9	19.1	19.1	0.0		8.4	18.9	22.6	22.6
HXZ4	6.0	8.1		24.6	13.2	16.8	26.6	33.0		34.6	45.2	54.0
BS6	25.7	25.7	2.7		35.1	35.1	8.4	8.4	0.0		24.1	24.1
FF5$_\text{C}$	2.1	25.9	40.3	103.2		25.1	0.0	18.1	14.0	31.1		18.1
FF6$_\text{C}$	0.6	0.6	15.7	62.4	0.0		0.0	0.0	2.4	11.0	0.0	

Table D.2: Ex post mean-variance efficient portfolios of models using cost-mitigated factors

For each of the factor models specified by the row headings, this table presents the weights of each factor in the portfolio consisting of the model's factors that maximizes the ex-post squared Sharpe ratio, SR^2. The factors are all versions that use the 20% banding cost-mitigation strategies described in Section 6.1. Panel A uses factor returns that ignore transaction costs. Panel B uses factor returns that account for transaction costs. The sample period is January 1972 to December 2017.

Panel A: Gross												
	Weights											
	MKT	SMB	HML	RMW	CMA	MOM	ME	ROE	IA	HML(m)	RMW_C	SR^2
FF5	16	9	1	34	41							1.26
FF6	16	8	5	26	33	12						1.83
HXZ4	15						12	29	43			2.15
BS6	14	12				11		25	18	19		2.61
$FF5_C$	16	12	-1		30						41	1.73
$FF6_C$	16	11	3		27	9					34	2.17

Panel B: Net												
	Weights											
	MKT	SMB	HML	RMW	CMA	MOM	ME	ROE	IA	HML(m)	RMW_C	SR^2
FF5	18	10	6	32	33							0.88
FF6	17	9	9	26	28	11						1.19
HXZ4	20						8	27	44			0.97
BS6	17	12				13		22	14	22		1.30
$FF5_C$	17	13	3		24						42	1.23
$FF6_C$	17	12	6		21	8					35	1.45

Table D.3: Ex post mean-variance efficient portfolios of factor models accounting for transaction costs incorporating trading diversification

For each of the factor models specified by the row headings, this table presents the weights of each factor in the portfolio consisting of the model's factors that maximizes the ex-post squared Sharpe ratio, SR^2. Portfolio returns account for transaction costs incorporating the trading diversification effect described in Section 6.2. The sample period is January 1972 to December 2017.

	Weights											
	MKT	SMB	HML	RMW	CMA	MOM	ME	ROE	IA	HML(m)	RMW_C	SR^2
FF5	19	11	7	31	31							0.96
FF6	19	10	9	28	27	7						1.05
HXZ4	18						14	40	28			0.79
BS6	16	11				11		24	20	18		1.03
$FF5_C$	17	14	4		20						45	1.41
$FF6_C$	17	13	6		18	4					41	1.47

References

Asness, C. S. and Frazzini, A. (2013), The devil in HML's details, *Journal of Portfolio Management* 39, 49–68.

Ball, R., Gerakos, J., Linnainmaa, J. T. and Nikolaev, V. (2016), Accruals, cash flows, and operating profitability in the cross section of stock returns, *Journal of Financial Economics* 121, 28–45.

Barillas, F., Kan, R., Robotti, C. and Shanken, J. (2019), Model comparison with sharpe ratios. Working paper.

Barillas, F. and Shanken, J. (2017), Which alpha?, *Review of Financial Studies* 30, 1316–1338.

Barillas, F. and Shanken, J. (2018), Comparing asset pricing models, *Journal of Finance* 73, 715–754.

Barroso, P. and Detzel, A. (2020), Do limits to arbitrage explain the benefits of volatility-managed portfolios?, *Journal of Financial Economics* (forthcoming).

Berk, J. B. and van Binsbergen, J. H. (2015), Measuring skill in the mutual fund industry, *Journal of Financial Economics* 118, 1 – 20.

Bryzgalova, S., Huang, J. and Julliard, C. (2020), Bayesian solutions for the factor zoo: We just ran two quadrillion models. Working paper.

Carhart, M. M. (1997), On persistence in mutual fund performance, *Journal of Finance* 52, 57–82.

Chen, A. and Velikov, M. (2019), Accounting for the anomaly zoo: a trading cost perspective, SSRN working paper.

Chen, A. and Zimmerman, T. (2018), Publication bias and the cross-section of stock returns, SSRN working paper.

Daniel, K., Hirshleifer, D. and Sun, L. (2019), Short- and long-horizon behavioral factors, *The Review of Financial Studies* (forthcoming.

DeMiguel, V., Garlappi, L. and Uppal, R. (2009), Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?, *Review of Financial Studies* 22, 1915–1953.

DeMiguel, V., Martín-Utrera, A., Nogales, F. J. and Uppal, R. (2020), A transaction-cost perspective on the multitude of firm characteristics, *Review of Financial Studies* 33, 2180–2222.

Detzel, A., Schaberl, P. and Strauss, J. (2019), Expected versus ex post profitability in the cross-section of industry returns, *Financial Management* 48, 505–536.

Detzel, A. and Strauss, J. (2018), Combination return forecasts and portfolio allocation with the cross-section of book-to-market ratios, *Review of Finance* 22, 1949–1973.

Fama, E. F. (1991), Efficient capital markets: II, *Journal of Finance* 46, 1575–1617.

Fama, E. F. and French, K. R. (1993), Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Fama, E. F. and French, K. R. (2015), A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.

Fama, E. F. and French, K. R. (2018), Choosing factors, *Journal of Financial Economics* 128, 234–252.

Feng, G., Giglio, S. and Xiu, D. (2020), Taming the factor zoo: A test of new factors, *Journal of Finance* 75, 1327–1370.

Ferson, W. E., Siegel, A. F. and Wang, J. L. (2019), Asymptotic variances for tests of portfolio efficiency and factor model comparisons with conditioning information. Working paper.

Gibbons, M. R., Ross, S. and Shanken, J. (1989), A test of the efficiency of a given portfolio, *Econometrica* 57, 1121–52.

Harvey, C. R. (2017), Presidential address: The scientific outlook in financial economics, *The Journal of Finance* 72, 1399–1440.

Harvey, C. R., Liu, Y. and Zhu, H. (2016), ...and the Cross-Section of Expected Returns, *The Review of Financial Studies* 29, 5–68.

Hasbrouck, J. (2009), Trading costs and returns for US equities: estimating effective costs from daily data, *Journal of Finance* 64, 1445–1477.

Hou, K., Mo, H., Xue, C. and Zhang, L. (2019), Which Factors?*, *Review of Finance* 23, 1–35.

Hou, K., Xue, C. and Zhang, L. (2015), Digesting anomalies: An investment approach, *Review of Financial Studies* 28, 650–705.

Kan, R., Wang, X. and Zheng, X. (2019), In-sample and out-of-sample sharpe ratios of multi-factor asset pricing models. Working paper.

Kozak, S., Nagel, S. and Santosh, S. (2019), Shrinking the cross-section, *Journal of Financial Economics* (forthcoming).

Lesmond, D. A., Schill, M. J. and Zhou, C. (2004), The illusory nature of momentum profits, *Journal of Financial Economics* 71, 349 – 380.

Lewellen, J., Nagel, S. and Shanken, J. (2010), A skeptical appraisal of asset pricing tests, *Journal of Financial Economics* 96, 175–194.

Li, S. A., DeMiguel, V. and Martín-Utrera, A. (2020), Which factors with price-impact costs? LBS Working paper.

Novy-Marx, R. and Velikov, M. (2016), A taxonomy of anomalies and their trading costs, *Review of Financial Studies* 29, 104–147.

Novy-Marx, R. and Velikov, M. (2019), Comparing cost-mitigation techniques, *Financial Analyst Journal* 75, 85 – 102.

Novy-Marx, R. and Velikov, M. (2020), Betting against betting against beta, *Journal of Financial Economics* (forthcoming).

Roll, R. (1984), A simple implicit measure of the effective bid-ask spread in an efficient market, *The Journal of Finance* 39, 1127–1139.

Ross, S. A. (1976), The arbitrage theory of capital asset pricing, *Journal of Economic Theory* 13, 341–360.

Stambaugh, R. F. and Yuan, Y. (2017), Mispricing factors, *Review of Financial Studies* 30, 1270–1315.