# Detecting Anomalies: The Relevance and Power of Standard Asset Pricing Tests*

Malcolm Baker, Patrick Luo, Ryan Taliaferro

July 12, 2018

**Abstract**

The two standard approaches for identifying capital market anomalies are cross-sectional coefficient tests, in the spirit of Fama and MacBeth (1973), and time-series intercept tests, in the spirit of Jensen (1968). A new signal can pass the first test, which we label a "score anomaly," it can pass the second test as a "factor anomaly," or it can pass both. We demonstrate the relevance of each to a mean-variance optimizing investor facing simple transaction costs that are constant across stocks. For a risk-neutral investor facing transaction costs, only score anomalies are relevant. For a risk-averse investor facing no transaction costs, only factor anomalies are relevant. In the more general case of risk aversion and transaction costs, both tests matter. In extensions, we derive modified versions of the basic tests that net out anomaly execution costs for situations where the investor faces capital constraints, a multi-period portfolio choice problem, or transaction costs that vary across stocks. Next, we measure the econometric power of the two tests. The relative power of time-series factor tests falls with the in-sample Sharpe ratio of the incumbent factor model, as in Shanken (1992). New factor anomalies can be successively harder to detect, leading to a lower natural limit on the number of anomalies that can be identified by time series tests. Meanwhile, for an investor facing transaction costs, where score anomalies are also applicable, there can be a higher natural limit on the number of anomalies that can be statistically validated as relevant.

---

*Contact: mbaker@hbs.edu. The authors thank participants at seminars at London Business School and the London School of Economics for helpful comments. Please do not quote without permission. The authors would like to thank Owen Lamont for helpful comments. Malcolm Baker is at the National Bureau of Economic Research and serves as a consultant to Acadian Asset Management, Patrick Luo was previously at Harvard Business School and is a data scientist at Farallon Capital Management, and Ryan Taliaferro is senior vice president at Acadian Asset Management. In addition, Malcolm Baker and Patrick Luo acknowledge support from the Division of Research at Harvard Business School. 

# 1 Introduction

Capital market anomalies fall into two categories. The first is when a security characteristic or signal, such as the ratio of a firm's book to market value or its recent change in share price, predicts future returns but is otherwise not obviously related to risk. The second is when a signal, such as market beta, is theoretically and empirically connected to portfolio risk but nonetheless does not predict returns.

Anomalies attract both academic and practical interest. The focus of academic asset pricing is rationalizing seemingly anomalous predictability, by redefining or expanding what an investor considers to be risk. The focus of academic behavioral finance is uncovering a parsimonious set of psychological or sociological biases and institutional frictions that break the standard link between risk and return. And, the focus of practical investment management is delivering products to investors with the aim of delivering positive risk-adjusted returns.

In this paper, our two-part goal is to examine the practical relevance and statistical power of standard asset pricing tests in identifying anomalies. The protagonist we have in mind is an investor who is seeking to make sensible security selection decisions using historical data as a guide. This is in the spirit of Brennan, Schwartz, and Lagnado (1997) and Campbell and Viceira (1999), who examine the consequences of return predictability for portfolio choice in partial equilibrium. While those papers focus on asset allocation across stocks and bonds, we focus on security selection, much like the classic analysis of Markowitz (1952) or more recently Garleanu and Pedersen (2013). We consider a security characteristic or signal to be a relevant anomaly if it has non-zero weight in our protagonist's selection decision.

The sheer number of potential anomalies accumulated over decades of research demands a certain degree of simplification. Some recent attempts at grander simplification include Fama and French (2008, 2015, 2016) and Stambaugh, Yu, and Yuan (2012). Broadly speaking, there are two standard asset pricing tests. The first identifies candidate anomalies in the first category, looking at return predictability without considering risk. Fama and French (1992) is the canonical citation. The main empirical tool is cross-sectional return prediction using security-level signals and a pooled estimation, typically with the procedure of Fama and MacBeth (1973). We call a candidate that passes this first test a "score anomaly." The second test determines whether a candidate anomaly adds to a set of existing return factors. Fama and French (1993) is the canonical citation. The main empirical tool is an intercept, or alpha, test in a time series return prediction using contemporaneous factor returns, typically with the procedure of Jensen (1968). We call a candidate that passes this second test a "factor anomaly."[1]

We begin with the question of relevance, and a simple approach, where our protagonist investor is a Markowitz-style mean-variance optimizer. When our investor is risk averse and faces no transaction costs or other frictions, only time-series tests and the factor anomalies that emerge from these tests are relevant. In that sense, we can think of Fama and French (1993) as catering to the needs of this type of investor. When our investor is risk neutral and faces a simple form of transaction costs, constant across securities, only cross-section tests and the score anomalies that emerge are relevant. Fama and French (1992) are catering to this type of investor. For a risk-averse investor facing simple transaction costs, both sets of anomalies are relevant, in the sense that our protagonist investor will not be satisfied using only those anomalies that emerge from Fama and French (1993) time series tests. Those anomalies left in the editing room of Fama and French (1992) are also relevant. The upshot is that academic research might consider either test to be sufficient to establish a new and relevant asset pricing anomaly.

A caveat is that this framework of relevance ignores practical differences across anomalies. Investors face transaction costs that differ across securities, differ with capacity constraints, and differ in multi-period portfolio choice. Unfortunately, the standard asset pricing tests in their simple form are no longer relevant, except in special cases. In principle, they can be replaced with intuitive modifications. Essentially, the simple returns and alphas from cross-section and time series tests can be replaced with returns that are adjusted for execution costs and our investor's specific level of assets under management. More ephemeral anomalies whose conditional score variance is higher among securities with high transaction costs are, all else equal, less relevant.

Having established relevance, we then turn the power of the two tests. The power of both tests

[1]A given signal can be both a score anomaly and a factor anomaly. For example, profitability is a score anomaly in Table IV of Fama and French (2008) and a factor anomaly in Table 6 of Fama and French (2015). A signal can pass the first test as a score anomaly but fail the second. For example, the ratio of book to market value is a score anomaly in Table IV of Fama and French (2008) but it is not an independent factor anomaly in Table 6 of Fama and French (2015). And, a signal that fails the first test can in principle still be a factor anomaly. Members of this group do not predict stock-level returns but do hedge contemporaneous returns on other factors. Market beta is a leading example. It does not predict returns in the cross section, but it qualifies as a factor anomaly in the sense that it has a non-zero intercept in factor regressions.

rises with the number of securities and the number of time periods and falls with idiosyncratic security variance and factor variance in predictable ways. In addition and importantly, factor tests have two additional terms. On the one hand, the power of a factor test increases when other factors are useful in reducing the residual variance of the test factor's time series returns. On the other hand, the power a factor falls as the in-sample Sharpe ratio of the incumbent factors rises, as in Shanken (1992). In this sense, there can be a higher chance of a false negative, or Type II error. Moreover, this tendency rises with the size of the incumbent model, because the in-sample Sharpe ratio is strictly increasing in the number of incumbent factors. Through this second channel, there is a lower natural limit to the number of new factor anomalies that can be identified. Intuitively, this is accentuated in small samples, where degrees of freedom are consumed in parameter estimation.

A final note is that the power of time series tests can be resurrected by shortening the return horizon. This is immediately obvious from the power formula that we derive. Power falls with the in-sample Sharpe ratio. The Sharpe ratio is itself mechanically increasing in the return horizon, because returns (the numerator of the ratio) rise linearly with horizon, while standard errors (the denominator) rise with the square root of horizon. This means asset pricing tests that rely on quarterly returns have much lower power than tests that rely on monthly horizons, and should be avoided if possible. And, asset pricing tests with daily return horizons resolve the problem of relative power: The daily Sharpe ratio is sufficiently small that the difference in power between time series and cross section tests becomes negligible. However, we stop short of recommending that the standard in asset pricing tests move from monthly to daily. Scholes and Williams (1977)and Liu and Strong (2008) suggest reasons why inferences might be biased in a daily analysis. The optimal approach trades off bias and power. In 30 portfolios from Ken French's data library, analyzing 10-day return horizon essentially solves the problem of bias, suggesting that tests that use longer return horizons for these anomalies are needlessly sacrificing power.

We make connections to several related papers along the way. Garleanu and Pedersen (2013) consider the sort of partial equilibrium analysis that we do here, but they do not consider the econometric relevance of their optimal portfolios. Instead, they take the return generating process as given. Moreover, they use one simplifying assumption - that transaction costs are proportional to risk - to come to an elegant closed-form solution, while we consider a range of less elegant assumptions about transaction costs. Like us, Hoberg and Welch (2009) consider Fama and French

(1992) style tests. Their focus, unlike ours, is the use in time series tests of optimized portfolios, whose returns are derived from cross-sectional regressions, versus sorted portfolios, which are favored by Fama and French (1993). A large number of papers, including Fama (1998), Mclean and Pontiff (2016), Harvey, Liu, and Zhu (2015), Bailey and López de Prado (2014), and Novy-marx (2016) consider the issue of data mining and Type I error, in identifying anomalies that do not really exist. While this is a serious problem, our focus is instead on power and Type II error, in failing to identify legitimate anomalies, especially in time series tests. Just as we do, Loughran and Ritter (2000) and Ang, Liu, and Schwarz (2008) consider the power of asset pricing tests. Loughran and Ritter (2000) focus on the effects of weighting on the power of tests, both in aggregating firms at a point in time as well as in aggregating test statistics by market capitalization. Ang, Liu, and Schwarz (2008) focus on the use of aggregation in estimating factor loadings, while we focus on a comparison of cross-section and time series tests. We aggregate firm-level returns into factor returns, and we abstract from weighting schemes and the estimation of right-hand-side variables - including factor loadings - in the cross section, and instead focus on the lost power that comes from estimating covariances in time series tests. Consistent with our logic, Lewellen (2015) finds strong predictive power in a model that uses coefficients from more powerful cross-sectional estimation, while Simin (2008) and others find much less predictive power using less powerful time series estimation.

The paper proceeds as follows. Section 2 develops the investor's security selection problem, and considers the relevance of the factor and score anomalies that emerge from standard asset pricing tests for portfolio choice. Section 3 derives the asymptotic and small-sample power of score and factor anomaly tests. Section 4 concludes.

## 2 The Relevance of Standard Asset Pricing Tests

There are three potential audiences for asset pricing tests. The first, rational asset pricing, considers anomalies to be a misspecification of the risks that are relevant to the representative investor. If a characteristic reliably predicts stock returns, it must be compensation for risk. The factor returns covary with some underlying state variable that drives investor utility. New anomalies, if they are deemed to be robust, are added to the set of known risk factors. With the presumption that risk covariances are at the root of all seeming anomalies, rational asset pricing has necessarily focused on time-series intercept tests. The second audience, behavioral finance, considers anomalies to be examples of mispricing, driven by some combination of less-than-fully-rational preferences and limits to arbitrage. The third audience, practitioners in investment management, considers anomalies to be potential sources of risk-adjusted return that can improve the welfare of their clients in partial equilibrium.

While our focus is on relevance to the third audience, it is worth saying a few words qualitatively about the second. A candidate anomaly that passes the cross-section test but not the time series test is arguably of academic relevance. In particular, the limits to arbitrage and so-called intermediary asset pricing stipulates that shocks to arbitrageur or intermediary capital can make the returns to seemingly unrelated anomalies correlated in the time series. If the goal is to understand investor preferences or beliefs, then a characteristic that is uniquely useful in explaining the cross-section of returns, but is spanned by other stronger anomalies in the times series is nonetheless relevant for behavioral finance. A full exposition of this argument is beyond the scope of this paper.

We apply the classic portfolio choice model of Markowitz (1952) to the problem facing the third audience - an investor who cares about single period portfolio returns and variances. Rather than attempting to characterize the general equilibrium in the spirit of Tobin (1958) or Sharpe (1964) or Lintner (1965) that arises if all investors were rational and had these preferences, we stay in partial equilibrium. We are interested in the case where active portfolio management can deliver superior investment decisions for our non-representative investor. This happens when our investor has a different view of risk and return from the representative investor, either because of differences in preferences or beliefs. It is worth noting that mean-variance portfolio choice is commonly used by practitioners. For example, the portfolio construction software developed by MSCI, Axioma, and Northfield all use some form of myopic mean-variance optimization, with constraints and non-linear transaction costs.

In this context, our definition of an anomaly is simple: It is a set of scores, for each security in the opportunity set, that is *relevant* for our investor's portfolio choice. If our investor can safely ignore a set of scores, there is no anomaly. If our investor chooses to use this set of scores in his decision making, then there is an anomaly. We build intuition in three steps. The first is the classic case where our investor is risk averse and faces no trading frictions. The second is where our investor becomes risk neutral but faces a simple form of transaction costs that are constant across securities.

And, the final step combines both risk aversion and transaction costs. These help establish the applicability of two standard asset pricing tests: the cross section Fama and MacBeth (1973) test popularized by the Fama and French (1992) assessment of anomalies; and the time series Jensen (1968) alpha test popularized with the introduction of the Fama and French (1993) three factor portfolio.

We also consider three extensions to the basic model in the Appendix A that allow: for transaction costs that vary across securities; for varying levels of assets under management; and, for dynamic trading of the sort in Garleanu and Pedersen (2013). These extensions drive a wedge between gross and net returns that varies across anomalies and thereby suggest straightforward modifications to the two standard asset pricing tests that have the effect of netting out execution costs. In some cases, these adjustments are dependent on the level of assets under management, making the relevance of a particular anomaly context dependent, which is why we analyze them as extensions.

## 2.1 The Return Generating Process: Scores and Factor Returns

We suppose that returns for $N$ securities follow a linear factor structure at discrete times $t \in \{0, 1, \ldots, T\}$.

$$\begin{gathered} \mathbf{r}_t = \mathbf{\Gamma}_t \mathbf{f}_t + \boldsymbol{\varepsilon}_t \\ \mathbf{f}_t \sim N(\boldsymbol{\mu}, \mathbf{\Sigma}) \quad \boldsymbol{\varepsilon}_t \sim N\left(0, \sigma^2 \mathbf{I}\right) \end{gathered} \tag{1}$$

The vector of individual security returns $\mathbf{r}$, measured in excess of a risk-free rate of return, is governed by a matrix $\mathbf{\Gamma}$ consisting of $K < N$ row vectors of scores $\boldsymbol{\gamma}'$ that vary across securities $i$ at a time $t$, and a vector of normally distributed factor returns $\mathbf{f}$ that vary over time but not across securities. The return of any security is equal to the sum product of its scores and corresponding factor returns plus a residual idiosyncratic return. The $K$ factor returns can be thought of as returns to portfolios of stocks that can be estimated with scores and observed returns, $\hat{\mathbf{f}}_t = \left(\mathbf{\Gamma}_t{}'\mathbf{\Gamma}_t\right)^{-1} \mathbf{\Gamma}_t{}'\mathbf{r}_t$. What we refer to as scores are sometimes called characteristics in the academic literature. The canonical Fama and French (2015) characteristics now include the ratio of book to market value, the firm's market capitalization, the annual rate of growth in assets, and operating profitability scaled by assets, each transformed into buckets with common scores to limit the effect of extreme scores. To this, many researchers add stock price momentum, typically measured as the most recent annual return excluding the most recent month.

The assumption of normality in Equation 1 and a constant investment opportunity set, with no time subscripts on $\boldsymbol{\mu}$ or $\boldsymbol{\Sigma}$ aligns with the single-period mean-variance portfolio choice problem that we evaluate in the next subsection. Unlike essentially all of our other assumptions about the return generating process, these two come at the expense of generality. No doubt, the investment opportunity set changes over time and some factor returns are not normally distributed, and partial equilibrium investors likely care about timing factor returns and about the higher moments of their portfolio returns.

Below, we will use examples with two factors at a time to build intuition, and will often leave off the subscript $t$ to simplify the notation:

$$r_i = \gamma_{a,i} f_a + \gamma_{b,i} f_b + \varepsilon_i \tag{2}$$

We can map this into Equation 1:

$$\mathbf{r} = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_N \end{bmatrix} \quad \boldsymbol{\Gamma} = \begin{bmatrix} \gamma_{a,1} & \gamma_{b,1} \\ \gamma_{a,2} & \gamma_{b,2} \\ \vdots & \vdots \\ \gamma_{a,N} & \gamma_{b,N} \end{bmatrix} \quad \mathbf{f} = \begin{bmatrix} f_a \\ f_b \end{bmatrix} \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix} \tag{3}$$

To simplify the analysis, we often define the scores in a particular way, roughly in the spirit of Fama and French, to make them analogous to portfolio strategies. The first column of scores $\gamma_a$ is equal to 1 for all firms. The rest are defined so as to sum to zero. This means that the first factor return $f_a$ will be the average or market return on all securities. Under the capital asset pricing model (CAPM), for example, where returns are governed by a single factor, the second column of scores $\gamma_b$ is the demeaned CAPM beta $\beta$, equal to the standard CAPM beta less 1, and the second factor return is equal to the first, $f_a = f_b = r_m$, so that $\mathbf{r} = \boldsymbol{\beta} r_m + \boldsymbol{\varepsilon}$.

We are not otherwise specifying the distribution of the scores $\Gamma$, so in principle this setup could accommodate sorted portfolios of the type in Fama and French (1993) or continuous variables of the type in Fama and French (1992). There is a large literature on the relative merits of sorted portfolios versus more continuous or optimized weights. Hoberg and Welch (2009) argue that test portfolios and factor portfolios are better constructed via optimization than via sorting. Daniel and Titman (1997) and Davis, Fama, and French (2000) further consider whether sorting by characteristics and covariances helps to resolve the age-old question of whether a firm characteristic that is correlated with future returns is a risk factor or mispricing. We side-step both of these issues, but our analysis is closer in spirit to Hoberg and Welch (2009). In addition to possible improvements in explanatory power that they document, we find linearly orthogonalized portfolios easier to work with analytically.

### 2.1.1 Exposures to Raw and Unit Factor Portfolios

A portfolio is defined by its vector of weights $\mathbf{w}$ on the available securities. It will be useful for us to characterize any portfolio's exposure to two sets of factor portfolios, denoted by the matrices $\mathbf{Q}_{raw}$ and $\mathbf{Q}_{unit}$. We refer to the matrix $\mathbf{Q}_{raw} = \mathbf{\Gamma}$ as the set of *raw* factor portfolios that convert the $K$ sets of scores directly into portfolio weights. In general, as we have defined the scores above, the first raw factor portfolio is the market portfolio, and the subsequent raw factor portfolios are dollar neutral portfolios that tilt toward firm characteristics, such as $\beta$. The weights in these raw portfolios can in principle be correlated in the cross section. (The canonical Fama and French firm characteristics are correlated, to some extent, in their final formulation.) We will also refer to the set of portfolios that are orthogonal, or cross-sectionally uncorrelated with all but one of the raw factor portfolios, with the remaining covariance designed to be exactly one. This is the set of *unit* factor portfolios that can be obtained by cross sectional regression of returns on scores, $\mathbf{Q}_{unit} = \mathbf{\Gamma}\left(\mathbf{\Gamma}'\mathbf{\Gamma}\right)^{-1}$.

Any portfolio $\mathbf{w}$ can be expressed as a linear combination of either *raw* or *unit* factor portfolios plus an orthogonal residual $\boldsymbol{\eta}$ using a multivariate regression of the portfolio weights on the matrix $\mathbf{Q}$. This results in a vector $\mathbf{e}\left(\mathbf{w}\right)$ of $K$ multivariate exposures to the factor portfolios $\mathbf{Q}$.

$$\mathbf{w} = \mathbf{Q}\mathbf{e} + \boldsymbol{\eta} \Rightarrow \mathbf{e}\left(\mathbf{w}\right) = \left(\mathbf{Q}'\mathbf{Q}\right)^{-1}\mathbf{Q}'\mathbf{w} \tag{4}$$

The function $\mathbf{e}_{raw}$ takes any set of portfolio weights $\mathbf{w}$ as an input and uses the matrix of raw factor portfolios $\mathbf{Q} = \mathbf{Q}_{raw}$, while the function $\mathbf{e}_{unit}$ uses the matrix of unit portfolio weights $\mathbf{Q} = \mathbf{Q}_{unit}$.

$$\begin{aligned} \mathbf{e}_{raw}(\mathbf{w}) &= (\mathbf{\Gamma}'\mathbf{\Gamma})^{-1}\mathbf{\Gamma}'\mathbf{w} \\ \mathbf{e}_{unit}(\mathbf{w}) &= \mathbf{\Gamma}'\mathbf{w} \end{aligned} \tag{5}$$

Note that the full set of unit factor exposures in the matrix $\mathbf{E}_{unit}$ of the set of unit factor portfolios $\mathbf{Q}_{unit}$ is, as designed, equal to the identity matrix. Each unit factor portfolio has exactly unit exposure to a single factor and zero to the rest. Meanwhile, the unit factor exposure of the raw factor portfolios has off diagonal elements that are not zero. Raw factor portfolios can in principle have different gross exposure and also have incidental correlations among each other.

$$\begin{aligned} \mathbf{E}_{unit}(\mathbf{Q}_{unit}) &= \mathbf{\Gamma}'\mathbf{Q}_{unit} = \mathbf{I} \\ \mathbf{E}_{unit}(\mathbf{Q}_{raw}) &= \mathbf{\Gamma}'\mathbf{Q}_{raw} = \mathbf{\Gamma}'\mathbf{\Gamma} \end{aligned} \tag{6}$$

### 2.1.2 Computing Portfolio Expected Return and Variance

A portfolio's realized return can be characterized as the product of its unit factor exposures and the realized factor returns plus a residual return. Its expected return is the product of its unit factor exposures and the expected factor returns. Portfolio variance can be computed analogously.

$$\begin{aligned} \mathrm{E}(\mathbf{r}'\mathbf{w}) &= \mathrm{E}(\mathbf{f}'\mathbf{\Gamma}'\mathbf{w}+\varepsilon'\mathbf{w}) = \mathrm{E}(\mathbf{f}'\mathbf{e}_{unit}(\mathbf{w})+\varepsilon'\mathbf{w}) = \boldsymbol{\mu}'\mathbf{e}_{unit}(\mathbf{w}) \\ \mathrm{var}(\mathbf{r}'\mathbf{w}) &= \mathrm{var}(\mathbf{f}'\mathbf{\Gamma}'\mathbf{w}+\varepsilon'\mathbf{w}) = \mathrm{var}(\mathbf{f}'\mathbf{e}_{unit}(\mathbf{w})+\varepsilon'\mathbf{w}) = \mathbf{e}_{unit}(\mathbf{w})'\mathbf{\Sigma}\mathbf{e}_{unit}(\mathbf{w})+\sigma^2\mathbf{w}'\mathbf{w} \end{aligned} \tag{7}$$

With the return generating process in Equation 1, both expected return and variance can be computed parsimoniously with the knowledge of factor exposures and the distributional properties of factor returns. This is because the residual variance will often be small for large and diversified portfolios as $N$ becomes large, but the risk from the factor covariance matrix remains.

$$\mathbf{e}_{unit}(\mathbf{w})'\mathbf{\Sigma}\mathbf{e}_{unit}(\mathbf{w})+\sigma^2\mathbf{w}'\mathbf{w} \xrightarrow{\mathbf{w}'\mathbf{w}\to 0} \mathbf{e}_{unit}(\mathbf{w})'\mathbf{\Sigma}\mathbf{e}_{unit}(\mathbf{w}) \tag{8}$$

It is important to note that the factor returns themselves are not necessarily uncorrelated, even though they are returns to unit factor portfolios. They have unique exposure to a single set of scores. But, it is quite possible, and often true in US data, that two unit factor portfolio returns will be correlated with each other in the factor covariance matrix $\mathbf{\Sigma}$.

## 2.2 The Investor's Security Selection Problem

We consider a single-period investor, without limits on leverage or short-selling constraints, who cares about mean and variance and knows the return generating process in Equation 1. In selecting a portfolio, our investor faces a simple form of quadratic transaction costs, which act as limits on position size:

$$\max_{\mathbf{w}} E\left(\mathbf{r}'\mathbf{w}\right) - \frac{\lambda}{2}\text{var}\left(\mathbf{r}'\mathbf{w}\right) - \frac{\theta}{2}\mathbf{w}'\mathbf{w} \tag{9}$$

or, substituting the return generating process as simplified in Equation 7:

$$\max_{\mathbf{w}} \boldsymbol{\mu}'\mathbf{e}_{unit}\left(\mathbf{w}\right) - \frac{\lambda}{2}\left(\mathbf{e}_{unit}\left(\mathbf{w}\right)'\Sigma\mathbf{e}_{unit}\left(\mathbf{w}\right) + \sigma^2\mathbf{w}'\mathbf{w}\right) - \frac{\theta}{2}\mathbf{w}'\mathbf{w} \tag{10}$$

and, substituting the definition of the unit exposures of a given portfolio shown in Equation 5:

$$\max_{\mathbf{w}} \left(\boldsymbol{\Gamma\mu}\right)'\mathbf{w} - \frac{1}{2}\mathbf{w}'\left(\lambda\boldsymbol{\Gamma\Sigma\Gamma}' + \mathbf{I}\left(\sigma^2 + \theta\right)\right)\mathbf{w} \tag{11}$$

We acknowledge that all of these modeling assumptions come at the expense of generality. Most investors care about more than just mean and variance, they face sundry portfolio constraints, they have the ability to trade dynamically, and dynamic trade leads to more complicated effects of transaction costs, changing scores $\boldsymbol{\Gamma}_t$, and changes in the investment opportunity set $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. We analyze some of these as extensions below.

We consider two special cases of this objective function. The first is where the aversion to risk $\lambda$ is equal to zero. In other words, the investor is risk neutral, but the non-zero transaction costs that he faces cause his optimization problem to remain convex. The second is where transaction costs $\theta$ are zero, but there is aversion to risk $\lambda$. This situation, where our investor can trade at no cost, is the classic problem in the academic literature on mean-variance optimization. Practically speaking, its outputs might apply approximately to an investor with low assets under management. For investors with higher levels of assets under management, variable costs of trade limit position sizes. Realistically, investors care about both execution costs and risk, but it is easier to build intuition for the two separate cases before we consider the general case.

In extensions in Appendix A, we first replace $\theta$ with a vector of trading costs $\boldsymbol{\theta}$ that vary across securities. We then consider limits on leverage and the resulting effects of assets under management on investors facing the vector of cost parameters $\boldsymbol{\theta}$. And we finally consider simple dynamics in the spirit of Garleanu and Pedersen (2013) to capture the extra costs of factors whose value decays over time. These three extensions highlight several intuitive notions of execution costs at the level of factor portfolios, which can then be neatly characterized in closed form solutions and examples. Rather than using the standard asset pricing tests applied to *gross* returns, the two asset pricing tests must be applied to *net of execution cost* returns that depend on our investor's specific circumstances.

## 2.3 Optimal Weights

The solution to the investor's problem in Equation 11 involves a tradeoff between risk, return, and execution costs. The optimal weights on each security are a function of security scores $\mathbf{\Gamma}$, expected factor returns $\boldsymbol{\mu}$, the covariance of factor returns $\mathbf{\Sigma}$, transaction costs $\theta$, the investor's risk aversion $\lambda$, and assets under management $A$. At optimal portfolio weights, the marginal benefit of incremental weight in each security is equal its marginal cost in the optimal portfolio:

$$\begin{aligned} \mathbf{\Gamma}\boldsymbol{\mu} &= \left(\lambda\mathbf{\Gamma}\mathbf{\Sigma}\mathbf{\Gamma}' + \mathbf{I}\left(\lambda\sigma^2 + \theta\right)\right)\mathbf{w}^* \\ \Rightarrow \mathbf{w}^* &= \left(\lambda\mathbf{\Gamma}\mathbf{\Sigma}\mathbf{\Gamma}' + \mathbf{I}\left(\lambda\sigma^2 + \theta\right)\right)^{-1}\mathbf{\Gamma}\boldsymbol{\mu} \end{aligned} \tag{12}$$

We analyze two special cases, when transaction costs are zero and when risk aversion is zero, and then proceed to the general case, over the next three subsections, before considering the case of transaction costs that vary across securities in Appendix A.

### 2.3.1 Risk Neutral, Constant Transaction Costs

First, we consider the simplest case, where $\lambda$ is equal to zero and risk considerations are unimportant. Our investor is interested in maximizing returns net of transaction costs. Then, the optimal weights from Equation 12 simplify to:

$$\mathbf{w}^*_{tc} = \frac{1}{\theta}\mathbf{\Gamma}\boldsymbol{\mu} \tag{13}$$

Intuitively, the optimal weight for an individual security is high when it scores well on factors that have high expected returns. To get more visibility into the optimal weights, we can compute the exposure $\boldsymbol{e}$ of this portfolio $\mathbf{w}_{tc}^*$ to the *raw* and *unit* factor portfolios using Equation 5.

$$\begin{aligned} \mathbf{e}_{raw}\left(\mathbf{w}_{tc}^*\right) = & \quad \tfrac{1}{\theta}\boldsymbol{\mu} \\ \mathbf{e}_{unit}\left(\mathbf{w}_{tc}^*\right) = & \quad \tfrac{1}{\theta}\boldsymbol{\Gamma}'\boldsymbol{\Gamma}\boldsymbol{\mu} \end{aligned} \tag{14}$$

The upshot is that our risk neutral investor's problem can be reduced constructing the raw factor portfolios $\mathbf{Q}_{raw} = \boldsymbol{\Gamma}$, learning the magnitude of the expected factor returns $\boldsymbol{\mu}$, and using this vector as weights on the $K$ raw factor portfolios, which are the columns of $\mathbf{Q}_{raw}$. If the second factor were to have a zero expected return $\mu_b = 0$, it can be ignored in the optimization problem, regardless of its risk properties. What is trivially absent is $\boldsymbol{\Sigma}$. If a factor has a positive expected return $\mu_b > 0$, but it has a zero alpha with respect to an existing factor $c$, the standard logic of Fama and French (1993) says this *is not* a distinct anomaly. If factor $b$ has a zero expected return $\mu_b = 0$, but it has a non-zero alpha, hedging an existing factor $c$ with positive expected return $\mu_c > 0$, the standard logic of Fama and French (1993) says this *is* a distinct anomaly. But, for an investor who cares only about minimizing transaction costs, factors are anomalies if and only if they have a non-zero expected return in the sense of $\mu_b$.

So, what to make of the exposure to unit portfolios in the second half of Equation 14? The unit exposures depend not only on expected factor returns but also on the correlation structure of scores. These are the unintended common risks of the raw portfolio exposures in the first half of Equation 14. These exposures to unit portfolios are interesting, but not relevant to the investor's optimization problem. For a risk neutral investor, the optimal portfolio inherits unintended but irrelevant risk exposures. Hedging these unintended risks requires transaction costs and is therefore suboptimal.

> **Portfolio Choice With Transaction Costs:** For a risk neutral investor, identifying anomalies relies only on procedures like Fama and MacBeth (1973), using the results of papers that are in the spirit of Fama and French (1992) to test the significance of $\boldsymbol{\mu}$:

$$\text{Cross Section Test: } \hat{\boldsymbol{\mu}} = \overline{\mathbf{f}} = \frac{1}{T}\sum_t \mathbf{f}_t = \frac{1}{T}\sum_t \left(\boldsymbol{\Gamma}_t'\boldsymbol{\Gamma}_t\right)^{-1}\boldsymbol{\Gamma}_t'\mathbf{r}_t = \mathbf{0} \tag{15}$$

If a given factor $\gamma_a$ passes this test of $\mu_a \neq 0$, we call it a score anomaly. Performing factor regressions in the spirit of Fama and French (1993) on the resulting factor portfolios will lead to mistakes in factor selection and the search for anomalies, excluding valuable anomalies and including apparent anomalies that are valuable only for their risk properties $\mathbf{\Sigma}$ and not their expected returns. For the risk neutral investor facing transaction costs, the search for anomalies starts and ends with score anomalies.

**Example 1**

Consider an example of two factors, call them CAPM beta and operating profitability. (We are ignoring the market portfolio for the moment to keep $K = 2$.) Suppose that these two factors have the following structure, so that operating profitability and beta have a negative correlation from a common component $c$, with the residual components $a$ and $b$ uncorrelated with each other.

$$\gamma_{a,i} = \frac{OP}{A}_i = c_i + a_i \tag{16}$$

$$\gamma_{b,i} = \beta_i - 1 = -c_i + b_i \tag{17}$$

Suppose the expected returns of the two factor portfolios are roughly equal to their average returns in US data, so that $\mu_a > 0$ and $\mu_b = 0$. The optimal stock weight is proportional to its own mean return:

$$w^*_{tc,i} = \frac{1}{\theta}\left(\gamma_{a,i}\mu_a + \gamma_{b,i}\mu_b\right) = \frac{1}{\theta}\gamma_{a,i}\mu_a \tag{18}$$

The optimal portfolio is a scaled version of the raw portfolio $a$ that tilts towards firms with high operating profits and away from firms with low operating profits. The absolute magnitude of the weights depends on the expected factor return $\mu_a$ net of transaction costs $\theta$. When the ratio of return to cost is high, the weights are correspondingly large. When the ratio of returns to cost is low, the weights are correspondingly small. This is in loose terms a reflection of the capacity of the strategy in light of execution costs.

No information about CAPM beta is needed to form the optimal portfolio. To this investor, CAPM beta is dead, and can be safely ignored, because it does not pass a significance test that rejects $\mu_b = 0$ in the sense of Fama and French (1992). But, the exposure of the optimal portfolio to the unit CAPM beta portfolio $b$ reveals that the optimal weights have an incidental exposure, which here comes from the common component $c$. For the risk neutral investor, this incidental exposure is irrelevant in setting weights. Our investor could neutralize the exposure to CAPM beta, but he chooses not to. The intuitive rationale is that hedging this zero return exposure raises transaction costs without increasing the investor's utility. The results of this choice can be seen in the raw and unit exposures of the optimal portfolio, using Equation 5 and using notation $\text{var}(a) = s_a^2$:

$$\begin{aligned} \mathbf{e}_{raw}(\mathbf{w}_{tc}^*) = & \quad \tfrac{1}{\theta}\boldsymbol{\mu} = & & & \begin{bmatrix} \tfrac{1}{\theta}\mu_a \\ 0 \end{bmatrix} \\ \mathbf{e}_{unit}(\mathbf{w}_{tc}^*) = & \quad \tfrac{1}{\theta}\boldsymbol{\Gamma}'\boldsymbol{\Gamma}\boldsymbol{\mu} = & \tfrac{1}{\theta}\begin{bmatrix} s_c^2 + s_a^2 & -s_c^2 \\ -s_c^2 & s_c^2 + s_b^2 \end{bmatrix}\begin{bmatrix} \mu_a \\ 0 \end{bmatrix} & = & \begin{bmatrix} \tfrac{1}{\theta}\left(s_c^2 + s_a^2\right)\mu_a \\ -\tfrac{1}{\theta}s_c^2\mu_a \end{bmatrix} \end{aligned} \tag{19}$$

The second entry in the unit portfolio exposure vector shows the negative beta tilt that comes incidentally from exploiting operating profitability. Meanwhile, the second entry in the raw portfolio exposure vector shows that market beta is not a score anomaly in this two factor example for a risk neutral investor.

### 2.3.2 Risk Averse, No Transaction Costs

Next, we consider the case where transaction costs $\theta$ are equal to zero but our investor is risk averse, so that $\lambda > 0$. Then, the optimal weights from Equation 12 simplify to:

$$\mathbf{w}_{ra}^* = \frac{1}{\lambda}\left(\boldsymbol{\Gamma}\boldsymbol{\Sigma}\boldsymbol{\Gamma}' + \sigma^2\mathbf{I}\right)^{-1}\boldsymbol{\Gamma}\boldsymbol{\mu} \tag{20}$$

This is the classic solution to mean-variance optimization, when the return generating process is expressed with a linear factor structure. Weights are increasing in the individual stock expected returns, which here can be expressed as the linear combination of factor scores and expected factor returns. and weights are decreasing in the individual stock contributions to risk, which here can be expressed as the product of factor scores and the covariance of factor returns $\boldsymbol{\Gamma}\boldsymbol{\Sigma}\boldsymbol{\Gamma}'$ plus idiosyncratic risk $\sigma^2$. To get more visibility into the optimal weights, we can compute the exposure $\boldsymbol{e}$ of this portfolio $\mathbf{w}_{ra}^*$ to the *unit* factor portfolio . This is easiest to do by rearranging the first order condition in Equation 20 and substituting the definition of exposure to the unit portfolio from Equation 5:

$$\boldsymbol{\Gamma}\boldsymbol{\mu} = \lambda\left(\boldsymbol{\Gamma}\boldsymbol{\Sigma}\boldsymbol{\Gamma}'\mathbf{w}_{ra}^* + \sigma^2\mathbf{w}_{ra}^*\right) = \lambda\boldsymbol{\Gamma}\boldsymbol{\Sigma}\mathbf{e}_{unit}(\mathbf{w}_{ra}^*) + \lambda\sigma^2\mathbf{w}_{ra}^* \tag{21}$$

We can further rearrange, and take limits as the number of stocks $N$ grows large. As this happens the weight on any one security becomes small, and both sides of the equation go towards zero, allowing us to derive a simple and intuitive expression for the exposure of the optimal portfolio $\mathbf{w}_{ra}^*$ , to the unit factor portfolio:

$$\begin{aligned} \boldsymbol{\Gamma}\left(\boldsymbol{\mu} - \lambda \boldsymbol{\Sigma} \mathbf{e}_{unit}\left(\mathbf{w}_{ra}^*\right)\right) = \lambda \sigma^2 \mathbf{w}_{ra}^* \to 0 \\ \Rightarrow \mathbf{e}_{unit}\left(\mathbf{w}_{ra}^*\right) = \tfrac{1}{\lambda} \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \end{aligned} \tag{22}$$

The upshot is that our risk averse investor's problem reduces to a mean-variance optimization of factor portfolios, when he can trade frictionlessly at $\theta = 0$ with the number of available stocks $N$ large relative to the number of factors $K$. It no longer suffices to learn $\boldsymbol{\mu}$. Now, the covariance properties of the factor portfolios $\boldsymbol{\Sigma}$ are also relevant. If a factor has a positive expected return $\mu_a > 0$, but it has a zero alpha with respect to an existing factor $b$, the standard logic of Fama and French (1993) says this *is not* a distinct anomaly, and indeed it is not for a risk-averse investor who faces no transaction costs. If factor $a$ has a zero expected return $\mu_a = 0$, but it has a non-zero alpha, hedging an existing factor $b$ with positive expected return $\mu_b > 0$, the standard logic of Fama and French (1993) says this *is* a distinct anomaly. And, for a risk averse investor who does not care about minimizing transaction costs, this is indeed a useful hedge.

**Portfolio Choice With Risk Aversion:** For a risk averse investor facing no transaction costs, the search for anomalies occurs in two steps. The first step is to use a procedure like Fama and MacBeth (1973) to estimate both $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ as:

$$\begin{aligned} \hat{\boldsymbol{\mu}} = \bar{\mathbf{f}} = \tfrac{1}{T} \textstyle\sum_t \mathbf{f}_t = \tfrac{1}{T} \sum_t \left(\boldsymbol{\Gamma}_t' \boldsymbol{\Gamma}_t\right)^{-1} \boldsymbol{\Gamma}_t' \mathbf{r}_t \\ \hat{\boldsymbol{\Sigma}} = \widehat{\text{var}}\left(\mathbf{f}_t\right) = \tfrac{1}{T} \textstyle\sum_t \left(\mathbf{f}_t - \bar{\mathbf{f}}\right)\left(\mathbf{f}_t - \bar{\mathbf{f}}\right)' \end{aligned} \tag{23}$$

The second step is to perform factor regressions in the spirit of Fama and French (1993) on the resulting factor portfolio returns, including those with zero means $\mu_a = 0$. This will lead to the elimination of factor portfolios with positive means but zero alphas, and lead to the resurrection of factor portfolios with zero means but non-zero alphas that come from their useful hedging properties. Alpha here and above refers to a Jensen (1968) alpha test. Sample averages are inserted into Equation 22, $\hat{\boldsymbol{\sigma}}$ refers to elements of $\hat{\boldsymbol{\Sigma}}$, and the null is that this first factor $b$ is irrelevant in the choice of portfolio weights. If under the null, $e_{unit,b}\left(\mathbf{w}_{ra}^*\right) = 0$, this implies:

$$\sum_{k \neq b} \hat{\sigma}_{bk} e_{unit,k} \left(\mathbf{w}_{ra}^{*}\right) = \hat{\mu}_{b} \tag{24}$$

Equation 24 is equivalent to regressing the time series of $f_b$ on the time series of the remaining factors excluding $b$, estimating the multivariate factor loadings $\beta$ and testing the significance of the intercept:

$$\text{Time Series Test: } \hat{\mu}_b - \sum_{k \neq b} \hat{\beta}_{bk} \hat{\mu}_k = 0 \tag{25}$$

This derivation is in Appendix B. This means that some factors with $\mu_b = 0$ can nonetheless be factor anomalies because of their covariance properties mean that $\mu_b - \beta_{ba}\mu_a - \beta_{bc}\mu_c + \cdots \neq 0$. Some factors with $\mu_b \neq 0$ may nonetheless not be factor anomalies, because of their covariance properties $\mu_b - \beta_{ba}\mu_a - \beta_{bc}\mu_c + \cdots = 0$. For the risk averse investor facing no transaction costs, the search for anomalies starts and ends with factor anomalies. In our nomenclature, these remain score anomalies, but they are not factor anomalies.

**Example 2**

Consider again two factors, as in Example 1, but the first is now the standard market portfolio, and the second is still the CAPM beta, as before, with the mean return on the market factor $\mu_a > 0$ and the mean return on the beta factor $\mu_b = 0$. Further, assume that the payoffs to the two factors are positively correlated, so that the off-diagonal elements of $\mathbf{\Sigma}$ are positive, so that $\sigma_{ab} > 0$, meaning that a portfolio that is long firms with high betas has relatively higher returns when the market also has relatively higher returns, as is true in US data. Plugging in to Equation 22:

$$\mathbf{e}_{unit} \left(\mathbf{w}_{ra}^{*}\right) = \begin{bmatrix} e_{unit,r_m} \\ e_{unit,\beta} \end{bmatrix} = \frac{1}{\lambda} \begin{bmatrix} \frac{\sigma_{bb}^2}{\left(\sigma_{aa}^2 \sigma_{bb}^2 - (\sigma_{ab})^2\right)} \mu_a \\ -\frac{\sigma_{ab}}{\left(\sigma_{aa}^2 \sigma_{bb}^2 - (\sigma_{ab})^2\right)} \mu_a \end{bmatrix} \tag{26}$$

The optimal strategy involves exposures to the market portfolio and the unit beta portfolios. The absolute magnitude of the exposure to the market portfolio is increasing in its expected factor return $\mu_a$ and decreasing in its expected favor risk $\sigma_{aa}^2$, so roughly speaking increasing in the Sharpe ratio of the market, and this exposure is further increased because its risk can be mitigated with a short position in the unit beta portfolio. Unlike the case of a risk neutral investor facing transaction costs, this exposure is worth hedging despite its zero mean return, because it lowers risk and is costless to execute. So, in that sense the early reports of the death of market beta are exaggerated. It is very much alive as a factor anomaly, relevant for a risk averse investor facing no transaction costs, in this two factor example.

**Example 3**

Consider again an example of two factors in Example 1, now call them the ratio of book to market equity and low asset growth. Suppose that the mean returns are positive $\mu_a, \mu_b > 0$ as they are in US data. Further, assume that the payoffs to the two factors are positively correlated, so that the off-diagonal elements of $\mathbf{\Sigma}$ are positive, so that $\sigma_{ab} > 0$, meaning that a portfolio that is long firms with high ratios of book to market equity has relatively higher returns when a portfolio that is long firms with low asset growth also has relatively higher returns. Now we have the more general version of the unit exposures from Equation 22:

$$\mathbf{e}_{unit}\left(\mathbf{w}_{ra}^{*}\right) = \begin{bmatrix} e_{unit,\frac{B}{M}} \\ e_{unit,\frac{-\Delta A}{A}} \end{bmatrix} = \frac{1}{\lambda}\begin{bmatrix} \frac{\sigma_{bb}^2}{\left(\sigma_{aa}^2\sigma_{bb}^2-(\sigma_{ab})^2\right)}\mu_a - \frac{\sigma_{ab}}{\left(\sigma_{aa}^2\sigma_{bb}^2-(\sigma_{ab})^2\right)}\mu_b \\ \frac{\sigma_{aa}^2}{\left(\sigma_{aa}^2\sigma_{bb}^2-(\sigma_{ab})^2\right)}\mu_b - \frac{\sigma_{ab}}{\left(\sigma_{aa}^2\sigma_{bb}^2-(\sigma_{ab})^2\right)}\mu_a \end{bmatrix} \tag{27}$$

It turns out that empirically, the first entry is indistinguishable from zero, because the alpha of the unit book to market equity portfolio when low asset growth is included as a reference portfolio is approximately zero: $\mu_a - \frac{\sigma_{ab}}{\sigma_{bb}^2}\mu_b \approx 0$. In that sense, we could substitute out information on book to market equity:

$$\mathbf{e}_{unit}\left(\mathbf{w}_{ra}^{*}\right) = \frac{1}{\lambda}\begin{bmatrix} 0 \\ \frac{1}{\sigma_{bb}^2}\mu_b \end{bmatrix} \tag{28}$$

So, the ratio of book to market equity is not a factor anomaly, and not relevant for a risk averse investor facing no transaction costs, in this two factor example. The only information that is needed is the unit low asset growth portfolio and its Sharpe ratio. Interestingly though, the ratio of book to market equity remains a score anomaly, relevant to an investor facing realistic transaction costs, as we see next.

### 2.3.3 Risk Averse, Constant Transaction Costs

The more general case involves both risk aversion and transaction costs, bringing us back to Equation 12. Intuitively, when transaction costs are small, the conclusions of the simple risk aversion case apply. Transaction costs act much like idiosyncratic risk, which is an issue when the investment opportunity set is small, but not when the number of securities $N$ is very large. However, there are reasons to believe that the effect of $\theta$ might be meaningful even when the effect of idiosyncratic $\sigma^2$ is small. When assets under management are large, even a small weight in a given stock might be large in comparison to its trading volume. Recall that Fama and MacBeth tests as in Equation 15 are necessary and sufficient for identifying anomalies in the presence of transaction costs alone. And Jensen's alpha tests as in Equation 25 are necessary and sufficient for identifying anomalies in the presence of risk aversion alone. When our investor is risk averse *and* faces transaction costs, anomalies of both types are relevant. We can start with a variant of Equation 22 where we assume that the considerations of idiosyncratic risk are second order but transaction costs remain first order, so that $\theta \ggg \lambda\sigma^2$:

$$\begin{aligned} \mathbf{\Gamma}\left(\boldsymbol{\mu} - \lambda\mathbf{\Sigma}\mathbf{e}_{unit}\left(\mathbf{w}^*\right)\right) = \left(\lambda\sigma^2 + \theta\right)\mathbf{w}^* \rightarrow \theta\mathbf{w}^* \\ \Rightarrow \mathbf{e}_{unit}\left(\mathbf{w}^*\right) = \tfrac{1}{\lambda}\mathbf{\Sigma}^{-1}\left(\boldsymbol{\mu} - 2\theta\mathbf{e}_{raw}\left(\mathbf{w}^*\right)\right) \end{aligned} \tag{29}$$

While it is obviously unappealing to have optimal weights and exposures on both sides of the equation, Equation 29 is a useful intermediate relationship. Note that the optimal *unit* exposures are driven by the standard risk and return considerations, but the return is haircut by the transaction costs associated with the *raw* exposures of the optimal weights. The intuitive appeal is that we can think of the risk averse investor optimizing over risk and *net of transaction cost* return.

We can eliminate the dependence using the definition of raw exposures from Equation 5:

$$\mathbf{e}_{unit}\left(\mathbf{w}^*\right) = \left[\lambda\mathbf{\Sigma} + \theta\left(\mathbf{\Gamma}'\mathbf{\Gamma}\right)^{-\mathbf{1}}\right]^{-1}\boldsymbol{\mu} \tag{30}$$

We examine the intuition in Equation 30 in the example below, but it is apparent that the Jensen's alpha test will no longer be necessary for a factor to be a relevant anomaly. And, in general, factors that pass either test will be worthy of consideration.

**Portfolio Choice With Risk Aversion and Transaction Costs:** For a risk averse investor facing transaction costs, the search for anomalies means finding factors that pass either the test in the spirit of Fama and MacBeth (1973) in Equation 15 or the test in the spirit of Jensen's alpha test in Equation 25.

**Example 4**

Consider the same setup as Example 3, with the ratio of book to market equity and low asset growth. Recall we found that the ratio of book to market was a score anomaly, because it has a mean return greater than zero $\mu_a > 0$, but not a factor anomaly, at least controlling for low asset growth, because its Jensen's alpha is zero, with $\mu_a - \frac{\sigma_{ab}}{\sigma_{bb}^2}\mu_b \approx 0$. Further, assume that book to market equity and low asset growth scores, $\gamma_a$ and $\gamma_b$ are constructed so as to be orthogonal to one another, with unit standard deviation, which has the notation benefit of making $\mathbf{\Gamma}'\mathbf{\Gamma} = N\mathbf{I}$. Note that the correlation of factor returns $\mathbf{\Sigma}$ can still be high, even if the scores are uncorrelated. Now, we start from the result in Example 2, Equation 28 and generalize the unit exposures to include transaction cost effects:

$$\begin{aligned} \mathbf{e}_{unit}\left(\mathbf{w}^*\right) &= \frac{1}{\lambda C}\begin{bmatrix} \sigma_{aa}^2 + \frac{N\theta}{\lambda} & \sigma_{ab} \\ \sigma_{ab} & \sigma_{bb}^2 + \frac{N\theta}{\lambda} \end{bmatrix}^{-1}\begin{bmatrix} \mu_a \\ \mu_b \end{bmatrix} \\ &= \frac{1}{\lambda C}\begin{bmatrix} \sigma_{bb}^2\left(\mu_a - \frac{\sigma_{ab}}{\sigma_{bb}^2}\mu_b\right) + \frac{N\theta}{\lambda}\mu_a \\ \sigma_{aa}^2\left(\mu_b - \frac{\sigma_{ab}}{\sigma_{aa}^2}\mu_a\right) + \frac{N\theta}{\lambda}\mu_b \end{bmatrix} \end{aligned} \tag{31}$$

We substitute $C \equiv \left(\left(\sigma_{aa}^2 + \frac{N\theta}{\lambda}\right)\left(\sigma_{bb}^2 + \frac{N\theta}{\lambda}\right) - \left(\sigma_{ab}\right)^2\right)$ to save space. The optimal exposure to the unit portfolios is a balance of two concerns. If the diagonal of factor risk $\mathbf{\Sigma}$ is high and risk aversion is high, then the solution looks like Equation 22, while if the transaction costs $\theta$ are high, then the solution looks like Equation 14. And, in general, the optimal solution is a blend of these two concerns: managing portfolio risk-adjusted return while keeping execution costs low. Even substituting in the implications of a Jensen's alpha of zero for the book to market factor portfolio, there is still positive portfolio exposure to this factor, because of its raw factor return $\mu_a$:

$$\mathbf{e}_{unit}\left(\mathbf{w}^*\right) = \frac{1}{\lambda C}\begin{bmatrix} \frac{N\theta}{\lambda}\mu_a \\ \left(\frac{\sigma_{aa}^2\sigma_{bb}^2 - (\sigma_{ab})^2}{\sigma_{bb}^2} + \frac{N\theta}{\lambda}\right)\mu_b \end{bmatrix} \tag{32}$$

So, the ratio of book to market equity is not an anomaly for a risk averse investor facing no transaction costs in this two factor example, but it is resurrected in the case involving transaction costs. Our investor tilts somewhat toward the lower alpha unit portfolio to access raw returns at lower cost. The general conclusion is that the investor will put weight on both factor and score anomalies.

### 2.4 Summary

Suppose a candidate factor appears: A firm characteristic that has the potential to predict its stock returns. Is it a relevant anomaly? The literature contains two standard asset pricing tests. The first uses Fama and MacBeth regressions of the sort in Equation 15, testing whether the mean of the new factor returns is equal to zero using a series of multivariate cross-sectional regressions that contain other factors known to predict the cross section of stock returns. We show that this test is always applicable for an investor facing a simple form of constant transaction costs, whether he is risk averse or not. The second uses Jensen's alpha tests of the sort in Equation 25. These test whether the intercept in a regression of the new factor portfolio returns, from the aforementioned cross-sectional regressions, on the portfolio returns of existing factors. We show that this test is always applicable for a risk averse investor, whether he faces transaction costs or not. As it turns out, both are applicable for the general case of a risk-averse investor facing a simple form of transaction costs.

Realistic transaction costs complicate these conclusions, but in an intuitive way. When transaction costs vary across stocks, when assets under management are substantial, and when dynamic trading considerations appear, each of tests should in principle be performed on factor returns that are net of execution costs, and gross of future persistence in returns. Tests of statistical significance in the cross section test are still valid in a number of special cases. Appendix A provides some illustrations of how this might be done in practice.

We now turn to the power of the two tests.

## 3 The Power of Standard Asset Pricing Tests

The previous section shows the relevance of both score and factor anomalies for portfolio choice. In this section, we turn to the econometrics of identifying anomalies. There are two standard tests in empirical asset pricing. Tests in the spirit of Fama and MacBeth (1973) and Fama and French

(1992) are cross-sectional and do not consider the covariance properties of factor portfolio returns. Tests in the spirit of Jensen (1968) and Fama and French (1993) use time-series data and focus on the covariance properties of the factor portfolios as well their means.

We take two approaches to estimating the power of these two tests. In the first, we compute the asymptotic power curve analytically. The score test is more powerful by a multiplier that is increasing in the Sharpe ratio of the factor portfolios. In the second, we simulate the data generating process in small samples, and compute a simulated power curve. The effect of the Sharpe ratio is magnified by a small sample estimation of the factor portfolio return covariances.

## 3.1 Asymptotic Power Curves

Section 2 considers the problem facing an investor who understands the return generating process in Equation 1. Realistically, the investor does not know the expected payoffs to factor portfolios $\boldsymbol{\mu}$ or their covariances $\boldsymbol{\Sigma}$. He must use data on scores and stock returns to estimate these. And, in practice, our investor also needs to consider whether market forces might make estimates irrelevant for future returns. For now, we assume that our investor differs from the representative investor in either preferences or beliefs, so that the history of stock returns can be used to produce reliable forecasts of the parameters in Equation 1.

### 3.1.1 Data Mining

This raises the issue of data mining, which can take two forms: selection and overfitting. Problems of selection stem from starting with $n$ candidate anomalies and choosing $k < n$ that work in the historical data. Problems of overfitting come from optimally weighting $n$ candidate anomalies using their in-sample performance into one aggregate blended superscore. In both cases, the likelihood of Type I error, where a firm characteristic is deemed spuriously to be an anomaly, is high. These issues are discussed in Mclean and Pontiff (2016), Harvey, Liu, and Zhu (2015), and Bailey and López de Prado (2014). For example, Mclean and Pontiff (2016) examine the efficacy of anomalies using realistic rolling estimations of factor average returns and the incremental effect of publicizing the finding. Simin (2008) and Levi and Welch (2014) find that rolling estimates of expected returns from the Fama-French three factor model do not work well at forecasting return realizations. Their findings point to researcher data mining. Novy-marx (2016) computes adjusted t-statistics as a function of $n$, $k$, and the nature of the overfitting procedure employed. Given the number of researchers focused on asset pricing and the academic and commercial incentives for documenting anomalies, the issue of Type I errors is paramount.

However, Lewellen (2015) finds that rolling estimates using characteristics as in our return generating process in Equation 1 produce a t-statistic greater than 10. Moreover, it is worth considering the power to detect anomalies in the first place, and the possibility of Type II errors, which is the focus of this paper. For example, it is possible to turn the argument in Novy-marx (2016) around. Suppose the objective was not to establish an anomaly but rather to overturn an existing anomaly. A similar logic then applies. By starting with $n$ potential controls or established anomalies and risk factors and choosing $k < n$ in the historical data, it is possible to lower the power a test of the risk anomaly. This is particularly true in time series tests, as we argue below, and Novy-marx (2012), who argues that the risk anomaly is subsumed by profitability and value, is an example of this approach.

In this paper, we sidestep the issue of data mining and imagine that there is an established set of anomalies to which a single new candidate may be added. We start with cross sectional tests, which are relevant for an investor facing material transaction costs.

#### 3.1.2 The Cross Section Test

The investor, or the econometrician, estimates Equation 1, where we assume that the first factor is the market portfolio, so that $\boldsymbol{\gamma}_a = \mathbf{1}$, with a series of cross sectional regressions:

$$\mathbf{r}_t \sim \boldsymbol{\Gamma}_t \mathbf{f}_t + \boldsymbol{\varepsilon}_t \Rightarrow \hat{\mathbf{f}}_t = \left(\boldsymbol{\Gamma}_t' \boldsymbol{\Gamma}_t\right)^{-1} \boldsymbol{\Gamma}_t' \mathbf{r}_t \tag{33}$$

The time series mean of the factor payoffs is the Fama and MacBeth estimator of the mean payoff $\boldsymbol{\mu}$:

$$\hat{\boldsymbol{\mu}} = \frac{1}{T} \sum_t \hat{\mathbf{f}}_t \tag{34}$$

In a Bayesian sense, all factors will be relevant in at least a small way, in the sense that the point estimate for the return $\mu_b$ on a factor $b$ will never be exactly zero. But, in a frequentist sense, there are some factors that cannot be statistically deemed relevant. This is the usual notion of a score anomaly: We must be able to say that the hypothesis that $\mu_b \neq 0$ is true with sufficient probability. The standard error of the estimate of $\mu_b$ is equal to:

$$\mathrm{se}\left(\hat{\mu}_b\right) = \frac{1}{\sqrt{T}}\sqrt{\sigma_{bb}^2 + \frac{1}{N^2}\sigma^2\sum_i\left(\boldsymbol{\delta}_b'\boldsymbol{\gamma}_i\right)^2} \tag{35}$$

Where $\sigma_{bb}^2$ is the variance of factor $b$ from the factor covariance matrix $\boldsymbol{\Sigma}$, $N$ is the number of firms in the cross section, and $\sigma^2$ is the idiosyncratic variance that is assumed for simplicity to be constant across firms. And, we define the vector $\boldsymbol{\delta}_b$ from the inverse of the matrix $\boldsymbol{\Gamma}'\boldsymbol{\Gamma}$ as follows:

$$N\left(\boldsymbol{\Gamma}'\boldsymbol{\Gamma}\right)^{-1} = \begin{bmatrix} \boldsymbol{\delta}_a & \boldsymbol{\delta}_b & \boldsymbol{\delta}_c & \cdots \end{bmatrix} \tag{36}$$

We derive the expression for standard error in Equation 35 in Appendix B. Because we estimate the time series mean of the regression coefficients, we lose 1 degree of freedom and the estimate of the asymptotic variance is

$$\widehat{\mathrm{se}}\left(\hat{\mu}_b\right) = \frac{1}{\sqrt{T-1}}\sqrt{\sigma_{bb}^2 + \frac{1}{N^2}\sigma^2\sum_i\left(\boldsymbol{\delta}_b'\boldsymbol{\gamma}_i\right)^2} \tag{37}$$

For a large cross section of stocks $N$, the effect of residual risk $\varepsilon$ becomes small in the second part of the expression under the radical. The estimate of the factor payoff in each period using only returns $\mathbf{r}_t$ and scores $\boldsymbol{\Gamma}_t$ at any given time $t$ becomes very precise. The error in the estimate of the mean then depends only on the number of time periods $T$. This is the asymptotic standard error in the sense that $\sigma_{bb}^2$ and $\sigma^2$ are not known to the investor or the econometrician in small samples. We consider the small sample properties with Monte Carlo simulations below. For now, we plot the power curve using Equation 35 in Figure 1, using

$$\begin{aligned} N &= [50; 100; \underline{250}; 1,000; 10,000] \\ T &= [50; 100; \underline{200}; 500; 1,000] \\ \sigma_\varepsilon &= [0; \underline{3.27}; 5; 6; 7] \\ \sigma_{bb} &= [0.5; 1; 2; \underline{3}; 4] \end{aligned} \tag{38}$$

Underlined values are the base case and represents by the blue line in the graph. Monthly idiosyncratic risk $\sigma_\varepsilon$ and factor risk $\sigma_{bb}$ are in percentage. We vary monthly factor return $\mu_b$ so that its annual Sharpe ratio ranges from 0 to 2. Annual Sharpe ratio is calculated as $SR = \frac{\mu_b}{\sigma_{bb}} \cdot \sqrt{12}$.

The power curves illustrate the probability that the null of zero is rejected given a variety of inputs for the true mean of factor return $\mu_b$. It shows the power our investor has to detect relevant score anomalies. The $y$-intercept is the size of the test. There is a 5% chance of rejecting the null of a zero factor return when it is truly zero. In these situations, the investor concludes that there is a score anomaly when in truth there is not one. The shape of the power curve is otherwise not terribly interesting, as it depends on the assumed distribution of factor scores, the number of separate periods $T$, the assumed factor variance $\sigma_{bb}^2$, and the assumed number of firms $N$, and idiosyncratic risk $\sigma_2$. These shift the power curve in intuitive ways. The power rises more steeply with more firms in Panel (a) and lower idiosyncratic variance in Panel (b). A larger number of firms in the cross-section helps to eliminate the effect of idiosyncratic risk on factor returns. A smaller amount of idiosyncratic risk has a similar effect, in that even a small number of firms deliver a pure factor return. In both cases though, the improvements in power are limited. Even an infinite number of firms does not lead to an extremely powerful test, capable of detecting small anomalies. The power also rises more steeply with more time periods in Panel (c) and lower factor variance in Panel (d). A larger number of time periods means that the mean factor return per period can be estimated with greater and greater accuracy, assuming there are no changes in the underlying return generating process. Similarly, power rises more quickly if the factor payoffs are very reliable, falling very close to the mean in every period. These are situations where a score anomaly can be reliably detected even when the true average return is quite small. All four panels show that economically large score anomalies are always detected even in modest time series, but these will be rare in competitive markets, so power is important.

We are more concerned with the relative power of the cross section and time series tests, which we turn to next, than we are about the other comparative statics, which will improve the power of both proportionally.

#### 3.1.3 The Time Series Test

We next move to the time series test, which is relevant to an investor who is risk averse. In this case, there is a second step after the estimation of factor payoffs in Equation 33. Practically, our investor is interested in whether a particular factor will have zero effect on his portfolio choice in

Equation 22. And, as we argue above, this is equivalent to the factor passing a Jensen (1968) alpha test in Equation 25. We leave off the hats on the factor returns in the regression of the returns of a given factor $b$ on the remaining factors other than $b$:

$$f_{b,t} \sim \alpha_b + \boldsymbol{\beta}_b' \mathbf{f}_{-b,t} + \epsilon_t \Rightarrow \hat{\alpha}_b = \bar{f}_b - \hat{\boldsymbol{\beta}}_b' \bar{\mathbf{f}}_{-b} = \hat{\mu}_b -' \hat{\boldsymbol{\beta}}_b' \hat{\boldsymbol{\mu}}_{-b} \tag{39}$$

We use the subscript to indicate the full set of factor returns $\mathbf{f}_{-b}$ or means of the factor returns $\boldsymbol{\mu}_{-b}$ to indicate the vectors that exclude the factor $b$. The factor loadings are the vector $\boldsymbol{\beta}_{-b}$, and the variable of interest is the factor-risk-adjusted return $\alpha_b$. This is an intercept test to see whether the factor return $\mu_b$ is large enough given its covariances with the other factors in the set of anomalies. Again, in a Bayesian sense, all factors will be relevant in some small way. In a frequentist sense, we are interested in ruling out factors that are statistically irrelevant. This is the usual sense of a factor anomaly, that we must be able to say that the hypothesis that $\alpha_b \neq 0$ is true with sufficient probability. The standard error of the estimate of $\alpha_b$ is equal to:

$$\begin{aligned} \text{se}\left(\hat{\alpha}_b\right) &= \frac{1}{\sqrt{T}} \sqrt{\sigma_\epsilon^2 \left(1 + \boldsymbol{\mu}_{-b}' \boldsymbol{\Sigma}_{-b}^{-1} \boldsymbol{\mu}_{-b}\right)} \\ &= \frac{1}{\sqrt{T}} \sqrt{\left(\sigma_{bb}^2 + \frac{1}{N^2} \sigma^2 \sum_i \left(\boldsymbol{\delta}_b' \boldsymbol{\gamma}_i\right)^2\right) \left(1 - R^2\right) \left(1 + SR^2\right)} \end{aligned} \tag{40}$$

We derive the expression for standard error in Equation 40 in Appendix B. Because in the time series regression we estimate the $K+1$ coefficients, we lose $K+1$ degrees of freedom and the estimate of asymptotic variance is

$$\begin{aligned} \text{se}\left(\hat{\alpha}_b\right) &= \frac{1}{\sqrt{T-K-1}} \sqrt{\sigma_\epsilon^2 \left(1 + \boldsymbol{\mu}_{-b}' \boldsymbol{\Sigma}_{-b}^{-1} \boldsymbol{\mu}_{-b}\right)} \\ &= \frac{1}{\sqrt{T-K-1}} \sqrt{\left(\sigma_{bb}^2 + \frac{1}{N^2} \sigma^2 \sum_i \left(\boldsymbol{\delta}_b' \boldsymbol{\gamma}_i\right)^2\right) \left(1 - R^2\right) \left(1 + SR^2\right)} \end{aligned} \tag{41}$$

For a large cross section of stocks $N$, the effect of residual risk $\varepsilon$ becomes small and the variance of the residual factor return risk $\epsilon$ is simply $\sigma_{bb}^2$. The estimate of the factor payoff in each period using only returns $\mathbf{r}_t$ and scores $\boldsymbol{\Gamma}_t$ at any given time $t$ becomes very precise. When the cross-section of firms $N$ is not so large, then the variance is greater, so that $\sigma_\epsilon^2 > \sigma_{bb}^2$, and equals the quantity under the radical in Equation 35, discussed above. In addition to these comparative statics, there are two additional drivers of the standard error of the time series test. The standard error now depends on the means and covariances of the *other* factor returns too. The first term in parentheses is an increase in power that comes from the fact that the other factors can in principle reduce the residual variation in the regression equation. The residual in the time series test is smaller than the residual in the cross section test by an amount equal to one minus the time series R-squared. The second term in parentheses is decrease in power, equal to one plus the in-sample maximum squared Sharpe Ratio (SR) of the other factor returns. This is the mean-variance optimal combination of the existing factors. When the existing factors are very powerful predictors of return, then the standard error in Equation 40 rises, as in Shanken (1992). It is harder to reject the null. This is the asymptotic standard error again in the sense that $\sigma^2_{bb}$, $\sigma^2$, $\boldsymbol{\mu}_{-b}$, and $\boldsymbol{\Sigma}_{-b}$ are not known to the investor or the econometrician in small samples. We consider the small sample properties with Monte Carlo simulations below. For now, we plot the power curve using Equation 40 in Figure 2, using the same parameters as in Figure 2 with one more parameter to vary:

$$SR = [0.25, \underline{0.42}, 0.50, 0.75, 1] \tag{42}$$

Underlined values are the base case and represents by the blue line in the graph.

The power curve exactly mirrors the results in Figure 1, but with the power shifted down by the in-sample Sharpe ratio. As before power rises more steeply with more firms and lower idiosyncratic variance. The gains from these two parameters are bounded. Power rises more quickly with more periods and lower factor variance. With a large number of periods and low factor variance, the time series test can detect small anomalies.

To these comparative statics, we now add the Sharpe ratio of the existing factors in the next section. While it is not immediately apparent in the comparison of Figure 1 and Figure 2, the time series tests are all shifted down somewhat.

#### 3.1.4 A Comparison of the Cross Section and Time Series Tests

It is immediately apparent that the standard error in Equation 40 contains a potential loss in power, when compared to the standard error in 35. If a new factor has no true connection to the time series payoffs of the incumbent set, so that the R-squared in 40 is zero, then the standard errors are related by the following formula:

$$\text{se}\left(\hat{\alpha_b}\right) = \text{se}\left(\hat{\mu_b}\right)\sqrt{\left(1 + \boldsymbol{\mu}'_{-b}\boldsymbol{\Sigma}^{-1}_{-b}\boldsymbol{\mu}_{-b}\right)} \tag{43}$$

Correspondingly, the relation between estimates of asymptotic standard error is

$$\hat{\text{se}}\left(\hat{\alpha_b}\right) = \hat{\text{se}}\left(\hat{\mu_b}\right)\sqrt{\frac{T-1}{T-K-1}}\sqrt{\left(1 + \boldsymbol{\mu}'_{-b}\boldsymbol{\Sigma}^{-1}_{-b}\boldsymbol{\mu}_{-b}\right)} \tag{44}$$

In other words, it is harder to find a new factor anomaly than it is to find a new score anomaly. We overlay the power curves in Figure 1 onto Figure 3 at various Sharpe ratios.

This has a nice intuition. When the predictive power of existing factors is large relative to the portfolio variance - the Sharpe ratio is high - the estimation of a covariances $\boldsymbol{\beta}_{-b}$ between the new factor $b$ and existing factors becomes a source of error that drives power down relative to the cross-section test. It is possible to connect the new factor to existing ones in a way that is spurious. For example, both momentum and CAPM beta suffered very poor returns in the market reversal of the spring of 2009, but are otherwise essentially uncorrelated. Similarly much of the overlap between CAPM beta and the ratio of price to book occurs in the late 1990s and early 2000s. If we consider the Sharpe ratio of momentum and the price-to-book to be high - as they are in US data - and for these to be a spurious correlations - which may be true - then it is easier to reject CAPM beta as a factor anomaly. It is this possibility which lowers the relative power of the test. In this sense, our argument is related to Ang, Liu, and Schwarz (2008). They focus on the loss of power that can come from aggregation. Forming portfolios improves the estimation of covariances but throws away information. In our context, it is the extra need - for risk averse investors - to compute covariances that diminishes their ability to find relevant anomalies.

#### 3.1.5 The Sharpe Ratio in US Data

This begs the question: Which line in Figure 3 in the comparison of cross section and time series tests is the relevant one? This depends on the size of the in-sample Sharpe ratio for some set of existing factors, like the Fama-French five factors. Power falls monotonically in two ways: as the factor set increases, and as the horizon rises, from daily to weekly to monthly to annual returns.

First, a larger factor set by definition means a higher in-sample Sharpe ratio. The Fama and French (2015) five-factor model, for example, will mechanically reject more potential factor anomalies than the Fama and French (1993) three-factor model. So, there is a practical limit on the number of factor anomalies that can be discovered. By contrast, the number of score anomalies that can be discovered is only limited by the number of stocks N and the number of periods T, but it is not otherwise constrained by the predictive power of existing cross-sectionally orthogonal anomalies.

Second, a property of the Sharpe ratio is that it rises as the return horizon increases. This is because it is the ratio of mean to standard deviation. While the mean increases linearly in $T$, the standard deviation increases linearly in $\sqrt{T}$.

The qualitative impact of these two effects is clear. We use US data to convert qualitative to quantitative effects on econometric power. The Sharpe ratio of the set of standard, existing factors is reasonably high, when we use monthly returns, as is common practice in the academic literature. The bold line in Figure 3 shows the loss in power of a test that uses a standard set of the Fama-French five factors, momentum, and short-term reversal and a monthly return horizon. The empirical moments of these portfolios over the period from 1963:07 to 2016:03 are shown in Table 1. In Table 1, we also include a Fama-French style portfolio using market beta, using the beta estimation approach of Frazzini and Pedersen (2014). This portfolio divides the CRSP universe into small and large, using the median size among NYSE stocks as the breakpoints and further divides small and large stocks into three terciles according to market beta, again using NYSE breakpoints.

The individual annual Sharpe ratios range from 0.29 to 0.56, and the in-sample optimal annual Sharpe ratio of the seven factor returns, excluding market beta, is 1.45. The corresponding quarterly Sharpe is 0.73, the corresponding monthly Sharpe is 0.42 which is what we show in bold in Figure 3, and the corresponding daily Sharpe is 0.09.

#### 3.1.6 Choosing Return Horizon

Plugging these Sharpe ratios into Equation 43, we can compute the loss in power. To fix ideas, we use $T = 200, N = 250, \sigma_{bb} = 3, \sigma_{\varepsilon} = 3.27$. For monthly tests, which is the standard in the literature and which we indicate in bold in Figure 3, the maximum loss in power is 22.4 percent for a new factor that has a true annual Sharpe ratio of 0.67 – in other words, a very strong anomaly. Annual return horizons are rarely used, likely because they involve an intuitively dramatic reduction in power, with the maximum loss at 73.0 percent. Quarterly returns are occasionally used. For example, see

Cederburg and O'Doherty (2016). Quarterly returns lie in between, with a substantial maximum loss at 45.1 percent. This is an effective way to stack the deck against rejecting the null. Meanwhile, daily returns largely solve the power problem with a loss in power of only 1.4 percent at maximum.

To provide an illustration, we compute alphas in Table 2 for six of the Fama-French style factor portfolios summarized in Table 1. We leave out reversal, which has a payoff that is very short-lived. We use monthly, weekly, and daily returns and the Fama-French five-factor model in Panels A, B, and C, respectively. Our prior from Equation 43 and the properties of the Sharpe ratio as horizon falls is that the daily tests will be the most powerful, delivering lower standard errors and higher t-statistics on average. Table 2 bears this out. We are focused on the first two columns, which display coefficient estimates, standard errors, and t-statistics for the intercept, or alpha of each portfolio. In the monthly tests in Panel A, the average standard error of the alpha estimate for these six portfolios, at 1.25 percent annualized, is 25 percent higher than in the daily tests, shown in Panel C. The average t-statistic, at 4.38, is 44 percent higher. Notably, the conclusion from Fama and French (2015) that the price-to-book ratio is subsumed by the other factors is reversed in daily data. HML has a t-statistic of 0 in monthly data and a t-statistic of 2.19 in daily data. It retains a statistically significant alpha of 1.96 percent annualized.

To illustrate the difference in power further, we repeat this exercise for the five-by-five portfolios from Ken French's data library that double sort on size and book-to-market, profitability, and investment. We focus on these three, because daily returns are available, and the sorting variables are updated monthly, making comparisons among the horizons valid. We focus on the top and bottom sets of five portfolios, where we expect to find alphas different from zero, or 30 in all. For each of these 30 portfolios, we conduct a time-series test using the Fama-French five-factor model as in Table 2, excluding the factor of interest, recording the absolute annualized alpha coefficient, the annualized standard error, the absolute t-statistic, and the p-value for the resulting alpha, using first daily returns, then weekly, and then monthly returns. We plot the cumulative distribution of these four values in the four panels of Figure 4. The absolute annualized coefficient ranges from 0.01 percent to 22.8 percent, the annualized standard error of the coefficient ranges from 0.09 percent to 0.51 percent, the absolute t-statistic ranges from zero to 25.3, and the p-value ranges from zero to 1.0. Daily alphas reject the null of zero alpha relative to the Fama-French 5-factor model 63.3 percent of the time, relative to the 5% size of the test. Meanwhile, weekly alphas reject 56.7 percent.

And, monthly alphas reject only 40.0 percent. There are far more novel asset pricing anomalies to be discovered in daily data.

Why not then simply use the more powerful daily, or better yet intraday, returns? There are two reasons why the literature has tended to use monthly returns, beyond the ease of computation. One is that higher frequency covariances may understate the true, tradeable covariances, because of asynchronous correlation. For example, if some stocks simply do not trade every day, or trade in a way that is slow to incorporate aggregate information, then the estimated betas to the Fama-French five factors will be biased downward in daily regressions. [MB FILL IN Citations] Rather than use monthly returns, Scholes and Williams (1977) recommend aggregating lagged and leading covariances, effectively moving toward weekly regressions. The other is that daily returns may understate or overstate the tradeable annual returns of a given anomaly, as emphasized by Liu and Strong (2008). This is a closely related point. Instead of asynchronous correlation between a given anomaly and the Fama-French portfolios, this is the autocorrelation of anomaly returns. Positive autocorrelation, or anomaly momentum, means that annualized monthly returns are higher than annualized daily returns. Negative anomaly autocorrelation, or anomaly reversal, means that annualized monthly returns are lower than annualized daily returns. Both of these effects can cause biased inference in identifying tradeable anomalies, moving from monthly to more powerful daily regressions. Both of these are reasons that the analysis of shorter horizons might bias inference, but it is important to note that shifting to much longer horizons comes at the expense of power.

Fortunately, it is possible to split these effects apart and make a modest suggestion for best practice – at least if the anomalies portfolio from Ken French's data library that we analyzed in Figure 4 are a useful guide. The p-value results in panel D of Figure 4 come from two separate effects. The p-value is higher because of higher power – that is the reduction in average annualized standard errors in Panel B. In addition, the p-value is higher because of potential bias from asynchronous correlations and factor momentum and reversal – that is reflected in the increase the average annualized absolute coefficient in Panel A. What is the right tradeoff between power and bias? To answer this question, we plot the coefficients and standard errors in Figure 5. We scale the annualized coefficient estimate for each horizon by the 50-day annualized coefficient estimate, on the argument that the bias arising from trading effects is negligible at that point. We scale the annualized standard error by the 1-day estimate, where power is maximized. Panel A of Figure 5 shows the averages of these scaled values across our 30 portfolios as a function of return horizon, from daily returns to overlapping 50-trading-day returns. We repeat the exercise using medians instead of averages in Panel B.

Note that the average (and median) coefficient decreases as we move from daily to ten days and then levels off. This in principle might reflect biased inference, though it could also come from a more accurate estimation of covariances. Meanwhile, the average (and median) standard error continues to rise from daily horizons to 50-day horizons and beyond. If the data from Ken French's data library are suggestive of typical anomalies, our analysis suggests that the problem of bias for these factors appears to be largely solved at 10 trading days, or two weeks, and completely solved by 20 days. At ten days, the maximum loss in power is 11.6 percent, which still represents a considerable loss of power in time series tests relative to cross-sectional tests, but it is more modest than in the standard practice of analyzing monthly returns, with a loss of 22.4 percent. In what follows, we continue to use monthly returns, as the standard in the literature, but an important conclusion is that, by shifting from monthly to two-week returns, there appears to be a free increase in the power of standard asset pricing tests.

### 3.2 Small Sample Power Curves

The power curves in Figure 3 assume that all of the distributional parameters in the return generating process in Equation 1 are known by the investor. In practice, they are not. Sample estimates must replace their respective population values. We estimate the small sample properties of Equations 35 and 40 by running Monte Carlo simulations.

In the base case, we consider $N = 250$ securities and $T = 50$ periods. We use the seven factor portfolios discussed in last section. We first draw factor coefficients $\Gamma_i$ from normal distribution $N(0,1)$ for each stock $i$. Then we draw factor returns from a multivariate normal distribution $N\left(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}\right)$ with mean and variance estimated from US market data. Finally, we draw the idiosyncratic return from a normal distribution $N\left(0, \hat{\sigma_\varepsilon^2}\right)$ with variance estimated from US market data $\hat{\sigma_\varepsilon} = 3.27\%$ and calculate return of each stock. We repeat this Monte-Carlo simulation for 2000 times. We plot the simulated power curves against their theoretical values in Figure 6 and confirm they are consistent with each other. Then we plot the theoretical small sample distribution in Figures 7, 8 and 9.

The patterns we see in the asymptotic power curves also hold in the small sample power curves. Power of both tests get stronger as there are more securities, longer time periods, smaller variances of factor and idiosyncratic returns. The number of time periods and the test factor's variance have a more significant effect on the power of both tests. The Sharpe ratio drives the wedge between power of the cross section and time series tests. The difference in power for small sample is 23.7% at $SR = 0.69$, which is larger than that for asymptotic result, 22.4%, because of additional loss of degrees of freedom in the time series test.

## 4 Conclusion

What is an anomaly? Empirical asset pricing papers that aim to establish an anomaly often rely on two types of tests. One is the cross-sectional test in the spirit of Fama and MacBeth (1973) and famously used in Fama and French (1992), and the other is the times series test in the spirit of Jensen (1968) and popularized by Fama and French (1993). We consider these tests from two different points of view: relevance and power. The cross-section test is relevant to a risk neutral investor facing a simple form of transaction costs. The time series test is relevant to a risk-averse investor facing no transaction costs. Meanwhile, both are relevant in the more general case of risk aversion and transaction costs. Next, we show that the time series test can be inherently lower powered. Given that most professional investors face meaningful transaction costs and most commercial portfolio optimizers target a mean-variance objective, we believe that a test that passes either of the two tests can be considered an anomaly - in the sense that it is practically relevant for a large class of investors. Viewed in this light, the literature on empirical asset pricing has the potential to identify a richer set of interesting anomalies than are contained in the Fama-French 3-Factor model or even the newer 5-factor model. Unlike the time series analysis in Fama and French (1993), the cross-sectional framework in Fama and French (1992) has a higher upper limit on the number of relevant and statistically significant factors.

## References


Ang, A., J. Liu, and K. Schwarz. 2008. Using Individual Stocks or Portfolios in Tests of Factor Models.

Bailey, D. H., and M. López de Prado. 2014. The Deflated Sharpe Ratio: Correcting for Selection Bias, Backtest Overfitting, and Non-Normality. *Journal of Portfolio Management* 40:94–107.

Brennan, M. J., E. S. Schwartz, and R. Lagnado. 1997. Strategic asset allocation. *Journal of Economic Dynamics and Control* 21:1377–1403.

Campbell, J. Y., and L. M. Viceira. 1999. Consumption and Portfolio Decisions when Expected Returns are Time Varying. *Quarterly Journal of Economics* 114:433–495.

Cederburg, S., and M. S. O'Doherty. 2016. Does It Pay to Bet Against Beta? On the Conditional Performance of the Beta Anomaly. *Journal of Finance* 71:737–774.

Cochrane, J. H. 2009. *Asset Pricing.* Princeton University Press.

Daniel, K., and S. Titman. 1997. Evidence on the Characteristics of Cross Sectional Variation in Stock Returns. *Journal of Finance* 52:1.

Davis, J. L., E. F. Fama, and K. R. French. 2000. Characteristics, covariances, and average returns: 1929 to 1997. *Journal of Finance* 55:389–406.

Fama, E. F. 1998. Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics* 49:283–306.

Fama, E. F., and K. R. French. 1992. The Cross-Section of Expected Stock Returns. *Journal of Finance* 47:427–465.

Fama, E. F., and K. R. French. 1993. Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics* 33:3–56.

Fama, E. F., and K. R. French. 2008. Dissecting anomalies. *Journal of Finance* 63:1653–1678.

Fama, E. F., and K. R. French. 2015. A Five-Factor Asset Pricing Model. *Journal of Financial Economics* 116:1–22.

Fama, E. F., and K. R. French. 2016. Dissecting Anomalies with a Five-Factor Model. *Review of Financial Studies* 29:69–103.

Fama, E. F., and J. D. MacBeth. 1973. Risk, Return, and Equilibrium: Empirical Tests. *Journal of Political Economy* 81:607–636.

Frazzini, A., and L. H. Pedersen. 2014. Betting against beta. *Journal of Financial Economics* 111:1–25.

Garleanu, N., and L. H. Pedersen. 2013. Dynamic Trading with Predictable Returns and Transaction Costs. *Journal of Finance* 68:2309–2340.

Harvey, C. R., Y. Liu, and H. Zhu. 2015. ... And the cross-section of expected returns. *Review of Financial Studies* 29:5–68.

Hoberg, G., and I. Welch. 2009. Optimized vs. Sort-Based Portfolios.

Jegadeesh, N., and S. Titman. 1993. Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *Journal of Finance* 48:65–91.

Jensen, M. C. 1968. The performance of mutual funds in the period 1945-1964. *Journal of Finance* 23:389–416.

Levi, Y., and I. Welch. 2014. Long-Term Capital Budgeting.

Lewellen, J. 2015. The Cross-section of Expected Stock Returns. *Critical Finance Review* 4:1–44.

Lintner, J. 1965. Security Prices, Risk, and Maximal Gains From Diversification. *Journal of Finance* 20:587–615.

Liu, W., and N. Strong. 2008. Biases in Decomposing Holding Period Portfolio Returns. *Review of Financial Studies* 44:0–31.

Loughran, T., and J. R. Ritter. 2000. Uniformly least powerful tests of market efficiency. *Journal of Financial Economics* 55:361–389.

Markowitz, H. 1952. Portfolio Selection. *Journal of Finance* 7:77–91.

Mclean, R. D., and J. Pontiff. 2016. Does Academic Research Destroy Stock Return Predictability? *Journal of Finance* 71:5–32.

Novy-marx, R. 2012. Quality Investing.

Novy-marx, R. 2016. Testing strategies based on multiple signals.

Scholes, M., and J. Williams. 1977. Estimating betas from nonsynchronous data. *Journal of financial economics* 5:309–327.

Shanken, J. 1992. On the Estimation of Beta-Pricing Models. *Review of Financial Studies* 5:1–33.

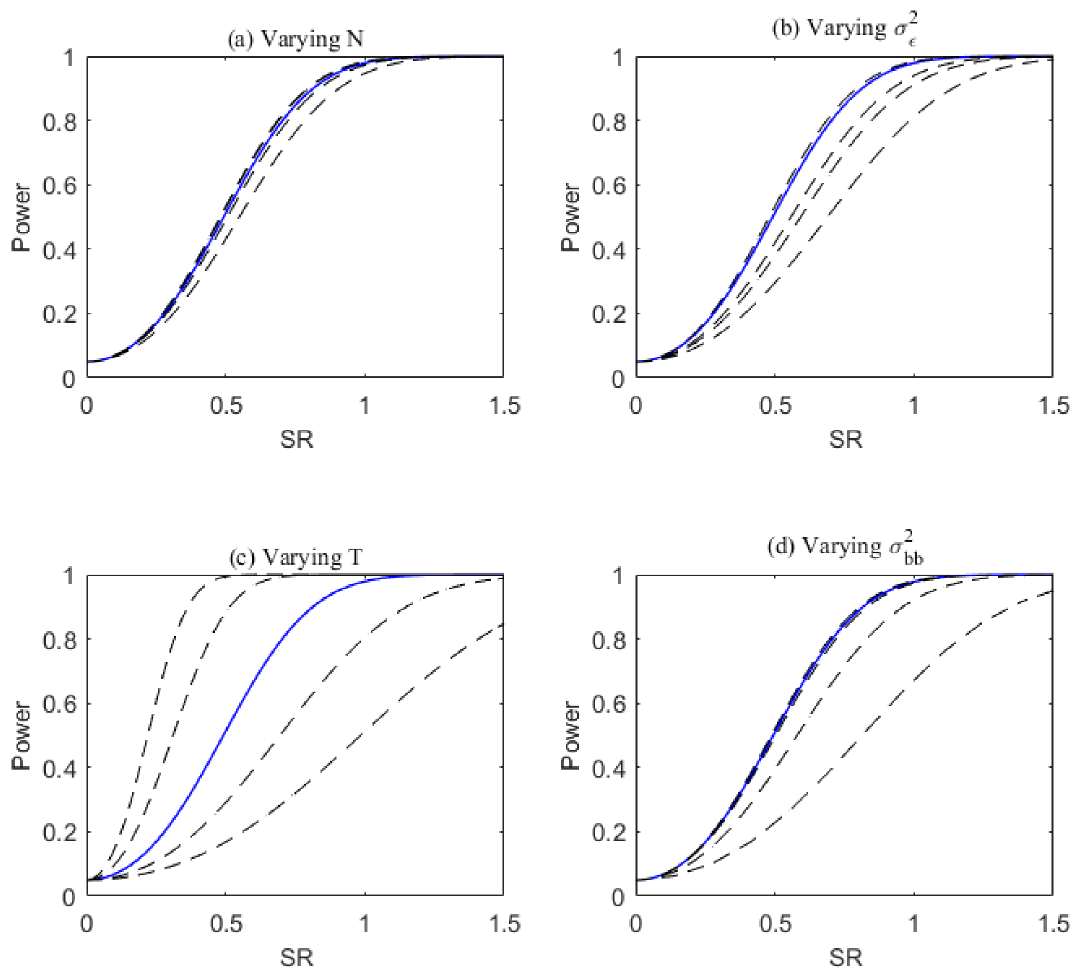
Sharpe, W. F. 1964. Capital Asset Prices: a Theory of Market Equilibrium Under Conditions of Risk. *Journal of Finance* 19:425–442.

Simin, T. 2008. The Poor Predictive Performance of Asset Pricing Models. *Journal of Financial and Quantitative Analysis* 43:355–380.

Stambaugh, R. F., J. Yu, and Y. Yuan. 2012. The short of it: Investor sentiment and anomalies. *Journal of Financial Economics* 104:288–302.

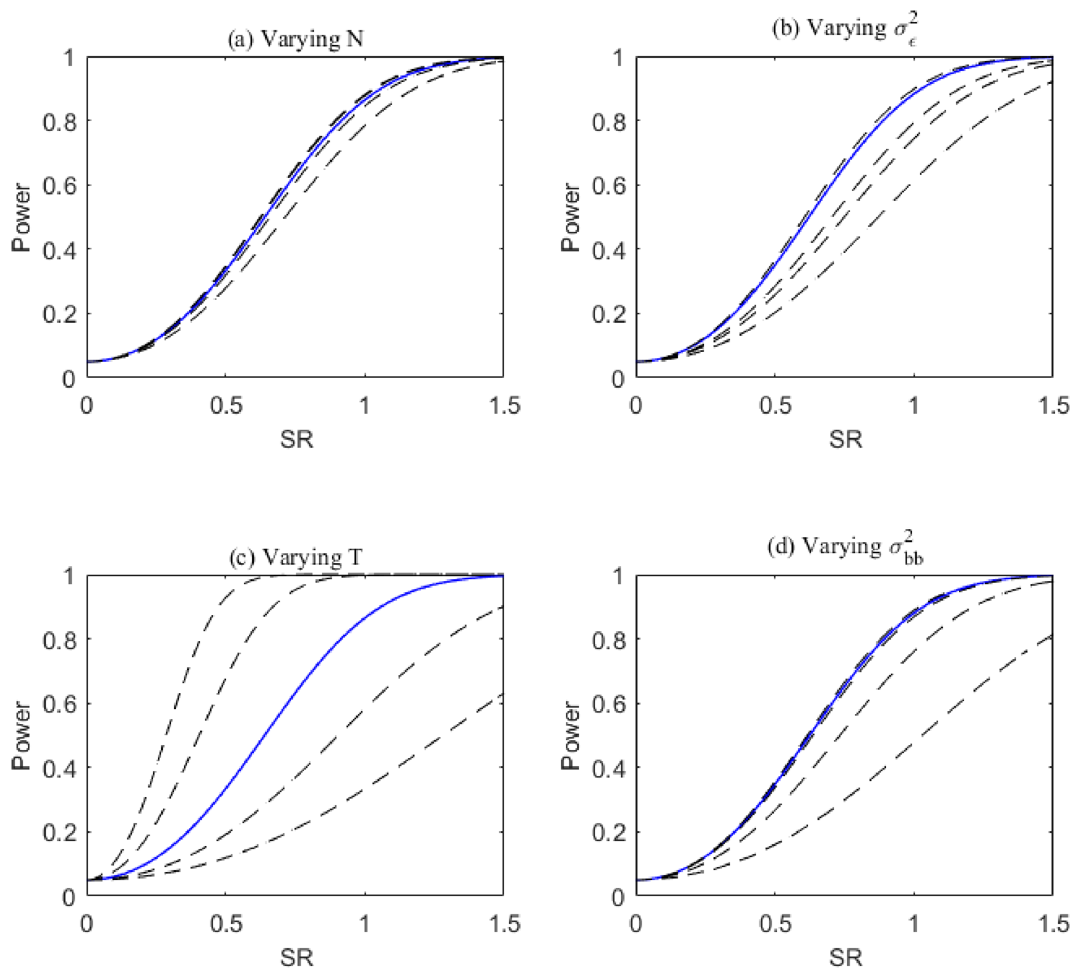
Tobin, J. 1958. Liquidity Preference as Behavior Towards Risk. *Review of Economic Studies* 25:65.

Figure 1: Asymptotic Power Curve of The Cross Section Test



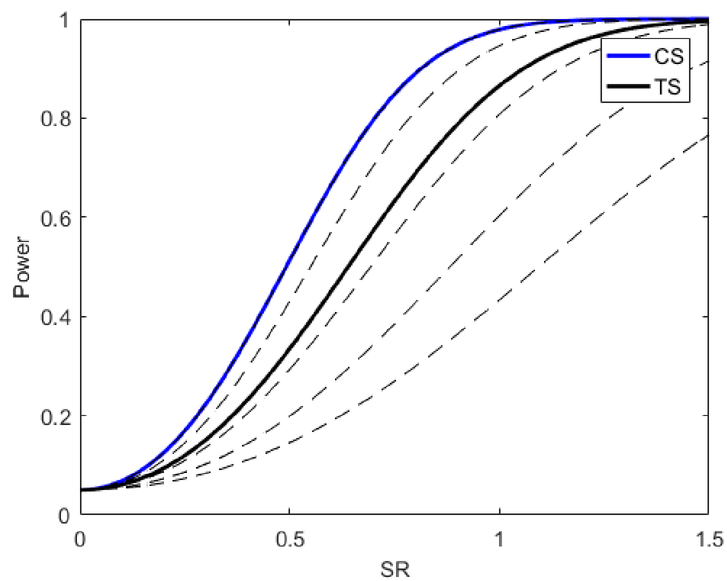

**Note**: The annual Sharpe ratio of the test factor is calculated as $SR = \frac{\mu_b}{\sigma_{bb}} \cdot \sqrt{12}$.

Figure 2: Asymptotic Power Curve of The Time Series Test




**Note**: The annual Sharpe ratio of the test factor is calculated as $SR = \frac{\mu_b}{\sigma_{bb}} \cdot \sqrt{12}$.

Figure 3: Asymptotic Power Curve of the CS and TS Test at Various Sharpe Ratios of the Incumbent Factor Model




**Note**: The annual Sharpe ratio of the test factor is calculated as $SR = \frac{\mu_b}{\sigma_{bb}} \cdot \sqrt{12}$.

Figure 4: Monthly, Weekly, and Daily TS Test Statistics

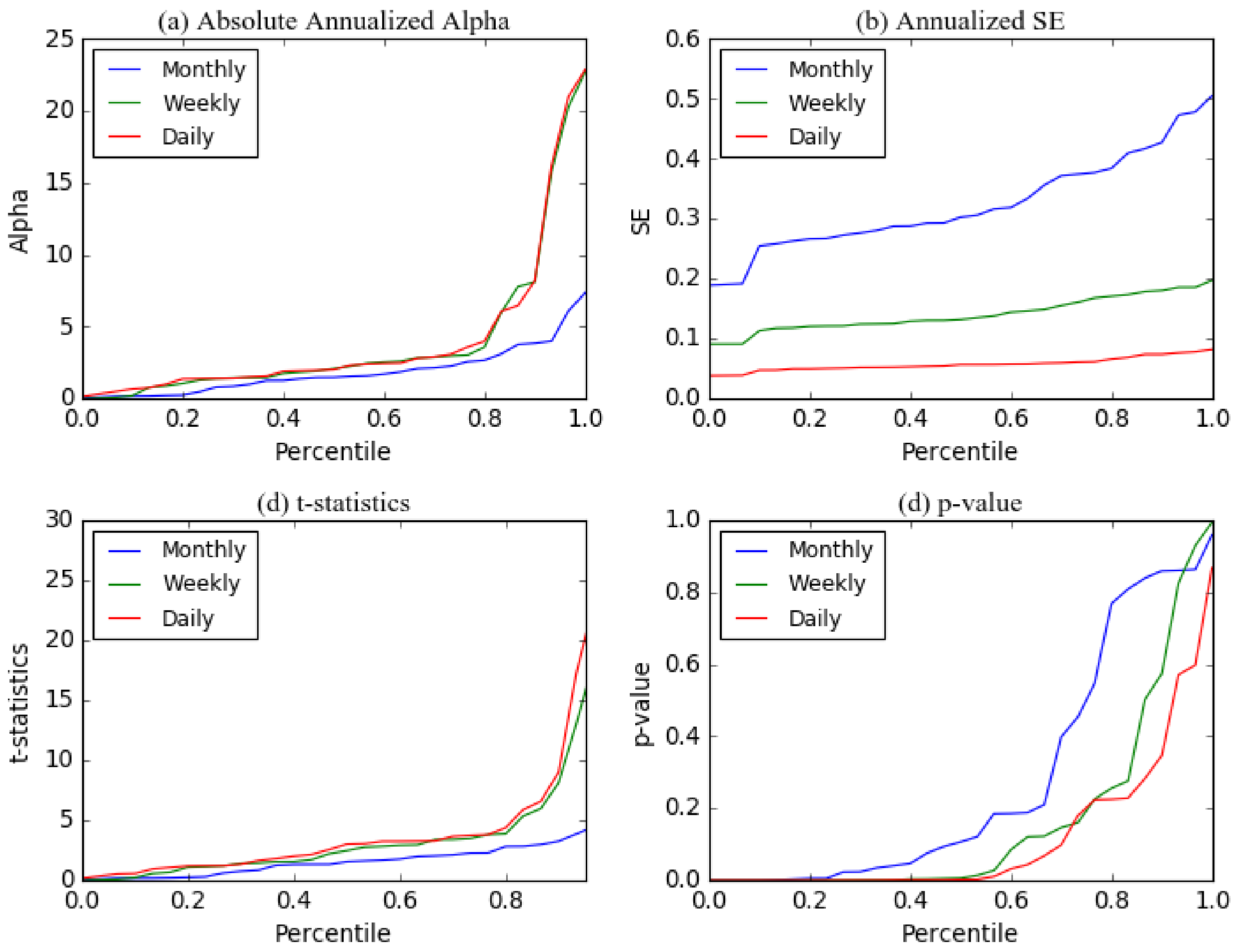


Figure 5: Bias and Power: Annualized Alphas By Return Horizon

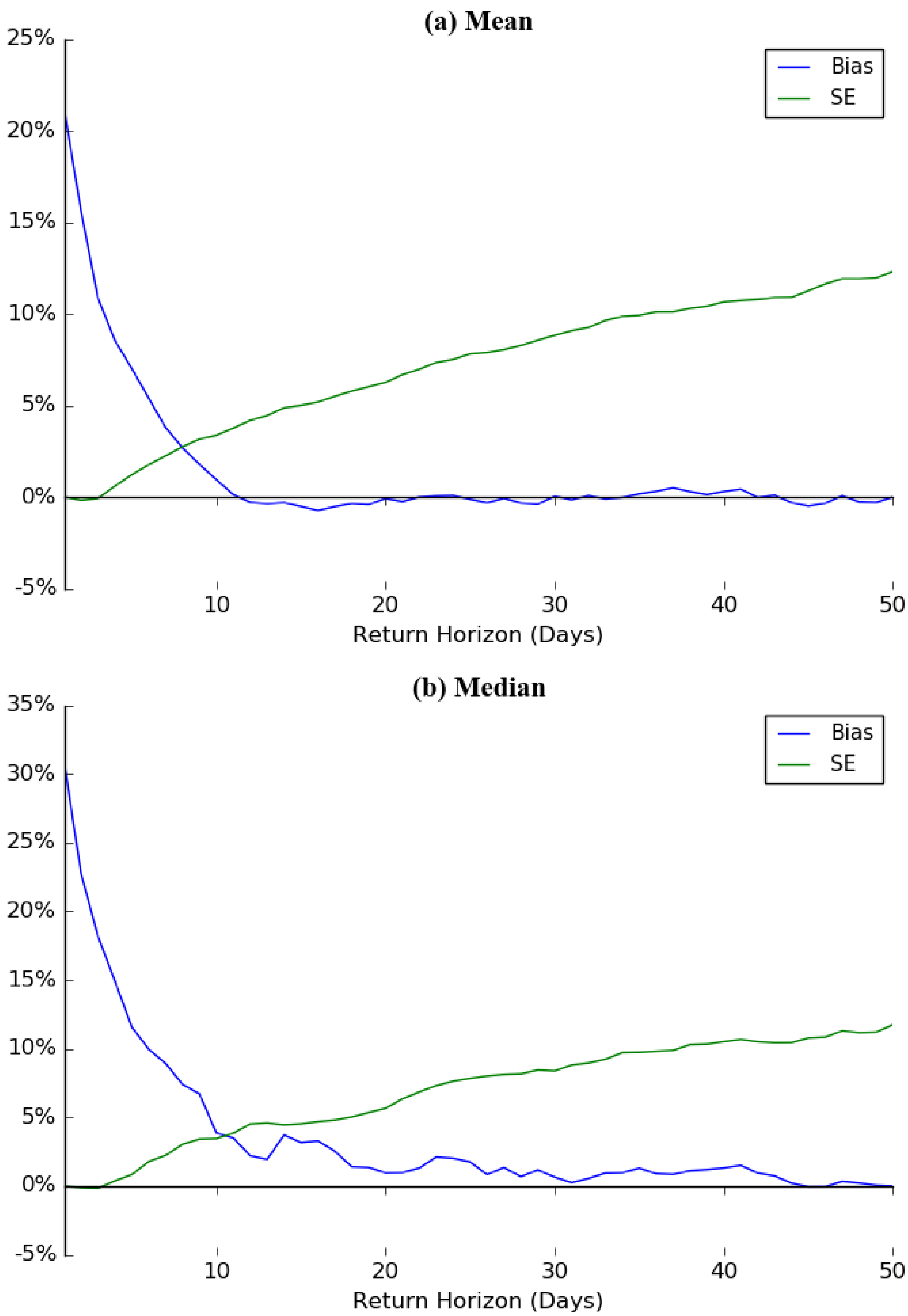
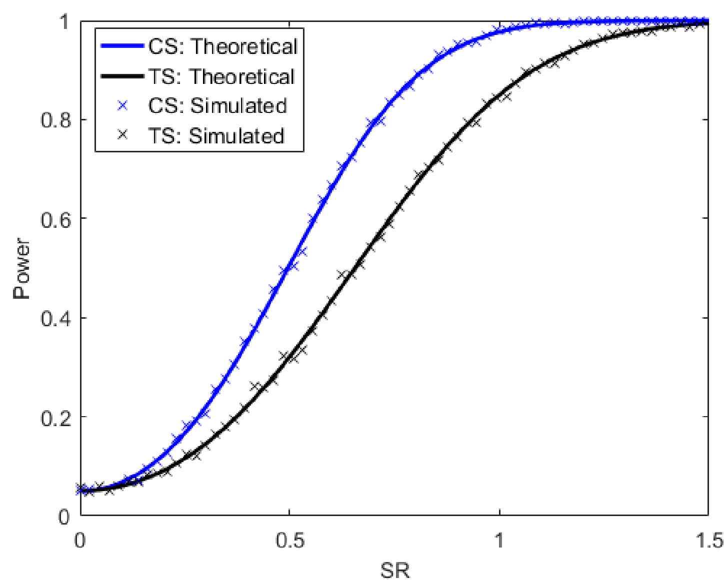
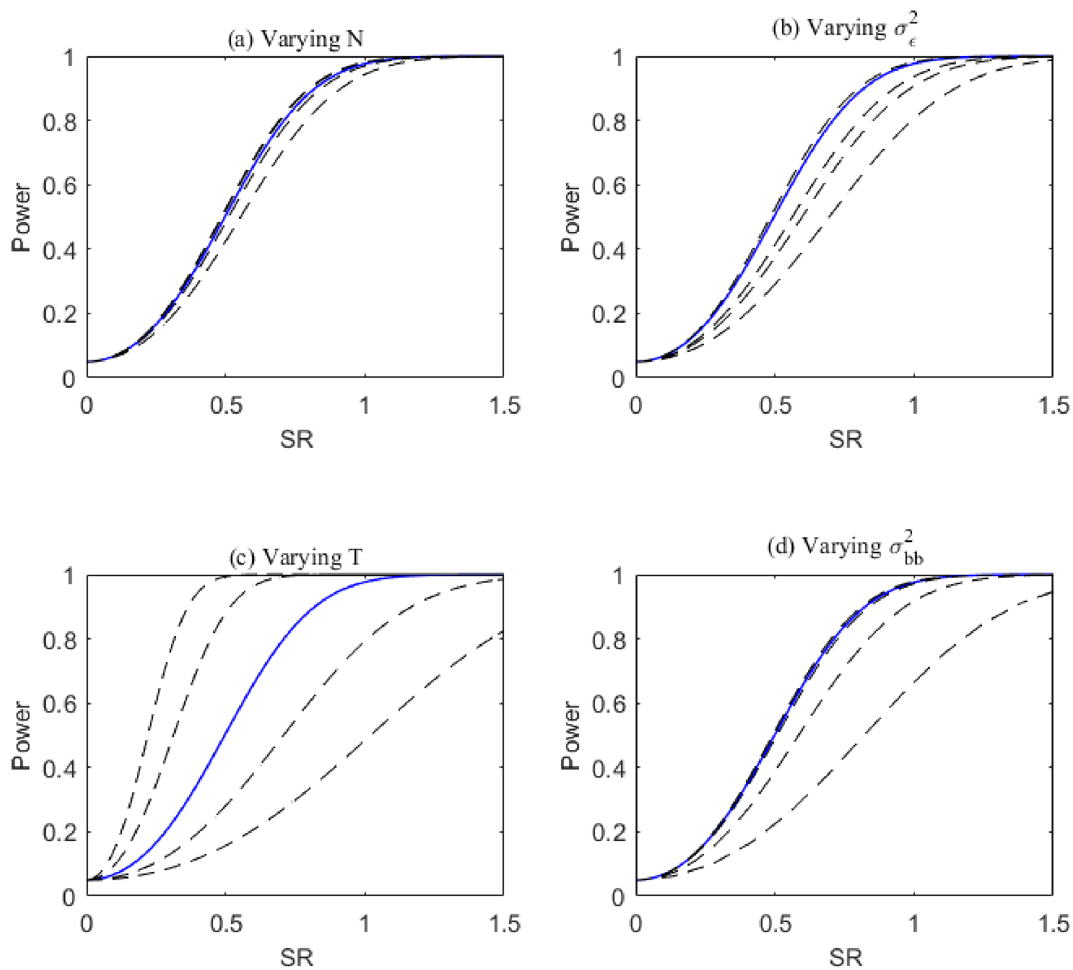


Figure 6: Small Sample and Simulated Power Curve of the CS and TS Test



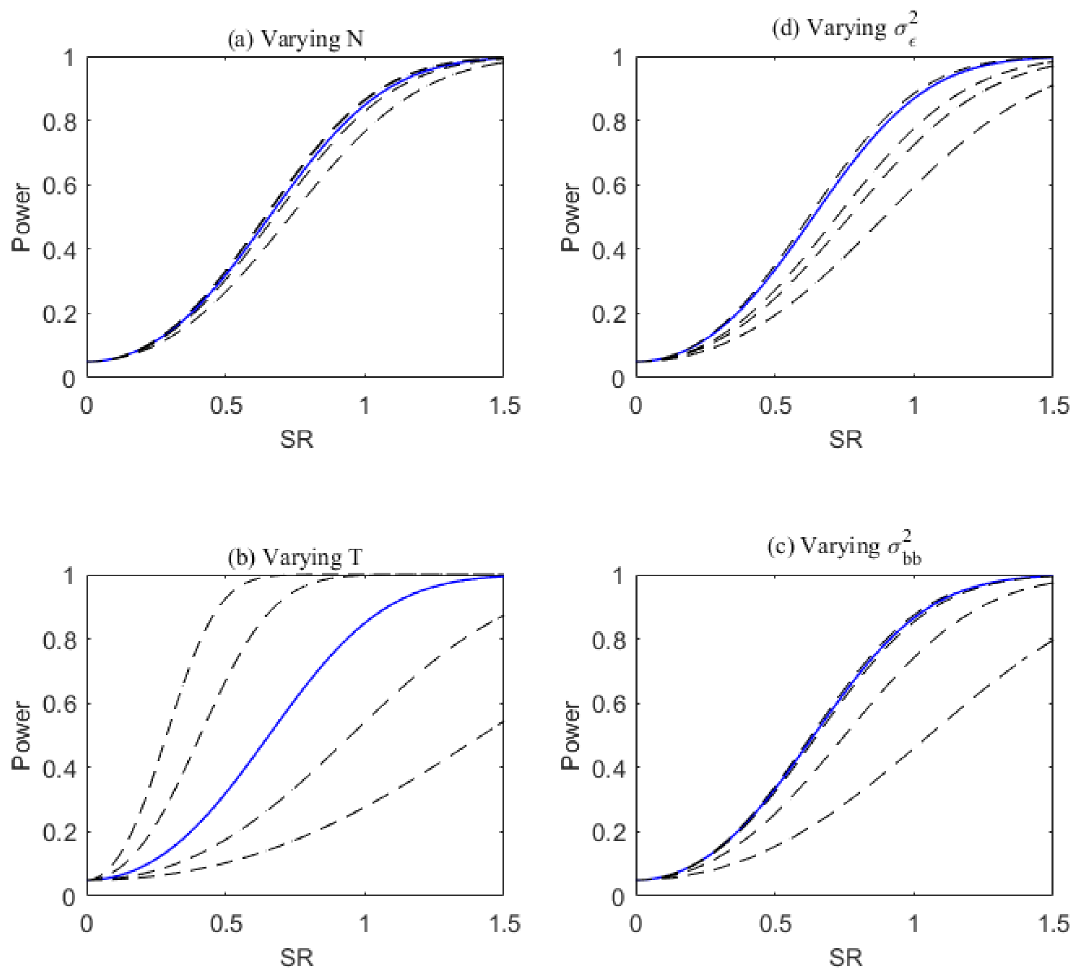

**Note**: The annual Sharpe ratio of the test factor is calculated as $SR = \frac{\mu_b}{\sigma_{bb}} \cdot \sqrt{12}$.

Figure 7: Small Sample Power Curve of the CS Test



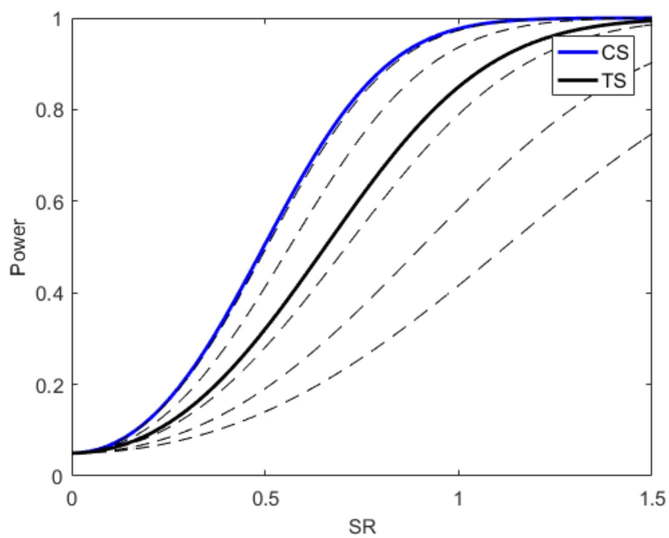

**Note**: The annual Sharpe ratio of the test factor is calculated as $SR = \frac{\mu_b}{\sigma_{bb}} \cdot \sqrt{12}$.

Figure 8: Small Sample Power Curve of the TS Test




**Note**: The annual Sharpe ratio of the test factor is calculated as $SR = \frac{\mu_b}{\sigma_{bb}} \cdot \sqrt{12}$.

Figure 9: Small Sample Power Curve of the CS and TS Test at Various Sharpe Ratios of the Incumbent Factor Model




**Note**: The annual Sharpe ratio of the test factor is calculated as $SR = \frac{\mu_b}{\sigma_{bb}} \cdot \sqrt{12}$.

Table 1: Empirical Moments for Factor Returns (1963:07 to 2016:03)

(a) Mean and variance

| | MKT | SMB | HML | RMW | CMA | MOM | STREV | RISK |
|---|---|---|---|---|---|---|---|---|
| Average | 0.50 | 0.25 | 0.34 | 0.25 | 0.31 | 0.69 | 0.48 | -0.22 |
| Annualized | 5.98 | 3.05 | 4.11 | 3.06 | 3.71 | 8.29 | 5.81 | -2.63 |
| SD | 4.45 | 3.06 | 2.85 | 2.12 | 2.01 | 4.24 | 3.13 | 4.65 |
| Annualized | 15.41 | 10.62 | 9.89 | 7.35 | 6.96 | 14.67 | 10.84 | 16.10 |
| Sharpe Ratio | 0.11 | 0.08 | 0.12 | 0.12 | 0.15 | 0.16 | 0.15 | -0.05 |
| Annualized | 0.39 | 0.29 | 0.42 | 0.42 | 0.53 | 0.56 | 0.54 | -0.16 |

(b) Covariance

| | MKT | SMB | HML | RMW | CMA | MOM | STREV | RISK |
|---|---|---|---|---|---|---|---|---|
| MKT | 19.75 | 3.82 | -3.77 | -1.93 | -3.46 | -2.44 | 3.96 | 14.79 |
| SMB | 3.82 | 9.38 | -1.01 | -2.35 | -0.70 | -0.30 | 1.51 | 3.60 |
| HML | -3.77 | -1.01 | 8.13 | 0.53 | 4.02 | -2.00 | -0.10 | -4.68 |
| RMW | -1.93 | -2.35 | 0.53 | 4.49 | -0.36 | 0.82 | -0.41 | -1.97 |
| CMA | -3.46 | -0.70 | 4.02 | -0.36 | 4.03 | -0.08 | -0.81 | -4.19 |
| MOM | -2.44 | -0.30 | -2.00 | 0.82 | -0.08 | 17.92 | -3.87 | -4.16 |
| STREV | 3.96 | 1.51 | -0.10 | -0.41 | -0.81 | -3.87 | 9.77 | 3.22 |
| RISK | 14.79 | 3.60 | -4.68 | -1.97 | -4.19 | -4.16 | 3.22 | 21.57 |

Table 2: Time Series Tests: Monthly, Weekly, Daily

(a) Monthly

| | Intercept | | Mkt_RF | | SMB | | HML | | RMW | | CMA | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | coef<br>t | se | coef<br>t | se | coef<br>t | se | coef<br>t | se | coef<br>t | se | coef<br>t | se |
| SMB | 3.65 | (1.36) | 0.13 | (0.03) | | | 0.06 | (0.05) | -0.42 | (0.05) | -0.12 | (0.08) |
| | [2.68] | | [4.55] | | | | [1.06] | | [-8.23] | | [-1.52] | |
| HML | 0.00 | (0.99) | 0.02 | (0.02) | 0.03 | (0.03) | | | 0.14 | (0.04) | 1.01 | (0.04) |
| | [0.00] | | [0.99] | | [1.06] | | | | [3.55] | | [23.44] | |
| RMW | 4.72 | (0.98) | -0.10 | (0.02) | -0.22 | (0.03) | 0.14 | (0.04) | | | -0.29 | (0.06) |
| | [4.82] | | [-5.13] | | [-8.23] | | [3.55] | | | | [-5.10] | |
| CMA | 2.74 | (0.66) | -0.11 | (0.01) | -0.03 | (0.02) | 0.46 | (0.02) | -0.13 | (0.03) | | |
| | [4.17] | | [-8.33] | | [-1.52] | | [23.44] | | [-5.10] | | | |
| MOM | 8.69 | (1.98) | -0.13 | (0.04) | 0.07 | (0.06) | -0.52 | (0.08) | 0.24 | (0.08) | 0.39 | (0.12) |
| | [4.38] | | [-3.19] | | [1.16] | | [-6.64] | | [3.02] | | [3.29] | |
| RISK | -3.40 | (1.54) | 0.63 | (0.03) | 0.04 | (0.04) | 0.05 | (0.06) | -0.21 | (0.06) | -0.53 | (0.09) |
| | [-2.20] | | [19.75] | | [0.86] | | [0.77] | | [-3.51] | | [-5.82] | |
| Average | 3.87 | (1.25) | 0.19 | (0.03) | 0.06 | (0.03) | 0.20 | (0.04) | 0.19 | (0.04) | 0.39 | (0.06) |
| | [3.04] | | [6.99] | | [2.14] | | [5.91] | | [3.90] | | [6.53] | |

(b) Weekly

| | Intercept | | Mkt_RF | | SMB | | HML | | RMW | | CMA | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | coef<br>t | se | coef<br>t | se | coef<br>t | se | coef<br>t | se | coef<br>t | se | coef<br>t | se |
| SMB | 3.63 | (1.13) | 0.00 | (0.01) | | | 0.01 | (0.02) | -0.41 | (0.02) | -0.07 | (0.03) |
| | [3.20] | | [-0.25] | | | | [0.28] | | [-17.24] | | [-2.00] | |
| HML | 1.54 | (0.97) | 0.00 | (0.01) | 0.00 | (0.02) | | | -0.08 | (0.02) | 0.84 | (0.02) |
| | [1.58] | | [0.31] | | [0.28] | | | | [-3.52] | | [36.07] | |
| RMW | 4.61 | (0.84) | -0.09 | (0.01) | -0.23 | (0.01) | -0.06 | (0.02) | | | -0.06 | (0.02) |
| | [5.47] | | [-10.79] | | [-17.24] | | [-3.52] | | | | [-2.37] | |
| CMA | 2.80 | (0.65) | -0.12 | (0.01) | -0.02 | (0.01) | 0.38 | (0.01) | -0.03 | (0.01) | | |
| | [4.30] | | [-20.04] | | [-2.00] | | [36.07] | | [-2.37] | | | |
| MOM | 8.23 | (1.69) | -0.07 | (0.02) | 0.05 | (0.03) | -0.64 | (0.03) | 0.13 | (0.04) | 0.60 | (0.05) |
| | [4.88] | | [-4.47] | | [1.75] | | [-19.61] | | [3.43] | | [12.29] | |
| RISK | -4.19 | (1.39) | 0.72 | (0.01) | 0.01 | (0.02) | 0.17 | (0.03) | -0.17 | (0.03) | -0.62 | (0.04) |
| | [-3.01] | | [53.18] | | [0.40] | | [6.25] | | [-5.62] | | [-15.43] | |
| Average | 4.17 | (1.11) | 0.17 | (0.01) | 0.05 | (0.02) | 0.21 | (0.02) | 0.14 | (0.02) | 0.36 | (0.03) |
| | [3.74] | | [14.84] | | [3.61] | | [10.96] | | [5.36] | | [11.36] | |

(c) Daily

| | Intercept | | Mkt_RF | | SMB | | HML | | RMW | | CMA | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | coef<br>t | se | coef<br>t | se | coef<br>t | se | coef<br>t | se | coef<br>t | se | coef<br>t | se |
| SMB | 4.30 | (1.05) | -0.11 | (0.00) | | | 0.04 | (0.01) | -0.46 | (0.01) | -0.07 | (0.01) |
| | [4.08] | | [-22.79] | | | | [4.32] | | [-39.87] | | [-4.71] | |
| HML | 1.96 | (0.90) | -0.01 | (0.00) | 0.03 | (0.01) | | | -0.12 | (0.01) | 0.76 | (0.01) |
| | [2.19] | | [-2.12] | | [4.32] | | | | [-11.34] | | [72.89] | |
| RMW | 4.44 | (0.74) | -0.10 | (0.00) | -0.23 | (0.01) | -0.08 | (0.01) | | | 0.04 | (0.01) |
| | [6.04] | | [-31.03] | | [-39.87] | | [-11.34] | | | | [4.03] | |
| CMA | 2.52 | (0.62) | -0.10 | (0.00) | -0.02 | (0.01) | 0.37 | (0.01) | 0.03 | (0.01) | | |
| | [4.05] | | [-36.69] | | [-4.71] | | [72.89] | | [4.03] | | | |
| MOM | 8.30 | (1.42) | -0.07 | (0.01) | 0.09 | (0.01) | -0.54 | (0.01) | 0.18 | (0.02) | 0.39 | (0.02) |
| | [5.85] | | [-10.87] | | [7.75] | | [-39.87] | | [10.94] | | [20.01] | |
| RISK | -5.24 | (1.30) | 0.84 | (0.01) | -0.09 | (0.01) | 0.22 | (0.01) | -0.14 | (0.02) | -0.56 | (0.02) |
| | [-4.05] | | [143.26] | | [-8.26] | | [17.70] | | [-9.52] | | [-31.12] | |
| Average | 4.46 | (1.00) | 0.20 | (0.00) | 0.08 | (0.01) | 0.21 | (0.01) | 0.16 | (0.01) | 0.30 | (0.01) |
| | [4.38] | | [41.13] | | [10.82] | | [24.35] | | [12.62] | | [22.13] | |

Table 3: Power Difference between the CS and TS Test

| | Asymptotic | | | Small Sample | | |
|---|---|---|---|---|---|---|
| | CS | TS | CS-TS | CS | TS | CS-TS |
| $SR = 0.25$ | | | | | | |
| Base | 17.3% | 12.2% | 5.1% | 17.0% | 11.8% | 5.2% |
| $N = 50$ | 14.8% | 10.7% | 4.1% | 14.6% | 10.4% | 4.2% |
| $N = 10,000$ | 17.9% | 12.5% | 5.4% | 17.6% | 12.1% | 5.5% |
| $\sigma_\varepsilon = 7$ | 17.9% | 13.0% | 5.0% | 17.6% | 12.5% | 5.1% |
| $\sigma_\varepsilon = 0$ | 11.2% | 8.8% | 2.3% | 11.0% | 8.6% | 2.4% |
| $T = 50$ | 8.0% | 6.8% | 1.2% | 7.7% | 6.4% | 1.3% |
| $T = 1,000$ | 61.8% | 41.2% | 20.6% | 61.6% | 40.9% | 20.8% |
| $SR = 0.50$ | | | | | | |
| Base | 52.5% | 34.3% | 18.2% | 51.8% | 32.9% | 18.9% |
| $N = 50$ | 44.2% | 28.6% | 15.6% | 43.6% | 27.4% | 16.1% |
| $N = 10,000$ | 54.5% | 35.7% | 18.8% | 53.8% | 34.3% | 19.5% |
| $\sigma_\varepsilon = 7$ | 54.6% | 37.5% | 17.1% | 53.9% | 36.0% | 17.9% |
| $\sigma_\varepsilon = 0$ | 30.4% | 20.8% | 9.6% | 29.9% | 19.9% | 9.9% |
| $T = 50$ | 17.3% | 12.2% | 5.1% | 16.1% | 10.5% | 5.6% |
| $T = 1,000$ | 99.5% | 93.5% | 6.0% | 99.5% | 93.3% | 6.2% |
| $SR = 1.00$ | | | | | | |
| Base | 97.7% | 86.0% | 11.7% | 97.5% | 84.4% | 13.1% |
| $N = 50$ | 94.4% | 77.8% | 16.5% | 94.0% | 76.0% | 18.1% |
| $N = 10,000$ | 98.2% | 87.6% | 10.6% | 98.0% | 86.1% | 11.9% |
| $\sigma_\varepsilon = 7$ | 98.2% | 89.4% | 8.8% | 98.0% | 88.0% | 10.1% |
| $\sigma_\varepsilon = 0$ | 80.6% | 60.8% | 19.9% | 80.1% | 58.8% | 21.3% |
| $T = 50$ | 50.6% | 33.0% | 17.6% | 47.9% | 27.4% | 20.5% |
| $T = 1,000$ | 100.0% | 100.0% | 0.0% | 100.0% | 100.0% | 0.0% |
| $SR = 1.50$ | | | | | | |
| Base | 100.0% | 99.6% | 0.4% | 100.0% | 99.4% | 0.6% |
| $N = 50$ | 100.0% | 98.5% | 1.5% | 100.0% | 98.0% | 1.9% |
| $N = 10,000$ | 100.0% | 99.7% | 0.3% | 100.0% | 99.6% | 0.4% |
| $\sigma_\varepsilon = 7$ | 100.0% | 99.8% | 0.2% | 100.0% | 99.7% | 0.3% |
| $\sigma_\varepsilon = 0$ | 99.0% | 92.2% | 6.8% | 98.8% | 91.0% | 7.9% |
| $T = 50$ | 84.8% | 63.2% | 21.6% | 82.6% | 54.5% | 28.1% |
| $T = 1,000$ | 100.0% | 100.0% | 0.0% | 100.0% | 100.0% | 0.0% |

**Note**: The annual Sharpe ratio of the test factor is calucated as $SR = \frac{\mu_b}{\sigma_{bb}} \cdot \sqrt{12}$.

# A Extensions to the Basic Model

## A.1 Variable Transaction Costs

So far, we have considered the case where every stock has the same level of transaction costs $\theta$. In reality, producing a factor portfolio with a given exposure to a particular set of scores varies in terms of its execution cost. This is in part because the extreme scores may tilt towards smaller, more volatile, or less liquid securities, which have a higher cost $\theta_i$ than those in the middle. The change in our investor's weights is slight, as we replace $\theta$ with a vector of transaction costs $\boldsymbol{\theta}$.

$$\mathbf{w}^*_{vtc} = \text{diag}\left(\boldsymbol{\theta}\right)^{-1} \boldsymbol{\Gamma} \boldsymbol{\mu} \tag{45}$$

The stock level weights are simple, and intuitive. Instead of scaling the expected return by a common transaction cost $\theta$, each stock has its own stock-specific transaction cost scalar $\theta_i$. In the case of simple transaction costs in Equation 14, the stock-level logic extends to factors. The exposure to all factor portfolios scales up or down with a single transaction cost parameter. The factor exposures with variable transaction costs are more complicated. To simplify notation, we assume again that the factors are defined to be orthogonal so that $\boldsymbol{\Gamma}'\boldsymbol{\Gamma} = N\mathbf{I}$ and the *raw* and *unit* exposures are the same. Also, we suppose for simplicity that all of the factors are structured so that the mean return $\boldsymbol{\mu}$ is positive. This is not necessary, but it serves to simplify the notation and develop intuition.

$$\begin{aligned} \mathbf{e}_{raw}\left(\mathbf{w}^*_{vtc}\right) = \mathbf{e}_{unit}\left(\mathbf{w}^*_{vtc}\right) &= \Gamma' \text{diag}\left(\boldsymbol{\theta}\right)^{-1} \boldsymbol{\Gamma}\boldsymbol{\mu} \\ &= \frac{1}{N} \begin{bmatrix} \sum_i \frac{\gamma_{ai}^2}{\theta_i} & \sum_i \frac{\gamma_{ai} \cdot \gamma_{bi}}{\theta_i} & \cdots \\ \sum_i \frac{\gamma_{bi} \cdot \gamma_{ai}}{\theta_i} & \sum_i \frac{\gamma_{bi}^2}{\theta_i} & \cdots \\ \vdots & \vdots & \end{bmatrix} \boldsymbol{\mu} \end{aligned} \tag{46}$$

Our investor chooses higher exposure to a factor when its return $\mu_a$ is high relative to its execution cost. Execution costs are low when extreme scores are negatively correlated with transaction costs. In other words, exposure is higher on a factor portfolio if the securities with the highest absolute scores, either positive or negative, so that $\gamma_{ai}^2$ is large, also happen to be easier to trade, with low $\theta_i$. There is a second, subtler force in the off-diagonal terms. Exposure is higher when the factor of interest has scores that are correlated with other high return $\mu_b$ factors, so that $\gamma_{ai} \cdot \gamma_{bi}$

is large, among stocks that are easier to trade, with low $\theta_i$. This is an echo of the unit portfolio exposures in the case of a risk neutral investor with constant transaction costs and correlated scores. Here the scores are by assumption uncorrelated, but incidental positive or negative exposures can nonetheless arise if there is a *conditional* positive or negative score correlation in the set of stocks that are cheaper to trade. This is one reason to prefer a single multi-factor portfolio optimization to a portfolio of factor portfolios, because it takes into account execution savings across factor portfolios.

The factor exposures can be further simplified in the case where the scores are also uncorrelated conditional on transaction costs, so that $E\left(\gamma_{ai} \cdot \gamma_{bi} | \boldsymbol{\theta}\right) = 0$. That eliminates any spillover from one factor exposure to another:

$$\mathbf{e}_{raw}\left(\mathbf{w}^*_{vtc}\right) = \mathbf{e}_{unit}\left(\mathbf{w}^*_{vtc}\right) = \begin{bmatrix} \frac{1}{N}\sum_i \frac{\gamma^2_{ai}}{\theta_i}\mu_a \\ \frac{1}{N}\sum_i \frac{\gamma^2_{bi}}{\theta_i}\mu_b \\ \vdots \end{bmatrix} \tag{47}$$

The conclusion is that our investor uses a simple formula to adjust expected returns for execution costs, which here are increasing in the correlation of trading costs with extreme scores. This will shift the tests in Equation 15 and 25, replacing the gross return with the net return above that adjusts for transaction costs. For the simple case of uncorrelated scores, the cross section test of statistical significance is unchanged.

**Example 5**

Consider an example of two factors, using low asset growth and operating profitability once again. Suppose that these two factors are constructed so that they have zero correlation, even conditional on transaction costs. Also assume that the extreme asset growth stocks, those that are growing very fast or contracting significantly, are less liquid and more costly to trade. Moreover, assume that firms with high operating profits are more liquid with lower cost to trade, so that $\sum \frac{\gamma^2_{bi}}{\theta_i} > \sum \frac{\gamma^2_{ai}}{\theta_i}$, which is roughly consistent with US data. The optimal exposures to the unit or raw factor portfolios, given that they are unconditionally uncorrelated, are:

$$\mathbf{e}_{raw}\left(\mathbf{w}^*_{vtc}\right) = \mathbf{e}_{unit}\left(\mathbf{w}^*_{vtc}\right) = \begin{bmatrix} e_{unit,\frac{-\Delta A}{A}} \\ e_{unit,\frac{OP}{A}} \end{bmatrix} = \begin{bmatrix} \frac{1}{N}\sum_i \frac{\gamma^2_{ai}}{\theta_i}\mu_a \\ \frac{1}{N}\sum_i \frac{\gamma^2_{bi}}{\theta_i}\mu_b \end{bmatrix} \tag{48}$$

The optimal portfolio has positive exposure to both the unit low asset growth portfolio and the unit operating profits portfolio. If they have the same gross factor portfolio return $\mu_a = \mu_b$, then unit exposure to the operating profits portfolio is higher because its execution costs are lower, making its net factor portfolio returns higher.

## A.2 Capacity Constraints

So far, we have sidestepped the issue of the capacity of the factor portfolios. Our investor has constant absolute risk aversion and no constraints on leverage, so there are no natural effects of assets under management, which we label $A$. A practical way of making capacity relevant is to add a few changes to the basic set up in Equations 1 and 11. We first imagine that returns in Equation 1 are defined relative to a benchmark, so that all of the factors have zero mean. Second, we suppose that our investor delegates his portfolio decision, while insisting on some fixed level of gross exposure, a fixed tracking error, or a minimum level of benchmark-adjusted return. For example, our investor might ask for a dollar neutral portfolio, where, for each dollar of equity, one dollar must be invested long and one dollar must be invested short. Or, as we do in Appendix B, we solve a typical case of fixed tracking error $\mathbf{w}'\text{var}\left(\mathbf{r}\right)\mathbf{w} = \sigma^2_T$. We proceed here with a further simplified, risk neutral case where the exposure constraint is expressed as $\mathbf{w}'\mathbf{w} = 1$.

$$\max\left(\boldsymbol{\mu}\boldsymbol{\Gamma}\right)'\mathbf{w} - \frac{A}{2}\mathbf{w}'\text{diag}\left(\boldsymbol{\theta}\right)\mathbf{w} \text{ s.t. } \mathbf{w}'\mathbf{w} = 1 \tag{49}$$

Our investor's new optimal weights a slight variation of the unconstrained version:

$$\mathbf{w}^*_{aum} = \left(A \cdot \text{diag}\left(\boldsymbol{\theta}\right) + 2C\left(A\right)\mathbf{I}\right)^{-1}\boldsymbol{\Gamma}\boldsymbol{\mu} \tag{50}$$

The Lagrange multiplier is $C\left(A\right)$, which satisfies $\left(\boldsymbol{\Gamma}\boldsymbol{\mu}\right)'\left(A\text{diag}\left(\boldsymbol{\theta}\right) + 2C\left(A\right)\mathbf{I}\right)^{-2}\left(\boldsymbol{\Gamma}\boldsymbol{\mu}\right) = 1$ and $\frac{dC}{dA} < 0$ as we show in Appendix B. Again, the weight on any given security is limited by its security specific transaction costs in the first term in parenthesis. Now, there is also a second consideration. We explore these effects of assets under management on the factor portfolio exposures. To simplify notation, we assume again that the factors are defined to be orthogonal so that $\mathbf{\Gamma}'\mathbf{\Gamma} = N\mathbf{I}$.

$$\begin{aligned}
\mathbf{e}_{raw}\left(\mathbf{w}^{*}_{aum}\right) = \mathbf{e}_{unit}\left(\mathbf{w}^{*}_{aum}\right) &= \frac{1}{N}\Gamma'\left(A \cdot \text{diag}\left(\boldsymbol{\theta}\right) + C\left(A\right)\mathbf{I}\right)^{-1}\mathbf{\Gamma}\boldsymbol{\mu} \\
&= \frac{1}{N}\begin{bmatrix} \sum_i \frac{\gamma_{ai}^2}{A\theta_i + C(A)} & \sum_i \frac{\gamma_{ai}\cdot\gamma_{bi}}{A\theta_i + C(A)} & \cdots \\ \sum_i \frac{\gamma_{bi}\cdot\gamma_{ai}}{A\theta_i + C(A)} & \sum_i \frac{\gamma_{bi}^2}{A\theta_i + C(A)} & \cdots \\ \vdots & \vdots & \end{bmatrix}\boldsymbol{\mu}
\end{aligned} \tag{51}$$

Our investor again chooses higher exposure to a factor when its return $\mu_a$ is high relative to its execution cost. Execution costs are low when extreme scores $\gamma_a^2$are negatively correlated with execution costs. But execution costs themselves are now a function of assets under management, with $\theta_i$ replaced by $A\theta_i + C\left(A\right)$. As before, the factor exposures can be further simplified in the case where the scores are also uncorrelated conditional on transaction costs, , so that $E\left(\gamma_{ai} \cdot \gamma_{bi}|\boldsymbol{\theta}\right) = 0$:

$$\mathbf{e}_{raw}\left(\mathbf{w}^{*}_{aum}\right) = \mathbf{e}_{unit}\left(\mathbf{w}^{*}_{aum}\right) = \begin{bmatrix} \frac{1}{N}\sum_i \frac{\gamma_{ai}^2}{A\theta_i + C(A)}\mu_a \\ \frac{1}{N}\sum_i \frac{\gamma_{bi}^2}{A\theta_i + C(A)}\mu_b \\ \vdots \end{bmatrix} \tag{52}$$

The key insight here is that $\frac{dC}{dA} < 0$ and $C$ does not vary across stocks $i$. So, for very small assets under management $A$, the factor allocation is directly proportional to return $\boldsymbol{\mu}$ as in Equation 14. The gross exposure constraint binds so quickly that transaction costs have no effect. As assets rise, the importance of transaction costs $\theta_i$ increases to the point that the allocations are proportional to those in Equation 47. The conclusion is that our investor uses a formula to adjust expected returns for execution costs, but this formula is highly dependent on assets under management, so the exact adjustment of the tests in Equation 15 and 25, which replace gross returns with net returns are context specific. An anomaly for one investor may not be an economically meaningful anomaly for another, because its execution costs are too high at the relevant level of assets $A$. Again, for the simple case of uncorrelated scores, the cross section test of statistical significance is unchanged.

### A.3 Dynamic Trading

So far, we have considered a static trading decision. A fully dynamic optimization with factor scores $\mathbf{\Gamma}$ that vary through time is beyond the scope of this paper. We analyze a very simple case where our risk neutral investor trades over a finite number of periods $T$. The assumption of risk neutrality is somewhat awkward, because our investor will simply accumulate positions over time, with the per period accumulation limited by transaction costs. But, this situation still provides the intuition that the investor in the first period will need to consider the returns not just in the next period but in subsequent periods, with a discount rate $\delta$. More persistent factors deserve greater weight, because they will generate returns over multiple future holding periods.

$$\max \sum_{t=0}^{T} \delta^t \left( (\boldsymbol{\mu}\boldsymbol{\Gamma}_t)' \mathbf{w}_t - \frac{1}{2} (\mathbf{w}_t - \mathbf{w}_{t-1})' \operatorname{diag}(\boldsymbol{\theta}) (\mathbf{w}_t - \mathbf{w}_{t-1}) \right) \tag{53}$$

For simplicity, we imagine that the factor payoffs and transaction costs are constant through time, but the factor scores vary according to an autoregressive process:

$$\boldsymbol{\Gamma}_t = \boldsymbol{\Gamma}_{t-1} \operatorname{diag}\left(\boldsymbol{\rho}_{K\times 1}\right)_{K\times K} + \boldsymbol{\Lambda}_t \tag{54}$$

This is straightforward to solve by backward induction, with the caveat that without a budget constraint or risk aversion the weights accumulate through time. Garleanu and Pedersen (2013) also arrive at a similarly interpretable solution with per period risk aversion by assuming that transaction costs are proportional to the stock level covariance matrix.

$$\mathbf{w}_0 = \operatorname{diag}(\boldsymbol{\theta})^{-1} \boldsymbol{\Gamma}_0 \left( \sum_{s=0}^{T} (\delta \operatorname{diag}(\rho))^s \right) \boldsymbol{\mu} \tag{55}$$

Again, we are interested in the resulting factor exposures of our investor's initial portfolio, and we make the same simplifying assumption that $\boldsymbol{\Gamma}'\boldsymbol{\Gamma} = N\mathbf{I}$ so that the *raw* and *unit* exposures are the same.

$$\begin{aligned} \mathbf{e}_{raw}\left(\mathbf{w}^*_{dt}\right) = \mathbf{e}_{unit}\left(\mathbf{w}^*_{dt}\right) &= \Gamma' \operatorname{diag}(\boldsymbol{\theta})^{-1} \boldsymbol{\Gamma} \left( \textstyle\sum_{\mathbf{s=0}}^{\mathbf{T}} (\delta \operatorname{diag}(\rho))^{\mathbf{s}} \right) \boldsymbol{\mu} \\ &= \frac{1}{N} \begin{bmatrix} \sum_i \frac{\gamma_{ai}^2}{\theta_i} & \sum_i \frac{\gamma_{ai}\cdot\gamma_{bi}}{\theta_i} & \cdots \\ \sum_i \frac{\gamma_{bi}\cdot\gamma_{ai}}{\theta_i} & \sum_i \frac{\gamma_{bi}^2}{\theta_i} & \cdots \\ \vdots & \vdots & \end{bmatrix} \begin{bmatrix} \frac{1}{1-\delta\rho_a}\mu_a \\ \frac{1}{1-\delta\rho_b}\mu_b \\ \vdots \end{bmatrix} \end{aligned} \tag{56}$$

Our investor chooses higher exposure to a factor when the full present value of its future returns $\frac{1}{1-\delta\rho_a}\mu_a$ is high relative to its execution cost. Execution costs are low as before, but now the benefits of trade extend beyond one period. The factor exposures can again be further simplified in the case where the scores are also uncorrelated conditional on transaction costs, so that $E\left(\gamma_{ai}\cdot\gamma_{bi}|\boldsymbol{\theta}\right) = 0$.

That eliminates any spillover from one factor exposure to another:

$$\mathbf{e}_{raw}\left(\mathbf{w}_{vtc}^{*}\right)=\mathbf{e}_{unit}\left(\mathbf{w}_{vtc}^{*}\right)=\begin{bmatrix}\frac{1}{N}\sum_{i}\frac{\gamma_{ai}^{2}}{\theta_{i}}\frac{1}{1-\delta\rho_{a}}\mu_{a}\\ \frac{1}{N}\sum_{i}\frac{\gamma_{bi}^{2}}{\theta_{i}}\frac{1}{1-\delta\rho_{b}}\mu_{b}\\ \vdots\end{bmatrix}\tag{57}$$

As an aside, the exposure of persistent factors increases through time, because the portfolio inherits the effects of past decisions. Nonetheless, the relevant information for our investor on a forward-looking is the execution-cost-adjusted and persistence-adjusted mean returns in Equation 57. Layering on capacity constraints does not produce nicely interpretable exposures, but the effect of combining capacity constraints with dynamic trading is to lower the exposure of costly-to-trade factors in particular when assets under management are high, and when these factors are less persistent. As in the previous two subsections, for the simple case of uncorrelated scores, the cross section test of statistical significance is unchanged.

**Example 6**

Consider an example of two factors, using operating profitability as before and using high frequency reversal instead of low asset growth. Jegadeesh and Titman (1993) among others have observed that the firms with the highest returns in the previous month have lower average returns in the month that follows. So, a stock with a high trailing one-month return might then have a low high frequency reversal score $\gamma_{a,i}$. Again, suppose that these two factors are constructed so that they have zero correlation with one another. For this particular example, we assume that the sets of stocks with high and low operating profitability are very persistent, so that there is a monthly mean persistence of $\rho_a = 0.98$ in factor returns on some initial set of scores $\boldsymbol{\gamma}_b$. And, we assume that high frequency reversal scores are essentially uncorrelated through time, with monthly persistence of $\rho_b = 0$. This is roughly consistent with US data. Finally, we use a monthly discount rate $\delta = 0.9$, meaning that the investor's portfolio might fully turnover about once per year. The dynamic target, optimal exposures to the unit or raw factor portfolios are:

$$\mathbf{e}_{raw}\left(\mathbf{w}_{vtc}^{*}\right)=\mathbf{e}_{unit}\left(\mathbf{w}_{vtc}^{*}\right)=\begin{bmatrix}e_{unit,r_{t-1}}\\ e_{unit,\frac{OP}{A}}\end{bmatrix}=\begin{bmatrix}\frac{1}{N}\sum_{i}\frac{\gamma_{ai}^{2}}{\theta_{i}}\left(\frac{1}{1-0.9\cdot0.98}\mu_{a}\right)\\ \frac{1}{N}\sum_{i}\frac{\gamma_{ai}^{2}}{\theta_{i}}\mu_{b}\end{bmatrix}\tag{58}$$

The optimal portfolio has positive exposure to both the unit high frequency reversal portfolio and the unit operating profits portfolio. If they have the same gross, per period factor portfolio return $\mu_a = \mu_b$, then exposure to operating profits is approximately 8.5 times higher because of its much higher persistence.

# B Derivations

## B.1 Derivation of Equation 25

Here, we would like to prove that Equation 25 is equivalent to Equation 24:

$$\sum_{k \neq a} \hat{\sigma}_{ak} e_{unit,k} \left( \mathbf{w}^*_{ra} \right) = \hat{\mu}_a \tag{59}$$

First, we substitute in the optimal unit exposure from Equation 22 and rearrange the terms, putting $\hat{\mu}_a$on the left-hand side.

$$\hat{\mu}_a = 2\lambda \left( \hat{\sigma}_{ab} e_b + \hat{\sigma}_{ac} e_c + \ldots \right) \tag{60}$$

Second, we note that Equation 25 can be rewritten as follows:

$$\hat{\mu}_a = \sum_{k \neq a} \hat{\beta}_{ak} \hat{\mu}_k = 0 \Rightarrow \hat{\mu}_a = \sum_{k \neq a} \hat{\beta}_{ak} \hat{\mu}_k = \sum_{k \neq a} \hat{\beta}_{ak} 2\lambda \left( \sum_{l=b} \hat{\sigma}_{kl} e_l \right) \tag{61}$$

Now, it remains to show that the right hand side of these two previous equations are the same. This will be true if the following holds for all factors $l$ not equal to the test factor $a$:

$$\forall l, \hat{\sigma}_{al} = \sum_{k \neq a} \hat{\beta}_{ak} \hat{\sigma}_{kl} \tag{62}$$

This identity is simply the identity that the covariance of the sum is equal to the sum of covariances:

$$\operatorname{cov}\left( f_a, f_l \right) = \operatorname{cov}\left( \sum_{k=b} \hat{\beta}_{ak} f_k, f_l \right) = \sum_{k=b} \hat{\beta}_{ak} \operatorname{cov}\left( f_k, f_l \right) \tag{63}$$

## B.2 Solution for Equation 49 with a Budget Constraint

Here, we would like to derive the optimal weights for an investor who faces variable transaction costs with variable assets under management. The general optimization problem is repeated from Equation 49 above:

$$\max\left(\boldsymbol{\mu}\boldsymbol{\Gamma}\right)'\mathbf{w}-\frac{A}{2}\mathbf{w}'\mathrm{diag}\left(\boldsymbol{\theta}\right)\mathbf{w}\text{ s.t. }\mathbf{w}'\mathbf{w}=1 \tag{64}$$

We solve the Lagrangian, where we use the parameter $C$ as the Lagrange multiplier, noting along the way that $C$ is a function of assets under management $A$,

$$\mathcal{L}=\left(\boldsymbol{\mu}\boldsymbol{\Gamma}\right)'\mathbf{w}-\frac{A}{2}\mathbf{w}'\mathrm{diag}\left(\boldsymbol{\theta}\right)\mathbf{w}-C\left(A\right)\big(\mathbf{w}'\mathbf{w}-1\big) \tag{65}$$

by taking the first order condition with respect to portfolio weights:

$$\frac{\partial\mathcal{L}}{\partial\mathbf{w}}=0 \tag{66}$$

$$\Rightarrow\mu\Gamma-\left(A\cdot\mathrm{diag}\left(\boldsymbol{\theta}\right)+2C\left(A\right)\mathbf{I}\right)\mathbf{w}=0 \tag{67}$$

This gives us a solution for$\mathbf{w}$:

$$\mathbf{w}=\left(A\cdot\mathrm{diag}\left(\boldsymbol{\theta}\right)+2C\left(A\right)\mathbf{I}\right)^{-1}\mu\Gamma \tag{68}$$

## B.3 Derivation of $\frac{\partial C(A)}{\partial A}<0$ in Equation 50

Here, we would like to show that the Lagrange multiplier, $C\left(A\right)$, from the previous section, and in Equation 50 is decreasing in assets under management, $A$. The FOC can rewritten as follows:

$$\left(\boldsymbol{\Gamma}\boldsymbol{\mu}\right)'\left(A\cdot\mathrm{diag}\left(\boldsymbol{\theta}\right)+2C\left(A\right)\mathbf{I}\right)^{-2}\left(\boldsymbol{\Gamma}\boldsymbol{\mu}\right)=1 \tag{69}$$

$$\Rightarrow\boldsymbol{\mu_R}'\left(A\cdot\mathrm{diag}\left(\boldsymbol{\theta}\right)+2C\left(A\right)\mathbf{I}\right)^{-2}\boldsymbol{\mu_R}=1 \tag{70}$$

Because $A\cdot\mathrm{diag}\left(\boldsymbol{\theta}\right)+2C\left(A\right)\mathbf{I}$ is an diagonal matrix, its inverse is still an diagonal matrix with inverses of corresponding diagonal elements. We can then express the squared inverse as follows:

$$(A \cdot \mathrm{diag}(\boldsymbol{\theta}) + 2C(A)\,\mathbf{I})^{-2} = \begin{bmatrix} \frac{1}{(A\theta_1+2C(A))^2} & & \\ & \frac{1}{(A\theta_2+2C(A))^2} & \\ & & \ddots \end{bmatrix} \tag{71}$$

Substituting for the squared inverse, simplifying, performing matrix multiplication, gives us the following first order condition:

$$\begin{bmatrix} \mu_{R,1} \\ \mu_{R,2} \\ \vdots \end{bmatrix}' \begin{bmatrix} \frac{1}{(A\theta_1+2C(A))^2} & & \\ & \frac{1}{(A\theta_2+2C(A))^2} & \\ & & \ddots \end{bmatrix} \begin{bmatrix} \mu_{R,1} \\ \mu_{R,2} \\ \vdots \end{bmatrix}' = 1 \tag{72}$$

$$\Rightarrow \begin{bmatrix} \frac{\mu_{R,1}}{(A\theta_1+2C(A))^2} & \frac{\mu_{R,2}}{(A\theta_2+2C(A))^2} & \cdots \end{bmatrix} \begin{bmatrix} \mu_{R,1} \\ \mu_{R,2} \\ \vdots \end{bmatrix}' = 1 \tag{73}$$

$$\Rightarrow \sum_{i=1}^{N} \frac{\mu_{R,i}^2}{(A\theta_i + 2C(A))^2} = 1 \tag{74}$$

Moving all terms to left hand side of the equation, we can now label the new FOC as $F(A)$:

$$F(A) = \sum_{i=1}^{N} \frac{\mu_{R,i}^2}{(A\theta_i + 2C(A))^2} - 1 = 0 \tag{75}$$

Finally, we apply the implicit function theorem and express total derivative in terms of partial derivatives,

$$\frac{dF}{dA} = 0 \tag{76}$$

$$\Rightarrow \frac{\partial F}{\partial A} + \frac{\partial F}{\partial C(A)} \frac{\partial C(A)}{\partial A} = 0 \tag{77}$$

and we rearrange terms, solving for $\frac{\partial C(A)}{\partial A}$:

$$\frac{\partial C(A)}{\partial A} = -\frac{\partial F/\partial A}{\partial F/\partial C(A)} \tag{78}$$

Because the numerator and denominator are both negative,

$$\frac{\partial F}{\partial A} = -2\sum_i \frac{\theta_i \mu_{R,i}^2}{\left(A\theta_i + 2C\left(A\right)\right)^3} < 0 \tag{79}$$

$$\frac{\partial F}{\partial C\left(A\right)} = -2\sum_i \frac{\mu_{R,i}^2}{\left(A\theta_i + 2C\left(A\right)\right)^3} < 0 \tag{80}$$

we have proven that $\frac{\partial C(A)}{\partial A} < 0$.

## B.4 Derivation of Equation 35 for the Cross Section Test

Here, we derive the standard error formula, Equation 35, for the cross section test. For the cross section test, we start by running a cross-sectional regression at each time $t$:

$$\mathbf{r}_t \sim \mathbf{\Gamma}_t \mathbf{f}_t + \boldsymbol{\varepsilon}_t \tag{81}$$

where

$$\mathbf{\Gamma'} = \begin{bmatrix} 1 & \cdots & 1 \\ \gamma_{a1} & \cdots & \gamma_{aN} \\ \gamma_{b1} & \cdots & \gamma_{bN} \\ \vdots & \vdots & \vdots \end{bmatrix} = [\boldsymbol{\gamma_1}, \cdots, \boldsymbol{\gamma_N}] \tag{82}$$

We obtain the regression coefficient at time $t$:

$$\hat{\mathbf{f}}_t = \left(\mathbf{\Gamma}_t' \mathbf{\Gamma}_t\right)^{-1} \mathbf{\Gamma}_t' \mathbf{r}_t \tag{83}$$

and we introduce the notation $\boldsymbol{\delta}$ to refer to the columns of $(\mathbf{\Gamma'\Gamma})^{-1}$:

$$\mathbf{\Gamma'\Gamma} = N \begin{bmatrix} 1 & \overline{\gamma_a} & \overline{\gamma_b} & \cdots \\ & \overline{\gamma_a^2} & \overline{\gamma_a \gamma_b} & \cdots \\ & & \overline{\gamma_b^2} & \cdots \\ & & & \ddots \end{bmatrix} \tag{84}$$

$$\left(\mathbf{\Gamma'\Gamma}\right)^{-1} = \frac{1}{N} \begin{bmatrix} \boldsymbol{\delta}_1 & \boldsymbol{\delta}_a & \boldsymbol{\delta}_b & \boldsymbol{\delta}_c & \cdots \end{bmatrix} \tag{85}$$

We first rewrite the regression coefficient $\hat{\mathbf{f}}_t$ by substituting for $\mathbf{r}_t$ with its definition:

$$\hat{\mathbf{f}}_t = \mathbf{f}_t + \left(\mathbf{\Gamma'\Gamma}\right)^{-1} \mathbf{\Gamma'} \boldsymbol{\varepsilon}_t \tag{86}$$

The term $\mathbf{\Gamma'}\boldsymbol{\varepsilon}_t$ can be rewritten as:

$$\mathbf{\Gamma}'\boldsymbol{\varepsilon}_t = \begin{bmatrix} \sum_i \varepsilon_{it} \\ \sum_i \gamma_{ai}\varepsilon_{it} \\ \sum_i \gamma_{bi}\varepsilon_{it} \\ \vdots \end{bmatrix} = \sum_i \begin{bmatrix} \varepsilon_{it} \\ \gamma_{ai}\varepsilon_{it} \\ \gamma_{bi}\varepsilon_{it} \\ \vdots \end{bmatrix} = \sum_i \boldsymbol{\gamma_i}\varepsilon_{it} \tag{87}$$

We then substitute this into the formula for $\hat{\mathbf{f}}_t$ and obtain the regression coefficient for a test factor which we label arbitrarily as $b$ as follows:

$$\hat{f}_{b,t} = f_{b,t} + \frac{1}{N}\boldsymbol{\delta}_b' \sum_i \boldsymbol{\gamma}_i \varepsilon_{it} \tag{88}$$

Now, we note that the test statistics for the cross section test is the time-series means of the regression coefficients is:

$$\hat{\boldsymbol{\mu}} = \frac{1}{T}\sum_t \hat{\mathbf{f}}_t \tag{89}$$

This means that the standard error of the test statistic for test factor $b$ is:

$$\text{se}\left(\hat{\mu}_b\right) = \frac{1}{\sqrt{T}}\sqrt{\sigma_{bb}^2 + \frac{1}{N^2}\sigma^2 \sum_i \left(\boldsymbol{\delta}_b'\boldsymbol{\gamma}_i\right)^2} \tag{90}$$

### B.5 GMM Derivation of Equation 40 for the Time Series Test

Here, we derive the standard error formula, Equation 40, for the time series test. We start by noting that the return generating process is as follows:

$$\mathbf{r}_t = \mathbf{\Gamma}_t \mathbf{f}_t + \boldsymbol{\varepsilon}_t \tag{91}$$

Next, we run a Jensen's alpha test using an arbitrary test factor $b$ by regressing its return on existing factors:

$$f_{b,t} \sim \alpha_b + \boldsymbol{\beta}_b' \mathbf{f}_{-b,t} + \epsilon_t \tag{92}$$

If the test factor's returns are orthogonal to all other factor returns, then $\alpha_b = \mu_b$, the mean return of the test factor portfolio from the return generating process. If the test factor's returns are not orthogonal and are instead fully spanned by the other factor returns, then $\alpha_b = 0$. Now, we note that the test statistic for the time series test is the intercept $\hat{\alpha}_b$ from this regression, where we substitute in population estimates for the parameters in the return generating process:

$$\hat{\alpha}_b = \bar{f}_b - \hat{\boldsymbol{\beta}}_b' \bar{\mathbf{f}}_{-b} = \hat{\mu}_b - \hat{\boldsymbol{\beta}}_b' \hat{\boldsymbol{\mu}}_{-b} \tag{93}$$

We then run a GMM estimation using the OLS moments to estimate coefficients, following the standard approach in Cochrane (2009, Ch 12):

$$g_T(\mathbf{b}) = E_T\left(\begin{bmatrix} \epsilon_t \\ f_{a,t}\epsilon_t \\ f_{c,t}\epsilon_t \\ \vdots \end{bmatrix}\right) = 0 \tag{94}$$

where we define $\mathbf{b}$ as the vector of parameters to be estimated:

$$\hat{\mathbf{b}} = \left[\hat{\alpha}, \hat{\beta}_{ba}, \hat{\beta}_{bc} \cdots\right]' \tag{95}$$

so that $\hat{\alpha}$ can be written as follows:

$$\hat{\alpha} = \overline{f_b} - \begin{bmatrix} \hat{\beta}_{ba} \\ \hat{\beta}_{bc} \\ \vdots \end{bmatrix}' \cdot \begin{bmatrix} \overline{f_a} \\ \overline{f_c} \\ \vdots \end{bmatrix} \tag{96}$$

Note that the GMM estimates here are the same as the OLS regression coefficients, and the full asymptotic joint distribution of the GMM estimates is as follows:

$$\sqrt{T}(\hat{\mathbf{b}} - \mathbf{b}) \to \mathcal{N}\left(0, (\mathbf{AD})^{-1}\mathbf{ASA}'(\mathbf{AD})^{-1\prime}\right) \tag{97}$$

where the matrices $\mathbf{A}$, $\mathbf{D}$, and $\mathbf{S}$ are defined next. Because $\mathbf{A}g_T(\hat{\mathbf{b}}) = 0$, $\mathbf{A} = \mathbf{I}$ and it drops out. Matrix $\mathbf{D}$, according to the GMM formula, is[2]:

$$\mathbf{D} \equiv \frac{\partial g_T(\mathbf{b})}{\partial \mathbf{b}} = -\boldsymbol{\Phi} = -\begin{bmatrix} 1 & \overline{\mathbf{f}_{-b}}' \\ \overline{\mathbf{f}_{-b}} & \hat{\boldsymbol{\Sigma}}_{-b} + \overline{\mathbf{f}_{-b}} \cdot \overline{\mathbf{f}_{-b}}' \end{bmatrix} \tag{98}$$

[2]From Cochrane (2012): precisely, $\mathbf{D}$ is defined as the population moment in the first equality, which we estimate in sample by the second equality

where we define $\overline{\mathbf{f}_{-b}}$ and $\hat{\mathbf{\Sigma}}_{-b}$ as follows:

$$\overline{\mathbf{f}_{-b}} = (\overline{f_a}, \overline{f_c}, \cdots)' \tag{99}$$

$$\hat{\mathbf{\Sigma}}_{-b} = \frac{1}{T}\sum_t \left(\mathbf{f}_{-b,t} - \overline{\mathbf{f}_{-b}}\right)\left(\mathbf{f}_{-b,t} - \overline{\mathbf{f}_{-b}}\right)' \tag{100}$$

The third matrix $\mathbf{S}$ is defined as:

$$\mathbf{S} \equiv \sum_{j=-\infty}^{\infty} E\left(\begin{bmatrix} \epsilon_t\epsilon_{t-j} & \epsilon_t\epsilon_{t-j}f_{a,t-j} & \epsilon_t\epsilon_{t-j}f_{c,t-j} & \cdots \\ \epsilon_t f_{a,t}\epsilon_{t-j} & \epsilon_t f_{a,t}\epsilon_{t-j}f_{a,t-j} & \epsilon_t f_{a,t}\epsilon_{t-j}f_{c,t-j} & \cdots \\ \epsilon_t f_{c,t}\epsilon_{t-j} & \epsilon_t f_{c,t}\epsilon_{t-j}f_{a,t} & \epsilon_t f_{c,t}\epsilon_{t-j}f_{c,t-j} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}\right) \tag{101}$$

Additionally, we assume that the residual terms are uncorrelated over time, are homoskedastic, and the factors other than the test factor $b$ is orthogonal to the residuals in the return generating process for factor $b$. This means that the matrix $\mathbf{S}$ can simplified to:

$$\mathbf{S} \equiv E\left(\begin{bmatrix} \epsilon_t^2 & \epsilon_t^2 f_{a,t} & \epsilon_t^2 f_{c,t} & \cdots \\ & \epsilon_t^2 f_{a,t}^2 & \epsilon_t^2 f_{a,t}f_{c,t} & \cdots \\ & & \epsilon_t^2 f_{c,t}^2 & \cdots \end{bmatrix}\right) = \mathbf{\Phi}E[\epsilon_t^2] \tag{102}$$

Note when the test factor portfolio return $f_b$ are orthogonal to the other factor returns, $E[\epsilon_t^2] = \sigma_{bb}^2$. Now, we can substitute the matrices $\mathbf{A}$, $\mathbf{D}$ and $\mathbf{S}$ into the asymptotic distribution and obtain the variance of the GMM estimate $\hat{\mathbf{b}}$:

$$Var(\hat{\mathbf{b}}) = \frac{1}{T}\mathbf{D}^{-1}\mathbf{S}\mathbf{D}^{-1\prime} = \frac{1}{T}\mathbf{\Phi}^{-1}\mathbf{\Phi}\sigma_\epsilon^2\mathbf{\Phi}^{-1} = \frac{1}{T}\mathbf{\Phi}^{-1}\sigma_\epsilon^2 \tag{103}$$

To obtain the variance of the test statistic $Var(\hat{\alpha})$, we need to calculate the top left corner of the matrix $Var(\hat{\mathbf{b}})$. First, we calculate the top left corner of $\mathbf{\Phi}^{-1}$. To do so, we perform a matrix inversion, which takes the following form:

$$\mathbf{\Phi}^{-1} \equiv \begin{bmatrix} \mathbf{C}_1^{-1} & \cdots \\ \cdots & \mathbf{C}_2^{-1} \end{bmatrix} \tag{104}$$

where, in our case, the upper left block is simply a scalar, though we continue to refer to it as the matrix $\mathbf{C}_1$:

$$\mathbf{C}_1 = 1 - \overline{\mathbf{f}_{-b}}'\left(\hat{\mathbf{\Sigma}}_{-b} + \overline{\mathbf{f}_{-b}} \cdot \overline{\mathbf{f}_{-b}}'\right)^{-1}\overline{\mathbf{f}_{-b}} \tag{105}$$

The inverse of $\left(\hat{\mathbf{\Sigma}}_{-b} + \overline{\mathbf{f}_{-b}} \cdot \overline{\mathbf{f}_{-b}}'\right)^{-1}$ can be rewritten as:

$$\hat{\boldsymbol{\Sigma}}_{-b}^{-1} - \hat{\boldsymbol{\Sigma}}_{-b}^{-1}\overline{\mathbf{f}_{-b}}\left(1 + \overline{\mathbf{f}_{-b}}'\hat{\boldsymbol{\Sigma}}_{-b}^{-1}\overline{\mathbf{f}_{-b}}\right)^{-1}\overline{\mathbf{f}_{-b}}'\hat{\boldsymbol{\Sigma}}_{-b}^{-1} \tag{106}$$

which we can then substitute into the formula for $\mathbf{C}_1$, simplify, and invert:

$$\mathbf{C}_1 = 1 - \frac{\overline{\mathbf{f}_{-b}}'\hat{\boldsymbol{\Sigma}}_{-b}^{-1}\overline{\mathbf{f}_{-b}}}{1 + \overline{\mathbf{f}_{-b}}'\hat{\boldsymbol{\Sigma}}_{-b}^{-1}\overline{\mathbf{f}_{-b}}} \tag{107}$$

$$= \frac{1}{1 + \overline{\mathbf{f}_{-b}}'\hat{\boldsymbol{\Sigma}}_{-b}^{-1}\overline{\mathbf{f}_{-b}}} \tag{108}$$

$$\Rightarrow \mathbf{C}_1^{-1} = 1 + \overline{\mathbf{f}_{-b}}'\hat{\boldsymbol{\Sigma}}_{-b}^{-1}\overline{\mathbf{f}_{-b}} \tag{109}$$

$$= 1 + \boldsymbol{\mu}'_{-b}\boldsymbol{\Sigma}_{-b}^{-1}\boldsymbol{\mu}_{-b} \tag{110}$$

Therefore, the standard error of the test statistics $\hat{\alpha}_b$ is:

$$\text{se}\left(\hat{\alpha}_b\right) = \sqrt{\frac{1}{T}\mathbf{C}_1^{-1}\sigma_\epsilon^2} \tag{111}$$

$$= \frac{1}{\sqrt{T}}\sqrt{\sigma_\epsilon^2\left(1 + \boldsymbol{\mu}'_{-b}\boldsymbol{\Sigma}_{-b}^{-1}\boldsymbol{\mu}_{-b}\right)} \tag{112}$$

From Appendix B.6, we derive the variance of the efficient portfolio for factor $b$ is $\sigma_{bb}^2 + \frac{1}{N^2}\sigma^2\sum_i\left(\boldsymbol{\delta}_b'\boldsymbol{\gamma}_i\right)^2$. Therefore, by decomposition of the total variance, we get:

$$\sigma_{bb}^2 + \frac{1}{N^2}\sigma^2\sum_i\left(\boldsymbol{\delta}_b'\boldsymbol{\gamma}_i\right)^2 = \boldsymbol{\beta}_b'\boldsymbol{\Sigma}_{-b}\boldsymbol{\beta}_b + \sigma_\epsilon^2 \tag{113}$$

$$\Rightarrow \sigma_\epsilon^2 = \left(1 - R^2\right)\left(\sigma_{bb}^2 + \frac{1}{N^2}\sigma^2\sum_i\left(\boldsymbol{\delta}_b'\boldsymbol{\gamma}_i\right)^2\right) \tag{114}$$

where $R^2$ corresponds the regression in Equation 92.

Plug it back in the Equation for $\text{se}\left(\hat{\alpha}_b\right)$ and get:

$$\text{se}\left(\hat{\alpha}_b\right) = \frac{1}{\sqrt{T}}\sqrt{(1-R^2)\left(\sigma_{bb}^2 + \tfrac{1}{N^2}\sigma^2\textstyle\sum_i\left(\boldsymbol{\delta}_b'\boldsymbol{\gamma}_i\right)^2\right)(1+SR^2)} \tag{115}$$

## B.6 Using an Efficient Factor Portfolio in Equation 40

Here, we derive the form of Equation 40 using an efficient portfolio for a test factor $b$. This is essentially the same as forming a portfolio for test factor $b$ using a cross section regression, as follows. We use optimization to construct a pure test factor portfolio $P$ that is dollar neutral, delivers unit exposure to the test factor of interest, zero exposure to all other factors, and otherwise minimizes idiosyncratic risk.

Let the weight in portfolio $P$ be $\mathbf{w}' = [w_1, w_2, \dots, w_N]$. The portfolio has the following return properties:

$$r_{Pt} = f_{bt} + \varepsilon_{Pt} \tag{116}$$

$$\mu_P = \mu_b \tag{117}$$

$$\sigma_P^2 = \sigma_{bb}^2 + \sigma_{P\varepsilon}^2 \tag{118}$$

We then minimize idiosyncratic risk:

$$\min \sigma_{P\varepsilon}^2 \tag{119}$$

subject to constraints of dollar neutrality, unit exposure to $b$ and zero exposure to all other factors:

$$\begin{cases} \sum_i w_i & = 0 \\ \sum_i w_i \gamma_{ai} & = 0 \\ \sum_i w_i \gamma_{bi} & = 1 \\ \sum_i w_i \gamma_{ci} & = 0 \\ \vdots & \end{cases} \tag{120}$$

The solution has the same form as the unit exposure portfolios $\mathbf{Q}_{unit} = \mathbf{\Gamma} \left(\mathbf{\Gamma}'\mathbf{\Gamma}\right)^{-1}$. And so, the portfolio variance can be written as:

$$\sigma_P^2 = \sigma_{bb}^2 + \sigma_{P\varepsilon}^2 = \sigma_{bb}^2 + \frac{1}{N^2}\sigma^2 \sum_i \left(\boldsymbol{\delta}_b'\boldsymbol{\gamma}_i\right)^2 \tag{121}$$

which is the same as the variance of the individual cross-sectional regression coefficient.