Pricing Without Mispricing

by*

Jianan Liu, Tobias J. Moskowitz, and Robert F. Stambaugh

August 30, 2021

Abstract

We investigate whether various asset pricing models could hold in an efficient market. Assuming decade-old information should be priced correctly, we test whether a model assigns zero alpha to investment strategies that use only such information. The CAPM passes this test, but prominent multifactor models do not. Multifactor betas may help capture expected returns on mispriced stocks, but persistence in those betas distorts the stocks' implied expected returns after prices correct. Such effects are strongest in large-cap stocks, whose multifactor betas are the most persistent. Hence, prominent multifactor models distort expected returns, absent mispricing, for the largest, most liquid stocks.

* Liu is at Shanghai Mingshi Investment Co, Ltd.; Moskowitz is at the Yale School of Management, Yale University, AQR Capital Management, and NBER; Stambaugh is at the Wharton School of the University of Pennsylvania and NBER. We are grateful for comments from Nick Barberis, Martijn Boons, Gene Fama, Lasse Pedersen, and Peter Nyberg. AQR Capital is a global investment firm that may or may not use the insights in this paper. The views expressed here are those of the authors and not necessarily those of AQR Capital.

1. Introduction

The joint-hypothesis problem posed by Fama (1970) is central to empirical investigations of market efficiency. Fama (1991) expresses the inherent challenge as follows:

> *We can only test whether information is properly reflected in prices in the context of a pricing model that defines the meaning of "properly."*

In this study, we ask which pricing model best defines "properly." Such a model offers the best benchmark for assessing market efficiency and gauging mispricing. Models that capture expected returns empirically could be identifying mispricing in an inefficient market or compensated risks in an efficient market.[1] Which pricing model, among prominent candidates, would best describe expected stock returns if there were no mispricing? That is the question addressed in this paper.

Tests of market efficiency seek to control for the equilibrium model of expected returns or identify situations where the underlying model differences out (such as examining deviations from the law of one price, e.g., Lamont and Thaler, 2003, Du, Tepper, and Verdelhan, 2018, Hu, Pan, and Wang, 2013). Conversely, we seek to avoid having market inefficiency confound tests of asset pricing models. To implement this idea empirically, we attempt to plausibly identify when mispricing is absent. Our key assumption is that any current mispricing of a stock gets corrected in less than ten years. Suppose a long-short spread based on decade-old information produces a significant alpha with respect to a given pricing model. That alpha does not reflect mispricing under our assumption, so we infer that the model does not describe expected stock returns in the absence of mispricing.

In many ways, finding settings where information must surely be reflected in prices is an easier task than identifying the right pricing model (or finding deviations from the law of one price). While our plausible assumption avoids mispricing, a drawback is low power. Tests where we can confidently rule out mispricing may also be tests that have low power to discriminate among pricing models. For instance, while information in decade-old characteristics should be reflected correctly in today's prices, many of those characteristics may have

[1] Prominent empirical models in asset pricing have received numerous interpretations from both sides of the market efficiency debate. An inexhaustive list of rational and behavioral theories includes Fama and French (1993, 1995, 2015), Gomes, Kogan, and Zhang (2003) , Zhang (2005), Li, Livdan, and Zhang (2009), Belo (2010), Li and Zhang (2010), Liu and Zhang (2008), Berk, Green, and Naik (1999), Johnson (2002), Sagi and Seasholes (2007), Liu, Whited, and Zhang (2009), and Hou, Xue, and Zhang (2015) on the rational, no mispricing side and Lakonishok, Shleifer, and Vishny (1994), Daniel and Titman (1997), Daniel, Hirshleifer, and Subrahmanyam (1998), Barberis, Shleifer, and Vishny (1998), Hong and Stein (1999), and Stambaugh and Yuan (2017) on the behavioral, mispricing side.

low correlation with the current values of multifactor betas that pricing models either omit or include. The betas of long-short spreads based on those lagged characteristics could then be essentially zero, offering little power to discriminate among models. Nevertheless, some characteristics provide adequate power. In particular, we find that one of the Stambaugh and Yuan (2017) composite mispricing measures, PERF, provides power to distinguish among prominent asset pricing models.

The five models we consider include the traditional CAPM of Sharpe (1964) and Lintner (1965) and four multifactor models: the three-factor model of Fama and French (1993), the four-factor model of Hou et al. (2015), the five-factor model of Fama and French (2015), and the CAPM augmented by the betting-against-beta factor of Frazzini and Pedersen (2014).[2] For each pricing model, our test's null hypothesis is that the model assigns zero alpha to investment strategies formed using only decade-old information. For a long-short strategy based on stocks' decade-old PERF scores, the CAPM passes our test but the above four multifactor models do not. In other words, our novel test prefers the CAPM to prominent multifactor models as the "proper" no-mispricing benchmark model for tests of market efficiency.

Our results admit a simple explanation. Observable characteristics can help identify mispricing of some stocks, and multifactor betas can help capture those stocks' expected returns. Some of those multifactor betas, however, can persist longer than the stocks' mispricing. If a multifactor beta gets included in a pricing model to help explain expected returns that in part reflect mispricing, then persistence in that beta injects unhelpful distortions into stocks' implied expected returns after the mispricing corrects. When a security's mispricing is more transitory than its multifactor betas, including factors that capture mispricing will distort the security's expected return absent mispricing. By excluding those multifactor betas, the CAPM avoids such effects. We find that PERF spreads have persistent multifactor betas that cause the multifactor models to assign positive alphas to spreads based on decade-old PERF scores.

We show that much of the persistence in the PERF spreads' multifactor betas can be explained by persistence in the measures of profitability, distress, and momentum that constitute PERF. Our test's power when based on PERF owes mainly to those underlying characteristics persisting longer than the associated mispricing. For example, a firm with

[2]A positive premium on the latter factor, theoretically motivated, implies positive (negative) CAPM alphas on assets with low (high) betas. Black (1972) and Fama (1976) provide earlier theoretical motivations for augmenting the CAPM with such a beta factor, and Black, Jensen, and Scholes (1972) and Fama and MacBeth (1973) provide related evidence.

a currently high PERF score (signaling overpricing) tends to remain unprofitable and distressed long thereafter. Previous studies link the predictive abilities of those characteristics to investor underreaction.[3] Suppose, for example, investors underreact to profitability information, paying too high a price for an unprofitable firm. We expect that the longer a firm has been unprofitable, the less likely are investors to underreact to currently low profitability. Consistent with that hypothesis, we find that the longer a stock has had its current PERF score, the less useful that score is in constructing a profitable long-short strategy today.

We find our test has little ability to discriminate among models when using spreads based on decade-old scores for MGMT, the other mispricing measure of Stambaugh and Yuan (2017). There is less persistence in those spreads' multifactor betas, as there is less persistence in MGMT's underlying stock characteristics. Consequently, our test lacks sufficient power. One property of MGMT scores, though, further supports the above underreaction scenario: In contrast to PERF scores, the MGMT scores that have persisted longest are no less useful in predicting returns than are the scores least like their previous values. Consistent with this result, unlike the PERF characteristics, the characteristics constituting MGMT, such as firms' issuance and asset growth, are not linked to underreaction by the literature.

We conduct our tests separately within different segments of firm size, for a number of reasons. Consider the largest firms, for example. Our assumption that mispricing disappears within 10 years seems especially plausible for the largest stocks, which are liquid and well-monitored. They should be the stocks least likely to have persistent market frictions impeding price correction over periods as long as a decade. Correctly pricing large stocks thus seems the purest test for any proper benchmark model to clear. Moreover, our test relies on decade-old information being related to current multifactor betas, and betas seem likely to be more stable over time for the largest firms. We empirically confirm that larger stocks have more persistent multifactor betas.

Consider instead the smallest stocks. They are probably subject to the greatest mispricing, thereby presenting salient targets for multifactor models aiming to fit average short-run returns. In this respect, small stocks could offer our test an especially informative sample for distinguishing such models from the proper no-mispricing benchmark. Offsetting this potential advantage, however, is that multifactor betas are least stable for small stocks, making the link between decade-old information and current betas substantially weaker. Because determinants of our test's informativeness likely differ across firm size, we allow each segment

[3]Although the literature provides empirical evidence linking these characteristics to underreaction, that evidence plays no role in how Stambaugh and Yuan (2017) group prominent anomalies into PERF versus their other mispricing score, MGMT.

to speak separately in the data.

We find that results differ meaningfully across firm size. Our test delivers its strongest message when based on the largest stocks in the economy. For stocks larger than the NYSE 70th percentile, and even for the largest 200 names, a long-short PERF strategy produces significant alpha that lasts for less than a year under the CAPM but persists for up to 10 years under prominent multifactor models. The latter result is striking, given the implausibility of decade-long mispricing afflicting the largest stocks in the economy. A more plausible explanation is that the multifactor models significantly distort what these stocks' expected returns would be in the absence of mispricing. In contrast, although we find greater apparent mispricing (larger initial alphas) among the smallest firms, the characteristics and betas of small firms change too rapidly to distinguish among competing models at longer lags of the mispricing measures—the low power problem. The fact that our results are driven by the largest firms provides compelling support for the notion that multifactor models pick up temporary mispricing that then distorts expected returns that should only reflect underlying risk premia. In other words, the largest firms would seem to present the strongest test of what model should be the proper benchmark for addressing questions about market efficiency (Fama, 1991). Our evidence indicates that the CAPM is a better model for this task than popular multifactor models.

Models failing our hurdle for a no-mispricing benchmark can perform well by the usual empirical metrics that assess a pricing model's ability to describe expected returns conditional on any available information. Those models potentially capture mispricing well, and some of them are even cast explicitly in the those terms (e.g., Stambaugh and Yuan (2017)). Such models can be quite useful in designing trading strategies, for example, but they are less useful for providing the benchmark model of expected returns to which "properly" applies in the sense of Fama (1991).

Our approach and evidence are novel, but they relate to results reported in recent studies by Cho and Polk (2020), Keloharju, Linnainmaa, and Nyberg (2020), and Yara, Boons, and Tamoni (2020). We discuss and interpret their results in the context of our tests. Our framework can reconcile and add perspective to those results, providing a unified explanation.

The rest of the paper is organized as follows. Section 2 motivates our test, discusses its power, describes our empirical setting, and presents the test results for various pricing models using spreads formed with decade-old variables. Section 3 provides insights into the test results by examining the resulting alphas and factor betas for the sequence of lags up to ten years. Section 4 examines the implications of persistent firm characteristics for factor betas as well as the underreaction scenario. Section 5 discusses our findings in the context of the related studies mentioned above, and Section 6 concludes.

2. A simple test of pricing without mispricing

In this section we introduce and apply our test of whether an asset pricing model is the proper benchmark that delivers expected returns in the absence of mispricing. We begin by motivating our test and discussing its inherent challenge in achieving power. We explain why large stocks are likely to provide our test with the greatest power, given their more persistent multifactor betas, which we document. We then apply the test to a variety of asset pricing models prominent in the literature. In addition to the traditional CAPM of Sharpe (1964) and Lintner (1965), we test the three-factor model of Fama and French (1993), the five-factor model of Fama and French (2015), the four-factor model of Hou, Xue, and Zhang (2015), and the CAPM augmented by the betting-against-beta (BAB) factor of Frazzini and Pedersen (2014).

2.1. Motivation and power

In the absence of mispricing, the proper benchmark asset pricing model should deliver zero alpha for any investment strategy. This implication motivates the canonical test of market efficiency, but it is plagued by the joint hypothesis problem identified by Fama (1970): any test of efficiency is a joint test of the pricing model. The proper model for testing efficiency is one that characterizes prices in an equilibrium wherein (public) information is correctly reflected in prices. Such an equilibrium may be just hypothetical in a market with noise traders or other frictions that hinder information from being incorporated immediately into prices. Nevertheless, the "proper" model is the one relevant to the joint hypothesis problem posed by Fama (1970). Under this view, non-zero alphas imply mispricing or an incorrect model, while zero alphas imply no mispricing or, again, an incorrect model. In the latter case, the incorrect model captures mispricing but not expected returns in the absence thereof. In other words, the model is not a useful empirical benchmark for gauging the extent of market inefficiencies.

Rather than investigate potential market inefficiencies, we explore the other side of the joint hypothesis. That is, we ask which asset pricing models best capture expected return in the absence of mispricing. To do so, we assume that decade-old information is correctly reflected in prices today, however inefficient the market may be in reflecting current information. Specifically, we test the abilities of various models to produce zero alpha for strategies based on decade-old stock characteristics. Such a test, we argue, is unlikely to be contaminated by information inefficiency.

The test's inherent challenge is achieving power. We seek to discriminate among pricing models that include or omit various factors. For many assets, however, current betas on those factors may have little or no relation to decade-old characteristics. Our test relies on finding assets and characteristics for which the relation between the assets' current betas and lagged characteristics is strong enough to provide power. We expect the greatest power to be offered by assets most likely to have stable betas.

To understand better the nature of our test and its potential power, let f_t denote the vector of returns in month t on the factor portfolios associated with a given asset pricing model. Consider the time-series regression,

$$r_t = \alpha + \beta' f_t + \epsilon_t, \tag{1}$$

where r_t is the excess return in month t on an investment strategy whose portfolio weights in month t are determined by sorting a given universe of assets, A, on the values in a vector x of a characteristic observed (publicly) in month $t - \tau$. The value of α depends on the identity of the characteristic, x, its lag length, τ, the set of factors, f, and the asset universe, A. We represent this dependence via the functional notation,

$$\alpha = \alpha(x, \tau, f, A). \tag{2}$$

Let f^* denote the factors in the proper pricing model that captures expected return in the absence of mispricing. Our test assumes

$$\alpha(x, 120, f^*, A) = 0 \tag{3}$$

for any characteristic x. If the market is efficient, then $\alpha(x, 1, f^*, A) = 0$ for any x. In an inefficient market, various choices of x can make $\alpha(x, 1, f^*, A) \neq 0$.

A model can fail as the no-mispricing benchmark by omitting a relevant factor or by including an irrelevant one. Omitting a relevant factor from a pricing model is a possibility long recognized in the literature. In an inefficient market, however, there is a distinct possibility that models are developed with factors that are irrelevant in the absence of mispricing. The factors in prominent pricing models are generally returns on portfolios formed using recent information, unavailable before month $t-1$. In an inefficient market, a set of such factors, $\tilde{f}$, can be useful in making

$$\alpha(x, 1, \tilde{f}, A) \approx 0 \tag{4}$$

for various choices of a characteristic, x. To reject such a model as being the correct no-mispricing benchmark, our test relies on detecting

$$\alpha(\hat{x}, 120, \tilde{f}, \hat{A}) \neq 0, \tag{5}$$

for a given asset universe, $A = \hat{A}$, and a given characteristic, $x = \hat{x}$.

Suppose that the model is not the no-mispricing benchmark because $\tilde{f}$ incorrectly includes or excludes a given factor, f_j. Let $\beta_j(\hat{x}, \tau, \tilde{f}, \hat{A})$ denote the factor's corresponding element of β in equation (1) if $\tilde{f}$ includes f_j. Otherwise, let $\beta_j(\hat{x}, \tau, \tilde{f}, \hat{A})$ denote the factor's slope coefficient in that regression when $\tilde{f}$ is augmented by f_j. A necessary condition for (5) to obtain, and thus for our test to have power, is that that

$$\beta_j(\hat{x}, 120, \tilde{f}, \hat{A}) \neq 0. \tag{6}$$

That nonzero beta in (6), when multiplied by the non-zero mean (factor premium) of f_j, then represents the component of expected return that is incorrectly included or excluded, resulting in the nonzero alpha in (5). The challenge faced by our test, noted earlier, is that for many asset universes and characteristics, there may be essentially no relation between individual assets' decade-old values in $\hat{x}$ and their current betas on f_j. In that scenario, a long-short strategy based on those values in $\hat{x}$ is likely to produce $\beta_j(\hat{x}, 120, \tilde{f}, \hat{A}) = 0$, the same as one would expect if the assets in the long and short legs were just randomly selected. Settings more likely to avoid this no-power scenario are those in which $\beta_j(\hat{x}, 1, \tilde{f}, \hat{A}) \neq 0$ and the individual assets' betas on f_j are stable over time. The latter condition, in turn, seems more likely when the selected universe, $\hat{A}$, comprises assets tending to have more stable factor betas generally.

2.2. Large stocks have the most stable betas

We expect larger stocks to have more stable betas. By virtue of their size, large firms are likely to have greater inertia, taking longer to change course in ways that would significantly impact their characteristics relevant to factor exposures. (Aircraft carriers turn more slowly than destroyers.)

Table 1 reports evidence consistent with this argument for betas with respect to the five factors in the model of Fama and French (2015). For each factor, we compute the rank correlation between stocks' estimated betas in months t and $t-120$ for all stocks that exist in both months. The beta estimate in a given month uses monthly returns in the prior 60 months. The table reports the average rank correlation across all months in our sample, which covers the period from January 1968 through December 2018. We compute these rank correlations within each of three size segments. Using NYSE market-cap percentiles as break points, and using market values at the end of month $t-120$, we form three subsamples consisting of all NYSE, AMEX, and NASDAQ stocks (i) above the 70th percentile, (ii) between the 70th and 20th percentiles, and (iii) below the 20th percentile. For labeling ease, we denote these segments as "large," "medium," and "small," while recognizing that other terms can be applied. For example, we sometimes use "microcaps" as an alternative label for the stocks below the 20th percentile, following Fama and French (2008).[4]

The results in Table 1 reveal that some factors generally have more stable betas than others. For example, MKT and SMB betas have rank correlations between 0.14 and 0.33, whereas the correlations for CMA betas are 0.03 or less in magnitude. For all factors, however, the rank correlation among large stocks exceeds that among the other two size segments. Moreover, aside from the tiny correlations for CMA betas, the rank correlation among medium-sized stocks exceeds that among small stocks for each of the other four factors. We see that beta stability increases in firm size, suggesting large stocks are likely to provide our test with the most power. Large stocks are also less volatile and more liquid, which reduces estimation error in their betas, which can also improve power.

2.3. Mispricing measures and test power

2.3.1. PERF and MGMT scores

Rather than entertain many separate stock characteristics, we rely on two measures that combine multiple characteristics. Specifically, we focus on PERF and MGMT, the two "mispricing" measures constructed by Stambaugh and Yuan (2017) using 11 stock characteristics corresponding to a set of prominent return anomalies analyzed previously by Stambaugh, Yu, and Yuan (2012, 2014, and 2015).[5] We follow this research design for two reasons: parsimony

[4] Stocks between the 70th and 20th NYSE percentiles essentially combine categories that are often denoted as "mid-cap" and "small-cap." In that respect, readers should not misinterpret our "small" labeling.

[5] The appendix included in Stambaugh and Yuan (2017) provides detailed descriptions of these 11 characteristics.

and transparency. First, Stambaugh and Yuan (2017) show that factors based on PERF and MGMT, when combined with market and size factors, produce jointly insignificant alphas for spreads formed using a large number of other stock characteristics beyond the set of 11. Second, with each of the two measures comprising a manageable number of characteristics, related behavioral hypotheses can be more easily identified, and we exploit that ability in analyzing our results.

Stambaugh and Yuan (2017) construct PERF and MGMT by separating the 11 characteristics into two clusters, with a cluster containing the characteristics most similar to each other. Similarity can be measured by either time-series correlations of anomaly returns or average cross-sectional correlations of characteristic rankings. As shown by Stambaugh and Yuan (2017), both methods produce the same two clusters of characteristics.

The characteristics in PERF are distress, O-score, momentum, gross profitability, and return on assets, while those in MGMT are net stock issues, composite equity issues, accruals, net operating assets, asset growth, and the ratio of investment to assets. Following Stambaugh and Yuan (2017), we compute a stock's PERF and MGMT scores in a given month by averaging the stock's rankings with respect to its available characteristics within each of the two clusters. Higher rankings for a characteristic correspond to lower abnormal returns, as identified by previous studies documenting the return anomalies. If a stock has fewer than three non-missing rankings within either cluster, its corresponding mispricing score (PERF or MGMT) for the month is missing.

2.3.2. Test power for PERF versus MGMT

At a given point in time, the stock characteristics underlying either PERF or MGMT can correlate cross-sectionally with the stocks' sensitivities to one or more of the factors in a multifactor pricing model. For such factors, the more stable are stocks' related characteristics, the more likely it is that a spread based on lagged values of those characteristics has current betas on those factors that differ between the spread's long and short legs. That is, the more likely is the spread to meet the condition in equation (6).

Figure 1 reports the persistence in PERF and MGMT scores for large, medium, and small cap stocks. Within each universe, we rank stocks on their PERF or MGMT scores and compute the average difference in scores between stocks in the top 30% versus the bottom 30%. We compute that average difference for scores lagged from 1 month ("year 0") to 120 months ("year 10"). The k-year persistence measure, displayed in the figure, is the ratio of the average score difference for the year-k lag to that for the year-0 lag.

Figure 1 shows that PERF scores exhibit substantially greater persistence than MGMT scores, for all three size categories. By the earlier reasoning, the multifactor betas on PERF spreads are more likely to be persistent and thereby satisfy the condition in equation (6), as compared to MGMT spreads. We also see in Figure 1 that the persistence in PERF scores is especially strong among large stocks. This observation that large firms have greater persistence in at least some of their characteristics seems consistent with the results presented earlier, showing multifactor betas are generally most stable for large stocks. Overall, given these findings, we expect PERF spreads among the largest stocks to provide our test with the most power. In a later section, we investigate carefully the role that persistence in characteristics and betas plays in our test results, which are presented next.

2.4. Test results

We apply our test to the CAPM, the Fama-French three- and five-factor models (denoted as FF3 and FF5), the Hou-Xue-Zhang four-factor model (denoted as Q4), and the CAPM augmented by the Frazzini-Pedersen BAB factor (denoted as CAPM+BAB).[6] For each of these five models, and for various market-cap subsamples, Table 2 reports the estimated alphas and associated t-statistics for spreads based on decade-old PERF or MGMT scores. To construct the spreads, in each month t, stocks are divided into the most underpriced 30% ($M1$), the middle 40% ($M2$), and the most overpriced 30% ($M3$) based on their PERF or MGMT scores in month $t-120$. The long-short spread is the $M1$ minus $M3$ return in month t, and the stocks in each leg are value weighted. This sorting and long-short construction is performed separately within each of the large, medium, and small segments of size described earlier, formed based on market values in month $t-120$. We also repeat the above within a "mega cap" subsample of the 200 largest stocks, as well as within the all-stock universe.

The most striking results in Table 2 occur when sorting on decade-old PERF scores, in Panel A. Across all size segments, spreads based on PERF produce relatively small CAPM alphas, between -14 and 21 basis points (bps) per month, with t-statistics of 1.55 or less in magnitude. In contrast, among large stocks, the same PERF spreads produce substantial and statistically significant alphas for all four of the multifactor models, between 29 and 46 bps, with t-statistics between 2.35 and 4.16. Even a PERF spread among just the mega caps produces equal or greater alphas for FF3, FF5, and Q4, with significant t-statistics between

[6]We download factors for the last four models from the respective authors' websites.

2.16 and 3.32, despite there being only 200 stocks in that category. For CAPM+BAB, the mega-cap alpha is 7 bps lower than among all large caps, with the t-statistic dropping from 2.71 to 1.81, but the mega-cap alpha is still a substantial 31 bps per month, versus the CAPM's alpha of 19 bps with a t-statistic of just 1.16.

Consistent with our prior that large stocks offer our test the most power, the large-stock segment is where we see differing test outcomes across models. For the small-stock segment, decade-old PERF scores produce nothing close to a rejection for any of the models, with t-statistics between -0.49 and -1.10. That is, microcaps do not allow our test to say anything about whether one model serves better than another as the proper no-mispricing benchmark. This result is unsurprising. Firms that were microcaps a decade ago, if still around today, are likely to have experienced significant changes in their fundamental characteristics, implying their decade-old PERF scores may have little relation to their current betas on any of the various models' factors. A sort on those PERF scores then produces essentially the same estimated alphas one would expect from a random sort—zero.

The corresponding results for medium-sized stocks lie between the two contrasting cases for large and small stocks, just as one would expect with a positive relation between firm size and test power. For medium stocks, the CAPM is again not rejected, producing an alpha of -9 bps with a t-statistic of -0.54, the smallest values in absolute magnitude among the four models. For the multifactor models, the results are mixed: FF5 and CAPM+BAB produce alphas of 26 and 28 bps, respectively, with t-statistics of 2.10 and 2.02, thus being rejected; Q4 produces nearly the same alpha, 25 bps, but with a marginal t-statistic of 1.64; FF3 is clearly not rejected, with an alpha of 10 bps and a t-statistic of 0.73. Thus, as compared to small stocks, stronger evidence against at least two of the multifactor models emerges among medium stocks, but the strongest evidence clearly resides with large stocks.

Microcaps account for roughly half of the total number of stocks in a given month, so given the above results for microcaps, including them when sorting on decade-old PERF scores is likely to shrink the multifactor alphas considerably. Indeed, when sorting within the all-stock universe, the resulting alphas with respect to FF5 and Q4 are just 17 and 18 basis points, with t-statistics of 1.62 and 1.54, which are smaller than what those models produce within the large and medium categories. For FF3 and CAPM+BAB, however, the all-stock alphas, though lower than the large-stock alphas, are nevertheless substantial, equal to 31 and 29 bps, with t-statistics of 2.86 and 2.19.

The results in Panel B of Table 2 reveal that our test cannot discriminate across models when based on MGMT scores. The alphas for all of the models, across all size categories, are no greater than 17 bps in magnitude, and the t-statistics are no greater than 1.54 in magnitude. MGMT scores, in contrast to PERF scores, exhibit less long-term persistence, and decade-old MGMT scores bear little relation to current estimated betas on the various model's factors. PERF scores exhibit considerable persistence, as shown earlier, and we later relate that property to persistence in the PERF spread's multifactor betas. Nevertheless, for those wishing to set aside this distinction between PERF and MGMT and view the earlier PERF results as simply "one out of two," we next conduct multiple-comparison versions of our test to assess whether rejection of the multifactor models based on our test is simply due to chance.

2.5. Multiple comparisons

The first four columns of results in Table 3 report p-values for a joint test of whether the PERF and MGMT alphas are jointly zero. This Wald test essentially corresponds to that of Gibbons, Ross, and Shanken (1989), modified to incorporate a heteroskedasticity- and autocorrelation-consistent covariance matrix, with p-values thus computed using the asymptotic chi-square distribution rather than the F-distribution. For large stocks, the CAPM produces a p-value of 7.5%, whereas the four multifactor models produce p-values less than 2%. Thus, even when ignoring the distinction between PERF and MGMT that is relevant to our test's power, a meaningful distinction emerges between the CAPM and the multifactor models. At a conventional 5% significance level, the joint test does not reject the CAPM as the appropriate no-mispricing benchmark model, but all four of the multifactor models are rejected. Among mega-cap stocks, the p-values for FF3, FF5, and Q4 are just slightly higher but still well below 5%. As with the PERF result in Table 2 for CAPM+BAB among mega caps, the corresponding result in Table 3 for that 200-stock category is again marginal, with a p-value of nearly 8%. In contrast, though, the CAPM's p-value among mega caps is 36%. For the medium and small categories, the joint test rejects none of the five models at conventional significance levels, which is not surprising given the results for these categories in Table 2.

The last two columns of Table 3 report p-values for tests that set aside the power-relevant distinctions between not only PERF and MGMT, but also between large and small firms. The test reported in the next-to-last column is the same as reported in the previous four columns but performed on the all-stock universe. The last column reports a test of whether both the PERF and MGMT alphas, in each of the three separate size categories (large, medium, and small), are all jointly equal to zero. In other words, this last test considers a six-element vector of alphas. Most of the p-values in both columns exceed conventional significance levels except those for FF3 and CAPM+BAB. For FF3, the p-values are 1.25% in the next-to-last column and 0.12% in the last column. For that popular, long-standing model, evidently the decade-old PERF scores of large stocks identify violations strong enough to survive dilution by including smaller stocks via either of the two underlying channels. For CAPM+BAB, the all-stocks p-value in the next-to-last column is 2.39%. The multifactor models' remaining p-values in the last two columns, between 5.7% and 13.1%, while exceeding conventional significance levels, are substantially smaller than those of the CAPM, which are 30% and 18.9%. Overall, these all-in multiple-comparison tests, which ignore power distinctions, echo the same message as earlier but with weaker statistical significance.

The results of our test among large stocks reveal what seems to be an economically significant shortcoming of popular multifactor models. These models fail significantly in capturing expected returns on strategies based on decade-old information. Our assumption that such information should be fully reflected in prices seems especially reasonable for the market's largest and most liquid stocks. The CAPM, in contrast, fares well in this regard, emerging as a better candidate for the no-mispricing benchmark model.

3. Alphas and betas across lags

To gain additional perspective, we compute the alphas and factor betas of portfolios formed by sorting on PERF and MGMT across various lag lengths, up to the ten years used in the test. We first examine the lag patterns of alphas within each of the size segments. We then decompose the large-stock alphas across lags, revealing which of PERF's factor betas exhibit the persistence underlying each multifactor model's rejection by our test.

3.1. Alphas and lag length

Figure 2 plots the alphas of long-short spreads sorted on PERF lagged 1, 12, 24, ..., 120 months, moving from left to right on the horizontal axis (labeled in years and denoting the 1-month lag as year 0). The top three panels plot the alphas under each model for each of the three size groups, and the bottom three panels plot their corresponding t-statistics (displaying reference lines at ± 1.65 and ± 1.96).

First, consider results for the large stocks, shown in the leftmost plots in Figure 2. When sorting on PERF scores in the most recent month, the $M1 - M3$ spread produces monthly alphas between 26 and 62 bps across the five models, all statistically significant (marginally so for Q4). These results echo those in Stambaugh and Yuan (2017), who sort stocks on the most recent mispricing score (combining PERF and MGMT) and find significant long-short alphas with respect to various prominent models. When PERF scores are instead lagged one year, however, we see that the resulting spread's CAPM alpha drops to just 6 bps, with an insignificant t-statistic of 0.42. In contrast, the spread's monthly alphas with respect to the other three models are between 29 and 44 bps, with t-statistics between 2.1 and 4.8. At longer lags, the CAPM alphas remain well smaller than those of the other models and never rise to a level that is statistically different from zero. In contrast, all four multifactor models produce positive and significant alphas that generally persist across the multi-year lags.

The results for medium-sized stocks, in the middle panels of Figure 2, are similar to those for large stocks in that the CAPM alphas lie consistently well below the multifactor alphas at longer lags. The statistical significance of the latter alphas is weaker, consistent with the lower power discussed earlier. For microcaps, in the rightmost plots, alphas for all models drop sharply as the lag in PERF scores lengthens. That pattern dominates any cross-model differences, which essentially converge to zero at the longest lags, consistent with microcaps having little long-run stability in multifactor betas and hence providing no power to discriminate across models.

Figure 3 plots alphas across lags for spreads based on MGMT scores. Recall from our test results that all models, in all size segments, deliver statistically insignificant alphas when MGMT scores are lagged ten years. Figure 3 reveals a general pattern of substantial declines in the alphas as the lag length increases from one month to a few years. The basic reason, noted earlier, is that factor betas for MGMT spreads are less persistent than those for PERF spreads. Therefore, the MGMT spread's long- and short-leg betas with respect to a given factor tend to converge more quickly to the same value as the lag lengthens. As a result, the MGMT spreads essentially provide no power for our test when sorting on decade-old scores.

Figure 4 shows results for the 200 largest stocks (mega caps) that correspond to the plots for large, medium, and small stocks displayed in Figures 2 and 3. As noted earlier, even when the PERF and MGMT spreads are formed just within the mega caps, our test produces results quite similar to those obtained for the broader large-cap segment. The plots in Figure 4 reveal that this strong similarity of results extends across all lags.

Even though the MGMT spreads provide our test with insufficient power to distinguish the various models' suitability as the no-mispricing benchmark, the MGMT spreads do help provide insight, along with PERF, into potential empirical motivations for multifactor pricing models. Specifically, consider the alphas for spreads formed based on the most recent scores (i.e., at the end of the previous month), corresponding to year 0 in the plots. In Figure 3, for example, note that the alphas based on the most recent MGMT scores are greatest for the CAPM across all size segments; the multifactor models, especially FF5 and Q4, reduce those alphas considerably. For PERF, the latter two multifactor models also reduce the year-0 CAPM alphas among the medium and small stocks, which are the segments in which year-0 CAPM alphas are the greatest.

Combined with our test results, the above year-0 observations suggest a narrative in which many apparent violations of the CAPM produced by year-0 (month $t-1$) sorting characteristics, such as PERF and MGMT, reflect short-term mispricing rather than omitted components of expected returns under the no-mispricing benchmark model. Multifactor models reduce apparent year-0 violations by partially capturing mispricing, but the models then distort expected returns relative to the no-mispricing benchmark. We next investigate sources of these distortions.

3.2. Decomposing alphas

When applied to large stocks, our test rejects each of the multifactor models as being the no-mispricing model. To understand better the source of these rejections, Tables 4 through 8 decompose the large-stock alphas for the PERF and MGMT spreads plotted in Figures 2 and 3. Specifically, the alpha at each lag is decomposed into the average return spread minus the product of the spread's estimated beta times the sample-average factor premium for each factor in the model. Panel A reports results for PERF, and Panel B reports results for MGMT. For example, the first row of Panel A of Table 4 decomposes the FF3 alpha based on PERF scores from the most recent month (year 0) as follows:

$$\underbrace{\text{raw excess return}}_{0.21} + \underbrace{(-\beta_{MKT}\overline{MKT})}_{0.14} + \underbrace{(-\beta_{SMB}\overline{SMB})}_{0.01} + \underbrace{(-\beta_{HML}\overline{HML})}_{0.26} = \underbrace{\alpha}_{0.62}$$

Note that the bulk of the alpha is coming from a negative exposure to HML that delivers an additional 26 bps above the 21 bps excess return. The alpha of 62 bps, with a t-statistic of 6.05, confirms the predictive ability of the PERF mispricing measure of Stambaugh and Yuan (2017).

When PERF is lagged 12 months or more, in the remaining rows of Panel A in Table 4, the average return spreads are generally quite small, all less than those at the one-month lag.

As the lag increases, the spread's market-exposure component shrinks slightly but remains fairly stable, declining from 0.14 at one month to 0.10 at ten years. The SMB component is consistently tiny. At all lags, the largest component of alpha comes from HML exposure, whose contribution is 26 bps at year 0 and declines only modestly to 20 bps at year 10. In other words, the PERF spread's negative beta on HML is very persistent. The persistent HML beta injects an unhelpful component of implied expected returns for the PERF spread well after any earlier mispricing associated with the stale PERF scores is likely to have disappeared. Any persistence in mispricing, especially among large and liquid stocks, is likely to be far less than the observed persistence in PERF's HML exposure. We see that the latter persistence accounts for the rejection of FF3 as the no-mispricing benchmark model in the sense of Fama (1970), based on our test using decade-old PERF scores.

Panel B of Table 4 reports the corresponding FF3 results for spreads based on lagged MGMT scores. The average return spread equals 37 bps at a one-month lag (year 0), declines to 14 bps at a three-year lag, and takes even smaller values at longer lags. The year-0 alpha of the spread is 27 bps (t-statistic: 3.63), and the year-1 alpha remains nearly that high at 25 bps (t-statistic: 3.09). At longer lags, however, the alpha is small and insignificant. The contributions to alpha from market and SMB exposures are consistently small. Most importantly, in contrast to the PERF results, the contribution from HML exposure shrinks substantially toward zero as the lag increases. Thus, as the lag in MGMT scores increases, the average spread return shrinks, but so does the contribution from HML exposure, resulting in no significant alpha. The shrinking exposure to HML indicates that the HML betas of the MGMT spread's long and short legs become similar within a few years. This lack of persistence in betas leaves our test based on decade-old MGMT scores with little power.

For the FF5 model, Table 5 presents the same alpha decomposition. The results based on PERF spreads, in Panel A, are very similar. The component of alpha arising from HML exposure is even somewhat larger than in FF3 (Table 4). Again, that exposure persists at long lags, injecting an unhelpful component of implied expected return well after any previous mispricing associated with the lagged PERF scores is likely to have been corrected. The negative alpha components reflecting exposures to the two additional factors, investment (CMA) and profitability (RMW), help produce lower alphas than does FF3. That result is consistent with those factors helping to capture mispricing. As the lag increases, however, PERF's exposures to CMA and RMW shrink and only modestly offset the positive contribution from the persistent HML exposure. The latter contribution, as with FF3, accounts for our test's rejecting FF5 as the no-mispricing benchmark model. For the spread based on decade-old PERF scores, the alpha is 29 bps per month with a t-statistic of 2.62.

Panel B of Table 5 reports results for MGMT spreads. As noted earlier, our test fails to reject FF5 when using MGMT, as the year-10 alpha is -12 bps with a t-statistic of -1.29. The results for the other lags in Panel B, however, offer a nuanced interpretation. Recall from Table 4 that the spread based on MGMT scores from the most recent month (year 0) produces a FF3 alpha of 27 bps, with a t-statistic of 3.63. That spread's alpha under FF5 drops to just 5 bps with a t-statistic of 0.76. At longer lags, though, the FF5 alphas turn negative and produce some relatively large t-statistics. For example, if we had designated seven years rather than ten in our test's no-mispricing assumption, the corresponding test would have produced a rejection for FF5 using MGMT (t-statistic: -2.47). The reason for the negative alphas at longer lags is that the MGMT spread's HML, CMA, and RMW exposures all make relatively persistent negative contributions to alpha that are not offset by MGMT's positive but progressively smaller average return spreads at longer lags. By year 10, the most important of the alpha contributions, reflecting exposure to the CMA investment factor, declines to less than 40% of its year-0 value. Therefore, insufficient beta stability, as discussed earlier, ultimately dooms our test based on decade-old MGMT scores. At the same time, somewhat more CMA beta stability would evidently have otherwise led to rejection. Also worth noting is that MGMT's year-0 alpha under FF3 (and under the CAPM) is the largest in absolute magnitude across all lags, whereas that same spread's tiny alpha under FF5 is the second smallest in absolute magnitude across all lags (and just 1 bp higher than the year-1 value). Viewed collectively, these results for MGMT at least suggest a scenario, as outlined earlier, in which adding investment and profitability factors helps capture mispricing and satisfy the condition in equation (4). Persistent exposures to those factors distort implied expected returns after the mispricing is corrected.

Table 6 reports the same alpha decomposition for the Q4 model of Hou, Xue, and Zhang (2015). This model does not contain the value factor, HML, which is the main factor causing distortions to expected returns in the FF3 and FF5 models. Instead of including HML, the Q4 model adds investment and profitability factors, IA and ROE, to the market and size factors. Panel A reports the results for the PERF spreads. As with FF3 and FF5, we again see a factor injecting an unhelpful positive component of alpha at longer lags. Unlike with those models, where the unhelpful factor is HML, here we see IA is the most troublesome factor for the Q4 model. The IA column in Panel A of Table 6 contains values very similar to those in the HML column in Panel A of Table 4 (in fact the year-0 and year-10 values are identical). As with FF5, PERF's exposure to the profitability factor makes a relatively persistent negative contribution to alpha that only partially offsets the unhelpful positive contribution from the most troublesome factor, IA. Thus, we arrive at the same explanation as to why our test rejects Q4 using PERF; the explanation here simply replaces HML with IA. We return a bit later to this swapping of roles between IA and HML as the troublesome factor causing distortions in expected returns.

Panel B of Table 6 suggests the same interpretation for the Q4 model discussed previously as that of Panel B of Table 5, for the FF5 model. The year-0 MGMT spread's alpha under Q4 is 5 bps, the same as the spread's FF5 alpha. At longer lags, alphas are again negative, larger in magnitude, and often with substantial t-statistics. As with FF5, the most important contributor to the negative alphas at longer lags is the investment factor, IA, and the pattern across years is similar: MGMT's IA exposure exhibits nontrivial persistence across lags, but not enough to produce a significant alpha at year 10. The IA contribution at year 10 is only 35% of its year-0 contribution.

Spreads based on lagged PERF scores exhibit persistent exposures to the value factor, HML, that largely account for the rejections of FF3 and FF5 as the no-mispricing benchmark model. Recall that the rejection of the Q4 model, however, owes largely to persistent exposures to Q4's investment factor, IA. The results in Table 7 reveal that this swapping of roles as the troublesome factor is due to the fact that Q4 does not include a value factor. To draw that conclusion, we augment the factors in Q4 with HML and decompose alphas with respect to that multifactor model. First, our test based on decade-old PERF scores among large stocks also rejects this model as the no-mispricing benchmark. Specifically, the year-10 alpha reported in Panel A of Table 7 is 36 bps with a t-statistic of 3.16. The largest contribution to that alpha is HML exposure, which is strongly persistent across all lags. The alpha contribution from IA exposure is consistently small. This dominance of HML exposure for the PERF spread therefore closely resembles that for FF5 in Panel A of Table 5. In other words, when comparing these results to those in Panel A of Table 6, we see that the IA exposure of the PERF spread simply serves as a stand-in for the spread's HML exposure when HML is absent, as in the Q4 model.

Panel B of Table 7 decomposes the MGMT-spread alphas with respect to the model that augments Q4 with the HML factor. These results are very similar to those for Q4 in Table 6. The alpha is very small for year-0 but is larger and negative at longer lags, often producing significant t-statistics. The year-10 value is again insignificant because the investment factor's large negative alpha contribution is not persistent enough as the lag increases.

Table 8 reports decompositions of large-stock alphas with respect to the CAPM augmented by the BAB factor of Frazzini and Pedersen (2014). Panel A presents results for the

PERF spreads. We see that persistent negative exposure to the BAB factor results in a persistent positive alpha contribution across lags. That contribution is the largest component of alpha at all lags beyond year 0. As with the other multifactor models, persistent exposures to a non-market factor continue to inject unhelpful components into implied expected returns well after mispricing related to stale PERF scores is likely to have been corrected.

Panel B of Table 8 reports corresponding results for MGMT-spread alphas. In this case, exposures to the BAB factor make a persistent contribution to alpha, but the contribution is negative and offsets positive contributions from average return and market exposure. As a result, the alpha at the ten-year lag is insignificant, with a t-statistic of -1.08. Again, spreads based on MGMT do not provide enough power to allow our test to distinguish among models as being the no-mispricing benchmark.

4. Persistence in characteristics

A spread based on decade-old PERF scores exhibits a very persistent HML beta, especially among large stocks. What accounts for that persistence? Is price correction why the spread's average return is less persistent? We next explore these two questions, focusing on the role of persistence in characteristics that constitute the PERF score.

4.1. HML beta and persistent characteristics

As noted earlier, a stock's PERF score is the average of its cross-sectional rankings (percentiles) on five characteristics: gross profitability, return on assets, distress, O-score, and momentum. To streamline our analysis, we average the first two characteristics to form a profitability score, and we average the second two to form a "financial health" score (the negative of the two distress measures). Figure 5 plots the average, across all sample months, of the current difference in average profitability scores (Panel A), financial health scores (Panel B), and momentum scores (Panel C) between the long and short legs of spreads formed on lagged PERF scores among large-cap stocks. The lags range from 1 month (0 years) to 10 years, and the shaded areas indicate plus/minus 1.96 standard errors.

As Figure 5 indicates, the spread in profitability scores between the PERF spread's long and short legs ($M1$ versus $M3$) declines slowly as the lag increases but remains significant even at the ten-year lag. In other words, the profitability difference between today's high-and low-PERF stocks tends to persist well into future years. This persistence in profitability is consistent with the low turnover of profitability-sorted portfolios found by Novy-Marx (2013) and Frazzini, Israel, and Moskowitz (2018). The second plot in Figure 5 shows that financial health is similarly persistent over a 10-year horizon. The third plot, however, shows that there is little persistence in momentum differences, which is consistent with the high turnover of momentum strategies (Frazzini, Israel, and Moskowitz (2018)).

The persistence in some of PERF's underlying characteristics can explain much of the persistence in PERF's HML beta. To show this, we first regress individual large stocks' current HML betas on the stocks' current profitability, financial health, and momentum scores. The HML betas are estimated in the FF5 model, using the most recent 60 months of return data. Table 9 reports the coefficients and t-statistics from this Fama-MacBeth regression. The HML beta exhibits a strongly significant negative relation to profitability (t-statistic: -8.87) and a marginally significant negative relation to financial health (t-statistic: -1.81). As observed above, the PERF spread's long-short differences in these characteristics are quite persistent.

We can assess the role of that persistence in capturing the persistence in PERF's HML beta. Specifically, for each characteristic, we multiply the coefficient in Table 9 by the PERF spread's corresponding long-short difference in the characteristic plotted in Figure 5. We then sum those products across characteristics to obtain the fitted HML beta of the PERF spread for each lag. Figure 6 plots those fitted betas along with the actual estimated betas. The fitted betas are uniformly smaller in absolute magnitude than actual betas, because the regression does not explain all cross-sectional variation in HML betas. Thus, the regression predicts less of a difference between the betas on the long and short legs than exists in the actual estimates. However, the fitted betas exhibit strong persistence, not as much as in the actual beta estimates, but a substantial amount nonetheless. In other words, persistence in PERF's characteristics accounts for much of the persistence in the HML beta of the PERF spread. Recall that the latter persistence lies at the heart of our test's rejection of the multifactor models that include the HML factor.

Given that the HML factor is a spread between high and low ratios of book to market (BM), and given that persistence in the HML betas on PERF spreads plays such a key role in our test's ability to distinguish among models, a natural question arises: Do spreads based on decade-old BM, rather than PERF, also allow our test to distinguish among models? The answer is no. Such spreads do not produce significant alphas with respect to any of the models considered. The reason is that the HML betas on BM spreads exhibit less persistence than do the HML betas on PERF spreads. In that respect, the HML betas on BM spreads have more in common with those on MGMT spreads, with the lack of persistence similarly depriving our test of sufficient power to distinguish among models.

At the root of this somewhat surprising result is that current BM is related less to past BM than to past PERF characteristics, especially profitability. Table 10 reports estimates from a multiple regression of large stocks' current BM values on decade-old values of BM, profitability, financial health, and momentum. The coefficient on past BM is insignificant, with a t-statistic of 1.22, but the coefficient on past profitability is strongly negative, with a t-statistic of -3.51. Recall from the earlier discussion that profitability is typically rather persistent. Our finding in Table 10 that firms with low past profitability tend to have high current BM is consistent with a scenario in which their past stock prices underreacted to the low profitability but eventually fell as profitability remained low. The same underreaction would make past BM, before that price decline, less useful in predicting current BM, thereby reducing the persistence in HML betas on BM spreads. We next examine the underreaction hypothesis more closely.

4.2. Underreaction and persistent characteristics

Our test rejects prominent multifactor models as being the no-mispricing benchmark. As discussed above, these rejections are consistent with the PERF characteristics, and thus the PERF spreads' multifactor betas, persisting longer than mispricing. In order for PERF to persist longer than whatever mispricing it identifies, PERF must be an imperfect indicator of mispricing. We next explore whether PERF's limitation as a mispricing measure is consistent with the likely source of mispricing it identifies.

PERF's underlying characteristics are profitability, financial health, and momentum. Previous studies provide evidence that the abilities of all three characteristics to predict stock returns reflect the market's underreaction to relevant information. For example, Bouchard, Krüger, Landier, and Thesmar (2019) link the profitability anomaly to underreaction; Dichev (1998) links the financial-health (distress) anomaly to underreaction; Da, Gurun, and Warachka (2014) link the momentum anomaly to underreaction. Moreover, Chen, He, Tao, and Yu (2020) directly link the PERF measure's predictive ability to underreaction.

If the mispricing identified by PERF reflects underreaction, then the degree of that mispricing should depend on how long a stock has had its current PERF score. As observed earlier, PERF scores are generally quite persistent, especially for large firms, but we expect any associated mispricing to be less persistent. For a stock whose PERF score has not changed in years, it seems unlikely that the market would still be underreacting to whatever information that PERF score initially represented. For a stock whose PERF score recently changed, underreaction to the underlying information reflected in that score is more plausible.

To examine this hypothesis, we distinguish PERF scores that have persisted at their current levels from PERF scores that have more recently changed. We first divide all large stocks into quintiles based on their PERF scores in the most recent month. Let $Q_{i,t}$ denote the stock's PERF quintile, with quintile 1 containing the most underpriced stocks and quintile 5 the most overpriced. Within PERF quintiles 1 and 5, we further divide stocks into quintiles, P1 through P5, based on the persistence of their current score over the past ten years. Specifically, for each stock i in month t, we compute

$$Persistence_{i,t} = [I(Q_{i,t} = 5) - I(Q_{i,t} = 1)] \times \sum_{\tau=13}^{120} \omega_\tau \left[M_{i,t-\tau} - 0.5\right], \tag{7}$$

where $I(Q_{i,t} = n)$ equals 1 if $Q_{i,t} = n$ and 0 otherwise, $\omega_\tau = [10 - int((\tau - 1)/12)]/9$, and $M_{i,t-\tau}$ is the percentile rank of stock i based on its PERF score in month $t - \tau$, with a high (low) rank corresponding to overpricing (underpricing). To provide some sense of the differences in the persistence of scores, P1 firms have on average been in their PERF-score quintile for around a year (12 months), while P5 firms on average have been in their respective PERF-score quintiles for about 8 out of the last 10 years.

For each persistence quintile, the first row of Table 11 reports the monthly CAPM alphas of long-short spreads based on PERF quintiles 5 minus 1. The underreaction hypothesis predicts that long-short spreads based on PERF scores should be the most profitable when formed with stocks whose PERF scores are the least persistent. Consistent with this hypothesis, the PERF-spread alphas are nearly monotonic across persistence quintiles. The spread based on the least persistent PERF scores has the highest alpha, equal to 89 bps per month with a t-statistic of 3.57. The spread based on the most persistent scores has an alpha of only 24 basis points with a t-statistic of 1.02. In other words, PERF scores that have remained relatively constant for a long time exhibit no significant ability to predict returns. The difference of 66 bps between the least and most persistent PERF quintiles has a t-statistic of 2.29. The last column shows that a PERF spread using the least persistent scores outperforms the PERF spread within the entire large-stock segment (regardless of score persistence) by 50 bps per month (t-statistic of 2.69). The results for the PERF spreads in Table 11 strongly support the underreaction hypothesis and demonstrate why mispricing is likely to be less persistent than PERF scores and their underlying stock characteristics.

The difference in persistence between mispricing and PERF scores among large stocks is essentially what enables our test to distinguish the CAPM from prominent multifactor models as the better candidate for the no-mispricing benchmark model. On the flip side, for the objective of finding profitable trading strategies, our results indicate that the stock characteristics underlying PERF, such as profitability and financial health, are primarily useful when dissimilar to their past values. When a firm's values of these characteristics have been similar for a long time, the characteristics lose their predictive ability for returns, consistent with our premise that mispricing does not last very long in competitive markets.

The second row of Table 11 repeats the above analysis using MGMT scores to identify mispricing. The literature generally does not advance underreaction as an explanation for the predictive abilities of the stock characteristics that constitute MGMT (net stock issues, composite equity issues, accruals, net operating assets, asset growth, and investment to assets). Consistent with that observation, we see no pattern in the MGMT-spread alphas across the persistence quintiles. The P1 minus P5 difference in alphas is -11 bps with a statistically insignificant t-statistic of -0.49. MGMT scores strongly predict returns, however, delivering a CAPM alpha for the MGMT spread within all large stocks of 64 bps per month (t-statistic of 5.75).

5. Reinterpreting related studies

When compared to prominent multifactor models, the CAPM emerges as the best choice for the benchmark model of proper pricing in the sense of Fama (1970). Our test reveals that the CAPM prices decade-old information better than prominent multifactor models. This finding resembles some other recent results from the literature, but we provide a rather different motivation and interpretation. Our framework can reconcile and provide some additional clarity to the other results, offering a unified explanation.

For example, Cho and Polk (2020) find that the CAPM describes the cross-section of prices better than it describes expected short-horizon returns. Their finding can be interpreted through the lens of our test. Price levels of stocks, viewed as discounted present values of long-term assets, are likely to depend more on discount rates applied in the long run than on short-run discount rates. If mispricing gets corrected in the short run, as we assume, then it affects expected returns (discount rates) only at short horizons. When abstracting from mispricing (conditioning only on decade-old information), we find that the CAPM best captures expected returns. Likewise, by looking at prices, Cho and Polk (2020) effectively mitigate the impact of mispricing distortions, present in short-horizon expected returns, thereby allowing the expected returns that are unaffected by mispricing, i.e., those prevailing in later periods, to have the dominant voice. The CAPM then fares well, consistent with our results. Cho and Polk (2020) do not interpret their result in this manner and instead leave it as a challenge for future work. Nor do they conduct tests to distinguish among various asset pricing models. We offer a novel test that distinguishes among models and provides a plausible interpretation of their results.

Keloharju, Linnainmaa, and Nyberg (2020) find that persistent differences in firm characteristics do not predict stock returns, a result that they argue is consistent with long-term expected returns not varying across stocks. While the authors admit a mispricing explanation, they largely interpret their findings as reflecting changes in firms' risks over time, with risks reverting toward common levels. We offer a different (and somewhat opposite) interpretation. That is, the reason persistent characteristics fail to have long-term return consequences is because they imperfectly identify mispricing, which dissipates faster than the characteristics change. The characteristics do not matter for expected returns absent mispricing (e.g., expected returns in more distant periods), consistent with our finding that the CAPM serves better than do multifactor models as the proper no-mispricing benchmark. Of course, separate from that point, CAPM betas could also revert to a common level (one), making long-run return forecasts equal across stocks, but our interpretation does not involve risks converging. In fact, as explained earlier, our test's power to discriminate among prominent asset pricing models hinges on stability in multifactor betas.

Yara, Boons, and Tamoni (2020) examine the returns to 56 characteristic-sorted portfolios up to five years after portfolio formation and compare the return differences between long-short portfolios sorted on the most recent information about a characteristic and lagged (up to five years) information about a characteristic. For about half of the cross-sectional characteristics they study, they find significant pricing errors between the recent and old characteristics that are not captured by existing asset pricing models, where 2/3 of those characteristics contain more return predictability from newer information and 1/3 contain more predictability from old information. Yara, Boons, and Tamoni (2020) argue that overfitting of short-term return predictability by multifactor models (e.g., Harvey and Liu, 2019) may explain these findings. However, the authors do not link their findings to asset pricing theory or explain why some characteristics benefit more from older versus newer information. They simply show that existing asset pricing models fail to capture characteristic-sorted portfolio returns at different horizons, where the CAPM does better at pricing older sorts, but worse at newer sorts, while popular mulitfactor models like the Fama and French five-factor model do better at pricing newer sorts than older sorts, all consistent with our results. Our study provides insights into *why* some characteristics produce alphas for longer periods than others, depending on which pricing model is used to compute alpha.

6. Conclusion

In an efficient market, public information is properly reflected in prices, but assessing efficiency rests on having a pricing model that defines "properly" (Fama, 1970, 1991). We investigate whether prominent asset pricing models appear suitable for that role. We assume prices properly reflect at least the information the market has had ten years to evaluate, whether or not the market is efficient. With this assumption, a model suitable as the no-mispricing benchmark should clear the seemingly modest hurdle of assigning zero alphas to long-short spreads based on decade-old information.

We find a number of prominent models fail that test, assigning significant alphas to spreads formed using ten-year lags of the Stambaugh and Yuan (2017) PERF mispricing score. The above models include the three- and five-factor models of Fama and French (1993, 2015), the four-factor model of Hou, Xue, and Zhang (2015), and the CAPM augmented with the betting-against-beta factor of Frazzini and Pedersen (2014). In contrast, the same long-short spreads do not produce significant alphas with respect to the traditional CAPM of Sharpe (1964) and Lintner (1965). Therefore, the CAPM emerges from this evaluation as the best candidate for the no-mispricing benchmark.

It seems reasonable that spreads based on decade-old information should receive zero alpha with respect to the no-mispricing benchmark. A model can fail to be the no-mispricing benchmark, assigning nonzero alphas to those spreads, by improperly including or excluding betas with respect to one or more factors. The main challenge faced by our test, however, is achieving power. For many stocks, current values of those betas may be unrelated to decade-old information. The above spreads then have essentially zero betas with respect to the improperly included or excluded factors, offering our test no power.

We show that large stocks have the most stable multifactor betas, thus offering our test the most power. Consistent with that evidence, the large-stock segment is where a ten-year-lag PERF spread provides the clearest rejections of the various multifactor models as being the no-mispricing benchmark. Our evidence is consistent with a scenario in which mispricing, due to underreaction that PERF imperfectly identifies, corrects faster than the PERF spread's multifactor betas decay toward zero. Those betas then persist in making unhelpful distortions to implied expected returns. The underreaction scenario is supported by previous evidence linking price underreaction to the characteristics that constitute PERF (i.e., firm profitability, financial health, and momentum). We find additional support for underreaction, focusing on large stocks, where the longer a PERF score remains unchanged, the less useful it is in predicting returns, consistent with the market not underreacting to information it has known for some time.

That large stocks play the strongest role in our test underscores the economic importance of the results. For these stocks that are the backbone of the US economy, prominent multifactor models evidently distort expected returns purged of mispricing. Mispricing of large stocks is likely short-lived, given their high visibility and liquidity, so long-run decisions using expected returns not purged of mispricing could create substantial economic consequences and capital misallocations.

The traditional CAPM fares well in our test, compared to prominent multifactor models. At the same time, our study is probably better viewed as saying more about the latter models than as issuing an unqualified endorsement of the CAPM as being the proper no-mispricing benchmark model. The relatively parsimonious set of models we consider is far from exhaustive. Factors such as firm size and liquidity, among many others, are reasonably entertained as belonging in the no-mispricing model. Identifying the right set of factors, let alone constructing them properly in a world that may include mispricing, must remain challenges for future research. In an initial attempt to compare the abilities of pricing models to serve as the no-mispricing benchmark, we believe the models we consider present a horserace with interesting entrants. We certainly acknowledge that there could be speedier horses out there. Seeking refinements of our approach that potentially offer more power also seems a worthy research objective.

Of course, another worthy and parallel objective for research in asset pricing is to continue building models that better describe actual expected returns, whether or not the prices determining those expected returns include mispricing. Although such models may be less useful for gauging the extent of market inefficiencies or understanding risk premia, they can be otherwise useful, such as in designing investment strategies.

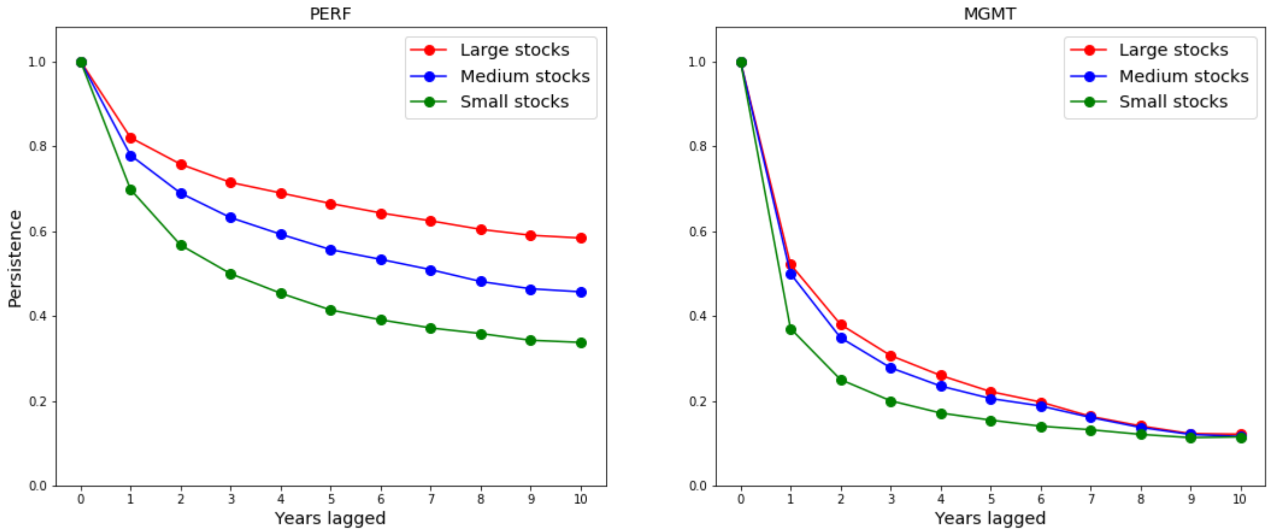

Figure 1. Persistence in PERF and MGMT scores. The figure displays the persistence in PERF and MGMT scores for large, medium, and small cap stocks. Within each universe, we rank stocks on their PERF or MGMT scores and compute the average difference in scores between stocks in the top 30% versus the bottom 30%. We compute that average difference for scores lagged from 1 month ("year 0") to 120 months ("year 10"). The k-year persistence measure, displayed in the figure, is the ratio of the average score difference for the year-k lag to that for the year-0 lag.

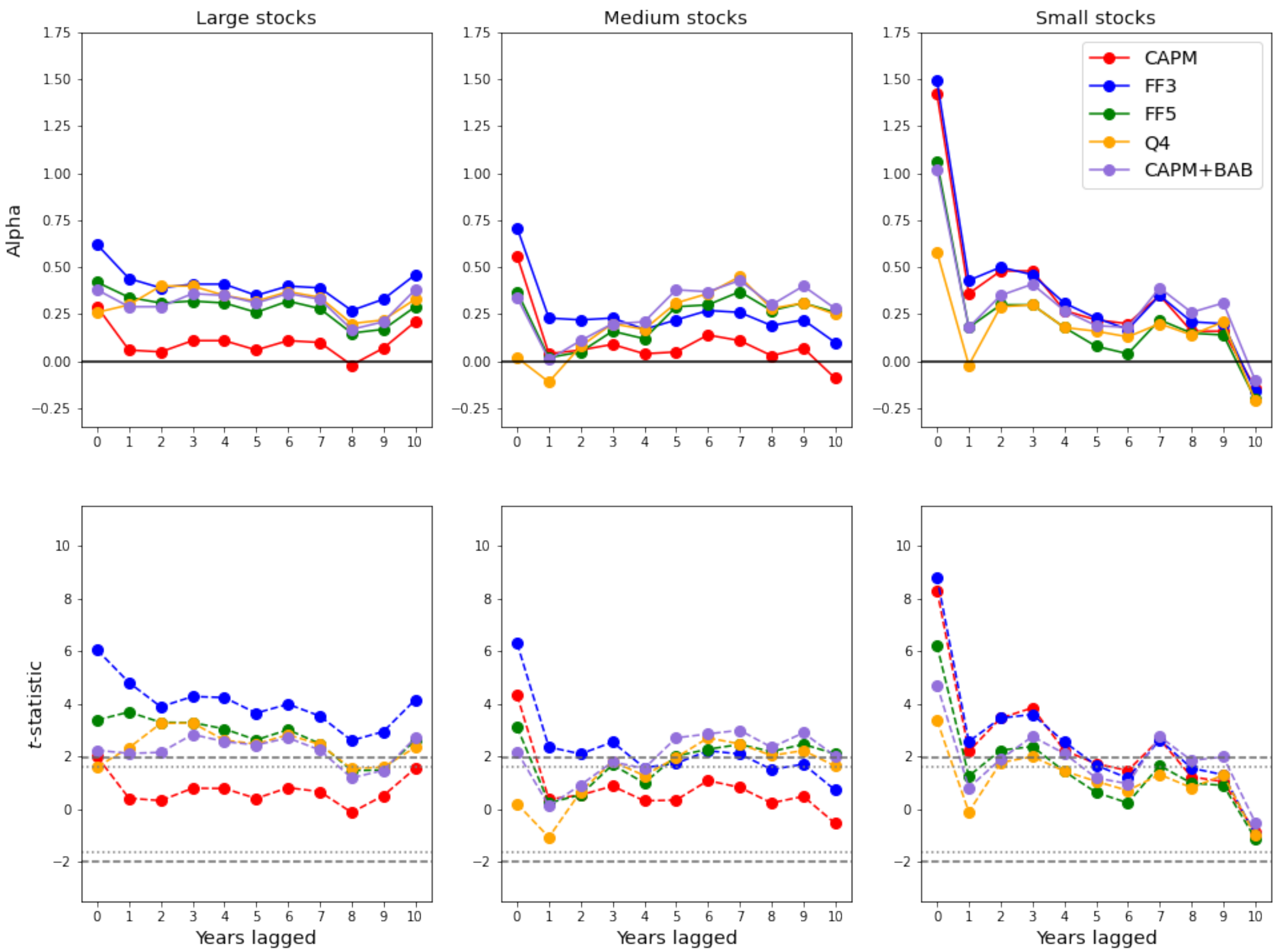

Figure 2. **Spreads formed with PERF scores.** The figure plots estimated alphas (top panels) and their t-statistics (bottom panels) for long-short spreads formed using lagged PERF scores. The long (short) leg is the value-weighted portfolio of stocks in the bottom (top) 30% of PERF scores, which are lagged from 1 month (year 0) up to 120 months (year 10). The PERF percentiles and spreads are constructed separately within three segments based on (lagged) market capitalization: large (left panels), medium(middle panels), and small (right panels), with the NYSE's 70th and 20th percentiles used as breakpoints. Alphas are computed with respect to six pricing models: CAPM, FF3, FF5, Q4, and CAPM+BAB. "BAB" denotes the betting-against-beta factor of Frazzini and Pedersen (2014). The lower subpanels include reference lines at t-statistics of ±1.65 and ±1.96. The sample period is from 1/1968 to 12/2018.

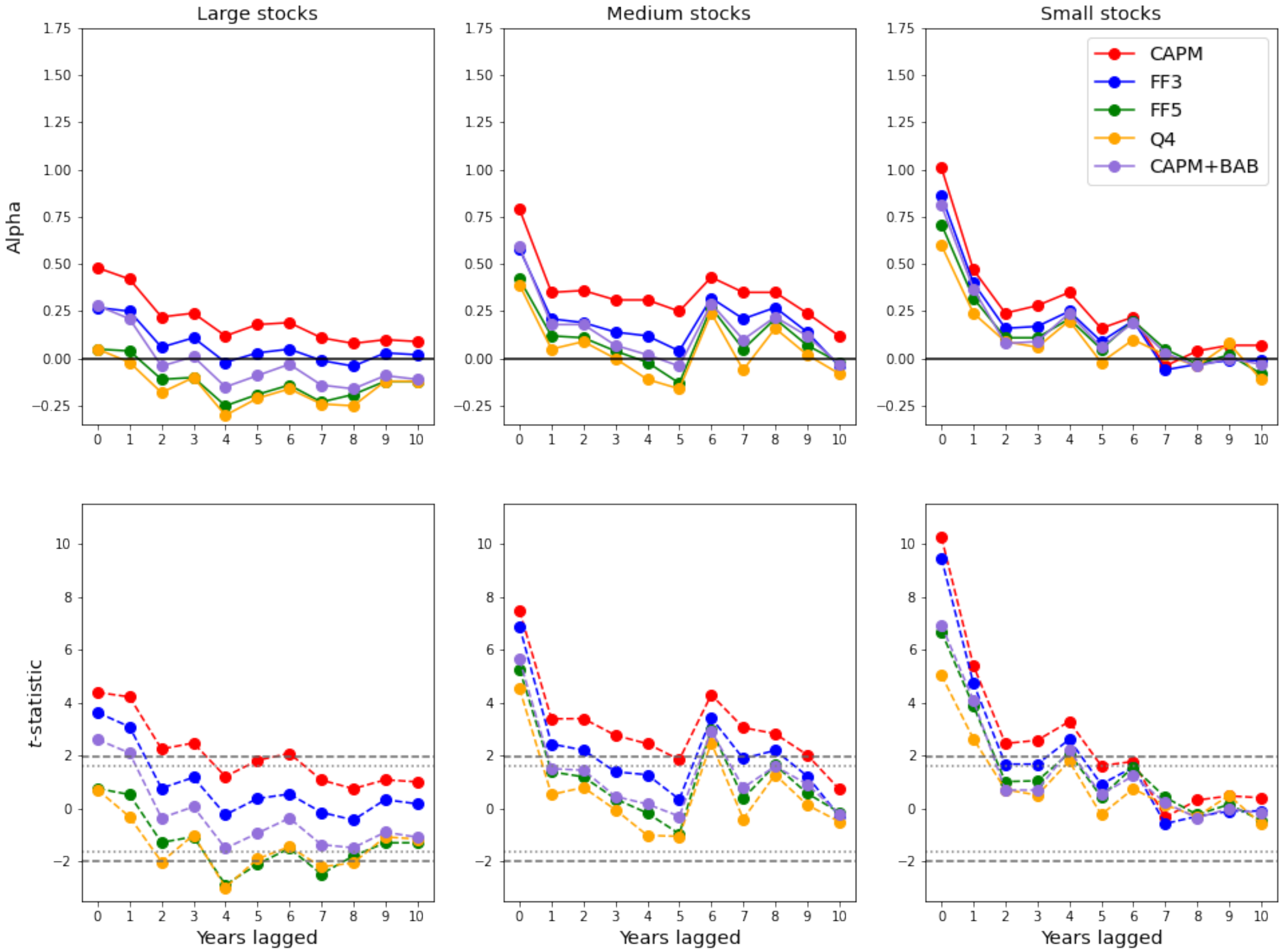

Figure 3. **Spreads formed with MGMT scores.** The figure plots estimated alphas (top panels) and their t-statistics (bottom panels) for long-short spreads formed using lagged MGMT scores. The long (short) leg is the value-weighted portfolio of stocks in the bottom (top) 30% of PERF scores, which are lagged from 1 month (year 0) up to 120 months (year 10). The MGMT percentiles and spreads are constructed separately within three segments based on (lagged) market capitalization: large (left panels), medium (middle panels), and small (right panels), with the NYSE's 70th and 20th percentiles used as breakpoints. Alphas are computed with respect to six pricing models: CAPM, FF3, FF5, Q4, and CAPM+BAB. "BAB" denotes the betting-against-beta factor of Frazzini and Pedersen (2014). The lower subpanels include reference lines at t-statistics of ± 1.65 and ± 1.96. The sample period is from 1/1968 to 12/2018.

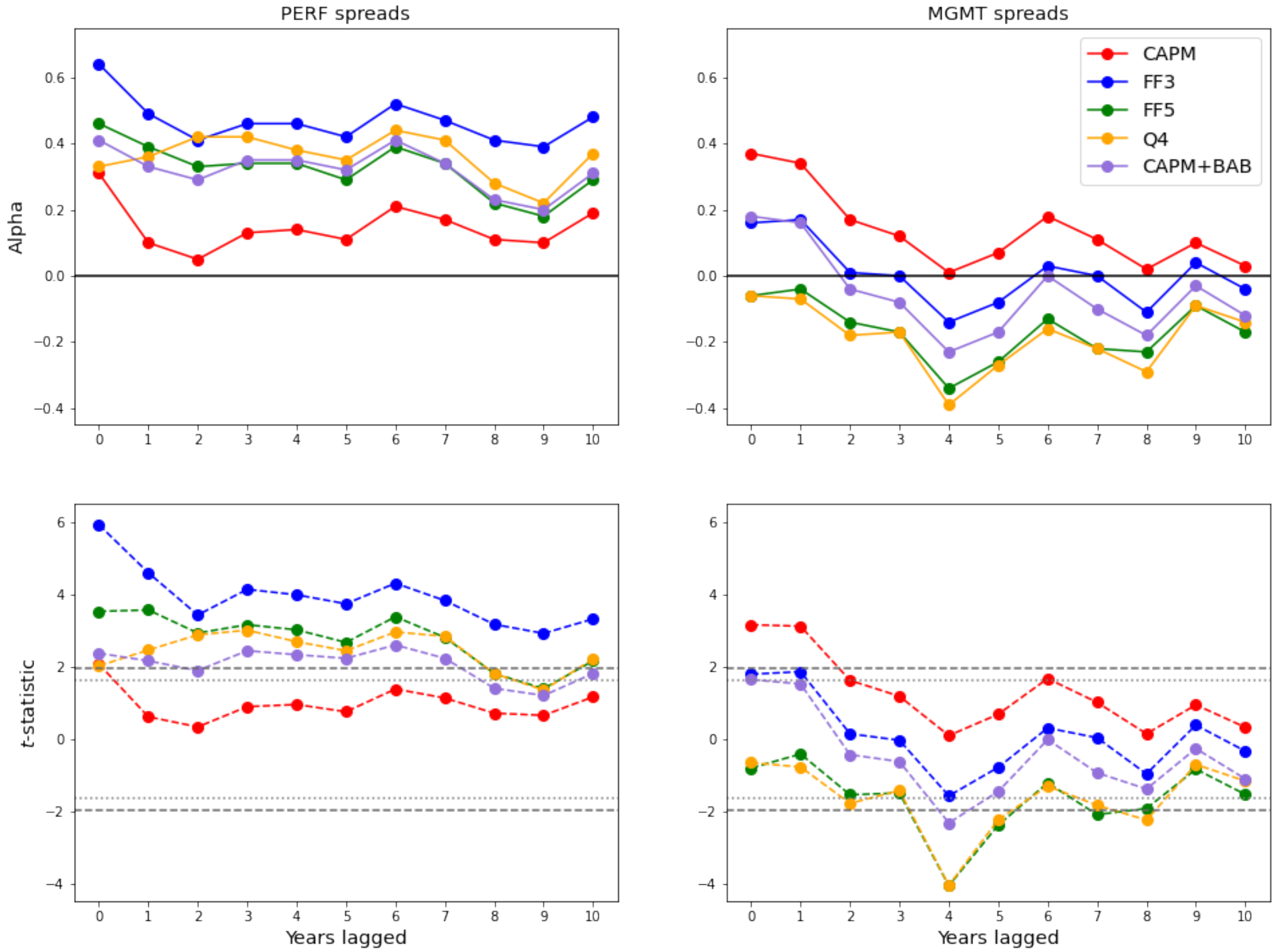

Figure 4. Spreads formed with PERF and MGMT scores among mega-caps. The figure plots estimated alphas (top panels) and their t-statistics (bottom panels) for long-short spreads formed using lagged PERF scores (left panels) and MGMT scores (right panels). The long (short) leg is the value-weighted portfolio of stocks in the bottom (top) 30% of scores, which are lagged from 1 month (year 0) up to 120 months (year 10). The percentiles and spreads are constructed within the sample of mega-cap stocks, consisting of the largest 200 stocks ranked by their most recent market capitalization. Alphas are computed with respect to six pricing models: CAPM, FF3, FF5, Q4, and CAPM+BAB. "BAB" denotes the betting-against-beta factor of Frazzini and Pedersen (2014). The lower subpanels include reference lines at t-statistics of ± 1.65 and ± 1.96. The sample period is from 1/1968 to 12/2018.

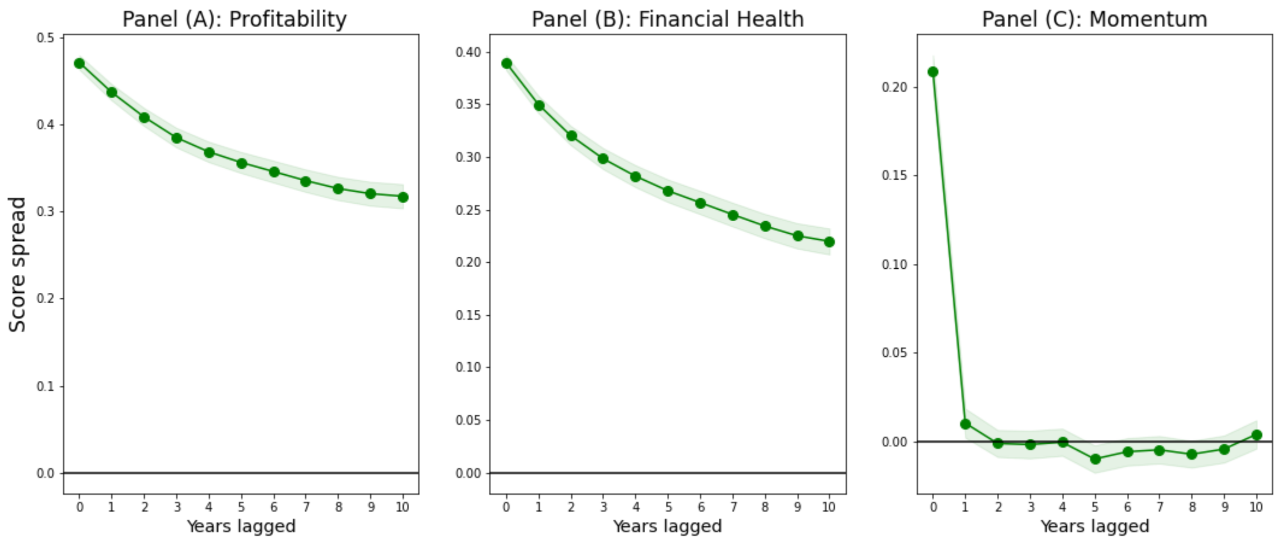

Figure 5. **Spreads in PERF characteristics.** Each panel plots the average difference in a characteristic's current score between the long and short legs of a spreads based on PERF scores lagged from 1 month (year 0) up to 120 months (year 10). The spreads contain large-cap stocks (exceeding the NYSE's 70th percentile), and each spread's long (short) leg contains stocks in the bottom (top) 30% of that sample's PERF scores lagged from 1 month (year 0) up to 120 months (year 10). The characteristics, which constitute the PERF measure, are profitability (left panel), financial health (middle panel), and momentum (right panel). A stock's profitability score is calculated as the simple average of the percentile rankings of two profitability measures, ROE and gross profitability, with high scores indicating high ROE and high gross profitability. A stock's financial health score is calculated as the simple average of the percentile rankings of two distress measures, failure probability and O-score, with high scores indicating low failure probability and low O-scores. A stock's momentum score is the percentile ranking of its previous 12-month return, omitting the most recent month. The shaded areas indicate 95% confidence intervals. The sample period is from 1/1968 to 12/2018.

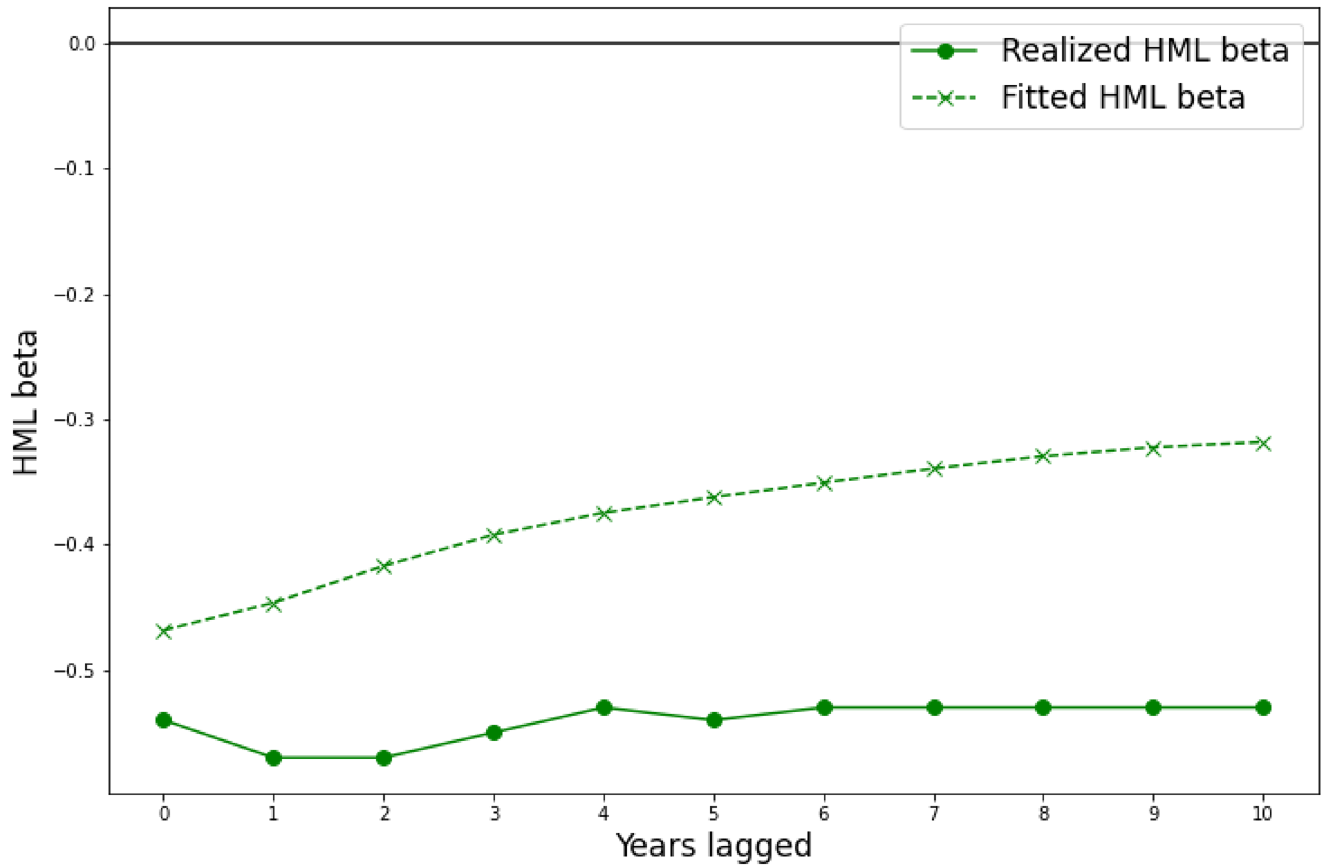

Figure 6. **Fitted and actual HML betas of PERF spreads.** The figure plots the realized (solid line) and fitted (dashed line) HML beta of long-short spreads formed using lagged PERF scores. The spreads contain large-cap stocks (exceeding the NYSE's 70th percentile), and each spread's long (short) leg contains stocks in the bottom (top) 30% of that sample's PERF scores lagged from 1 month (year 0) up to 120 months (year 10). The realized beta is the coefficient in a time-series regression using the most recent 60 months. The fitted beta combines the characteristic spreads in Figure 5 with the regression coefficients in Table 9. The sample period is from 1/1968 to 12/2018.

Table 1
Correlations between current and decade-old factor betas

The table reports rank correlations of stocks' factor betas in the current month with those ten years ago. The factors are those in the model of Fama and French (2015), and factor betas are estimated over the most recent 60 months. The rank correlations are computed separately within three segments based on (lagged) market capitalization—large, medium, and small—with the NYSE's 70th and 20th percentiles used as breakpoints. The rank correlations are averaged across all months from 1/1973 to 12/2018.

Factor	Large	Medium	Small
MKT	0.27	0.23	0.14
SMB	0.33	0.19	0.15
HML	0.17	0.12	0.04
CMA	0.03	-0.02	-0.01
RMW	0.10	0.06	0.05

Table 2
Pricing tests using decade-old information

The table reports estimated monthly alphas (in percent) and t-statistics (in parentheses) for spreads between value-weighted portfolios of stocks in the bottom and top 30% of stocks sorted by PERF scores (Panel A) and MGMT scores (Panel B), both lagged 10 years. Results are shown for the total stock universe as well as mega-cap stocks (largest 200) and three market-cap segments formed using NYSE percentiles as breakpoints: large (above 70th), medium (70th to 20th), and small (below 20th). The models tested are the CAPM of Sharpe (1964) and Lintner (1965), the three- and five-factor models of Fama and French (1993, 2015), denoted FF3 and FF5, the four-factor model of Hou, Xue, and Zhang (2015), denoted Q4, and CAPM+BAB, where "BAB" denotes the betting-against-beta factor of Frazzini and Pedersen (2014). The sample period is 1/1968 to 12/2018.

	Mega-cap	Large	Medium	Small	All stocks
	Panel A: Spreads based on PERF				
CAPM	0.19	0.21	−0.09	−0.14	0.12
	(1.16)	(1.55)	(−0.54)	(−0.87)	(0.99)
FF3	0.48	0.46	0.10	−0.16	0.31
	(3.32)	(4.16)	(0.73)	(−0.96)	(2.86)
FF5	0.29	0.29	0.26	−0.20	0.17
	(2.16)	(2.62)	(2.10)	(−1.10)	(1.62)
Q4	0.37	0.33	0.25	−0.21	0.18
	(2.21)	(2.35)	(1.64)	(−0.99)	(1.54)
CAPM+BAB	0.31	0.38	0.28	−0.10	0.29
	(1.81)	(2.71)	(2.02)	(−0.49)	(2.19)
	Panel B: Spreads based on MGMT				
CAPM	0.03	0.09	0.12	0.07	0.08
	(0.32)	(1.01)	(0.74)	(0.41)	(0.86)
FF3	−0.04	0.02	−0.04	−0.01	−0.01
	(−0.34)	(0.18)	(−0.28)	(−0.09)	(−0.09)
FF5	−0.17	−0.12	−0.02	−0.08	−0.14
	(−1.54)	(−1.29)	(−0.14)	(−0.45)	(−1.43)
Q4	−0.14	−0.12	−0.08	−0.11	−0.14
	(−1.17)	(−1.15)	(−0.49)	(−0.58)	(−1.34)
CAPM+BAB	−0.12	−0.11	−0.03	−0.03	−0.13
	(−1.11)	(−1.08)	(−0.19)	(−0.17)	(−1.33)

Table 3
Multiple comparison tests

The table entries are p-values. The first five columns test joint equality to zero of the alphas in Panels A and B of Table 2 for the corresponding size segment. The last column tests joint equality to zero of the six element vector containing the alphas in Panels A and B of Table 2 for each of the large, medium, and small size segments. The test is that of Gibbons, Ross, and Shanken (1989) but with a heteroskedasticity- and autocorrelation-consistent covariance matrix

	Mega-cap	Large	Medium	Small	All stocks	Large/Medium/Small
CAPM	0.3581	0.0753	0.6050	0.6727	0.2999	0.1892
FF3	0.0006	< 0.0001	0.7237	0.6363	0.0125	0.0012
FF5	0.0285	0.0126	0.1434	0.3957	0.1102	0.0570
Q4	0.0233	0.0197	0.2124	0.3404	0.1309	0.1005
CAPM+BAB	0.0777	0.0070	0.1270	0.8145	0.0239	0.0701

Table 4
Decomposing Fama-French three-factor alphas

The table decomposes alphas with respect to the three-factor model of Fama and French (1993) for spreads formed using PERF scores (Panel A) and MGMT scores (Panel B). The spreads contain large-cap stocks (exceeding the NYSE's 70th percentile), and each spread's long (short) leg contains stocks in the bottom (top) 30% of that sample's scores lagged from 1 month (year 0) up to 120 months (year 10). Summing, within a row, the *ExRet* column (spread's average return) and the subsequent columns of $-\beta\times$factor mean (contributions of factor exposures) gives the α column (spread's monthly percent alpha). The last column reports the alpha's t-statistic. The sample period is from 1/1968 to 12/2018.

Year	*ExRet*	$-\beta_{MKT}\overline{MKT}$	$-\beta_{SMB}\overline{SMB}$	$-\beta_{HML}\overline{HML}$	α	$t(\alpha)$
			Panel A: Spreads based on PERF			
0	0.21	0.14	0.01	0.26	0.62	6.05
1	0.02	0.11	0.01	0.29	0.44	4.81
2	0.01	0.10	0.01	0.27	0.39	3.89
3	0.08	0.08	0.01	0.24	0.41	4.28
4	0.08	0.08	0.01	0.24	0.41	4.24
5	0.03	0.09	0.00	0.23	0.35	3.65
6	0.08	0.09	0.00	0.23	0.40	3.99
7	0.07	0.08	0.01	0.23	0.39	3.54
8	−0.05	0.08	0.01	0.23	0.27	2.61
9	0.02	0.09	0.01	0.20	0.33	2.95
10	0.15	0.10	0.01	0.20	0.46	4.16
			Panel B: Spreads based on MGMT			
0	0.37	0.05	0.02	−0.17	0.27	3.63
1	0.32	0.05	0.02	−0.14	0.25	3.09
2	0.12	0.05	0.03	−0.13	0.06	0.77
3	0.14	0.06	0.02	−0.11	0.11	1.18
4	0.02	0.05	0.02	−0.11	−0.02	−0.23
5	0.09	0.04	0.02	−0.12	0.03	0.38
6	0.11	0.04	0.01	−0.11	0.05	0.55
7	0.04	0.03	0.02	−0.10	−0.01	−0.14
8	0.02	0.01	0.02	−0.10	−0.04	−0.43
9	0.05	0.02	0.02	−0.06	0.03	0.34
10	0.03	0.03	0.02	−0.06	0.02	0.18

Table 5
Decomposing Fama-French five-factor alphas

The table decomposes alphas with respect to the five-factor model of Fama and French (2015) for spreads formed using PERF scores (Panel A) and MGMT scores (Panel B). The spreads contain large-cap stocks (exceeding the NYSE's 70th percentile), and each spread's long (short) leg contains stocks in the bottom (top) 30% of that sample's scores lagged from 1 month (year 0) up to 120 months (year 10). Summing, within a row, the *ExRet* column (spread's average return) and the subsequent columns of $-\beta\times$factor mean (contributions of factor exposures) gives the α column (spread's monthly percent alpha). The last column reports the alpha's t-statistic. The sample period is from 1/1968 to 12/2018.

Year	*ExRet*	$-\beta_{MKT}\overline{MKT}$	$-\beta_{SMB}\overline{SMB}$	$-\beta_{HML}\overline{HML}$	$-\beta_{CMA}\overline{CMA}$	$-\beta_{RMW}\overline{RMW}$	α	$t(\alpha)$
				Panel A: Spreads based on PERF				
0	0.21	0.11	−0.00	0.31	−0.10	−0.10	0.42	3.39
1	0.02	0.10	0.01	0.30	−0.01	−0.07	0.34	3.69
2	0.01	0.09	0.01	0.26	0.01	−0.07	0.31	3.30
3	0.08	0.07	0.01	0.26	−0.05	−0.05	0.32	3.29
4	0.08	0.07	0.01	0.27	−0.07	−0.05	0.31	3.05
5	0.03	0.07	0.00	0.26	−0.06	−0.04	0.26	2.63
6	0.08	0.07	0.01	0.27	−0.08	−0.02	0.32	3.02
7	0.07	0.06	0.01	0.27	−0.09	−0.04	0.28	2.49
8	−0.05	0.06	0.01	0.27	−0.09	−0.05	0.15	1.47
9	0.02	0.06	0.00	0.25	−0.09	−0.08	0.17	1.49
10	0.15	0.08	−0.00	0.25	−0.11	−0.08	0.29	2.62
				Panel B: Spreads based on MGMT				
0	0.37	0.01	0.02	−0.07	−0.23	−0.05	0.05	0.76
1	0.32	0.01	0.02	−0.04	−0.21	−0.06	0.04	0.53
2	0.12	0.02	0.02	−0.07	−0.15	−0.06	−0.11	−1.28
3	0.14	0.03	0.01	−0.03	−0.17	−0.07	−0.10	−1.06
4	0.02	0.02	0.01	−0.04	−0.16	−0.09	−0.25	−2.88
5	0.09	0.01	0.01	−0.05	−0.16	−0.08	−0.19	−2.10
6	0.11	0.01	−0.00	−0.05	−0.14	−0.07	−0.14	−1.50
7	0.04	−0.00	0.00	−0.02	−0.17	−0.08	−0.23	−2.47
8	0.02	−0.01	0.01	−0.06	−0.10	−0.06	−0.19	−1.79
9	0.05	0.00	0.01	−0.02	−0.10	−0.07	−0.12	−1.29
10	0.03	0.01	0.01	−0.03	−0.09	−0.06	−0.12	−1.29

Table 6
Decomposing Hou-Xue-Zhang Q4 alphas

The table decomposes alphas with respect to the four-factor "Q4" model of Hou, Xue, and Zhang (2015) for spreads formed using PERF scores (Panel A) and MGMT scores (Panel B). The spreads contain large-cap stocks (exceeding the NYSE's 70th percentile), and each spread's long (short) leg contains stocks in the bottom (top) 30% of that sample's scores lagged from 1 month (year 0) up to 120 months (year 10). Summing, within a row, the *ExRet* column (spread's average return) and the subsequent columns of $-\beta \times$ factor mean (contributions of factor exposures) gives the α column (spread's monthly percent alpha). The last column reports the alpha's t-statistic. The sample period is from 1/1968 to 12/2018.

Year	*ExRet*	$-\beta_{MKT}\overline{MKT}$	$-\beta_{SMB}\overline{SMB}$	$-\beta_{ROE}\overline{ROE}$	$-\beta_{IA}\overline{IA}$	α	$t(\alpha)$
			Panel A: Spreads based on PERF				
0	0.21	0.09	−0.01	−0.31	0.26	0.26	1.61
1	0.02	0.08	0.03	−0.19	0.36	0.30	2.33
2	0.01	0.08	0.04	−0.10	0.37	0.40	3.26
3	0.08	0.07	0.03	−0.07	0.29	0.40	3.28
4	0.08	0.06	0.02	−0.09	0.27	0.35	2.62
5	0.03	0.07	0.02	−0.07	0.27	0.32	2.48
6	0.08	0.07	0.02	−0.06	0.26	0.37	2.83
7	0.07	0.06	0.03	−0.07	0.26	0.34	2.46
8	−0.05	0.06	0.04	−0.09	0.25	0.20	1.54
9	0.02	0.07	0.02	−0.11	0.22	0.22	1.60
10	0.15	0.08	0.02	−0.11	0.20	0.33	2.35
			Panel B: Spreads based on MGMT				
0	0.37	0.03	0.03	0.00	−0.37	0.05	0.71
1	0.32	0.02	0.02	−0.06	−0.33	−0.02	−0.31
2	0.12	0.02	0.03	−0.07	−0.29	−0.18	−2.01
3	0.14	0.04	0.02	−0.05	−0.25	−0.10	−1.00
4	0.02	0.03	0.01	−0.11	−0.26	−0.30	−2.98
5	0.09	0.02	0.01	−0.08	−0.26	−0.21	−1.90
6	0.11	0.02	−0.00	−0.06	−0.23	−0.16	−1.44
7	0.04	0.01	0.01	−0.07	−0.24	−0.24	−2.20
8	0.02	0.00	0.01	−0.08	−0.20	−0.25	−2.05
9	0.05	0.01	0.02	−0.07	−0.13	−0.12	−1.07
10	0.03	0.02	0.02	−0.06	−0.13	−0.12	−1.15

Table 7
Decomposing alphas with respect to Q4 plus HML

The table decomposes alphas with respect to the four-factor "Q4" model of Hou, Xue, and Zhang (2015), augmented by the HML factor of Fama and French (2015), for spreads formed using PERF scores (Panel A) and MGMT scores (Panel B). The spreads contain large-cap stocks (exceeding the NYSE's 70th percentile), and each spread's long (short) leg contains stocks in the bottom (top) 30% of that sample's scores lagged from 1 month (year 0) up to 120 months (year 10). Summing, within a row, the *ExRet* column (spread's average return) and the subsequent columns of $-\beta \times$ factor mean (contributions of factor exposures) gives the α column (spread's monthly percent alpha). The last column reports the alpha's t-statistic. The sample period is from 1/1968 to 12/2018.

Year	*ExRet*	$-\beta_{MKT}\overline{MKT}$	$-\beta_{SMB}\overline{SMB}$	$-\beta_{ROE}\overline{ROE}$	$-\beta_{IA}\overline{IA}$	$-\beta_{HML}\overline{HML}$	α	$t(\alpha)$
				Panel A: Spreads based on PERF				
0	0.21	0.11	−0.01	−0.24	−0.02	0.24	0.29	2.27
1	0.02	0.09	0.02	−0.11	0.06	0.25	0.34	3.59
2	0.01	0.10	0.04	−0.04	0.11	0.21	0.43	4.50
3	0.08	0.08	0.03	−0.00	0.03	0.22	0.44	4.64
4	0.08	0.07	0.02	−0.02	−0.01	0.24	0.38	3.72
5	0.03	0.08	0.02	0.00	−0.00	0.23	0.35	3.60
6	0.08	0.08	0.02	0.01	−0.01	0.23	0.41	3.87
7	0.07	0.07	0.03	−0.01	−0.01	0.23	0.38	3.29
8	−0.05	0.07	0.03	−0.03	−0.01	0.22	0.23	2.10
9	0.02	0.08	0.02	−0.05	−0.02	0.20	0.25	2.11
10	0.15	0.09	0.01	−0.05	−0.05	0.21	0.36	3.16
				Panel B: Spreads based on MGMT				
0	0.37	0.03	0.03	−0.03	−0.26	−0.09	0.04	0.57
1	0.32	0.02	0.02	−0.07	−0.25	−0.06	−0.03	−0.45
2	0.12	0.02	0.03	−0.09	−0.19	−0.08	−0.20	−2.35
3	0.14	0.04	0.02	−0.07	−0.18	−0.06	−0.11	−1.11
4	0.02	0.02	0.02	−0.13	−0.16	−0.08	−0.31	−3.37
5	0.09	0.02	0.01	−0.10	−0.16	−0.08	−0.22	−2.14
6	0.11	0.02	0.00	−0.08	−0.13	−0.08	−0.17	−1.65
7	0.04	0.01	0.01	−0.08	−0.17	−0.06	−0.25	−2.34
8	0.02	−0.00	0.01	−0.11	−0.10	−0.09	−0.26	−2.34
9	0.05	0.01	0.02	−0.09	−0.06	−0.06	−0.13	−1.17
10	0.03	0.02	0.02	−0.08	−0.05	−0.06	−0.13	−1.29

Table 8
Decomposing alphas with respect to the CAPM plus BAB

The table decomposes alphas with respect to the CAPM, augmented by the betting-against-beta (BAB) factor of Frazzini and Pedersen (2014), for spreads formed using PERF scores (Panel A) and MGMT scores (Panel B). The spreads contain large-cap stocks (exceeding the NYSE's 70th percentile), and each spread's long (short) leg contains stocks in the bottom (top) 30% of that sample's scores lagged from 1 month (year 0) up to 120 months (year 10). Summing, within a row, the *ExRet* column (spread's average return) and the subsequent columns of $-\beta \times$ factor mean (contributions of factor exposures) gives the α column (spread's monthly percent alpha). The last column reports the alpha's t-statistic. The sample period is from 1/1968 to 12/2018.

Year	$ExRet$	$-\beta_{MKT}\overline{MKT}$	$-\beta_{BAB}\overline{BAB}$	α	$t(\alpha)$
		Panel A: Spreads based on PERF			
0	0.21	0.08	0.08	0.38	2.24
1	0.02	0.05	0.22	0.29	2.12
2	0.01	0.05	0.24	0.29	2.16
3	0.08	0.04	0.24	0.36	2.84
4	0.08	0.04	0.23	0.35	2.57
5	0.03	0.04	0.24	0.31	2.44
6	0.08	0.04	0.24	0.36	2.71
7	0.07	0.04	0.22	0.33	2.25
8	−0.05	0.04	0.18	0.17	1.19
9	0.02	0.05	0.14	0.21	1.48
10	0.15	0.06	0.16	0.38	2.71
		Panel B: Spreads based on MGMT			
0	0.37	0.11	−0.20	0.28	2.61
1	0.32	0.09	−0.20	0.21	2.10
2	0.12	0.09	−0.24	−0.04	−0.35
3	0.14	0.09	−0.22	0.01	0.07
4	0.02	0.09	−0.26	−0.15	−1.50
5	0.09	0.08	−0.26	−0.09	−0.93
6	0.11	0.07	−0.21	−0.03	−0.36
7	0.04	0.06	−0.23	−0.14	−1.36
8	0.02	0.05	−0.23	−0.16	−1.48
9	0.05	0.05	−0.19	−0.09	−0.89
10	0.03	0.06	−0.19	−0.11	−1.08

Table 9
Large stocks' HML betas regressed on their PERF characteristics

The table reports the coefficient estimates from a monthly Fama-Macbeth regression of large stocks' HML betas on their scores for the characteristics that constitute PERF: profitability, financial health, and momentum. A stock's profitability score is calculated as the simple average of its percentile rankings for two profitability measures, ROE and gross profitability (high scores indicating high ROE and high gross profitability). The financial-health score is the simple average of the percentile rankings of two distress measures, failure probability and O-scores (high scores indicating low failure probability and low O-scores). The momentum score is the percentile ranking of past 12-month returns, omitting the most recent month. HML betas are estimated based on the five-factor model of Fama and French (2015), using the prior 60 months of monthly data. The sample period is from 1/1968 to 12/2018.

Characteristic	Coefficient
Profitability	−0.0087
	(−8.87)
Financial health	−0.0019
	(−1.81)
Momentum	0.0008
	(0.70)

Table 10
Large stocks' book-to-market ratios regressed on decade-old PERF characteristics and book-to-market ratios

The table reports the coefficient estimates and t-statistics (in parentheses) from a monthly Fama-Macbeth regression of large stocks' book-to-market (BM) ratios on values 120 months earlier of BM ratios and scores for the characteristics that constitute PERF: profitability, financial health, and momentum. A stock's profitability score is the simple average of its percentile rankings for two profitability measures, ROE and gross profitability (high scores indicating high ROE and high gross profitability). The financial-health score is the simple average of the percentile rankings of two distress measures, failure probability and O-scores (high scores indicating low failure probability and low O-scores). The momentum score is the percentile ranking of past 12-month returns, omitting the most recent month. The sample period is from 1/1968 to 12/2018.

Characteristic	Coefficient
Profitability	−0.0082
	(−3.51)
Financial health	0.0017
	(2.06)
Momentum	0.0003
	(0.35)
Book-to-market	0.1539
	(1.22)

Table 11
Effects on alpha of persistence in mispricing scores

The table reports the CAPM alphas for spreads based on PERF and MGMT scores, conditional on the degree to which the stocks' current scores have persisted. The universe consists of large stocks (above the NYSE's 70th percentile) and is first divided into quintiles based on the most recent PERF scores or MGMT scores. Within the top and bottom quintiles of those mispricing scores, we rank stocks based on $Persistence_{i,t}$, which measures the degree to which a stock's score in past years has been similar to the current level. Specifically, with the past window truncated at 10 years,

$$Persistence_{i,t} = [I(Q_{i,t} = 5) - I(Q_{i,t} = 1)] \times \sum_{\tau=13}^{120} \omega_\tau [M_{i,t-\tau} - 0.5],$$

where $Q_{i,t}$ is stock i's quintile of PERF or MGMT score in month t, $I(Q_{i,t} = k)$ equals 1 if $Q_{i,t} = k$ and 0 otherwise, $\omega_\tau = [10 - int((\tau - 1)/12)]/9$, and $M_{i,t-\tau}$ is the rank percentile of stock i based on its PERF or MGMT score in month $t - \tau$. For different quintiles of $Persistence_{i,t}$, the table reports the monthly percent CAPM alpha for PERF and MGMT spreads (20% most underpriced minus 20% most overpriced), with t-statistics in parentheses. The sample period is from 1/1968 to 12/2018.

	Among stocks whose current scores have persisted					Least minus	All	Least minus
	Least	Next 20%	Next 20%	Next 20%	Most	most	stocks	all
PERF spreads	0.89	0.65	0.44	0.19	0.24	0.66	0.40	0.50
	(3.57)	(2.81)	(2.03)	(0.94)	(1.02)	(2.29)	(2.22)	(2.69)
MGMT spreads	0.50	0.85	0.86	0.64	0.62	−0.11	0.64	−0.14
	(3.19)	(4.82)	(4.89)	(3.84)	(3.36)	(−0.49)	(5.75)	(−0.91)

References

Barberis, N., Shleifer, A., Vishny, R., 1998. A model of investor sentiment. Journal of Financial Economics 49, 307–343.

Belo, F., 2010. Production-based measures of risk for asset pricing. Journal of Monetary Economics 57, 146–163.

Berk, J. B., Green, R. C., Naik, V., 1999. Optimal investment, growth options, and security returns. Journal of finance 54, 1553–1607.

Black, F., 1972. Capital market equilibrium with restricted borrowing. Journal of Business 45, 444–455.

Black, F., Jensen, M. C., Scholes, M. S., 1972. The capital asset pricing model: Some empirical tests. In: Jensen, M. C. (ed.), *Studies in the Theory of Capital Markets*, Praeger, New York, pp. 79–121.

Bouchard, J.-P., Krüger, P., Landier, A., Thesmar, D., 2019. Sticky expectations and the profitability anomaly. Journal of Finance 74, 639–674.

Chen, X., He, W., Tao, L., Yu, J., 2020. Media coverage and underreaction-related anomalies. Working paper, Southwestern University of Finance and Economics, University of International Business and Economics, and Tsinghua University .

Cho, T., Polk, C., 2020. Asset pricing with price levels. Working paper, London School of Economics .

Da, Z., Gurun, U. G., Warachka, M., 2014. Frog in the pan: Continuous information and momentum. Review of Financial Studies 27, 2171–2218.

Daniel, K., Hirshleifer, D., Subrahmanyam, A., 1998. Investor psychology and security market under-and overreactions. Journal of Finance 53, 1839–1885.

Daniel, K., Titman, S., 1997. Evidence on the characteristics of cross sectional variation in stock returns. Journal of Finance 52, 1–33.

Dichev, I. D., 1998. Is the risk of bankruptcy a systematic risk? Journal of Finance 53, 1131–1147.

Du, W., Tepper, A., Verdelhan, A., 2018. Deviations from covered interest rate parity. Journal of Finance 73, 915–957.

Fama, E. F., 1970. Efficient capital markets: A review of theory and empirical work. Journal of Finance 25, 383–417.

Fama, E. F., 1976. Foundations of Finance. Basic Books, New York.

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

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

Fama, E. F., French, K. R., 1995. Size and book-to-market factors in earnings and returns. Journal of Finance 50, 131–155.

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

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

Fama, E. F., MacBeth, J. D., 1973. Risk, return, and equilibrium: Empirical tests. Journal of Political Economy 81, 607–636.

Frazzini, A., Israel, R., Moskowitz, T. J., 2018. Trading costs. Working paper, AQR Capital Management and Yale University .

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

Gibbons, M. R., Ross, S. A., Shanken, J., 1989. A test of the efficiency of a given portfolio. Econometrica pp. 1121–1152.

Gomes, J., Kogan, L., Zhang, L., 2003. Equilibrium cross section of returns. Journal of Political Economy 111, 693–732.

Harvey, C. R., Liu, Y., 2019. A census of the factor zoo. Working paper, Duke University and Purdue University .

Hong, H., Stein, J. C., 1999. A unified theory of underreaction, momentum trading, and overreaction in asset markets. Journal of Finance 54, 2143–2184.

Hou, K., Xue, C., Zhang, L., 2015. Digesting anomalies: an investment approach. Review of Financial Studies 28, 650–705.

Hu, G. X., Pan, J., Wang, J., 2013. Noise as information for illiquidity. Journal of Finance 68, 2341–2382.

Johnson, T. C., 2002. Rational momentum effects. Journal of Finance 57, 585–608.

Keloharju, M., Linnainmaa, J. T., Nyberg, P., 2020. Long-term discount rates do not vary across firms. Journal of Financial Economics, forthcoming.

Lakonishok, J., Shleifer, A., Vishny, R. W., 1994. Contrarian investment, extrapolation, and risk. Journal of Finance 49, 1541–1578.

Lamont, O. A., Thaler, R. H., 2003. Can the market add and subtract? Mispricing in tech stock carve-outs. Journal of Political Economy 111, 227–268.

Li, D., Zhang, L., 2010. Does q-theory with investment frictions explain anomalies in the cross section of returns? Journal of Financial Economics 98, 297–314.

Li, E. X., Livdan, D., Zhang, L., 2009. Anomalies. Review of Financial Studies 22, 4301–4334.

Lintner, J., 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. Review of Economics and Statistics 47, 13–37.

Liu, L. X., Whited, T. M., Zhang, L., 2009. Investment-based expected stock returns. Journal of Political Economy 117, 1105–1139.

Liu, L. X., Zhang, L., 2008. Momentum profits, factor pricing, and macroeconomic risk. Review of Financial Studies 21, 2417–2448.

Novy-Marx, R., 2013. The other side of value: The gross profitability premium. Journal of Financial Economics 108, 1–28.

Sagi, J. S., Seasholes, M. S., 2007. Firm-specific attributes and the cross-section of momentum. Journal of Financial Economics 84, 389–434.

Sharpe, W. F., 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. Journal of Finance 19, 425–442.

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

Stambaugh, R. F., Yu, J., Yuan, Y., 2014. The long of it: Odds that investor sentiment spuriously predicts anomaly returns. Journal of Financial Economics 114, 613–619.

Stambaugh, R. F., Yu, J., Yuan, Y., 2015. Arbitrage asymmetry and the idiosyncratic volatility puzzle. Journal of Finance 70, 1903–1948.

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

Yara, F. B., Boons, M., Tamoni, A., 2020. New and old sorts: Implications for asset pricing. Working paper, Nova SBE, Tilburg University, Rutgers Business School .

Zhang, L., 2005. The value premium. Journal of Finance 60, 67–103.