Mispricing Factors

Robert F. Stambaugh

The Wharton School, University of Pennsylvania and NBER

Yu Yuan

Shanghai Advanced Institute of Finance, Shanghai Jiao Tong University and Wharton Financial Insitutions Center

A four-factor model with two "mispricing" factors, in addition to market and size factors, accommodates a large set of anomalies better than notable four- and five-factor alternative models. Moreover, our size factor reveals a small-firm premium nearly twice usual estimates. The mispricing factors aggregate information across 11 prominent anomalies by averaging rankings within two clusters exhibiting the greatest return co-movement. Investor sentiment predicts the mispricing factors, especially their short legs, consistent with a mispricing interpretation and the asymmetry in ease of buying versus shorting. A three-factor model with a single mispricing factor also performs well, especially in Bayesian model comparisons.

Modern finance has long valued models relating expected returns to factor sensitivities. A virtue of such models is parsimony. Once factors are constructed, the only additional data required to compute implied expected returns in standard applications are the historical returns on the assets being analyzed. Moreover, the number of factors has typically been small. For many years only a single market factor was popular, following the CAPM of Sharpe (1964) and Lintner (1965). Fama and French (1993) spurred widespread use of three factors, motivated by violations of the single-factor CAPM related to firm size and value-versus-growth measures.

Numerous studies have identified anomalies that violate the three-factor model, but only occasionally have anomalies contended for status as additional factors, given the virtue of parsimony in a factor model.[1] Given the proliferation of anomalies, however, the need for an alternative factor model that can accommodate more anomalies has become increasingly clear. Two additional factors have recently received significant attention. Hou, Xue, and Zhang (2015a) propose a four-factor model that combines market and size factors with two new factors based on investment and profitability. Fama and French (2015) add somewhat different versions of investment and profitability factors to their earlier three-factor model (Fama and French (1993)), creating a five-factor model. Both studies provide theoretical motivations for why these factors contain information about expected return: Hou, Xue, and Zhang (2015a) rely on an investment-based pricing model, while Fama and French (2015) invoke comparative statics of a present-value relation. At the same time, it should be noted that both investment and profitability are two of the numerous anomalies documented earlier in the literature.[2] In subsequent studies, Fama and French (2016) and Hou, Xue, and Zhang (2015b) examine their models' abilities to explain other anomalies.

We take a different approach to factor construction. Instead of having a factor correspond to a single anomaly, we combine the information in multiple anomalies. Clearly an important dimension on which a parsimonious factor model is judged is its ability to accommodate a wide range of anomalies. Our approach exploits that range when forming the factors. Rather than construct a factor using stocks' rankings on a single anomaly variable, such as investment, we construct a factor by averaging rankings across multiple anomalies. By averaging, we aim to achieve a less noisy measure of a stock's mispricing, thereby identifying

---

[1] A notable example subsequent to Fama and French (1993) is the momentum anomaly documented by Jegadeesh and Titman (1993), which motivates the frequently used momentum factor proposed by Carhart (1997).

[2] Titman, Wei, and Xie (2004) and Xing (2008) show that high investment predicts abnormally low returns, while Fama and French (2006), Chen, Novy-Marx, and Zhang (2010), and Novy-Marx (2013) show that high profitability predicts abnormally high returns.

more precisely which stocks to long and which stocks to short when constructing a factor that can better accommodate anomalies reflecting mispricing.

We apply our approach by constructing two factors from the set of 11 prominent anomalies examined by Stambaugh, Yu, and Yuan (2012, 2014, 2015). In constructing the factors, we average rankings within two clusters of anomalies formed by grouping together the anomalies exhibiting the greatest similarity. We can measure similarity either by time-series correlations of anomalies' long-short return spreads or by cross-sectional correlations of stocks' rankings on the anomaly variables. Both measures yield the same two clusters of anomalies. The two mispricing factors are then combined with market and size factors to obtain a four-factor model.

Our model's overall ability to accommodate anomalies exceeds that of both the four-factor model of Hou, Xue, and Zhang (2015a) and the five-factor model of Fama and French (2015). This conclusion obtains not only within the set of anomalies used to construct the factors but also for the substantially larger set of 73 anomalies examined previously by Hou, Xue, and Zhang (2015a, 2015b). For example, when applied to the 51 of those anomalies having data over our entire sample period, the Gibbons-Ross-Shanken (1989) test of whether all the anomalies' alphas equal zero produces a p-value of 0.10 for our model compared to 0.003 or less for these four- and five-factor alternative models. Our model also performs better than these alternatives when the models are judged by their abilities to explain each other's factors. As discussed by Barillas and Shanken (2015a, 2015b), judging factor models this way is implied by standard model-comparison procedures, under both frequentist and Bayesian approaches. We apply both approaches in our comparisons.

We also construct a three-factor model by replacing our two mispricing factors with a single factor that averages rankings across the entire set of 11 anomalies, rather than within two clusters in that set. When models are again judged by their abilities to explain each other's factors, this three-factor model outperforms the four-factor model of Hou, Xue, and Zhang (2015a) and the five-factor model of Fama and French (2015). It also outperforms the latter model in explaining anomalies.

Our size factor is constructed using stocks least likely to be mispriced, as identified by the measures used to construct our mispricing factors. Our resulting SMB delivers a small-firm premium of 46 bps per month over our 1967–2013 sample period, nearly twice the premium of 25 bps implied by the familiar SMB factor in the Fama-French three-factor model. Consistent with mispricing exerting less effect on our size factor, the investor sentiment index of Baker and Wurgler (2006) exhibits significant ability to predict the Fama-

2

French SMB but not our SMB.

The basic concepts motivating our approach are that anomalies in part reflect mispricing and that mispricing has common components across stocks. Both concepts are consistent with previous evidence. As we discuss, a large empirical literature links anomalies to mispricing, and numerous studies find pervasive effects often characterized as investor sentiment. By combining information across anomalies, we aim to construct factors capturing common elements of mispricing. Consistent with this intent, we find that investor sentiment predicts our mispricing factors, especially their short legs. The stronger predictability of the short legs is consistent with asymmetry in the ease of buying versus shorting (e.g., Stambaugh, Yu, and Yuan (2012)).

Factor models can be useful whether expected returns reflect risk or mispricing. Factors can capture systematic risks for which investors require compensation, or they can capture common sources of mispricing, such as market-wide investor sentiment. This point is emphasized, for example, by Hirshleifer and Jiang (2010) and Kozak, Nagel, and Santosh (2015). Moreover, there need not be a clean distinction between mispricing and risk compensation as alternative motivations for factor models of expected return. For example, DeLong, Shleifer, Summers, and Waldman (1990) explain how fluctuations in market-wide "noise-trader" sentiment create an additional source of systematic risk for which rational traders require compensation.

When expected returns reflect mispricing and not just compensation for systematic risks, some of the mispricing may not be driven by pervasive sentiment factors but may instead be asset specific, as discussed for example by Daniel and Titman (1997). In that sense the concept of "mispricing factors" potentially embeds some inconsistency. On the other hand, previous studies discussed below do find that mispricing appears to exhibit commonality across stocks. The extent to which our factors help describe expected returns is an empirical question. A parsimonious factor model that outperforms feasible alternatives seems useful from a practical perspective, as no model can be entirely correct.

One practical use of factor models, in addition to explaining expected returns, is to capture systematic time-series variation in realized returns. We also examine the extent to which our mispricing factors can perform this role as compared to the factors in the alternative models we consider. Our results indicate that the ability of mispricing factors to explain expected returns better (i.e., to accommodate amomalies better) does not come at the cost of sacrificing ability to capture return variance.

The remainder of the paper proceeds as follows. Section 1 briefly discusses the considerable evidence linking anomalies to mispricing and to pervasive sentiment effects. Section 2 explains the construction of our two mispricing factors and our size factor and examines their empirical properties. The resulting four-factor model is compared to notable alternative factor models in Section 3. Section 4 considers a model with just a single mispricing factor. Section 5 illustrates a shared limitation of the factor models, showing how they can seem to explain the idiosyncratic volatility puzzle if the role of mispricing is not considered. Section 6 reviews our conclusions.

1. Anomalies, Mispricing, and Sentiment

Much of the return-anomaly literature, too extensive for us to survey comprehensively, points to mispricing as being at least partially responsible for the documented anomalous returns. We base our mispricing factors on a prominent subset of the many anomalies reported in the literature, and, within this subset, studies containing mispricing interpretations include Ritter (1991) for net stock issues, Daniel and Titman (2006) for composite equity issues, Sloan (1996) for accruals, Hirshleifer, Hou, Teoh, and Zhang (2004) for net operating assets, Cooper, Gulen, and Schill (2008) for asset growth, Titman, Wei, and Xie (2004) for investment-to-assets, Campbell, Hilscher, and Szilagyi (2008) for financial distress, Jegadeesh and Titman (1993) for momentum, and Wang and Yu (2013) for profitability anomalies including return on assets and gross profitability. A mispricing interpretation of anomalies is also consistent with the evidence of McLean and Pontiff (2015), who observe that following an anomaly's academic publication, there is greater trading activity in the anomaly portfolios, and anomaly profits decline.

Idiosyncratic volatility (IVOL) represents risk that deters price-correcting arbitrage. This concept is advanced, for example, by DeLong, Shleifer, Summers, and Waldman (1990), Pontiff (1996), Shleifer and Vishny (1997), and Stambaugh, Yu, and Yuan (2015). One should therefore expect stronger anomaly returns among stocks with higher IVOL. Studies finding that various return anomalies are indeed stronger among high-IVOL stocks include Pontiff (1996) for closed-end fund discounts, Wurgler and Zhuravskaya (2002) for index inclusions, Mendenhall (2004) for post-earnings announcement drift, Ali, Hwang, and Trombley (2003) for the value premium, Zhang (2006) for momentum, Mashruwala, Rajgopal, and Shevlin (2006) for accruals, Scruggs (2007) for "Siamese twin" stocks, Ben-David and Roulstone (2010) for insider trades and share repurchases, McLean (2010) for long-term reversal, Li

4

and Zhang (2010) for asset growth and investment to assets, Larrain and Varas (2013) for equity issuance, and Wang and Yu (2013) for return on assets.

As explained by Stambaugh, Yu, and Yuan (2015), if there is less arbitrage capital available to short overpriced stocks than to purchase underpriced stocks, then the effect of IVOL should be larger among overpriced stocks. Jin (2013) examines ten anomaly long-short spreads and finds all to be more profitable among high-IVOL stocks than among low-IVOL stocks, and this difference is attributable primarily to the short leg of each spread. Stambaugh, Yu, and Yuan (2015) find, consistent with arbitrage risk and mispricing, that the IVOL-return relation is negative among overpriced stocks but positive among underpriced stocks, with mispricing determined by combining the same 11 return anomalies used in this study. Moreover, those authors find that the negative IVOL-return relation among overpriced stocks is stronger than the positive relation among underpriced stocks, consistent with the arbitrage asymmetry in buying versus shorting.

When mispricing is present, stocks that are more difficult to short should also be those for which overpricing is less easily corrected. Evidence that short-leg profits of anomaly long-short spreads are indeed greater among stocks with greater shorting impediments is provided by Nagel (2005), Hirshleifer, Teoh, and Yu (2011), Avramov, Chordia, Jostova, and Philipov (2013), Drechsler and Drechsler (2014), and Stambaugh, Yu, and Yuan (2015). The last study also shows that the negative IVOL-return relation among overpriced stocks is stronger among stocks less easily shorted.

Evidence consistent with a common sentiment-related component of mispricing is provided, for example, by Baker and Wurgler (2006) and Stambaugh, Yu, and Yuan (2012).[3] The latter study finds that the short-leg returns for long-short spreads associated with each of 11 anomalies we use in this study are significantly lower following a high level of investor sentiment as measured by the Baker-Wurgler sentiment index. Stambaugh, Yu, and Yuan (2015) find that the negative (positive) IVOL-return relation among overpriced (underpriced) stocks is stronger following a high (low) level of the Baker-Wurgler index, consistent with arbitrage risk deterring the correction of sentiment-related mispricing.

This study's objective is not to make a case for the presence of mispricing in the stock market. For that we rely on the previous literature discussed above. We do, however, provide two novel results with regard to the role of investor sentiment, as will be discussed later. First, investor sentiment predicts our mispricing factors, particularly their short (overpriced)

---

[3]Baker, Wurgler, and Yuan (2012) find that sentiment-related effects similar to those documented in the U.S. by Baker and Wurgler (2006) also occur in a number of other countries.

legs. Second, unlike the size factor constructed by Fama and French (1993), our size factor—constructed to be less contaminated by mispricing—is not predicted by sentiment.

2. Anomalies and Factors

Our objective is to explore parsimonious factor models that include factors combining information from a range of anomalies. We first construct a four-factor model that includes two mispricing factors along with market and size factors. Later we consider a three-factor model with just a single mispricing factor.

The first factor in our four-factor model is the excess value-weighted market return, standard in essentially all factor models with pre-specified factors. Constructing the remaining three factors—a size factor and two mispricing factors—involves averaging stocks' rankings with respect to various anomalies. We use the same 11 anomalies analyzed by Stambaugh, Yu, and Yuan (2012, 2014, 2015). While the number of anomalies used to construct the factors could be expanded, we use this previously specified set to alleviate concerns that a different set was chosen to yield especially favorable results for this study. The Appendix provides brief descriptions of the 11 anomalies: net stock issues, composite equity issues, accruals, net operating assets, asset growth, investment-to-assets, distress, O-score, momentum, gross profitability, and return on assets.

Rather than constructing a five-factor model by adding our two mispricing factors to the three factors of Fama and French (1993), we opt for only four factors. That is, we do not include a book-to-market factor and instead include only a size factor in addition to the market and our mispricing factors. Our motivation here is parsimony and long-standing evidence that firm size is related not only to average return but also to a number of other stock characteristics such as volatility, liquidity, and sensitivities to macroeconomic conditions.[4] Not including a factor based on book-to-market reflects the literature's less settled view of that variable's role and importance. As we report later, our mispricing factors price the book-to-market factor, suggesting our decision to exclude a book-to-market factor is reasonable.

---

[4]For example, see Banz (1981) on average return, Amihud and Mendelson (1989) on volatility and liquidity, and Chan, Chen, and Hsieh (1985) on sensitivities to macroeconomic conditions.

2.1. The Mispricing Factors

We construct factors based on averages of stocks' anomaly rankings. This approach is easily motivated. Let α denote a non-zero vector of alphas for a universe of stocks with respect to a benchmark factor, P.[5] That is, α is the intercept vector in the multivariate regression

$$r_t = \alpha + \beta r_{P,t} + \epsilon_t, \tag{1}$$

where r_t is the vector containing the stocks' excess returns in period t, and $r_{P,t}$ is the benchmark's excess return. Consider a factor Q with return $r_{Q,t} = w'r_t$, where w is a weight vector. Suppose that adding $r_{Q,t}$ to the right-hand side of (1) leaves no remaining alphas. That is, the resulting regression becomes

$$r_t = \Theta \begin{bmatrix} r_{P,t} \\ r_{Q,t} \end{bmatrix} + \eta_t. \tag{2}$$

If this additional factor can also produce a covariance matrix for η_t of the form $\sigma^2 I$, then setting w proportional to α produces the desired Q, as shown by MacKinlay and Pástor (2000).[6] In other words, the additional factor that completes the pricing job is constructed by going long stocks with positive alphas and short stocks with negative alphas. Our approach to this long-short construction of Q essentially treats each stock's cross-sectional ranking with respect to an anomaly as a noisy proxy for the stock's alpha ranking. Some of that noise is diversified away by averaging rankings across anomalies, thereby more precisely indicating which stocks to buy and which stocks to short when constructing the factor. Excluding stocks from the middle of the average-ranking distribution, as we do, further increases the likelihood of making correct long/short classifications when constructing the factor. We term the resulting factor corresponding to Q a "mispricing" factor, reflecting our view, discussed earlier, that mispricing is an important element of anomaly-related alphas.

Our approach based on averaging anomaly rankings stands in contrast to previous approaches that construct a factor by ranking on a single variable that initially gained attention as a return anomaly. If such a variable is uniquely motivated as capturing either a systematic-risk sensitivity or mispricing, then our approach simply contaminates that variable with extraneous information. On the other hand, if that variable is not so uniquely valuable, then our approach can work better. Our empirical results support the latter scenario.

---

[5] Assuming a single-factor benchmark is essentially without loss of generality, as P can be viewed as the maximum-Sharpe-ratio combination of multiple factors.

[6] Those authors show that when the unique portfolio Z that is orthogonal to P and produces zero alphas also produces a scalar covariance matrix of residuals, then Z is a combination of P and a portfolio whose weights are proportional to α, as are the weights in Q. (See in particular their equation 26 on page 891.) Regressing r_t on $r_{P,t}$ and $r_{Z,t}$ therefore produces the same residuals and (zero) intercept vector as does regressing r_t on $r_{P,t}$ and $r_{Q,t}$.

As noted earlier, we construct mispricing factors by averaging rankings within the set of 11 prominent anomalies examined by Stambaugh, Yu, and Yuan (2012, 2014, 2015). The initial step in constructing two mispricing factors is to separate the 11 anomalies into two clusters, with a cluster containing the anomalies most similar to each other. Similarity can be measured by either of two methods, using either time-series correlations of anomaly returns or average cross-sectional correlations of anomaly rankings. Both methods produce the same two clusters of anomalies.

In the first method, for each anomaly i we compute the spread, $R_{i,t}$, between the value-weighted returns in month t on stocks in the first and tenth NYSE deciles of the ranking variable in a sort at the end of month $t-1$ of all NYSE/AMEX/NASDAQ stocks with share prices greater than \$5, where the ordering produces a positive estimated intercept in the regression

$$R_{i,t} = \alpha_i + b_i MKT_t + c_i SMB_t + u_{i,t}, \tag{3}$$

and MKT_t and SMB_t are the market and size factors constructed by Fama and French (1993).[7] Next we compute the correlation matrix of the estimated residuals in equation (3). Our sample period runs from January 1967 through December 2013, except data for the distress anomaly begin in October 1974, and data for the return-on-assets anomaly begin in November 1971. To deal with the heterogeneous starting dates, we compute the correlation matrix using the maximum likelihood estimator analyzed by Stambaugh (1997). Using this correlation matrix, we form two clusters by applying the same procedure as Ahn, Conrad, and Dittmar (2009), who combine a correlation-based distance measure with the clustering method of Ward (1963).[8]

In the second method, we compute the z-score of each stock's ranking percentile for each anomaly and then compute the cross-sectional correlations between the z-scores for all available pairs of the 11 anomalies. This procedure gives a set of correlations each month, and we average these correlations across the months in our sample period. The resulting 11×11 matrix of average correlations is then used to form two clusters using the same

---

[7]For the anomaly variables requiring Compustat data from annual financial statements, we require at least a four-month gap between the end of month $t-1$ and the end of the fiscal year. When using quarterly reported earnings, we use the most recent data for which the reporting date provided by Compustat (item RDQ) precedes the end of month $t-1$. When using quarterly items reported from the balance sheet, we use those reported for the quarter prior to quarter used for reported earnings. The latter treatment allows for the fact that a significant number of firms do not include include balance-sheet information with earnings announcements and only later release it in 10-Q filings (see Chen, DeFond, and Park (2002)). For anomalies requiring return and market capitalization, we use data recorded for month $t-1$ and earlier, as reported by CRSP.

[8]Using the version of SMB we construct later in subsection 2.2, instead of the Fama-French version of SMB, does not change any of our cluster-identification results.

procedure applied above to the correlation matrix of long-short returns.

The first cluster of anomalies includes net stock issues, composite equity issues, accruals, net operating assets, asset growth, and investment to assets. These six anomaly variables all represent quantities that firms' managements can affect rather directly. Thus, we denote the factor arising from this cluster as $MGMT$. (The factor construction is described below.) The second cluster includes distress, O-score, momentum, gross profitability, and return on assets. These five anomaly variables are related more to performance and less directly controlled by management, so we denote the factor arising from this cluster as $PERF$. Although we assign names to the clusters, we do not suggest that a cluster reflects a single behavioral story. For example, within the $MGMT$ cluster, equity issuance anomalies could reflect managerial action triggered by mispricing (e.g., Daniel and Titman (2006)), while asset growth and investment anomalies could reflect mispricing triggered by managerial action (e.g., Cooper, Gulen, and Schill (2008)). Moreover, even within the issuance anomalies, multiple effects could be at work, as the results of Greenwood and Hanson (2012) suggest. While there may exist unifying behavioral themes underlying the identities of our clusters, discovering such a framework is beyond the scope of our study.

We next average a stock's rankings with respect to the available anomaly measures within each of the two clusters. Thus, each month a stock has two composite mispricing measures, $P1$ and $P2$. Our averaging of anomaly rankings closely follows the approach of Stambaugh, Yu, and Yuan (2015), who construct a single composite mispricing measure by averaging across all 11 anomalies.[9] As in that study, we equally weight a stock's rankings across anomalies—a weighting that is simple, transparent, and not sample-dependent. As discussed earlier, the rationale for averaging is that, through diversification, a stock's average rank yields a less noisy measure of its mispricing than does its rank with respect to any single anomaly. The evidence suggests that such diversification is effective. As observed by Stambaugh, Yu, and Yuan (2015), the spread between the alphas for portfolios of stocks in the top and bottom deciles of the average ranking across the 11 anomalies is nearly twice the average across those anomalies of the spread between the top- and bottom-decile alphas of portfolios formed using an individual anomaly (with alphas computed using the three-factor model of Fama and French (1993)). We verify a similar result in our sample: The former spread is 95 basis points per month while the latter spread is 53 basis points, and the difference of 42 basis points has a t-statistic of 3.80.

---

[9]Stambaugh, Yu, and Yuan (2015) also report a robustness exercise that employs a clustering approach similar to that reported above.

We construct the mispricing factors by applying a 2×3 sorting procedure resembling that of Fama and French (2015). The approach in that study generalizes the approach in Fama and French (1993), and a similar procedure is applied in Hou, Xue, and Zhang (2015a). Specifically, each month we sort NYSE, AMEX, and NASDAQ stocks (excluding those with prices less than \$5) by size (equity market capitalization) and split them into two groups using the NYSE median size as the breakpoint. Independently, we sort all stocks by $P1$ and assign them to three groups using as breakpoints the 20th and 80th percentiles of the combined NYSE, AMEX, and NASDAQ universe. We similarly assign stocks to three groups according to sorts on $P2$. To construct the first mispricing factor, $MGMT$, we compute value-weighted returns on each of the four portfolios formed by the intersection of the two size categories with the top and bottom categories for $P1$. The value of $MGMT$ for a given month is then the simple average of the returns on the two low-$P1$ portfolios (underpriced stocks) minus the average of the returns on the two high-$P1$ portfolios (overpriced stocks). The second mispricing factor, $PERF$, is similarly constructed from the low- and high-$P2$ portfolios.

The persistence of the measures used to construct our mispricing factors is similar to that of measures used to form other familiar factors. A simple gauge of persistence is the time-series average of the cross-sectional correlation between a given measure's rankings in adjacent months. This average correlation equals 0.955 and 0.965 for the composite mispricing measures used to construct $MGMT$ and $PERF$. The measures used to form the book-to-market, investment, and profitability factors in Fama and French (2015) have average rank correlations of 0.983, 0.943, and 0.981, respectively. Hou, Xue, and Zhang (2015a) construct essentially the same investment factor, while their somewhat different profitability factor uses a measure whose average rank correlation is 0.883. In comparison, market capitalization of equity, used to construct the size factors in all of the above models, has an average rank correlation of 0.996.

One might note that for the breakpoints of $P1$ and $P2$, we use the 20th and 80th percentiles of the NYSE/AMEX/NASDAQ, rather than the 30th and 70th percentiles of the NYSE, used by the studies cited above that apply a similar procedure to different variables. These modifications reflect the notion that relative mispricing in the cross-section is likely to be more a property of the extremes than of the middle. Stambaugh, Yu, and Yuan (2015) find, for example, that the negative (positive) effects of idiosyncratic volatility for overpriced (underpriced) stocks are consistent with the role of arbitrage risk deterring the correction of mispricing, and those authors show that such effects occur primarily in the extremes of a composite mispricing measure and are stronger for smaller stocks. Subsection 3.4 explains

that our main results are robust to the various deviations we take from the more conventional factor-construction methodology tracing to Fama and French (1993). The online appendix reports detailed results of those robustness checks.

Table 1 presents means, standard deviations, and correlations for monthly series of the four factors in our model. (The construction of our size factor, SMB, is explained below.) We see that the two mispricing factors, $MGMT$ and $PERF$, have zero correlation with each other (to two digits) in our overall 1976–2013 sample period. That is, the clustering procedure, coupled with the averaging of individual anomaly rankings, essentially produces two orthogonal factors.

2.2. The Size Factor

When constructing our size factor, we depart more significantly from the approach in Fama and French (2015) and other studies cited above. The stocks we use to form the size factor in a given month are the stocks not used in forming either of the mispricing factors. Specifically, to construct our size factor, SMB (small minus big—that notation we keep), we compute the return on the small-cap leg as the value-weighted portfolio of stocks present in the intersection of both small-cap middle groups when sorting on $P1$ and $P2$. Similarly, the large-cap leg is the value-weighted portfolio of stocks in the intersection of the large-cap middle groups in the sorts on the mispricing measures. The value of SMB in a given month is the return on the small-cap leg minus the large-cap return.

Each 2×3 sort on size and one of the mispricing measures produces six categories, so in total twelve categories result from the sorts using each of the two mispricing measures. If we were to follow the more familiar approach of Fama and French (2015) and others, we would compute SMB as the simple average of the value-weighted returns on the six small-cap portfolios minus the corresponding average of returns on the six large-cap portfolios. By averaging across the three mispricing categories, that approach would seek to neutralize the effects of mispricing when computing the size factor. The problem is that such a neutralization can be thwarted by arbitrage asymmetry—a greater ability or willingness to buy than to short for many investors. With such asymmetry, the mispricing within the overpriced category is likely to be more severe than the mispricing within the underpriced category. Moreover, this asymmetry is likely to be greater for small stocks than for large ones, given that small stocks present potential arbitrageurs with greater risk (e.g., idiosyncratic volatil-

11

ity).[10] Thus, simply averaging across mispricing categories would not neutralize the effects of mispricing, and the resulting SMB would have an overpricing bias. This bias is a concern not just when sorting on our mispricing measures but when sorting on any measure that is potentially associated with mispricing. Some studies argue that book-to-market, for example, contains a mispricing effect (e.g., Lakonishok, Shleifer, and Vishny (1994)), so one might raise a similar concern in the context of the version of SMB computed by Fama and French (1993). By instead computing SMB using stocks only from the middle of our mispricing sorts, avoiding the extremes, we aim to reduce this effect of arbitrage asymmetry.

Consistent with the above argument, our approach delivers a small-cap premium that significantly exceeds not only the value produced by the above alternative method but also the small-cap premium implied by the version of SMB in the three-factor model of Fama and French (1993). For our sample period of January 1967 through December 2013, our SMB factor has an average of 46 bps per month. In contrast, the alternative method discussed above gives an SMB with an average of 28 bps, close to the average of 25 bps for the three-factor Fama-French version of SMB. The differences between our estimated small-cap premium and these alternatives are significant not only statistically (t-statistics: 3.99 and 4.19) but economically as well, indicating a size premium that is nearly twice that implied by the familiar Fama-French version of SMB. This result is similar to the conclusion of Asness, Frazzini, Israel, Moskowitz, and Pedersen (2015), who find that the size premium becomes substantially greater when controlling for other stock characteristics potentially associated with mispricing. Those authors conclude that explaining a significant size premium presents a challenge to asset pricing theory. Such a challenge is beyond the scope of our study as well. Even though the size premium is a fundamentally important quantity, our comparison below of factor models' abilities to explain anomalies is not sensitive to the method used to construct the size factor. (We present further discussion and evidence of this point in subsection 3.4 and in the online appendix.)

Our procedure also appears to have minimal effect on the distribution of firm sizes used in computing SMB. For example, if we first compute the value-weighted average of log size (with size in $1000) for the six small-cap portfolios described above in the more familiar approach, and we then take the simple average of those six values (analogously to what is done with returns), the result is 12.28. If we instead compute the value-weighted average of log size for the firms in the small-cap leg of our SMB, the result is 12.31, nearly identical. The same comparison for large firms gives 15.81 for the more familiar approach versus 15.83 for the firms in the large-cap leg of our SMB.

---

[10]See Stambaugh, Yu, and Yuan (2015) for supporting evidence.

One might ask whether the same approach we take in constructing the size factor—excluding stocks more likely to be mispriced—matters for constructing other factors as well. For example, one could follow this approach when constructing a book-to-market factor. We explore this question for that factor in particular and do find a substantial effect. Specifically, we separate stocks into six groups, following the same procedure as Fama and French (1993), but then before computing value-weighted returns within each group and forming the HML factor, we delete the stocks in the top 20% and bottom 20% of either of our mispricing measures. This additional step renders the value premium 40 percent smaller and statistically insignificant.[11]

2.3. Factor Betas, Arbitrage Asymmetry, and Sentiment Effects

Table 2 gives parameter estimates from our four-factor model for the individual long-short strategies based on the anomaly measures used above as well as book-to-market. Panel A contains the alphas and factor sensitivities ("betas") of the long-short spreads between the value-weighted portfolios of stocks in the long leg (bottom decile) and short leg (top decile). Panel B gives corresponding estimates for the long legs, and Panel C reports estimates for the short legs. The breakpoints are based on NYSE deciles, but all NYSE/AMEX/NASDAQ stocks with share prices of at least \$5 are included.[12] For anomalies in the first cluster, the long-short betas on the first mispricing factor, $MGMT$, are positive with t-statistics between 6.09 and 18.12, whereas the same anomalies' long-short betas on the second factor, $PERF$, are uniformly lower and have t-statistics of mixed signs that average just 1.27. Similarly, for anomalies in the second cluster, the long-short betas on $PERF$ are positive with t-statistics between 5.02 and 24.10, while the betas on $MGMT$ have mixed-sign t-statistics averaging -0.17. These results confirm that averaging anomaly rankings within a cluster produces a factor that captures common variation in returns for the anomalies in that cluster. Not surprisingly, for each anomaly with respect to its corresponding factor, the short-leg beta is significantly negative and the long-leg beta is significantly positive, with the long leg for accruals being the only exception.

Also observe in Table 2 that the short-leg betas are generally larger in absolute magnitude than their long-leg counterparts. With the first-cluster anomalies, for example, the average short-leg $MGMT$ beta is -0.46, whereas the average long-leg $MGMT$ beta is 0.20. Similarly,

---

[11]For our sample period, the monthly average of this alternative book-to-market factor is 0.22% with a t-statistic of 1.81, whereas the Fama-French HML factor has an average of 0.37% with a t-statistic of 3.01.

[12]NYSE breakpoints are also used, for example, by Fama and French (2016) and Hou, Xue, and Zhang (2015a).

for the second-cluster anomalies, the short-leg $PERF$ betas average -0.49 as compared to 0.30 for the long legs. If the factors indeed capture systematic components of mispricing, a greater short-leg sensitivity is consistent with the arbitrage asymmetry discussed above. This arbitrage asymmetry leaves more uncorrected overpricing than uncorrected underpricing, implying greater sensitivity to systematic mispricing for overpriced (short-leg) stocks than for uncerpriced (long-leg) stocks.

Arbitrage asymmetry is also consistent with the relation between investor sentiment and anomaly returns. For each of the anomalies we use to construct our factors, Stambaugh, Yu, and Yuan (2012) observe that the short leg of the long-short anomaly spread is significantly more profitable following high investor sentiment, whereas the long-leg profits are less sensitive to sentiment. We observe similar sentiment effects for our mispricing factors. Table 3 reports the results of regressing each factor as well as its long and short legs on the previous month's level of the investor sentiment index of Baker and Wurgler (2006).[13] For the two mispricing factors, $MGMT$ and $PERF$, the slope coefficients on both the long and short legs are uniformly negative, consistent with sentiment effects, but the slopes for the short legs are two to three times larger in magnitude. The short-leg coefficients for the two factors are nearly identical, as are the t-statistics of -2.06 and -2.05. The long-leg t-statistics, in contrast, are just -0.98 and -1.29.

The stronger sentiment effects for the short legs of $MGMT$ and $PERF$ can be understood in the context of the study by Stambaugh, Yu, and Yuan (2012), who find that the short-leg returns of each of the same 11 anomalies used here are significantly lower following high sentiment. Unlike that study's long-short spreads, $MGMT$ and $PERF$ reflect average anomaly rankings and include more than just the top and bottom deciles, but it is not surprising our results are nevertheless similar. As that study explains, given that many investors are less willing or able to short stocks than to buy them, overpricing resulting from high investor sentiment gets corrected less by arbitrage than does underpricing resulting from low sentiment. Sentiment therefore exhibits a stronger relation to short-leg anomaly returns than to long-leg returns. The significantly positive t-statistics in Table 3 for the sentiment sensitivity of each long-short difference (i.e., each mispricing factor) confirm the greater sentiment effect on the short-leg returns. Overall, the long-short asymmetry in factor betas (Table 2) and sentiment effects (Table 3) is consistent with a mispricing interpretation of our factors.

---

[13]Because investor sentiment can be correlated with economic conditions (e.g., investors can be excessively optimistic when times are good), we use the raw sentiment index produced by Baker and Wurgler (2006) rather than version they orthogonalize with respect to macro factors.

Sentiment does not exhibit much ability to predict our size factor. In Table 3, the t-statistic is -1.60 for the slope coefficient when regressing the long-short spread (SMB) on lagged sentiment, and the t-statistics for the long and short legs (small and large firms) are -1.72 and -1.17. If sentiment affects prices, then periods of high (low) sentiment are likely to be followed by especially low (high) returns on overpriced (underpriced) stocks, especially among smaller stocks, which are likely to be more susceptible to mispricing. Baker and Wurgler (2006) report evidence consistent with this hypothesis, which implies a negative relation between lagged sentiment and the return on a spread that is long small stocks and short large stocks—if mispriced stocks are included, especially in the small-stock leg. The lack of a significant relation between our SMB factor and sentiment suggests some success in our attempt to avoid mispriced stocks when constructing the factor. In contrast, for example, sentiment does exhibit a significant ability to predict the familiar SMB factor from the three-factor model of Fama and French (1993). The slope coefficient is nearly 50% greater in magnitude (-0.32 versus -0.22) and has a t-statistic of -2.31. In fact, the t-statistic for the difference in slopes of -0.10 is -1.68, which is significant at the 5% level for the one-tailed test implied by the alternative hypothesis that our SMB factor is less affected by mispricing.

Our labeling of $MGMT$ and $PERF$ as "mispricing" factors is not what distinguishes them from factors in other models. For example, the investment and profitability factors in the models of Fama and French (2015) and Hou, Xue, and Zhang (2015a) can also reflect mispricing. While both of those studies provide models linking investment and profitability to expected return, their models do not distinguish rational risk-based compensation versus mispricing as the source of expected return. The latter could well be at work: when the short-leg returns of those factors (two from each study) are regressed on lagged investor sentiment, the t-statistics lie between -1.93 and -2.36, consistent with both the results and mispricing interpretation in Stambaugh, Yu, and Yuan (2012). As explained earlier, what instead distinguishes our factors is that they are based on combining the information in multiple anomalies, as opposed to being single-anomaly factors.

3. Comparing Factor Models

Fama and French (2016) explore the ability of the five-factor model of Fama and French (2015) to accommodate various return anomalies. Hou, Xue, and Zhang (2015b) compare that model to the four-factor model of Hou, Xue, and Zhang (2015a) by investigating the two

models' abilities to explain a range of anomalies. We evaluate our four-factor model relative to both of those models, also including the three-factor model of Fama and French (1993) in the comparison as a familiar benchmark.[14] In subsection 3.1, we compare the models' relative abilities to explain a range of individual anomalies, both the set of 12 anomalies examined in Table 2 as well as the substantially wider set of 73 anomalies analyzed by Hou, Xue, and Zhang (2015a, 2015b). Subsection 3.2 then reports pairwise model comparisons that evaluate each model's ability to explain factors present in another. Subsection 3.3 compares models using Bayesian posterior model probabilities. Results of robustness investigations are summarized in subsection 3.4 (with details reported in the online appendix). Subsection 3.5 compares the abilities of the factor models to explain return variance for a variety of stock portfolios.

3.1. Comparing models' abilities to explain anomalies

Table 4 reports alphas from the various factor models for each of the 11 anomalies used in our factors plus book-to-market. For convenience, we denote the factor models as follows:

FF-3: three-factor model of Fama and French (1993)

FF-5: five-factor model of Fama and French (2015)

q-4: four-factor "q-factor" model of Hou, Xue, and Zhang (2015a)

M-4: four-factor mispricing-factor model introduced here

For each anomaly, we construct the difference between the value-weighted monthly return on stocks ranked in the bottom decile and the return on those in the top decile. (The highest rank corresponds to the lowest three-factor Fama-French (1993) alpha.) We then use each long-short return as the dependent variable in 12 regressions of the form

$$R_{i,t} = \alpha_i + \sum_{j=1}^{K} \beta_{i,j} F_{j,t} + u_{i,t}, \tag{4}$$

where the $F_{j,t}$'s are the K factors in a given model. Panel A reports the estimated α_i's for each model. Also reported (first column) are the averages of the $R_{i,t}$'s. Panel B reports the corresponding t-statistics.

The alternative models—FF-3, FF-5, and q-4—exhibit at best only modest ability to accommodate the anomalies. Consistent with having been identified as anomalies with respect to the FF-3 model, the first 11 anomalies in Table 4 (i.e., excluding book-to-market)

---

[14]We are grateful to all of these authors for providing time series of their factors.

produce FF-3 alphas that are significant both economically and statistically. The monthly alphas for those anomalies range from 0.32% (asset growth) to 1.59% (momentum), and the t-statistics range from 2.83 to 5.70. Model FF-5 lowers all but one of the FF-3 alphas for those anomalies, but only the alpha for asset growth—essentially the investment factor in FF-5—drops to insignificance (0.06%, t-statistic: 0.58). Ten alphas remain economically and statistically significant, ranging from 0.32% (net stock issues) to 1.35% (momentum), with the t-statistics ranging from 2.29 to 4.12. Model q-4 does a somewhat better job than FF-5. Asset growth—also essentially the investment factor in q-4—is similarly accommodated, while the alphas on three additional anomalies—distress, momentum, and return on assets—drop to levels insignificant from at least a statistical perspective (t-statistics ranging from 0.72 to 1.40). At the same time, though, seven anomalies have both economically and statistically significant q-4 alphas ranging from 0.32% (investment to assets) to 0.65% (accruals), with t-statistics ranging from 2.50 to 4.30.

Model M-4, true to its intent, does the best job of accommodating the anomalies. Of the nine positive M-4 alphas, all but one are lower than any of the corresponding alphas for the other models. The sole exception is return on assets, for which model q-4 produces a smaller alpha (0.10% versus 0.27%)—unsurprising given that model q-4 includes a profitability factor. Only two of the M-4 t-statistics exceed 2.0 (a third has a t-statistic of 1.90). The alphas for asset growth and distress flip to negative values in model M-4 (with t-statistics of -1.96 and -1.03).

Table 5 compares the models on several measures that summarize abilities to accommodate the set of anomaly long-short spreads: average absolute alpha, average absolute t-statistic of alpha, the number of anomalies for which the model produces the lowest absolute alpha among the four models being compared, and the Gibbons, Ross, and Shanken (1989) "GRS" test of whether all alphas equal zero.[15] Panel A reports these measures for the set of 12 anomalies examined above. Because two of the anomaly series start at later dates than the others, as noted earlier, we compute two versions of the GRS test. The first, denoted GRS_{10}, uses the ten anomalies with full-length histories. The second, GRS_{12},

---

[15]For T time-series observations on the N long-short spreads and the K factors, define the multivariate regression $R = X\Theta + U$, where R and U are $T \times N$, X is $T \times (K+1)$ and Θ is $(K+1) \times N$. The first column of X contains ones, and the remaining K columns contain the factors. The least-squares estimator is $\hat{\Theta} = (X'X)^{-1}X'R$, where the first row of $\hat{\Theta}$, transposed to a column vector, is the $N \times 1$ vector $\hat{\alpha}$. Compute the unbiased residual covariance-matrix estimator, $\hat{\Sigma} = \frac{1}{T-K-1}(R - X\hat{\Theta})'(R - X\hat{\Theta})$, and let $\omega_{1,1}$ denote the 1,1 element of $(X'X)^{-1}$. Then

$$F = \frac{T-N-K}{N(T-K-1)}\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}/\omega_{1,1}$$

has an F distribution with degrees of freedom N and $T - N - K$.

uses all of the anomalies for the shorter sample period with complete data on all 12. The relative performance of the models is consistent across all of the summary measures.[16] For each measure, we see M-4 performs best, followed in decreasing order of performance by q-4, FF-5, and FF-3. (The values in the first column correspond to a zero-factor model, with alphas equal to average excess returns.) The average absolute alpha of 0.18% for M-4 is about half of the next best value of 0.34%, achieved by q-4, and slightly more than a third of the 0.45% value for FF-5. The average absolute t-statistics follow a similar pattern, with M-4 having an average of only 1.29, compared to values of 2.34 and 2.93 for q-4 and FF-5. For nine of the anomalies, model M-4 achieves the lowest absolute alpha, compared to two anomalies for model q-4, one for FF-5, and none for FF-3. The GRS tests, if judged by the p-values, deliver perhaps the sharpest differences between M-4 and the other models. For example, for M-4 the GRS_{10} test produces a p-value of 0.05. In other words, at a significance level of 5% or less, the test does not reject the hypothesis that all ten full-sample anomalies are accommodated by this four-factor model. In contrast, the corresponding p-value is only 0.00000001 for q-4 and just 0.0000000007 for FF-5. For the GRS_{12} test the M-4 p-value is 0.03, but the p-values for q-4 and FF-5 are just 0.000008 and 0.000006.

We next examine a substantially larger set of anomalies. One reason for doing so recognizes the potential advantage that M-4 may have when pricing a set of 12 anomalies of which 11 are used to construct the model's factors. The factors in the models to which we compare M-4 are also constructed from essentially that same set of 12 anomalies, but fewer of the anomalies are used. Thus, M-4 may have a relative advantage in pricing the 12 anomalies analogous the advantage enjoyed by the FF-3 factors when pricing the set of 25 size- and book-to-market-sorted portfolios in Fama and French (1993). The latter advantage is analyzed by Lewellen, Nagel, and Shanken (2010).

Panels B and C of Table 5 report the same measures as Panel A but for the 73 anomalies examined by Hou, Xue, and Zhang (2015a, 2015b). Those authors construct two sets of long-short returns for each anomaly. The first set, analyzed in Hou, Xue, and Zhang (2015a), uses NYSE deciles as the breakpoints for allocating stocks in forming value-weighted portfolios. The second set, constructed in Hou, Xue, and Zhang (2015b), excludes stocks with market capitalizations below the NYSE 20th percentile and then uses deciles of the remaining

---

[16]The consistency in results, here and later, for comparisons of $|\alpha|$ as well as the t and F statistics reflects the similarity in the various models' abilities to explain time-series variance of anomaly returns. Otherwise, models that explain substantially greater variance could produce smaller alpha magnitudes accompanied by stronger rejections (e.g., Fama and French (1993)).

NYSE/AMEX/NASDAQ universe to form equally weighted portfolios.[17]

Panel B of Table 5 reports results for the first set of long-short spreads for the 73 anomalies. Here again we compute two GRS tests, one with the 51 anomalies whose data begin by January 1967, and another with the 72 anomalies with data beginning by February 1986.[18] The relative performance of the three models is the same as for the smaller set of anomalies analyzed in Panel A: model M-4 performs best, followed in order by q-4, FF-5, and FF-3. While the margin between M-4 and q-4 narrows somewhat, M-4 again has a smaller absolute alpha (0.18 versus 0.20) and a smaller absolute t-statistic (0.99 versus 1.15), and M-4 produces lower absolute alphas for nearly twice as many anomalies (37 versus 19). Model M-4 again does better on the GRS test as well. For the test with the set of full-sample anomalies, GRS_{51} delivers p-value for model M-4 of 0.10, failing to reject the hypothesis that all 51 anomalies are accommodated by the model. In contrast, the corresponding q-4 p-value is 0.003.

Panel C of Table 5 reports results using the second set of 73 long-short spreads. Models q-4 and M-4 are closer here, but model M-4 again produces the lowest average absolute alpha (0.22 versus 0.23) and average absolute t-statistic (1.38 versus 1.44), and it achieves a lower absolute alpha on more anomalies (32 versus 23). The GRS statistics are very close, with each of models q-4 and M-4 doing slightly better on one of the two. Both models again enjoy substantial margins over FF-5 and FF-3.

Given that model q-4 is the closest competitor to model M-4 in Table 5, we also consider a modified challenge. Specifically, we reduce the set of 73 anomalies by excluding those most highly correlated with the factors in these two models. For each of four factors—those in models q-4 and M-4 other than the market and size factors—the five anomalies whose long-short returns are most highly correlated with the factor are eliminated. This procedure could eliminate up to 20 anomalies, but somewhat fewer are actually eliminated due to some overlap across factors. (The anomalies eliminated are detailed in the Appendix.)

When using the first set of 73 long-short returns analyzed in Panel B of Table 5, 16 anomalies are eliminated, and the results using the remaining 57 are reported in Panel A of Table 6. For the second set of long-short returns analyzed in Panel C of Table 5, 19 anomalies are eliminated, and Panel B of Table 6 presents results for the remaining 54 anomalies. The

---

[17]The Appendix lists the 73 anomalies. We are grateful to the authors for generously providing us with both sets of these data.

[18]The reason for using only 72 anomalies instead of 73 is that the data for one anomaly, corporate governance (G), are available only from September 1990 through December 2006, so we exclude it to avoid substantially shortening the sample period for the GRS tests.

results in Table 6 deliver essentially the same message as those in Table 5, though the margin of M-4 over q-4 increases a bit. In Panel A of Table 6, as compared to Panel B of Table 5, the gap widens for the average absolute t-statistic. In addition, M-4 produces p-values for the GRS statistic of 0.13 and 0.01, whereas those for q-4 are less than 0.01. In the head-to-head comparison of q-4 and M-4 shown in the last row, M-4 produces the smallest alpha for 36 of the 57 anomalies, compared to 21 for q-4.

For the second set of 73 anomalies, in Panel B of Table 6, as compared to Panel C of Table 5, the gaps for the average absolute alpha and t-statistic widen slightly, while the GRS statistics for q-4 and M-4 are again close, with the latter slightly better on the set of 40 anomalies with longer histories and the former slightly better on the set of 53 with shorter histories. In the comparison of q-4 to M-4 in the last row, M-4 produces the smallest alpha for 33 of the 54 anomalies, compared to 21 for q-4. As before, the margin of M-4 over q-4 is generally smaller than the margin of q-4 over FF-5 and FF-3.

Two additional four-factor models are used in a number of studies. The model of Carhart (1997), MOM-4, adds a momentum factor to FF-3, while the model of Pástor and Stambaugh (2003), LIQ-4, adds a liquidity factor. Including a liquidity factor barely improves, at best, the three-factor model's ability to explain anomalies. This outcome is perhaps unsurprising, as LIQ-4 is the only model of the six considered here whose additional factor is not formed by ranking on a characteristic producing a return anomaly in an earlier study.[19] The improvement over FF-3 produced by MOM-4 is greater than what LIQ-4 produces, but the ability to explain anomalies falls short of that for M-4, q-4, and, generally, FF-5. The results with MOM-4 and LIQ-4 are reported in the online appendix.

3.2. Comparing models' abilities to explain each other's factors

We next investigate whether the factors unique to one model produce non-zero alphas with respect to another model. That is, we explore the extent to which one model can price the factors of the other. Panel A of Table 7 reports the alphas and corresponding t-statistics for (i) the FF-5 factors HML (book-to-market), RMW (profitability), and CMA (investment), (ii) the q-4 factors I/A (investment) and ROE (profitability), and (iii) the M-4 factors $MGMT$ and $PERF$. For each of these three sets, Panel B reports the GRS statistic and

---

[19]The Pástor-Stambaugh (2003) factor is constructed by ranking stocks on their betas with respect to a market-wide liquidity measure. Such betas were previously unexamined in the literature, and as Pástor and Stambaugh explain, they are quite distinct, both conceptually and empirically, from measures of individual stock liquidity. The latter have also been related to average returns (e.g., Amihud and Mendelson (1989)).

p-value testing whether all of the alphas with respect to an alternative model jointly equal zero.

Both models q-4 and M-4 appear to price fairly well the three FF-5 factors—HML, RMW, and CMA. Those factor produce M-4 alphas of 0.11 or less in absolute magnitude, with t-statistics of 1.35 or less in magnitude. The GRS p-value equals 0.58 for the test of whether all three M-4 alphas for HML, RMW, and CMA equal zero. The q-4 model does even better with the FF-5 factors, producing alphas less than 0.04% in magnitude and a GRS p-value of 0.87. This latter result seems not so surprising, given that q-4 contains just different versions of the profitability and investment factors in FF-5.

The FF-5 model fails to price the factors of model M-4 and, somewhat surprisingly, even the factors of model q-4.[20] The FF-5 alphas for the two q-4 factors (I/A and ROE) are 0.12% and 0.45%, with t-statistics of 3.48 and 5.53; the GRS p-value for the joint test is less than 10^{-8}. The FF-5 alphas for the two M-4 factors are even larger, 0.33% and 0.64%, with t-statistics of 4.93 and 4.17; the GRS p-value for the joint test is less than 10^{-10}. Overall, the FF-5 model fares least well in the comparisons reported in Table 7.

The comparison of models q-4 and M-4 in Table 7 provides a more even match in which M-4 finishes modestly ahead. Model M-4 is unable to price the ROE factor of model q-4; the alpha estimate is 0.36% with a t-statistic of 4.00. This significant alpha for the profitability factor of model q-4 seems consistent with the fact that the ROA profitability anomaly is one of the few anomalies in Table 4 not well accommodated by model M-4. The other q-4 factor seems not to present a big problem for M-4, as the alpha estimate for I/A is only 0.09% with a t-statistic of 1.57. In contrast, neither factor of model M-4 is priced by model q-4. The alpha estimates are 0.36% and 0.35%, with t-statistics of 4.54 and 2.24. The GRS tests confirm that neither q-4 nor M-4 can price both factors of the other model, as the p-values are small. At the same time, the rejection is more extreme for model q-4, given its GRS statistic is 72% larger (and the degrees of freedom are the same in both tests).

In sum, model FF-5 accommodates neither the factors of model q-4 nor those of model M-4. In contrast, both of those models appear to price the FF-5 factors, in that the latter factors do not have significant alphas with respect to either model. Models q-4 and M-4 are more evenly matched, as each fails to price all of the other's factors. Model M-4 appears to have the edge, though, in that it accommodates one of the two factors in model q-4 fairly well, whereas the latter model fails to price either of the factors in model M-4.

---

[20]Hou, Xue, and Zhang (2015b) also report that model FF-5 fails to price the q-4 factors.

3.3. Bayesian model comparisons

The analysis in Table 7 discussed above takes a frequentist approach in comparing models' abilities to explain each other's factors. Another approach to making this comparison is Bayesian. The results of such a comparison also favor model M-4 over the alternative factor models.

Suppose we compare two models, M_1 and M_2, and before observing the data we assign probabilities $p(M_1)$ and $p(M_2)$ to each model being the right one, with $p(M_1) + p(M_2) = 1$. After observing the data, D, the posterior probability of model i is given by

$$p(M_i|D) = \frac{p(M_i) \cdot p(D|M_i)}{p(M_1) \cdot p(D|M_1) + p(M_2) \cdot p(D|M_2)}, \tag{5}$$

where model i's marginal likelihood is given by

$$p(D|M_i) = \int_{\theta_i} p(\theta_i)p(D|\theta_i)d\theta_i, \tag{6}$$

$p(\theta_i)$ is the prior distribution for model i's parameters, and $p(D|\theta_i)$ is the likelihood function for model i. As shown by Barillas and Shanken (2015a), when D includes observations of the factors in both models (including the market) as well as a common set of "test" assets, the latter drop out of the computation in (5). Moreover, that study also shows that when $p(\theta_i)$ follows a form as in Pástor and Stambaugh (2000), then $p(D|M_i)$ can be computed analytically. The key feature of the prior, $p(\theta_i)$, is that it is informative about how large a Sharpe ratio can be produced by combining a given set of assets, in this case the assets represented by the model's factors. Specifically, the prior implies a value for the expected maximum squared Sharpe ratio, relative to the (observed) Sharpe ratio of the market. We use the Barillas and Shanken (2015a) analytical results here to compute posterior model probabilities in the above two-way model comparison. Prior model probabilities of each model are set to one-half.[21]

Panel A of Figure 1 displays posterior probabilities in a comparison of model M-4 to model FF-5. The value on the horizontal axis is the square root of the prior expected maximum squared Sharpe ratio achievable by combining the model's factors, $[\mathrm{E}_{\text{prior}}\{S_{MAX}^2\}]^{1/2}$, divided by the observed Sharpe ratio of the market, S_{MKT}. We see in Panel A that when

---

[21]We do not simultaneously compare three or more models, because assigning prior probabilities to models with differing degrees of similarity becomes more complicated. For example, assigning prior probabilities of 1/3 to each of models FF-5, q-4, and M-4 seems unreasonable, in that FF-5 and q-4 are more similar to each other than to M-4, as both FF-5 and q-4 include profitability and investment factors. For a category-based approach to such a problem, see Barillas and Shanken (2015a).

this Sharpe-ratio multiplier is only 1.01 or so, corresponding to a prior expectation that the market's Sharpe ratio can be improved only very modestly, the data favor FF-5. The model probabilities are about equal for a multiplier of 1.05, and then the probability of M-4 rises steeply for higher values, to nearly 1.0 for multipliers of 1.1 or higher. In other words, when the prior admits more than very modest improvement over the market's Sharpe ratio, the data strongly favor M-4 over FF-5.

The posterior probabilities in the comparison of models M-4 and q-4 are reported in Panel B of Figure 1. Here again we see that for smaller values of the Sharpe-ratio multiplier, model M-4 is less favored than the alternative. The probabilities cross at a multiplier between 1.1 and 1.2, and then the probability of M-4 increases to near 1.0 as the multiplier increases to 2.0 or more. In this Bayesian comparison, as in the earlier comparisons, model q-4 again fares better than model FF-5 when compared to model M-4, but the latter is strongly favored to either alternative model if prior beliefs admit a reasonable chance of achieving a substantially higher Sharpe ratio than that of the market index.

The reason that the other models are favored over M-4 when the Sharpe-ratio multiplier is low is essentially that the maximum Sharpe ratio produced by the M-4 factors is higher than for the other models. The Sharpe ratio of the market in our sample period is 0.11, and a multiplier of, say, 1.2 corresponds to an expected maximum Sharpe ratio of roughly just 0.13. In contrast, the sample produces a maximum Sharpe ratio for the M-4 factors equal to 0.49, higher than the maximum Sharpe ratios of 0.35 and 0.45 produced by the FF-5 and q-4 factors, respectively. Given that the sample maximum Sharpe ratio for M-4 is in greatest conflict with a value on the order of 0.13, that model receives the lowest posterior probability when the prior favors such a low maximum Sharpe ratio.

3.4. Robustness

As discussed earlier, a number of the methodological choices we make when constructing our factors deviate somewhat from conventions that originate with Fama and French (1993) and are adopted in later studies such as Fama and French (2015) and Hou, Xue, and Zhang (2015a). None of these choices are crucial to our model's relative performance in explaining anomalies and the other models' factors. For example, because we believe that the extremes of our mispricing measure best identify mispricing, we use breakpoints of 20% and 80% rather than the conventional 30% and 70%. If we recompute Tables 5, 6, and 7 using 30% and 70%, however, our conclusions remain unchanged. Model M-4 maintains its edge over

model q-4 and continues to outperform model FF-5 by a substantial margin. The same holds true if we apply the 20% and 80% breakpoints to just the NYSE universe, instead of to the NYSE/AMEX/NASDAQ universe that covers a broader range of the mispricing scores. The results in Tables 5, 6, and 7 are also not affected much by replacing our SMB, which uses only stocks in the middle category of the mispricing-score rankings, with the more standard version that averages returns across all three categories. If we make all of the above changes simultaneously, again the model comparisons do not change materially. The online appendix reports all of these results.

In cases where annual data are used to construct factors in model FF-3 of Fama and French (1993), model FF-5 of Fama and French (2015), and model q-4 of Hou, Xue, and Zhang (2015a), the authors of those studies require a gap after the end of the fiscal year of at least 6 months and potentially up to 18 months. When using annual data, we instead require a gap of at least 4 months, following precedent in the accounting literature (e.g., Hirshleifer, Hou, Teoh, and Zhang (2004)). Thus, one potential source of performance differences between our model, M-4, and models FF-5 and q-4 is this difference in timing used to contruct factors. To explore this possibility, we reconstruct the factors SMB and HML in model FF-3, SMB, HML, RMW, and CMA in model FF-5, and I/A in model q-4, which rely on annual data, by imposing the minimum gap of 4 months for information from annual statements, while using immediate prior-month values of market capitalization. Again our key results—Tables 5, 6, and 7—are not very sensitive to modifying models FF-5 and q-4 in that fashion. The online appendix reports these results as well.

We also investigate the stability of our results by splitting our overall 47-year period into two subperiods: 1967–1990 (24 years) and 1991–2013 (23 years). First, each of the 11 anomalies we use to construct the factors produces positive long-short FF-3 alphas in both subperiods. In the earlier subperiod, the t-statistics for the long-short alphas range from 1.5 to 5.4 across the 11 anomalies and average 3.4. In the later subperiod, the t-statistics range from 1.2 to 5.0 and again average 3.4. Second, in each subperiod, the assignment of anomalies to the two clusters is identical to that of the overall period, using either of the two methods described earlier. Finally, the superiority of model M-4, as indicated by the results in Tables 5 and 7, is supported consistently across both subperiods. The online appendix reports the results of these model comparisons in each subperiod.

3.5. Factors and Return Variance

Factor models are generally viewed as useful not just for explaining expected returns. An additional role of factor models is to capture systematic time-series variation in realized returns. We briefly explore the extent to which mispricing factors can perform this role as compared to the factors in the alternative models we consider. For various sets of assets, we compute the R-squared in a regression of each asset's monthly return on the factors for a given model. We then average those R-squared values across the assets.

Table 8 reports the average R-squared values obtained for five different sets of assets: 30 industry portfolios plus four sets of 25 portfolios formed with independent two-way 5×5 sorts of stocks on size and either book-to-market (B/M); the 11-anomaly composite mispricing measure used by Stambaugh, Yu, and Yuan (2015) (and used below to form the factor in model M-3); market beta; and return volatility.[22] We estimate the regressions for six factor models: one with just the excess market return (MKT); one with MKT plus the size-factor (SMB) from the three-factor model (FF-3) of Fama and French (1993); the latter three-factor model; the five-factor model (FF-5) of Fama and French (2015); the four-factor model (q-4) of Hou, Xue, and Zhang (2015a); and model M-4. The factor betas on the assets are treated as constant over the sample period, so we probably understate the fraction of variance a given set of factors could explain if the betas were modeled as time-varying. Such a generalization lies beyond our study. We suggest the simple exercise reported briefly here nevertheless provides some insight into the relative abilities of the factor models to capture return variance.

We see in the first row of Table 8 that the market factor on average explains 59.5% of the variance of monthly industry returns. Adding a size factor increases that average R-squared only slightly, to 61.1%, and the other factor models further increase the R-squared only modestly. Model M-4 explains 63.3%, very close to the values of 63.6% for model q-4 and 63.5% for model FF-3. Thus, the model that does best in explaining anomalies—model M-4—does not appear to do so by sacrificing much if any ability to capture variance of industry returns.

For the remaining four sets of assets, the most salient difference from the industry results is that the size factor improves the R-squared substantially over that obtained with just a single market factor. This outcome is not surprising, as size is one of the sorting variables

---

[22]We thank Ken French for providing on his website the returns for the 30 industry portfolios as well as the size-B/M, size-beta, and size-volatility portfolios.

used to form these sets of portfolios. Including the size factor adds roughly 8% to 10% to the R-squared produced by just the market factor. The further increases produced by including the additional factors in the other factor models are generally more modest, and the relative ranking of the R-squared values depends on which sorting variable is used in addition to size. For B/M, models FF-3 and FF-5 enjoy an advantage, which is not surprising given that B/M is used to form a factor in those models. For a similar reason, model M-4 enjoys an advantage when the composite mispricing measure is used to sort. Overall the message delivered here is similar to that for industry portfolios: Mispricing factors generally stack up well against the alternative models considered when judged by abilities to capture return variance.

4.   A Three-Factor Model

In this section we consider an even more parsimonious factor model, denoted as M-3, that replaces the two mispricing factors in M-4 with a single mispricing factor, UMO (underpriced minus overpriced). We construct this single factor in a manner similar to the two factors it replaces. The main difference is that, instead of separating the 11 anomalies into two clusters, we simply average each stock's rankings across all 11 anomalies. We then use this mispricing measure, along with firm size, to construct a size factor and a mispricing factor, applying a 2×3 sorting procedure similar to that used by Fama and French (1993) to construct SMB and HML, but with the same deviations in methodology explained earlier. That is, SMB is constructed using only stocks in the middle mispricing category, as opposed to averaging across all three categories. Also, the mispricing-measure breakpoints are the 20th and 80th percentiles of NYSE/AMEX/NASDAQ, rather than the 30th and 70th percentiles of the NYSE. None of the results are sensitive to these deviations, however.

4.1.   Comparing abilities to explain anomalies

Table 9 reports measures summarizing model M-3's ability to accommodate anomalies. Panels A, B, and C of Table 9 correspond to the same anomaly samples reported in the corresponding panels of Table 5, and the summary measures reported in Table 9 for model M-3 are directly comparable to those reported in Table 5 for models FF-3, FF-5, q-4, and M-4.

First observe that the three-factor model M-3 uniformly dominates the five-factor model FF-5 in all three panels of Table 9 for all four summary measures: average $|\alpha|$, average

$|t|$, and the two GRS statistics for the shorter and longer sample periods. FF-5 in turn uniformly dominates FF-3, so the contrast between the three-factor models, FF-3 and M-3, is especially strong. For example, for FF-3, the values of the average $|\alpha|$ in the three panels in Table 5 are 0.67, 0.44, and 0.53, while, for M-3, the corresponding values in Table 9 are 0.28, 0.24, and 0.30. Comparisons of the average $|t|$ and GRS statistics reveal similarly sharp differences between FF-3 and M-3. In addition to the results reported in Table 9, we find that M-3 produces a smaller $|\alpha|$ than does FF-3 for 11 of the 12 anomalies in Panel A, 55 of the 73 anomalies in Panel B, and 58 of the 73 anomalies in Panel C. It appears that essentially replacing the book-to-market factor in FF-3 with a composite mispricing factor delivers a three-factor model with substantially greater ability to accommodate a wide range of anomalies.

The comparison of q-4 and M-3 in terms of their abilities to explain amomalies does not reveal a clear winner. On one hand, M-3 does better by producing p-values for five out of six of the GRS tests in Table 9 that are higher than the corresponding p-values produced by q-4 in Table 5. In addition, M-3 produces lower average $|\alpha|$ and lower average $|t|$ for the 12 anomalies in Panel A of those tables. On the other hand, q-4 produces lower average $|\alpha|$ and lower average $|t|$ for the larger set of 73 anomalies, reported in Panels B and C.

Model M-4 generally appears to explain anomalies better than model M-3. For all measures except the GRS tests in Panel C, where M-3 holds a slight edge, M-4 performs better than M-3. Overall, among all of the models we consider, M-4 generally exhibits the greatest ability to explain anomalies.

4.2. Comparing abilities to explain other models' factors

Table 10 repeats the same analysis reported earlier in Table 7, except that model M-3, with its single factor, UMO, replaces model M-4 and its two factors, $MGMT$ and $PERF$. The most striking results in Table 10 are the large UMO alphas obtained under both FF-5 and q-4. Observe in Panel A that the UMO alpha under FF-5 is 77 basis points per month with a t-statistic of 6.88, and the UMO alpha under q-4 is 60 basis points with a t-statistic of 5.08. Given these t-statistics, the p-values for the corresponding GRS test statistics in Panel B, less than 5×10^{-10}, are very small.[23] Overall, models FF-5 and q-4 fail rather strongly to price the single mispricing factor in model M-3.

---

[23]The F-statistics in Panel B for these single-asset versions of the GRS test do not exactly equal the square of the t-statistics in Panel A, because the latter are computed using the heteroskedasticity correction of White (1980), whereas the GRS F-test assumes homoskedastic disturbances.

Model M-3 does a comparatively better job of handling the factors in FF-5 and q-4. In Panel A of Table 10, the alpha of HML under M-3 is 34 basis points with a t-statistic of 2.53, while the other two FF-5 factors, RMW and CMA, have insignificant alphas of 13 basis points or less. In Panel B, the GRS joint test of zero alphas for HML, RMW, and CMA yields a p-value of 0.06, insignificant by usual standards. Model M-3 has somewhat more trouble pricing the factors in q-4 than those in FF-5, as is the case for model M-4, noted earlier. Although the alpha for ROE under M-3 is just 17 basis points with a t-statistic of 1.60, the alpha for I/A is 25 basis points with a t-statistic of 2.87, and the GRS joint test of both alphas being zero gives a p-value of just 4.1×10^{-4}. Observe, however, that none of the alphas for another model's factors are even half as large as that model's alpha for UMO.

Table 11 reports a comparison of the abilities of models M-3 and M-4 to explain each other's factors. We see that neither model does well in this regard. In Panel A, the alpha of UMO under model M-4 is 21 basis points with a t-statistic of 4.76, while the alphas of $MGMT$ and $PERF$ under model M-3 are 33 and -43 basis points, with t-statistics of 3.18 and -3.63. In Panel B, the p-values of both GRS tests are small, though the one for model M-3 is somewhat larger. In sum, even though the non-nested models M-3 and M-4 both contain mispricing factors, neither model subsumes the other.

The Bayesian model comparisons displayed earlier in Figure 1 are repeated in Figure 2, with model M-3 replacing model M-4 in the comparisons to models FF-5 and q-4. When either of the latter two models are compared to M-3, the posterior probability for model M-3 is nearly one for all priors except those that admit extremely small potential improvements over the market's Sharpe ratio. Moreover, in this Bayesian model comparison, model M-3 dominates model M-4 with a plot (not displayed) nearly identical to that displayed in both panels of Figure 2. Playing a key role in the superior performance of model M-3 in these comparisons is that the three factors of M-3 combine to give a maximum Sharpe ratio of 0.50, higher than the earlier-reported values of 0.35, 0.45, and 0.49 for models FF-5, q-4, and M-4, respectively. The slightly higher Sharpe ratio achieved by M-3 as compared to M-4 is consistent with the noise-diversification rationale for averaging anomaly rankings, discussed earlier. Averaging across all 11 anomalies (as in M-3) eliminates more noise than averaging separately within the clusters used to form the two mispricing factors in M-4. We see that even though the weights on the latter two factors are optimized within the sample when computing the maximum Sharpe ratio for M-4, the stronger diversification benefit of averaging across all 11 anomalies allows M-3 to achieve a slightly higher Sharpe ratio.

When judged simply by abilities to price other models' factors, model M-3 dominates in

the head-to-head comparisons with each of the other models. In contrast, model M-4 does best in accommodating a wide range of anomalies. As with our findings for model M-4, our conclusions regarding model M-3 are generally supported within each half of the overall period. The online appendix reports the analyses in Tables 9, 10, and 11 for each subperiod.

5.   Arbitrage Risk and the Factor Models

Idiosyncratic volatility (IVOL) represents risk for arbitrageurs seeking to exploit mispricing (e.g., Pontiff (2006)). IVOLs relation to average return also presents an anomaly, in that stocks with high (low) IVOL have negative (positive) alphas with respect to the CAPM of Sharpe (1964) and Lintner (1965) as well as the three-factor model of Fama and French (1993). (See, for example, Ang, Hodrick, Xing, and Zhang (2006).) Some of the models with additional factors appear to explain this IVOL anomaly, but as we show below, such results prove illusory when recognizing that IVOL represents arbitrage risk.

As explained by Stambaugh, Yu, and Yuan (2015), since higher IVOL implies greater arbitrage risk, mispricing should get corrected less among stocks with high IVOL. Among overpriced stocks, the relation between IVOL and alpha should therefore be negative, as arbitrage eliminates less overpricing in high-IVOL stocks. Among underpriced stocks, the relation between IVOL and alpha should be positive, as less underpricing is eliminated in high-IVOL stocks. With arbitrage asymmetry, however, the negative relation among overpriced stocks should be stronger, resulting in the overall negative relation between alpha and IVOL—the IVOL anomaly. Stambaugh, Yu, and Yuan (2015) report results consistent with these predictions. Panel A of Table 12 essentially repeats their analysis with just a slightly longer sample period. As in that study, stocks are sorted independently with respect to IVOL and a mispricing measure that averages the rankings for the 11 anomalies.[24] We obtain similar results. The last row of Panel A displays the familiar negative relation between IVOL and FF-3 alpha, and the negative IVOL effect among overpriced stocks is stronger than the positive relation among underpriced stocks.

Panel B of Table 12 repeats the analysis, using the same mispricing and IVOL rankings as in Panel A, but with alphas computed with respect to FF-5. The negative IVOL-alpha relation among the overpriced stocks is stronger than the positive relation among the under-

---

[24]We compute individual stock IVOL, following Ang, Hodrick, Xing, and Zhang (2006), as the standard deviation of the most recent month's daily benchmark-adjusted returns. The latter returns are computed as the residuals in a regression of each stock's daily return on the three factors in model FF-3.

priced stocks. Thus the overall IVOL-alpha relation in the bottom row of the panel is again negative, though it is substantially weaker than in Panel A. In the latter case the difference between the highest and lowest IVOL portfolio alphas is -81 bps per month (t-statistic: -6.04), as compared to the corresponding value in Panel B of -33 bps (t-statistic: -2.69).

Panels C and D of Table 12 repeat the same analysis for models q-4 and M-4. Both of these models appear to explain the IVOL anomaly, given the results in the last row of each panel. For model q-4, the difference between the highest and lowest IVOL portfolio alphas is -23 bps per month, about one-third smaller than that for FF-5, and the t-statistic of -1.58 is insignificant at conventional levels. For model M-4, reported in Panel D, the highest-versus-lowest IVOL difference is -12 bps, and the t-statistic is just -0.79. We can see, however, that these results in the last row of each panel obscure important IVOL effects. For both models, the relation between IVOL and alpha is strongly negative (positive) among the overpriced (underpriced) stocks. The insignificance of the overall IVOL-alpha relation simply reflects a lower degree of asymmetry in the strength of the IVOL effects in the overpriced and underpriced segments.

This lower asymmetry in the IVOL effects for models q-4 and M-4 is consistent with the earlier reported results indicating that these models have the greatest overall abilities to capture mispricing reflected in anomalies. A model with a greater ability to capture mispricing is likely to be better at capturing mispricing where it tends to be most severe—among high-IVOL overpriced stocks. Observe in Table 12 that the monthly alpha for the high-IVOL overpriced portfolio increases monotonically from -1.87% to -0.96% when moving across the models in Panels A through D. This 91 basis point difference, much larger than any of the other cross-model differences in Table 12, substantially weakens the negative IVOL effect in overpriced stocks when moving from model FF-3 to model M-4. With this weaker (but still significant) negative IVOL effect, there is less asymmetry in the opposing IVOL effects for underpriced and overpriced stocks.

6. Conclusions

Popular factors are generally constructed by ranking stocks on characteristics that initially gained the finance profession's attention by producing return anomalies. We also construct factors based on documented anomalies, but with a key difference. Rather than have each factor correspond to a single anomaly, as is typical, we construct factors by combining stocks' rankings with respect to 11 prominent anomalies. Specifically, we form two factors

by averaging rankings within two clusters of anomalies whose long-short profits exhibit the greatest co-movement. We denote the resulting long-short return spreads as "mispricing" factors, motivated by evidence that anomalies in part reflect mispricing and possess common sentiment effects. Both of the mispricing factors exhibit a significant relation to lagged investor sentiment.

We combine the two mispricing factors with market and size factors to produce a four-factor model. We find that this four-factor model's ability to accommodate a wide range of anomalies exceeds that of both the four-factor model of Hou, Xue, and Zhang (2015a) and the five-factor model of Fama and French (2015). Our four-factor model also dominates those models when judging models by their abilities to price each other's factors, using both frequentist and Bayesian methods. If our two new factors are replaced by a single factor that simply averages rankings across our entire set of anomalies, rather than within two clusters, the resulting three-factor model also performs well. It explains anomalies better than models with more factors, especially the Fama-French five-factor model, and it actually performs best of all when comparing models' abilities to price each other's factors.

The typical approach to constructing a size factor using a two-way sort is to average across the categories of the other sorting variable, which in our case is a mispricing measure. We instead construct the size factor using stocks in neither extreme of a mispricing measure, in order to be less susceptible to asymmetric degrees of overpricing versus underpricing. The resulting size factor reveals a substantially larger small-firm premium than usual estimates.

Mispricing factors can reflect common elements of mispricing, such as market-wide sentiment effects, but a parsimonious factor model is challenged to fully explain expected returns when mispricing is present. For example, factor models have difficulty accommodating the role idiosyncratic volatility (IVOL) plays as risk that deters price-correcting arbitrage. Although the well known IVOL anomaly of Ang, Hodrick, Xing, and Zhang (2006) appears to be explained by some of the factor models we consider, such success is misleading. Alphas with respect to these factor models exhibit opposite and offsetting relations to IVOL, depending on whether the mispricing that the models fail to capture—the alphas—reflect relative underpricing or overpricing. In this sense, while parsimonious factor models are appealing and useful, there are limits to their abilities to explain expected returns. Nevertheless, among such models, those with mispricing factors appear to have greater ability than prominent alternatives.

Table 1

Factor Summary Statistics

The table reports summary statistics for the monthly observations of the factors in the M-4 model. The sample period is from January 1967 through December 2013 (564 months).

			Correlations			
Factor	Mean(%)	Std. Dev.(%)	$MGMT$	$PERF$	SMB	MKT
$MGMT$	0.62	2.93	1	0.00	-0.30	-0.55
$PERF$	0.70	3.83	0.00	1	-0.06	-0.25
SMB	0.46	2.90	-0.30	-0.06	1	0.26
MKT	0.51	4.60	-0.55	-0.25	0.26	1

Table 2
Factor Loadings and Alphas of Anomaly Strategies
Under the Mispricing-Factor Model

The table reports the model's factor loadings and monthly alphas (in percent) for 12 anomalies. For each anomaly, the regression estimated is

$$R_t = \alpha + \beta_{MKT}MKT_t + \beta_{SMB}SMB_t + \beta_{MGMT}MGMT_t + \beta_{PERF}PERF_t + \epsilon_t,$$

where R_t is the return in month t on the anomaly's long-leg, short-leg, or long-short spread, MKT_t is the excess market return, SMB_t is the model's size factor, and $MGMT_t$ and $PERF_t$ are the mispricing factors. The long leg of an anomaly is the value-weighted portfolio of stocks in the lowest decile of the anomaly measure, and the short leg contains the stocks in the highest decile, where a high value of the measure is associated with lower return. The breakpoints are based on NYSE deciles, but all NYSE/AMEX/NASDAQ stocks with share prices of at least \$5 are included. Data for the distress anomaly begin in October 1974, and data for the return-on-assets anomaly begin in November 1971. The other ten anomalies begin at the start of the sample period, which is from January 1967 through December 2013 (564 months). All t-statistics are based on the heteroskedasticity-consistent standard errors of White (1980).

Anomaly	α	β_{MKT}	β_{SMB}	β_{MGMT}	β_{PERF}	t_α	t_{MKT}	t_{SMB}	t_{MGMT}	t_{PERF}
Panel A: Long-Short Spreads										
First Cluster (used to construct mispricing factor MGMT)										
Net stock issues	0.06	0.02	-0.14	0.63	0.22	0.70	0.86	-3.64	17.21	8.85
Composite equity issues	0.07	-0.07	-0.06	0.85	0.05	0.70	-2.23	-1.22	18.12	1.72
Accruals	0.31	0.00	-0.28	0.38	0.02	2.08	0.12	-5.23	6.09	0.48
Net operating assets	0.22	0.11	-0.05	0.46	-0.01	1.70	2.72	-0.73	8.54	-0.28
Asset growth	-0.22	0.04	0.33	0.94	-0.02	-1.96	1.31	7.42	15.99	-0.54
Investment-to-assets	0.06	0.03	0.25	0.64	-0.09	0.54	1.12	5.40	11.83	-2.61
Average	0.08	0.02	0.01	0.65	0.03	0.63	0.65	0.33	12.96	1.27
Second Cluster (used to construct mispricing factor PERF)										
Distress	-0.16	-0.29	-0.35	0.31	1.17	-1.03	-7.13	-3.89	3.96	24.10
O-score	0.35	-0.15	-0.73	-0.09	0.23	2.42	-3.86	-14.43	-1.39	5.02
Momentum	0.12	0.12	0.16	0.25	1.21	0.47	1.78	1.44	1.71	12.15
Gross profitability	0.11	-0.14	-0.05	-0.32	0.66	0.92	-4.45	-1.29	-6.08	18.04
Return on assets	0.27	-0.02	-0.38	0.06	0.66	1.90	-0.41	-5.49	0.95	13.21
Average	0.14	-0.10	-0.27	0.04	0.79	0.94	-2.81	-4.73	-0.17	14.50
Book-to-market	-0.17	0.09	0.54	0.89	-0.35	-1.10	2.50	9.21	12.70	-7.52

33

Anomaly	α	β_{MKT}	β_{SMB}	β_{MGMT}	β_{PERF}	t_α	t_{MKT}	t_{SMB}	t_{MGMT}	t_{PERF}
			Panel B: Long Legs							
First Cluster (used to construct mispricing factor MGMT)										
Net stock issues	-0.06	1.04	-0.03	0.31	0.06	-1.39	83.49	-1.22	14.29	4.18
Composite equity issues	-0.01	0.98	0.05	0.48	-0.06	-0.18	33.81	0.98	11.13	-2.31
Accruals	0.19	1.05	0.02	-0.22	0.08	1.74	34.37	0.44	-4.64	2.37
Net operating assets	0.13	1.08	0.03	0.09	-0.08	1.44	44.02	0.86	2.49	-2.92
Asset growth	-0.15	1.11	0.35	0.34	-0.01	-1.58	43.33	11.25	6.99	-0.18
Investment-to-assets	-0.01	1.07	0.35	0.17	0.02	-0.13	50.21	10.80	5.52	0.83
Average	0.02	1.06	0.13	0.20	0.00	-0.02	48.21	3.85	5.96	0.33
Second Cluster (used to construct mispricing factor PERF)										
Distress	-0.22	0.98	0.11	0.04	0.39	-2.32	37.65	1.79	0.78	12.27
O-score	0.18	0.94	-0.12	-0.32	0.14	2.04	43.07	-3.38	-6.62	4.26
Momentum	0.10	1.15	0.38	-0.18	0.46	0.77	31.69	6.59	-2.40	9.48
Gross profitability	0.05	0.97	-0.01	-0.01	0.24	0.52	36.95	-0.28	-0.29	7.14
Return on assets	0.14	1.01	-0.06	-0.25	0.27	1.92	53.19	-2.03	-7.25	10.94
Average	0.05	1.01	0.06	-0.14	0.30	0.59	40.51	0.54	-3.16	8.82
Book-to-market	-0.10	1.09	0.38	0.54	-0.17	-0.91	38.00	8.42	10.38	-4.95
			Panel C: Short Legs							
First Cluster (used to construct mispricing factor MGMT)										
Net stock issues	-0.13	1.02	0.11	-0.32	-0.16	-1.52	46.95	3.62	-8.82	-6.62
Composite equity issues	-0.09	1.05	0.11	-0.37	-0.11	-1.26	58.06	3.80	-11.80	-5.11
Accruals	-0.11	1.04	0.30	-0.61	0.06	-1.35	47.35	8.75	-16.11	2.43
Net operating assets	-0.09	0.97	0.08	-0.37	-0.07	-1.20	43.72	2.34	-11.45	-2.80
Asset growth	0.07	1.07	0.02	-0.60	0.01	1.05	55.02	0.60	-20.20	0.77
Investment-to-assets	-0.07	1.04	0.10	-0.47	0.11	-0.75	41.33	2.90	-9.14	3.21
Average	-0.07	1.03	0.12	-0.46	-0.03	-0.84	48.74	3.67	-12.92	-1.35
Second Cluster (used to construct mispricing factor PERF)										
Distress	-0.06	1.28	0.46	-0.27	-0.78	-0.52	39.36	8.75	-4.67	-23.10
O-score	-0.17	1.09	0.62	-0.23	-0.10	-1.80	40.74	13.34	-5.14	-3.05
Momentum	-0.02	1.03	0.21	-0.43	-0.75	-0.14	24.38	3.04	-4.91	-12.39
Gross profitability	-0.06	1.10	0.04	0.31	-0.42	-0.56	38.35	1.25	5.94	-13.00
Return on assets	-0.13	1.03	0.32	-0.32	-0.39	-0.99	28.53	4.69	-5.06	-8.69
Average	-0.09	1.11	0.33	-0.19	-0.49	-0.80	34.27	6.21	-2.77	-12.05
Book-to-market	0.07	1.00	-0.16	-0.35	0.18	0.92	51.01	-6.31	-11.45	7.87

34

Table 3
Investor Sentiment and the Factors

The table reports estimates of b in the regression

$$R_t \;=\; a + bS_{t-1} + u_t,$$

where R_t is the excess return in month t on either the long leg, the short leg, or the long-short difference for each of the factors (MKT, SMB, $MGMT$ and $PERF$), and S_{t-1} is the previous month's level of the investor-sentiment index of Baker and Wurgler (2006). All t-statistics are based on the heteroskedasticity-consistent standard errors of White (1980). The sample period is from January 1967 through December 2010 (528 months).

Factor	Long Leg		Short Leg		Long$-$Short	
	$\hat{b}$	t-stat.	$\hat{b}$	t-stat.	$\hat{b}$	t-stat.
MKT	-	-	-	-	-0.32	-1.37
SMB	-0.49	-1.72	-0.27	-1.17	-0.22	-1.60
$MGMT$	-0.22	-0.98	-0.66	-2.06	0.44	2.81
$PERF$	-0.31	-1.29	-0.67	-2.05	0.36	2.02

Table 4
Anomaly Alphas Under Different Factor Models

For long-short returns corresponding to 12 anomalies, the table reports information about alphas computed under four different factor models: the three-factor model of Fama and French (1993), denoted FF-3; the five-factor model of Fama and French (2015), denoted FF-5; the four-factor model of Hou, Xue, and Zhang (2015a), denoted q-4; and the four-factor mispricing-factor model introduced in this study, denoted M-4. Also reported are the average unadjusted return spreads (the alphas in a model with no factors). In constructing the long-short spreads, the long leg is the value-weighted portfolio of stocks in the lowest decile of the anomaly measure, and the short leg contains the stocks in the highest decile, where a high value of the measure is associated with lower return. The breakpoints are based on NYSE deciles, but all NYSE/AMEX/NASDAQ stocks with share prices of at least \$5 are included. Panel A reports the monthly alphas (in percent); Panel B reports their heteroskedasticity-consistent t-statistics based on White (1980). The sample period is from January 1967 through December 2013 (564 months).

Anomaly	Unadjusted	FF-3	FF-5	q-4	M-4
Panel A: Alphas					
Net stock issues	0.56	0.66	0.32	0.37	0.06
Composite equity issues	0.58	0.54	0.34	0.51	0.07
Accruals	0.43	0.51	0.56	0.65	0.31
Net operating assets	0.53	0.53	0.50	0.43	0.22
Asset growth	0.52	0.32	0.06	0.08	-0.22
Investment-to-assets	0.53	0.42	0.35	0.32	0.06
Distress	0.44	1.21	0.62	0.20	-0.16
O-score	0.05	0.49	0.45	0.47	0.35
Momentum	1.26	1.59	1.35	0.48	0.12
Gross profitability	0.28	0.69	0.35	0.39	0.11
Return on assets	0.58	0.91	0.43	0.10	0.27
Book-to-market	0.43	-0.20	-0.14	-0.03	-0.17
Panel B: t-statistics					
Net stock issues	4.77	6.60	3.42	3.54	0.71
Composite equity issues	3.88	4.93	2.94	4.10	0.70
Accruals	2.95	3.61	3.94	4.30	2.08
Net operating assets	4.32	4.10	3.63	3.03	1.70
Asset growth	3.69	2.83	0.58	0.72	-1.96
Investment-to-assets	4.28	3.48	3.04	2.72	0.54
Distress	1.54	5.03	2.29	0.78	-1.03
O-score	0.30	4.28	3.92	3.89	2.42
Momentum	4.58	5.70	4.12	1.40	0.47
Gross profitability	1.79	5.22	2.78	2.50	0.92
Return on assets	3.18	5.52	3.13	0.85	1.90
Book-to-market	2.39	-1.99	-1.33	-0.19	-1.10

36

Table 5
Summary Measures of Models' Abilities to Explain Anomalies

The table reports measures that summarize the degree to which anomalies produce alpha under four different factor models: the three-factor model of Fama and French (1993), denoted FF-3; the five-factor model of Fama and French (2015), denoted FF-5; the four-factor model of Hou, Xue, and Zhang (2015a), denoted q-4; and the four-factor mispricing-factor model introduced in this study, denoted M-4. Also reported are measures based on the average unadjusted return spreads (the alphas in a model with no factors). For each model, the table reports the average absolute alpha, average absolute t-statistic, the F-statistic and associated p-value for the "GRS" test of Gibbons, Ross, and Shanken (1989), and the number of anomalies for which the model produces the smallest absolute alpha among the models being compared in the table. In Panel A, two versions of the GRS test are reported. Data for the distress anomaly begin in October 1974, and data for the return-on-assets anomaly begin in November 1971. The other ten anomalies begin at the start of the sample period, which is from January 1967 through December 2013 (564 months). GRS_{10} tests whether the ten alphas for the full-sample anomalies equal zero, while GRS_{12} uses all 12 anomalies and begins the sample in October 1974. Panels B and C also report two versions of the GRS test: GRS_{51} uses 51 anomalies whose data are available by January 1967, while GRS_{72} uses those anomalies plus 21 others whose data are available by February 1986.

Measure	Unadjusted	FF-3	FF-5	q-4	M-4
Panel A: 12 Anomalies, value-weighted, NYSE deciles					
Average $|\alpha|$	0.52	0.67	0.45	0.34	0.18
Average $|t|$	3.14	4.44	2.93	2.34	1.29
GRS_{10}	6.89	10.10	6.71	5.99	1.84
p_{10}	3.4×10^{-10}	1.1×10^{-15}	6.9×10^{-10}	1.2×10^{-8}	0.05
GRS_{12}	6.16	7.71	4.17	3.95	1.88
p_{12}	4.5×10^{-10}	4.2×10^{-13}	3.1×10^{-6}	8.3×10^{-6}	0.03
Number of $min|\alpha|$	-	0	1	2	9
Panel B: 73 Anomalies, value-weighted, NYSE deciles					
Average $|\alpha|$	0.39	0.44	0.30	0.20	0.18
Average $|t|$	2.14	2.74	1.77	1.15	0.99
GRS_{51}	2.74	2.60	1.91	1.68	1.28
p_{51}	9.3×10^{-9}	6.5×10^{-8}	2.9×10^{-4}	3.2×10^{-3}	0.10
GRS_{72}	2.23	2.10	1.79	1.78	1.54
p_{72}	2.2×10^{-6}	1.3×10^{-5}	5.3×10^{-4}	5.8×10^{-4}	8.1×10^{-3}
Number of $min|\alpha|$	-	7	10	19	37
Panel C: 73 Anomalies, equally weighted, NYSE/AMEX/NASDAQ deciles					
Average $|\alpha|$	0.50	0.53	0.35	0.23	0.22
Average $|t|$	3.02	3.72	2.41	1.44	1.38
GRS_{51}	5.85	6.31	5.15	4.19	4.17
p_{51}	2.7×10^{-27}	6.2×10^{-30}	4.3×10^{-23}	2.3×10^{-17}	3.1×10^{-17}
GRS_{72}	2.95	3.25	2.64	2.41	2.68
p_{72}	1.4×10^{-10}	3.0×10^{-12}	1.1×10^{-8}	2.3×10^{-7}	6.2×10^{-9}
Number of $min|\alpha|$	-	7	11	23	32

Table 6
Summary Measures of Models' Abilities to Explain Anomalies Less
Correlated with Factors in Models q-4 and M-4

The table reports measures that summarize the degree to which anomalies produce alpha under four different factor models: the three-factor model of Fama and French (1993), denoted FF-3; the five-factor model of Fama and French (2015), denoted FF-5; the four-factor model of Hou, Xue, and Zhang (2015a), denoted q-4; and the four-factor mispricing-factor model introduced in this study, denoted M-4. Also reported are measures based on the average unadjusted return spreads (the alphas in a model with no factors). For each model, the table reports the average absolute alpha, average absolute t-statistic, the F-statistic and associated p-value for the "GRS" test of Gibbons, Ross, and Shanken (1989), and the number of anomalies for which the model produces the smallest absolute alpha among the models being compared in the table. Panels A and B correspond to Panels B and C of Table 5, except the set of anomalies is reduced: For each of four factors—those in models q-4 and M-4 other than the market and size factors—the five anomalies (of the 73) whose long-short returns are most highly correlated with the factor are eliminated. This procedure leaves 57 anomalies in Panel A and 54 in Panel B. The sample period is from January 1967 through December 2013 (564 months). Two versions of the GRS test are reported: GRS_{41} in Panel A (or GRS_{40} in Panel B), uses the 41 (or 40) anomalies whose data are available by January 1967, while GRS_{56} in Panel A (or GRS_{53} in Panel B) uses the remaining anomalies whose data are available by February 1986.

Measure	Unadjusted	FF-3	FF-5	q-4	M-4
Panel A: 57 Anomalies, value-weighted, NYSE deciles					
Average $|\alpha|$	0.37	0.39	0.27	0.19	0.17
Average $|t|$	2.14	2.60	1.71	1.15	0.97
GRS_{41}	3.01	2.81	1.95	1.63	1.26
p_{41}	6.7×10^{-9}	6.1×10^{-8}	5.8×10^{-4}	9.3×10^{-3}	0.13
GRS_{56}	2.29	2.16	1.73	1.64	1.53
p_{56}	5.5×10^{-6}	2.5×10^{-5}	2.3×10^{-3}	5.2×10^{-3}	0.01
Number of $min|\alpha|$	-	4	7	15	31
Number of $min|\alpha|$, q-4 vs. M-4	-	-	-	21	36
Panel B: 54 Anomalies, equally weighted, NYSE/AMEX/NASDAQ deciles					
Average $|\alpha|$	0.45	0.46	0.32	0.25	0.23
Average $|t|$	2.94	3.53	2.41	1.61	1.47
GRS_{40}	6.38	7.03	5.78	4.69	4.55
p_{40}	7.3×10^{-26}	5.0×10^{-29}	8.3×10^{-23}	2.9×10^{-17}	1.6×10^{-16}
GRS_{53}	3.36	4.07	3.12	2.82	3.03
p_{53}	3.6×10^{-11}	1.0×10^{-14}	6.7×10^{-10}	$2.0E \times 10^{-8}$	1.8×10^{-9}
Number of $min|\alpha|$	-	6	9	16	23
Number of $min|\alpha|$, q-4 vs. M-4	-	-	-	21	33

Table 7

Abilities of Models FF-5, q-4, and M-4 to Explain Each Other's Factors

Panel A reports a factor's estimated monthly alpha (in percent) with respect to each of the other models (with White (1980) heteroskedasticity-consistent t-statistics in parentheses). Panel B computes the Gibbons-Ross-Shanken (1989) F-test of whether a given model produces zero alphas for the factors of an alternative model (with p-values in parentheses). The factors whose alphas are tested are those other than a model's market and size factors. The models considered are the five-factor model of Fama and French (2015), denoted FF-5, which includes the factors HML, RMW, and CMA; the four-factor model of Hou, Xue, and Zhang (2015a), denoted q-4, which includes the factors I/A and ROE; and the four-factor mispricing-factor model, denoted M-4, which includes the factors $MGMT$ and $PERF$. The sample period is from January 1967 through December 2013 (564 months).

| | Alpha computed with respect to model | | |
Factors	FF-5	q-4	M-4
Panel A: Alpha (t-statistic)			
Factors in FF-5			
HML	-	0.04	-0.03
		(0.43)	(-0.28)
RMW	-	0.04	0.11
		(0.55)	(1.35)
CMA	-	0.02	-0.03
		(0.47)	(-0.56)
Factors in q-4			
I/A	0.12	-	0.09
	(3.48)		(1.57)
ROE	0.45	-	0.36
	(5.53)		(4.00)
Factors in M-4			
$MGMT$	0.33	0.36	-
	(4.93)	(4.54)	
$PERF$	0.64	0.35	-
	(4.17)	(2.24)	
Panel B: GRS F-statistic (p-value)			
HML, RMW, CMA	-	0.23	0.65
		(0.87)	(0.58)
$I/A, ROE$	19.06	-	9.12
	(9.8×10^{-9})		(1.3×10^{-4})
$MGMT, PERF$	25.35	15.66	-
	(2.9×10^{-11})	(2.4×10^{-7})	

Table 8
Percent of Return Variance Explained by Factor Models

The table reports the average R-squared in regressions of monthly asset returns on factors. Regressions are estimated for seven factor models: one with just the excess market return (MKT); one with MKT plus a size-factor (SMB); the three-factor model of Fama and French (1993), denoted FF-3; the five-factor model of Fama and French (2015), denoted FF-5; the four-factor model of Hou, Xue, and Zhang (2015a), denoted q-4; and the four-factor mispricing-factor model introduced in this study, denoted M-4. The sets of assets consist of 30 industry portfolios plus four sets of 25 portfolios formed with independent two-way 5×5 sorts on size and either book-to-market (B/M), the 11-anomaly composite mispricing measure, market beta, or return volatility. The sample period is from January 1967 through December 2013 (564 months).

Number of Assets	Description	Factor Model					
		MKT	MKT & SMB	FF-3	FF-5	q-4	M-4
30	Industry Portfolios	59.5	61.1	63.5	66.1	63.6	63.3
25	Size-B/M Portfolios	74.9	85.3	91.6	92.1	88.4	88.1
25	Size-Mispricing Portfolios	81.2	89.8	90.8	91.9	91.2	92.3
25	Size-Beta Portfolios	75.6	86.0	88.8	89.8	88.1	87.7
25	Size-Volatility Portfolios	75.3	83.2	87.1	89.4	86.4	86.6

40

Table 9
**Summary Measures of the Abilities of a Model with a Single Mispricing Factor
to Explain Anomalies**

The table reports summary statistics for alphas computed under a model that combines a market and size factor
with a single mispricing factor, denoted model M-3. The mispricing factor is constructed using a composite
mispricing measure that averages stocks' rankings across 11 anomalies. The table reports the model's average
absolute monthly (percent) alpha, average absolute t-statistic, and the F-statistic and associated p-value for
the "GRS" test of Gibbons, Ross, and Shanken (1989). In Panel A, two versions of the GRS test are reported.
Data for the distress anomaly begin in October 1974, and data for the return-on-assets anomaly begin in
November 1971. The other ten anomalies begin at the start of the sample period, which is from January
1967 through December 2013 (564 months). GRS_{10} tests whether the ten alphas for the full-sample anomalies
equal zero, while GRS_{12} uses all 12 anomalies and begins the sample in October 1974. Panels B and C also
report two versions of the GRS test: GRS_{51} uses 51 anomalies whose data are available by January 1967,
while GRS_{72} use those anomalies plus 21 others whose data are available by February 1986.

Panel A: 12 Anomalies, value-weighted, NYSE deciles

Average $|\alpha|$	0.28
Average $|t|$	1.78
GRS_{10}	2.50
p_{10}	6.1×10^{-3}
GRS_{12}	3.03
p_{12}	4.2×10^{-4}

Panel B: 73 Anomalies, value-weighted, NYSE deciles

Average $|\alpha|$	0.24
Average $|t|$	1.17
GRS_{51}	1.46
p_{51}	0.02
GRS_{72}	1.67
p_{72}	1.9×10^{-3}

Panel C: 73 Anomalies, equally weighted,
NYSE/AMEX/NASDAQ deciles

Average $|\alpha|$	0.30
Average $|t|$	1.61
GRS_{51}	4.09
p_{51}	9.7×10^{-17}
GRS_{72}	2.48
p_{72}	8.9×10^{-8}

Table 10

Abilities of Models FF-5, q-4, and M-3 to Explain Each Other's Factors

Panel A reports a factor's estimated monthly alpha (in percent) with respect to each of the other models (with White (1980) heteroskedasticity-consistent t-statistics in parentheses). Panel B computes the Gibbons-Ross-Shanken (1989) F-test of whether a given model produces zero alphas for the factors of an alternative model (with p-values in parentheses). The factors whose alphas are tested are those other than a model's market and size factors. The models considered are the five-factor model of Fama and French (2015), denoted FF-5, which includes the factors HML, RMW, and CMA; the four-factor model of Hou, Xue, and Zhang (2015a), denoted q-4, which includes the factors I/A and ROE; and the three-factor mispricing factor model, denoted M-3, which includes the factor UMO. The sample period is from January 1967 through December 2013 (564 months).

	Alpha computed with respect to model		
Factors	FF-5	q-4	M-3
Panel A: Alpha (t-statistic)			
Factors in FF-5			
HML	-	0.04	0.34
		(0.43)	(2.53)
RMW	-	0.04	0.08
		(0.55)	(0.90)
CMA	-	0.02	0.13
		(0.47)	(1.56)
Factors in q-4			
I/A	0.12	-	0.25
	(3.48)		(2.87)
ROE	0.45	-	0.17
	(5.53)		(1.60)
Factor in M-3			
UMO	0.77	0.60	-
	(6.88)	(5.08)	
Panel B: GRS F-statistic (p-value)			
HML, RMW, CMA	-	0.23	2.44
		(0.87)	(0.06)
I/A, ROE	19.06	-	7.89
	(9.8×10^{-9})		(4.1×10^{-4})
UMO	67.40	38.98	-
	(1.5×10^{-15})	(4.9×10^{-10})	

Table 11
Table 11
Abilities of Models M-3 and M-4 to Explain Each Other's Factors

Panel A reports a factor's estimated monthly alpha (in percent) with respect to each of the other models (with White (1980) heteroskedasticity-consistent t-statistics in parentheses). Panel B computes the Gibbons-Ross-Shanken (1989) F-test of whether a given model produces zero alphas for the factors of an alternative model (with p-values in parentheses). The factors whose alphas are tested are those other than a model's market and size factors. The models considered are the three-factor mispricing factor model, denoted M-3, which includes the factor UMO, and the four-factor mispricing-factor model, denoted M-4, which includes the factors $MGMT$ and $PERF$. The sample period is from January 1967 through December 2013 (564 months).

	Alpha computed with respect to model	
Factors	M-3	M-4
Panel A: Alpha (t-statistic)		
Factor in M-3		
UMO	-	0.21
		(4.76)
Factors in M-4		
$MGMT$	0.33	-
	(3.18)	
$PERF$	-0.43	-
	(-3.63)	
Panel B: GRS F-statistic (p-value)		
UMO	-	21.97
		(3.5×10^{-6})
$MGMT$, $PERF$	6.67	-
	(1.4×10^{-3})	

Table 12
Idiosyncratic Volatility Effects in Underpriced versus Overpriced Stocks

This table reports, for each of four models, the monthly percentage alphas on value-weighted portfolios formed by an independent 5×5 sort on the composite mispricing measure—the average ranking across 11 anomalies—and IVOL—the previous month's daily volatility of residuals from the three-factor model of Fama and French (1993). The most overpriced stocks are those in the top 20% of the mispricing measure, while the most underpriced are those in the bottom 20%. Panel A reports alphas with respect to the three-factor model of Fama and French (1993), model FF-3; Panel B reports alphas for the five-factor model of Fama and French (2015), model FF-5; Panel C reports alphas for the four-factor model of Hou, Xing, and Zhang (2015a), model q-4; Panel D reports alphas for the four-factor mispricing-factor model, model M-4. Heteroskedasticity-consistent t-statistics based on White (1980) are shown in parentheses. The sample period is from January 1967 through December 2013 (564 months).

	Highest IVOL	Next 20%	Next 20%	Next 20%	Lowest IVOL	Highest −Lowest
Panel A: FF-3 Alpha						
Most Overpriced	-1.87	-0.92	-0.77	-0.51	-0.23	-1.64
	(-12.04)	(-6.89)	(-5.42)	(-4.13)	(-1.85)	(-8.46)
Most Underpriced	0.45	0.61	0.50	0.35	0.15	0.30
	(2.86)	(4.56)	(4.95)	(4.3)	(2.18)	(1.70)
All Stocks	-0.72	-0.12	-0.02	0.02	0.10	-0.81
	(-6.52)	(-1.61)	(-0.38)	(0.46)	(2.31)	(-6.04)
Panel B: FF-5 Alpha						
Most Overpriced	-1.40	-0.65	-0.55	-0.47	-0.17	-1.23
	(-8.71)	(-4.63)	(-3.85)	(-3.54)	(-1.27)	(-6.15)
Most Underpriced	0.57	0.61	0.39	0.17	-0.07	0.64
	(3.44)	(4.03)	(3.75)	(1.99)	(-1.02)	(3.61)
All Stocks	-0.36	0.01	-0.01	-0.08	-0.03	-0.33
	(-3.55)	(0.19)	(-0.11)	(-1.44)	(-0.70)	(-2.69)
Panel C: q-4 Alpha						
Most Overpriced	-1.25	-0.44	-0.55	-0.45	-0.14	-1.11
	(-7.57)	(-3.11)	(-3.38)	(-2.8)	(-0.85)	(-4.88)
Most Underpriced	0.47	0.59	0.32	0.07	-0.14	0.61
	(2.43)	(3.56)	(2.93)	(0.84)	(-1.65)	(2.96)
All Stocks	-0.28	0.09	0.01	-0.13	-0.05	-0.23
	(-2.40)	(1.13)	(0.13)	(-2.00)	(-1.05)	(-1.58)
Panel D: M-4 Alpha						
Most Overpriced	-0.96	-0.20	-0.16	-0.11	0.08	-1.04
	(-6.33)	(-1.61)	(-1.05)	(-0.88)	(0.57)	(-4.40)
Most Underpriced	0.36	0.28	0.23	0.00	-0.26	0.62
	(2.08)	(1.93)	(2.16)	(-0.05)	(-3.66)	(3.17)
All Stocks	-0.21	0.13	0.08	-0.06	-0.09	-0.12
	(-1.75)	(1.60)	(1.22)	(-0.91)	(-1.80)	(-0.79)

44
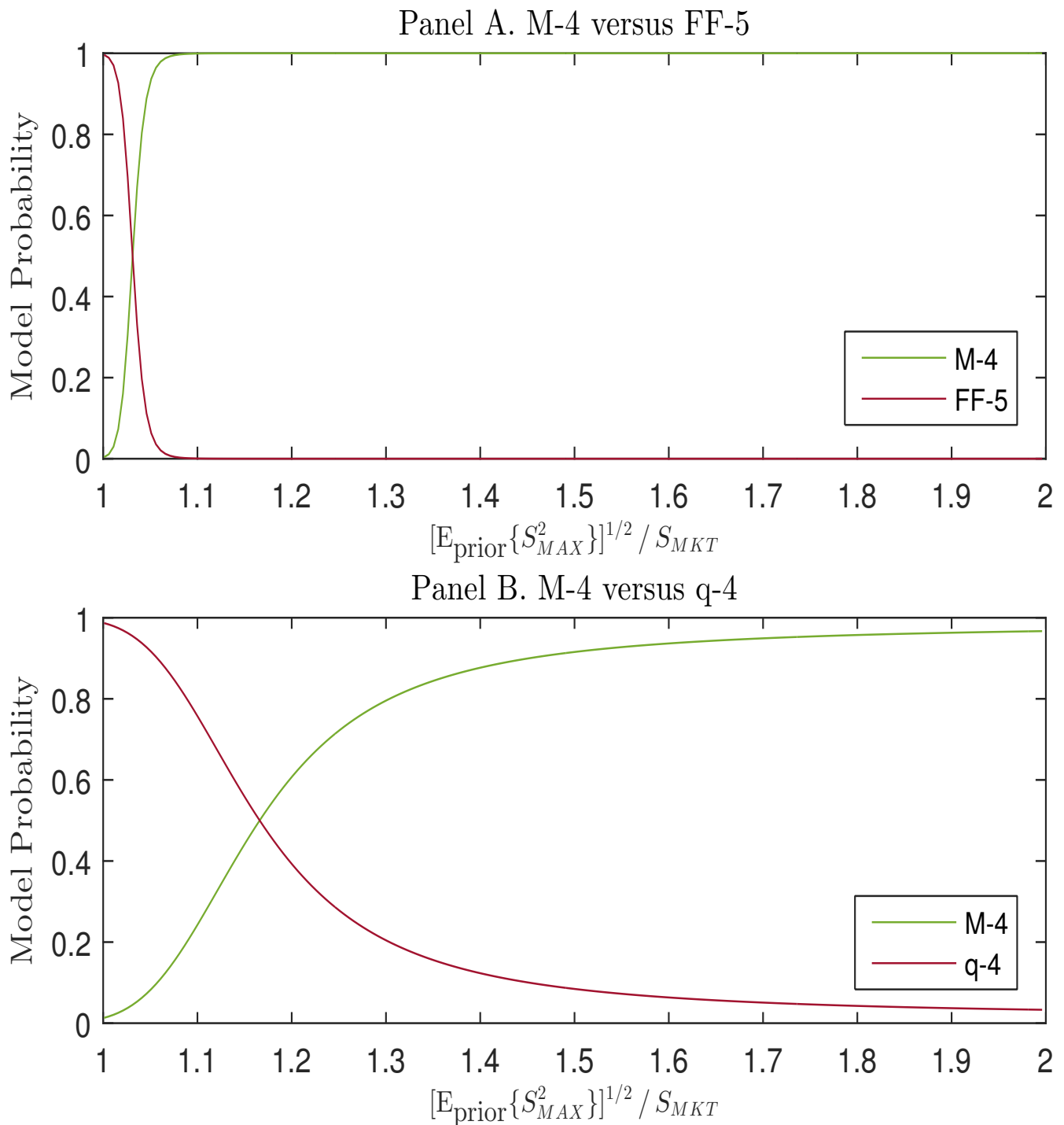
Figure 1. Model probabilities comparing model M-4 to models FF-5 and q-4. The figure displays Bayesian posterior model probabilities for two-way model comparisons. The value on the horizontal axis is the square root of the prior expected maximum Sharpe ratio achievable by combining the model's factors, divided by the observed Sharpe ratio of the market. Prior model probabilities are equal. Panel A compares model M-4 to the five-factor model of Fama and French (2015), model FF-5. Panel B compares M-4 to the four-factor model of Hou, Xing, and Zhang (2015a), model q-4. The sample period is from January 1967 through December 2013 (564 months)
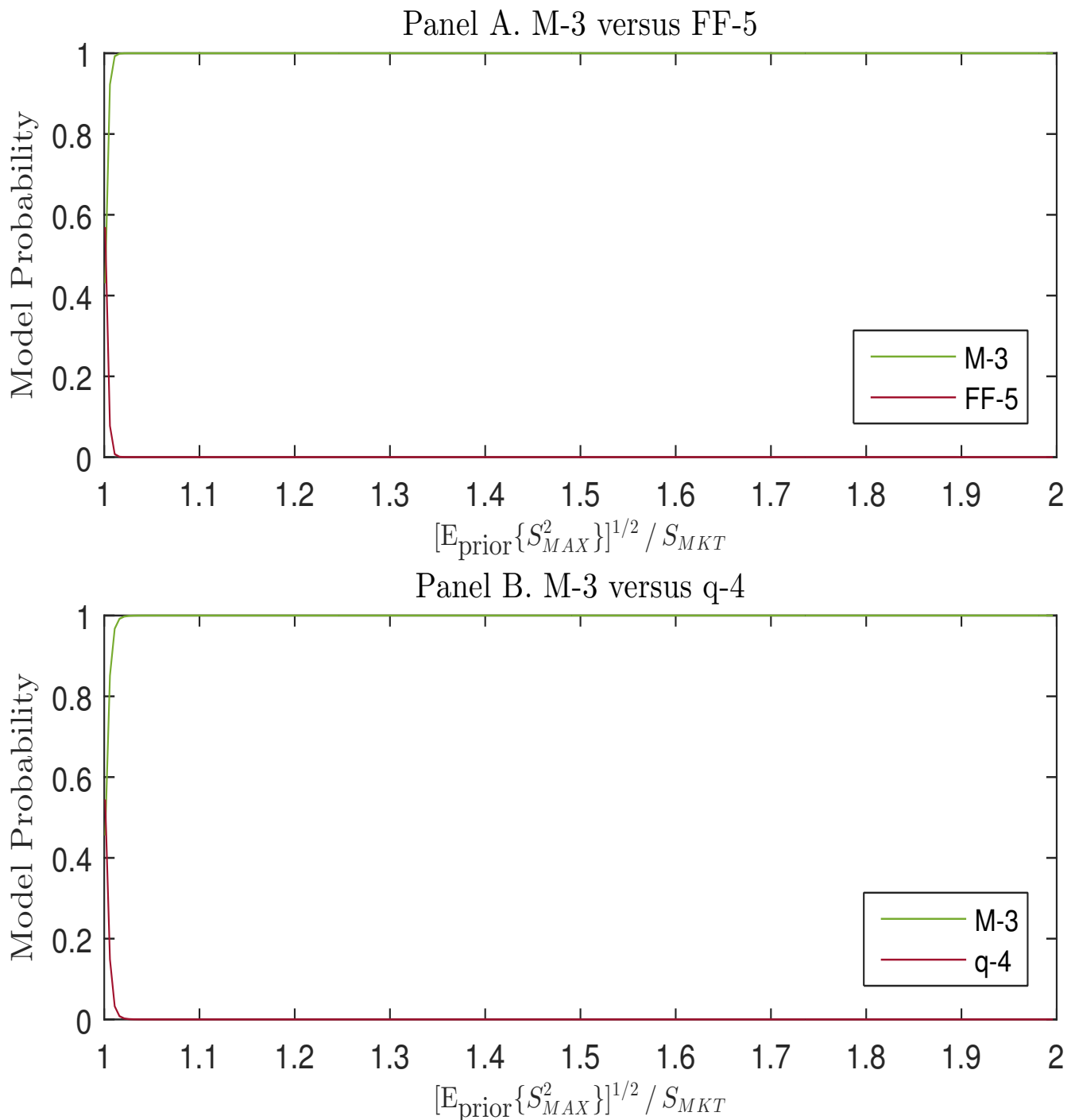
Figure 2. Model probabilities comparing model M-3 to models FF-5 and q-4. The figure displays Bayesian posterior model probabilities for two-way model comparisons. The value on the horizontal axis is the square root of the prior expected maximum Sharpe ratio achievable by combining the model's factors, divided by the observed Sharpe ratio of the market. Prior model probabilities are equal. Panel A compares model M-3 to the five-factor model of Fama and French (2015), model FF-5. Panel B compares M-3 to the four-factor model of Hou, Xing, and Zhang (2015a), model q-4. The sample period is from January 1967 through December 2013 (564 months)

46

APPENDIX

The 11 Anomalies Used to Construct the Mispricing Factors

Below we detail the construction of the anomaly measures used to construct mispricing scores and form anomaly portfolios and mispricing factors. The anomaly measures are computed at the end of each month. We exclude stocks with share prices less than \$5, primarily to avoid microstructure effects, and we use ordinary common shares (CRSP codes 10 and 11). The anomaly portfolios are constructed using NYSE deciles as breakpoints. When constructing the mispricing factors, we require that a stock have non-missing values at the end of month $t-1$ for at least three of the (five or six) anomalies in a cluster in order to be included in that cluster's factor. Also, in order for an anomaly to be included in the mispricing measure for its cluster at the end of month $t-1$, we require at least 30 stocks to have non-missing values for that anomaly. The values computed at the end of month $t-1$ for each anomaly are constructed as follows:

1. *Net Stock Issues:* The stock issuing market has long been viewed as producing an anomaly arising from sentiment-driven mispricing: smart managers issue shares when sentiment-driven traders push prices to overvalued levels. Ritter (1991) and Loughran and Ritter (1995) show that, in post-issue years, equity issuers underperform matching nonissuers with similar characteristics. Motivated by this evidence, Fama and French (2008) show that net stock issues and subsequent returns are negatively correlated. Following Fama and French (2008), we measure net issuance as the annual log change in split-adjusted shares outstanding. Split-adjusted shares equal shares outstanding (Compustat annual item CSHO) times the adjustment factor (Compustat annual item ADJEX_C). The most recent reporting year used is the one that ends (according to item DATADATE) at least four months before the end of month $t-1$. Stocks with negative net issues are assigned to decile 1, and those with zero net issues are assigned to decile 2. The remaining stocks are assigned to the remaining eight deciles using NYSE breakpoints.

2. *Composite Equity Issues:* Daniel and Titman (2006) find that issuers underperform nonissuers using a measure they denote as composite equity issuance, defined as the growth in the firm's total market value of equity minus (i.e., not attributable to) the stock's rate of return. We compute this measure by subtracting the 12-month cumulative stock return from the 12-month growth in equity market capitalization. We lag the quantity four months, to make its timing more coincident with the above measure of net stock issues.

3. *Accruals:* Sloan (1996) shows that firms with high accruals earn abnormally lower average returns than firms with low accruals, and he suggests that investors overestimate the persis-

47

tence of the accrual component of earnings when forming earnings expectations. Following Sloan (1996), we measure total accruals as the annual change in noncash working capital minus depreciation and amortization expense (Compustat annual item DP), divided by average total assets (item AT) for the previous two fiscal years. Noncash working capital is computed as the change in current assets (item ACT) minus the change in cash and short-term investment (item CHE), minus the change in current liabilities (item DLC), plus the change in debt included in current liabilities (item LCT), plus the change in income taxes payable (item TXP). The most recent reporting year used is the one that ends (according to item DATADATE) at least four months before the end of month $t - 1$.

4. *Net Operating Assets:* Hirshleifer, Hou, Teoh, and Zhang (2004) find that net operating assets, defined as the difference on the balance sheet between all operating assets and all operating liabilities, scaled by total assets, is a strong negative predictor of long-run stock returns. The authors suggest that investors with limited attention tend to focus on accounting profitability, neglecting information about cash profitability, in which case net operating assets (equivalently measured as the cumulative difference between operating income and free cash flow) captures such a bias. Following equations equations (4), (5), and (6) of that study, we measure net operation assets as operating assets minus operating liabilities, divided by lagged total assets (Compustat annual item AT). Operating assets equal total assets (item AT) minus cash and short-term investment (item CHE). Operating liabilities equal total assets minus debt included in current liabilities (item DLC), minus long-term debt (item DLTT), minus common equity (item CEQ), minus minority interests (item MIB), minus preferred stocks (item PSTK). (The last two items are zero if missing.) The most recent reporting year used is the one that ends (according to item DATADATE) at least four months before the end of month $t - 1$.

5. *Asset Growth:* Cooper, Gulen, and Schill (2008) find that companies that grow their total assets more earn lower subsequent returns. They suggest that this phenomenon is due to investors' initial overreaction to changes in future business prospects implied by asset expansions. Asset growth is measured as the growth rate of total assets in the previous fiscal year. Following that study, we measure asset growth as the most recent year-over-year annual growth rate of total assets (Compustat annual item AT). The most recent reporting year used is the one that ends (according to item DATADATE) at least four months before the end of month $t - 1$.

6. *Investment-to-Assets:* Titman, Wei, and Xie (2004) and Xing (2008) show that higher past investment predicts abnormally lower future returns. Titman, Wei, and Xie (2004) attribute this anomaly to investors' initial underreaction to overinvestment caused by managers'

empire-building behavior. Here, investment to assets is measured as the annual change in gross property, plant, and equipment, plus the annual change in inventories, scaled by lagged book value of assets. Following the above studies, we compute investment-to-assets as the changes in gross property, plant, and equipment (Compustat annual item PPEGT) plus changes in inventory (item INVT), divided by lagged total assets (item AT). The most recent reporting year used is the one that ends (according to item DATADATE) at least four months before the end of month $t - 1$.

7. *Distress:* Financial distress is often invoked to explain otherwise anomalous patterns in the cross-section of stock returns. However, Campbell, Hilscher, and Szilagyi (2008) find that firms with high failure probability have lower rather than higher subsequent returns. The authors suggest that their finding is a challenge to standard models of rational asset pricing. Failure probability is estimated with a dynamic logit model that uses several equity market variables, such as stock price, book-to-market, stock volatility, size relative to the S&P 500, and cumulative excess return relative to the S&P 500. Specifically, using the above study's equations (2) and (3) along with its Table IV (12-month column), we compute the distress anomaly measure—failure probability—as

$$
\begin{aligned}
\pi \;=\; & -20.26\,NIMTAAVG + 1.42\,TLMTA - 7.13\,EXRETAVG + 1.41\,SIGMA \\
& -0.045\,RSIZE - 2.13\,CASHMTA + 0.075\,MB - 0.058\,PRICE - 9.16,
\end{aligned}
$$

where

$$
\begin{aligned}
NIMTAAVG_{t-1,t-12} \;&=\; \frac{1-\phi^3}{1-\phi^{12}}(NIMTA_{t-1,t-3} + \cdots + \phi^9 NIMTA_{t-10,t-12}) \\
EXRETAVG_{t-1,t-12} \;&=\; \frac{1-\phi}{1-\phi^{12}}(EXRET_{t-1} + \cdots + \phi^{11} EXRET_{t-12}),
\end{aligned}
$$

and $\phi = 2^{-1/3}$. *NIMTA* is net income (Compustat quarterly item NIQ) divided by firm scale, where the latter is computed as the sum of total liabilities (item LTQ) and market equity capitalization (data from CRSP). $EXRET_s$ is the stock's monthly log return in month s minus the log return on the S&P500 index. Missing values for *NIMTA* and *EXRET* are replaced by those quantities' cross-sectional means. *TLMTA* equals total liabilities divided by firm scale. *SIGMA* is the stock's daily standard deviation for the most recent three months, expressed on an annualized basis. At least five non-zero daily returns are required. *RSIZE* is the log of the ratio of the stock's market capitalization to that of the S&P500 index. *CASHMTA* equals cash and short-term investment (item CHEQ) divided by firm scale. *MB* is the market-to-book ratio. Following Campbell, Hilscher, and Szilagyi (2008), we increase book equity by 10% of the difference between market equity and book equity. If the resulting value of book equity is negative, then book equity is set to $1. *PRICE* is the log of the share

price, truncated above at \$15. All explanatory variables except *PRICE* are winsorized above and below at the 5% level in the cross section. CRSP based variables, *EXRETAVG*, *SIGMA*, *RSIZE* and *PRICE* are for month $t-1$. *NIQ* is for the most recent quarter for which the reporting date provided by Compustat (item RDQ) precedes the end of month $t-1$, whereas the items requiring information from the balance sheet (*LTQ*, *CHEQ* and *MB*) are for the prior quarter.

8. *O-score:* This distress measure, from Ohlson (1980), predicts returns in a manner similar to the measure above. It is the probability of bankruptcy estimated in a static model using accounting variables. Following Ohlson (1980), we construct it as:

$$
\begin{aligned}
O &= -0.407\,SIZE + 6.03\,TLTA - 1.43\,WCTA + 0.076\,CLCA - 1.72\,OENEG \\
&= -2.37\,NITA - 1.83\,FUTL + 0.285\,INTWO - 0.521\,CHIN - 1.32,
\end{aligned}
$$

where *SIZE* is the log of total assets (Compustat annual item AT), *TLTA* is the book value of debt (item DLC plus item DLTT) divided by total assets, *WCTA* is working capital (item ACT minus item LCT) divided by total assets, *CLCA* is current liabilities (item LCT) divided by current assets (item ACT), *ONEEG* is 1 if total liabilities (item LT) exceed total assets and is zero otherwise, *NITA* is net income (item NI) divided by total assets, *FUTL* is funds provided by operations (item PI) divided by total liabilities, *INTWO* is equal to 1 if net income (item NI) is negative for the last 2 years and zero otherwise, *CHIN* is $(NI_j - NI_{j-1})/(|NI_j| + |NI_{j-1}|)$, in which NI_j is the income (item NI) for year j, which is the most recent reporting year that ends (according to item DATADATE) at least four months before the end of month $t-1$.

9. *Momentum:* The momentum effect, discovered by Jegadeesh and Titman (1993), is one of the most robust anomalies in asset pricing. It refers to the phenomenon whereby high (low) past recent recent returns forecast high (low) future returns. The momentum ranking at the end of month t-1 uses the cumulative returns from month t-12 to month t-2. This is the choice of ranking variable used by Carhart (1997) to construct the widely used momentum factor.

10. *Gross Profitability Premium:* Novy-Marx (2013) shows that sorting on the ratio of gross profit to assets creates abnormal benchmark-adjusted returns, with more profitable firms having higher returns than less profitable ones. He argues that gross profit is the cleanest accounting measure of true economic profitability. The farther down the income statement one goes, the more polluted profitability measures become, and the less related they are to true economic profitability. Following that study, we measure gross profitability as total revenue (Compustat annual item REVT) minus the cost of goods sold (item COGS), divided

by current total assets (item AT). The most recent reporting year used is the one that ends (according to item DATADATE) at least four months before the end of month $t-1$.

11. *Return on Assets:* Fama and French (2006) find that more profitable firms have higher expected returns than less profitable firms. Chen, Novy-Marx, and Zhang (2010) show that firms with higher past return on assets earn abnormally higher subsequent returns. Return on assets is measured as the ratio of quarterly earnings to last quarter's assets. Wang and Yu (2013) find that the anomaly exists primarily among firms with high arbitrage costs and high information uncertainty, suggesting that mispricing is a culprit. Following Chen, Novy-Marx, and Zhang (2010), we compute return on assets as income before extraordinary items (Compustat quarterly item IBQ) divided by the previous quarter's total assets (item ATQ). Income is for the most recent quarter for which the reporting date provided by Compustat (item RDQ) precedes the end of month $t-1$, and assets are for the prior quarter.

The Larger Set of 73 Anomalies

We list here the 73 anomalies for which long-short returns were generously provided by the authors of Hou, Xue, and Zhang (2015a, 2015b). See those studies for detailed explanations and references. The anomalies are grouped into six categories:

1. *Momentum*
   SUE-1: Earnings surprise (1-month holding period)
   SUE-6: Earnings surprise (6-month holding period)
   Abr-1: Cumulative abnormal stock returns around earnings announcements
         (1-month holding period)
   Abr-6: Cumulative abnormal stock returns around earnings announcements
         (6-month holding period)
   RE-1: Revisions in analysts earnings forecasts (1-month holding period)
   RE-6: Revisions in analysts earnings forecasts (6-month holding period)
   R6-1: Price momentum (6-month prior returns, 1-month holding period)
   R6-6: Price momentum (6-month prior returns, 6-month holding period)
   R11-1: Price momentum, (11-month prior returns, 1-month holding period)
   I-Mom: Industry momentum

2. *Value-versus-growth*
   B/M: Book-to-market equity
   A/ME: Market Leverage
   Rev: Reversal
   E/P: Earnings-to-price

EF/P: Analysts earnings forecasts-to-price

CF/P: Cash flow-to-price

D/P: Dividend yield

O/P: Payout yield

NO/P: Net payout yield

SG: Sales growth

LTG: Long-term growth forecasts of analysts

Dur: Equity duration

3. *Investment*

ACI: Abnormal corporate investment

I/A: Investment-to-assets

NOA: Net operating assets

ΔPI/A: Changes in property, plant, and equipment plus changes in inventory scaled by assets

IG: Investment growth

NSI: Net stock issues

CEI: Composite issuance

NXF: Net external financing

IvG: Inventory growth

IvC: Inventory changes

OA: Operating accruals

TA: Total accruals

POA: Percent operating accruals

PTA: Percent total accruals

4. *Profitability*

ROE: Return on equity

ROA: Return on assets

RNA: Return on net operating assets

PM: Profit margin

ATO: Asset turnover

CTO: Capital turnover

GP/A: Gross profits-to-assets

F: *F*-score

TES: Tax expense surprise

TI/BI: Taxable income-to-book income

RS: Revenue surprise

NEI: Number of consecutive quarters with earnings increases

FP: Failure probability

O: O-score

5. *Intangibles*

OC/A: Organizational capital-to-assets

BC/A: Brand capital-to-assets

Ad/M: Advertisement expense-to-market

RD/S: R&D-to-sales

RD/M: R&D-to-market

RC/A: R&D capital-to-assets

H/N: Hiring rate

OL: Operating leverage

G: Corporate governance

AccQ: Accrual quality

6. *Trading frictions*

ME: Market equity

Ivol: Idiosyncratic volatility

Tvol: Total volatility

Svol: Systematic volatility

MDR: Maximum daily return

β: Market beta

D-β: Dimsons beta

S-Rev: Short-term reversal

Disp: Dispersion of analysts earnings forecasts

Turn: Share turnover

1/P: 1/share price

Dvol: Dollar trading volume

Illiq: Illiquidity as absolute return-to-volume

Anomalies Eliminated in Producing Table 6

We eliminate the five anomalies from the above set of 73 that are most highly correlated with each factor other than the market and size factors, in both models q-4 and M-4. The two relevant factors in model q-4 are investment (I/A) and profitability (ROE), while in model M-4 they are the mispricing factors, $MGMT$ and $PERF$. In Panel A of Table 6, the anomalies that each factor eliminates are as follows:

53

I/A: CEI, H/N, I/A, LTG, SG

ROE: Disp, FP, 1/P, ROA, ROE

MGMT: CEI, LTG, NO/P, O/P, Turn

PERF: FP, R11-1, R6-1, R6-6, ROA

In Panel B of Table 6, the anomalies eliminated:

I/A: A/ME, B/M, H/N, I/A, SG

ROE: Disp, FP, NEI, ROA, ROE

MGMT: CEI, LTG, NO/P, NXF, O/P

PERF: FP, 1/P, R11-1, R6-1, R6-6

REFERENCES

Ahn, Dong-Hyun, Jennifer Conrad, and Robert F. Dittmar, 2009, Basis assets, *Review of Financial Studies* 22, 5133–5174.

Ali, Ashiq, Lee-Seok Hwang, and Mark A. Trombley, 2003, Arbitrage risk and the book-to-market anomaly, *Journal of Financial Economics* 69, 355–373.

Amihud, Yakov and Haim Mendelson, 1989, The effects of beta, bid-ask spread, residual risk, and size on stock returns, *Journal of Finance* 44, 479–486.

Ang, Andrew, Robert J. Hodrick, Yuhang Xing, and Xiaoyan Zhang, 2006, The cross-section of volatility and expected returns, *Journal of Finance* 51, 259–299.

Asness, Cliff, Andrea Frazzini, Ronen Israel, Tobias Moskowitz, and Lasse H. Pedersen, 2015, Size matters, if you control your junk, Working paper, AQR Capital Management, University of Chicago, and Copenhagen School of Business.

Avramov, Doron, Tarun Chordia, Gergana Jostova, and Alexander Philipov, 2013, Anomalies and financial distress, *Journal of Financial Economics* 108, 139–159.

Baker, Malcolm, and Jeffrey Wurgler, 2006, Investor sentiment and the cross-section of stock returns, *Journal of Finance* 61, 1645–1680.

Baker, Malcolm, Jeffrey Wurgler, and Yu Yuan, 2012, Global, local, and contagious investor sentiment, *Journal of Financial Economics* 104, 272–287.

Banz, Rolf W., 1981, The relationship between return and market value of common stocks, *Journal of Financial Economics* 9, 3–18.

Barillas, Francisco, and Jay Shanken, 2015a, Comparing asset pricing models, Working paper, Emory University.

Barillas, Francisco, and Jay Shanken, 2015b, Which alpha?, Working paper, Emory University.

Ben-David, Itzhak, and Darren Roulstone, 2010, Idiosyncratic risk and corporate transac-

tions, Working paper, Ohio State University.

Campbell, John Y., Jens Hilscher, and Jan Szilagyi, 2008, In search of distress risk, *Journal of Finance* 63, 2899–2939.

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

Chan, K.C., Nai-fu Chen, and David A. Hsieh, 1985, An exploratory investigation of the firm size effect, *Journal of Financial Economics* 14, 451–471.

Chen, Long, Robert Novy-Marx, and Lu Zhang, 2010, An alternative three-factor model, Working paper, CKGSB, University of Rochester, and Ohio State University.

Chen, Shuping, Mark L. DeFond, and Chul W. Park, 2002, Voluntary disclosure of balance sheet information in quarterly earnings announcements, *Journal of Accounting and Economics* 33, 229–251.

Cooper, Michael J., Huseyin Gulen, and Michael J. Schill, 2008, Asset growth and the cross-section of stock returns, *Journal of Finance* 63, 1609–1652.

Daniel, Kent D., and Sheridan Titman, 1997, Evidence on the characteristics of cross sectional variation in stock returns, *Journal of Finance* 52, 1–33.

Daniel, Kent D., and Sheridan Titman, 2006, Market reactions to tangible and intangible information, *Journal of Finance* 61, 1605–1643.

DeLong, Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert Waldmann, 1990, Noise trader risk in financial markets, *Journal of Political Economy* 90, 703–738.

Drechsler, Itamar and Qingyi Freda Drechsler, 2014, The shorting premium and asset pricing anomalies, Working paper, New York University and Wharton Research Data Services.

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

Fama, Eugene F., and Kenneth French, 2006, Profitability, investment, and average returns, *Journal of Financial Economics* 82, 491–518.

Fama, Eugene F., and Kenneth French, 2008, Dissecting anomalies, *Journal of Finance* 63, 1653–1678.

Fama, Eugene F., and Kenneth French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 16, 1–22.

Fama, Eugene F., and Kenneth French, 2016, Dissecting anomalies with a five-factor model, *Review of Financial Studies* 29, 69–103.

Gibbons, Michael R., Stephen A. Ross, and Jay Shanken, 1989, A test of the efficiency of a given portfolio, *Econometrica* 57, 1121–1152.

Greenwood, Robin and Samuel G. Hanson, 2012, Share issuance and factor timing, *Journal of Finance* 67, 761–798.

Hirshleifer, David, Kewei Hou, Siew Hong Teoh, and Yinglei Zhang, 2004, Do investors overvalue firms with bloated balance sheets? *Journal of Accounting and Economics* 38, 297–331.

Hirsheifer, David, Siew Hong Toeh, and Jeff Jiewei Yu, 2011, Short arbitrage, return asymmetry, and the accrual anomaly, *Review of Financial Studies* 24, 2429–2461.

Hirshleifer, David, and Danling Jiang, 2010, A financing-based misvaluation factor and the cross-section of expected returns, *Review of Financial Studies* 23, 3401–3436.

Hou, Kewei, Chen Xue, and Lu Zhang, 2015a, Digesting anomalies: An investment approach, *Review of Financial Studies* 28, 650–705.

Hou, Kewei, Chen Xue, and Lu Zhang, 2015b, A comparison of new factor models, Working paper, Ohio State University, University of Cincinnati, and Ohio State University.

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for market efficiency, *Journal of Finance* 48, 65–91.

Jin, Lucy, 2013, Idiosyncratic volatility, arbitrage risk, and anomaly returns, PhD thesis, University of Pennsylvania.

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh 2015, Interpreting factor models, Working paper, University of Michigan, University of Michigan, and University of Maryland.

Lakonishok, Josef, Andrei Shleifer, and Robert W. Vishny, 1994, Contrarian investment, extrapolation, and risk, *Journal of Finance* 49, 1541–1578.

Larrain, Borja, and Felipe Varas, 2013, Equity issues and return volatility, *Review of Finance* 17, 767–808.

Lewellen, Jonathan, Stefan Nagel, and Jay Shanken, 2010, A skeptical appraisal of asset pricing tests, *Journal of Financial Economics* 96, 175–194.

Li, Dongmei, and Lu Zhang, 2010, Does Q-theory with investment frictions explain anomalies in the cross-section of returns? *Journal of Financial Economics* 98, 297–314.

Li, Xi and Rodney N. Sullivan, 2011, The limits to arbitrage revisited: The accrual and asset growth anomalies, *Financial Analysts Journal* 67(4), 50–66.

Lintner, John, 1965, Security prices, risk, and maximal gains from diversification, *Journal of Finance* 20, 587–615.

Loughran, Tim, and Jay R. Ritter, 1995, The new issues puzzle, *Journal of Finance* 50, 23–51.

MacKinlay, A. Craig, and ĽubošPástor, 2000, Asset pricing models: Implications for expected returns and portfolio selection, *Review of Financial Studies* 13, 883–916.

Mashruwala, Christina, Shivaram Rajgopal, and Terry Shevlin, 2006, Why is the accrual anomaly not arbitraged away? The role of idiosyncratic risk and transaction costs, *Journal of Accounting and Economics* 42,3–33.

McLean, R. David, 2010, Idiosyncratic risk, long-term reversal, and momentum, *Journal of Financial and Quantitative Analysis* 45, 883–906.

McLean, R. David, and Jeffrey Pontiff, 2015, Does academic research destroy stock return predictability?, forthcoming in *Journal of Finance*.

Mendenhall, Richard R., 2004, Risk and post-earnings-announcement drift, *Journal of Business* 77, 875-894.

Nagel, Stefan, 2005, Short sales, institutional investors and the cross-section of stock returns, *Journal of Financial Economics* 78, 277–309.

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

Ohlson, James A., 1980, Financial ratios and the probabilistic prediction of bankruptcy, *Journal of Accounting Research* 18, 109–131.

Pástor, Ľuboš, and Robert F. Stambaugh, 2000, Comparing asset pricing models: an investment perspective, *Journal of Financial Economics* 56, 335-381.

Pástor, Ľuboš, and Robert F. Stambaugh, 2003, Liquidity risk and expected stock returns, *Journal of Political Economy* 111, 642–685.

Pincus, Morton, Shivaram Rajgopal, and Mohan Venkatachalam, 2007, The Accrual Anomaly: International Evidence, *The Accounting Review* 82, 169–203.

Pontiff, Jeffrey, 1996, Costly arbitrage: Evidence from closed-end funds, *Quarterly Journal of Economics* 111, 1135–1152.

Pontiff, Jeffrey, 2006, Costly arbitrage and the myth of idiosyncratic risk, *Journal of Accounting and Economics* 42, 35–52.

Ritter, Jay R., 1991, The long-run performance of initial public offerings, *Journal of Finance* 46, 3–27.

Scruggs, John T., 2007, Noise trader risk: Evidence from the Siamese twins, *Journal of Financial Markets* 10, 76–105.

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

Shleifer, Andrei and Robert Vishny, 1997, The limits of arbitrage, *Journal of Finance*, 52, 35–55.

Sloan, Richard G., 1996, Do stock prices fully reflect information in accruals and cash flows about future earnings? *The Accounting Review* 71, 289–315.

Stambaugh, R.F., 1997, Analyzing investments whose histories differ in length. Journal of Financial Economics 45, 285–331.

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

Stambaugh, Robert F., Jianfeng Yu, and Yu Yuan, 2014, The long of it: Odds that investor sentiment spuriously predicts anomaly returns, *Journal of Financial Economics* 114, 613–619.

Stambaugh, Robert F., Jianfeng Yu, and Yu Yuan, 2015, Arbitrage asymmetry and the idiosyncratic volatility puzzle, *Journal of Finance* 70, 1903–1948.

Titman, Sheridan, K. C. John Wei, and Feixue Xie, 2004, Capital investments and stock returns, *Journal of Financial and Quantitative Analysis* 39, 677–700.

Wang, Huijun, and Jianfeng Yu, 2013, Dissecting the profitability premium, Working paper, University of Minnesota.

Ward, Joe H., 1963, Hierarchical grouping to optimize an objective function, *Journal of the American Statistical Association* 58, 236–244.

White, Halbert, 1980, A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, *Econometrica* 48, 817–838.

Wurgler, Jeffrey, and Ekaterina Zhuravskaya, 2002, Does arbitrage flatten demand curves for stocks? *Journal of Business* 75, 583–608.

Xing, Yuhang, 2008, Interpreting the value effect through the Q-theory: An empirical investigation, *Review of Financial Studies* 21, 1767–1795.

Zhang, Frank, 2006, Information uncertainty and stock returns, *Journal of Finance* 61, 105–136.

Online Appendix for "Mispricing Factors"

by*

Robert F. Stambaugh and Yu Yuan

This appendix contains the robustness results discussed in subsection 4.4 as well as the results for the models adding a momentum or liquidity factor, as discussed at the end of subsection 4.1. Tables B1 through B15 recompute Tables 5, 6, and 7 with the following modifications to the original procedure:

1. Reconstruct the M-4 factors using mispricing-measure breakpoints of 30% and 70% instead of 20% and 80% (Tables B1, B5, and B9).

2. Reconstruct the M-4 factors using mispricing-measure breakpoints for the NYSE instead of NYSE/AMEX/NASDAQ (Tables B2, B6, and B10).

3. Reconstruct the M-4 SMB factor by averaging across the three mispricing categories instead of using only stocks in the middle category (Table B3, B7, and B11).

4. Incorporate all of the above simultaneously (Tables B4, B8, and B12).

5. Reconstruct annual-data factors in models FF-3, FF-5, and q-4 with a minimum gap of 4 months after the end of the fiscal year instead of a gap of 6 to 18 months (Tables B13–B15).

Tables B16 through B25 recompute Tables 5, 7, 9, 10 and 11 in each of two subperiods: 1967–1990 (24 years) and 1991–2013 (23 years). Table B26 reports results when models MOM-4 and LIQ-4 are added to Table 5. Figures B1 through B3 report rolling subperiod comparisons of anomalies' average absolute alphas obtained under model M-4 versus models FF-5 and q-4.

Table B1
Summary Measures of Models' Abilities to Explain Anomalies,
Factors Constructed Using 30% and 70% Breakpoints

This table recomputes Table 5 with the factors constructed using mispricing-measure breakpoints of 30% and 70% instead of 20% and 80%. The table reports measures that summarize the degree to which anomalies produce alpha under four different factor models: the three-factor model of Fama and French (1993), denoted FF-3; the five-factor model of Fama and French (2015), denoted FF-5; the four-factor model of Hou, Xue, and Zhang (2015a), denoted q-4; and the four-factor mispricing-factor model introduced in this study, denoted M-4. Also reported are measures based on the average unadjusted return spreads (the alphas in a model with no factors). For each model, the table reports the average absolute alpha, average absolute t-statistic, the F-statistic and associated p-value for the "GRS" test of Gibbons, Ross, and Shanken (1989), and the number of anomalies for which the model produces the smallest absolute alpha among the models being compared in the table. In Panel A, two versions of the GRS test are reported. Data for the distress anomaly begin in October 1974, and data for the return-on-assets anomaly begin in November 1971. The other ten anomalies begin at the start of the sample period, which is from January 1967 through December 2013 (564 months). GRS_{10} tests whether the ten alphas for the full-sample anomalies equal zero, while GRS_{12} uses all 12 anomalies and begins the sample in October 1974. Panels B and C also report two versions of the GRS test: GRS_{51} uses 51 anomalies whose data are available by January 1967, while GRS_{72} use those anomalies plus 21 others whose data are available by February 1986.

Measure	Unadjusted	FF-3	FF-5	q-4	M-4
Panel A: 12 Anomalies, value-weighted, NYSE deciles					
Average $|\alpha|$	0.52	0.67	0.45	0.34	0.16
Average $|t|$	3.14	4.44	2.93	2.34	1.18
GRS_{10}	6.89	10.10	6.71	5.99	1.84
p_{10}	3.4×10^{-10}	1.1×10^{-15}	6.9×10^{-10}	1.2×10^{-8}	0.05
GRS_{12}	6.16	7.71	4.17	3.95	1.73
p_{12}	4.5×10^{-10}	4.2×10^{-13}	3.1×10^{-6}	8.3×10^{-6}	0.06
Number of $min|\alpha|$	-	0	1	2	9
Panel B: 73 Anomalies, value-weighted, NYSE deciles					
Average $|\alpha|$	0.39	0.44	0.30	0.20	0.18
Average $|t|$	2.14	2.74	1.77	1.15	0.99
GRS_{51}	2.74	2.60	1.91	1.68	1.34
p_{51}	9.3×10^{-9}	6.5×10^{-8}	2.9×10^{-4}	3.2×10^{-3}	0.06
GRS_{72}	2.23	2.10	1.79	1.78	1.60
p_{72}	2.2×10^{-6}	1.3×10^{-5}	5.3×10^{-4}	5.8×10^{-4}	4.1×10^{-3}
Number of $min|\alpha|$	-	10	10	23	30
Panel C: 73 Anomalies, equally weighted, NYSE/AMEX/NASDAQ deciles					
Average $|\alpha|$	0.50	0.53	0.35	0.23	0.21
Average $|t|$	3.02	3.72	2.41	1.44	1.30
GRS_{51}	5.85	6.31	5.15	4.19	4.11
p_{51}	2.7×10^{-27}	6.2×10^{-30}	4.3×10^{-23}	2.3×10^{-17}	6.7×10^{-17}
GRS_{72}	2.95	3.25	2.64	2.41	2.45
p_{72}	1.4×10^{-10}	3.0×10^{-12}	1.1×10^{-8}	2.3×10^{-7}	1.3×10^{-7}
Number of $min|\alpha|$	-	7	13	23	30

2

Summary Measures of Models' Abilities to Explain Anomalies,
Factors Constructed Using NYSE Breakpoints

This table recomputes Table 5 with the factors constructed using mispricing-measure breakpoints for the NYSE instead of NYSE/AMEX/NASDAQ. The table reports measures that summarize the degree to which anomalies produce alpha under four different factor models: the three-factor model of Fama and French (1993), denoted FF-3; the five-factor model of Fama and French (2015), denoted FF-5; the four-factor model of Hou, Xue, and Zhang (2015a), denoted q-4; and the four-factor mispricing-factor model introduced in this study, denoted M-4. Also reported are measures based on the average unadjusted return spreads (the alphas in a model with no factors). For each model, the table reports the average absolute alpha, average absolute t-statistic, the F-statistic and associated p-value for the "GRS" test of Gibbons, Ross, and Shanken (1989), and the number of anomalies for which the model produces the smallest absolute alpha among the models being compared in the table. In Panel A, two versions of the GRS test are reported. Data for the distress anomaly begin in October 1974, and data for the return-on-assets anomaly begin in November 1971. The other ten anomalies begin at the start of the sample period, which is from January 1967 through December 2013 (564 months). GRS_{10} tests whether the ten alphas for the full-sample anomalies equal zero, while GRS_{12} uses all 12 anomalies and begins the sample in October 1974. Panels B and C also report two versions of the GRS test: GRS_{51} uses 51 anomalies whose data are available by January 1967, while GRS_{72} use those anomalies plus 21 others whose data are available by February 1986.

Measure	Unadjusted	FF-3	FF-5	q-4	M-4
Panel A: 12 Anomalies, value-weighted, NYSE deciles					
Average $\|\alpha\|$	0.52	0.67	0.45	0.34	0.15
Average $\|t\|$	3.14	4.44	2.93	2.34	1.13
GRS_{10}	6.89	10.10	6.71	5.99	1.80
p_{10}	3.4×10^{-10}	1.1×10^{-15}	6.9×10^{-10}	1.2×10^{-8}	0.06
GRS_{12}	6.16	7.71	4.17	3.95	1.81
p_{12}	4.5×10^{-10}	4.2×10^{-13}	3.1×10^{-6}	8.3×10^{-6}	0.04
Number of $min\|\alpha\|$	-	0	1	2	9
Panel B: 73 Anomalies, value-weighted, NYSE deciles					
Average $\|\alpha\|$	0.39	0.44	0.30	0.20	0.18
Average $\|t\|$	2.14	2.74	1.77	1.15	0.98
GRS_{51}	2.74	2.60	1.91	1.68	1.24
p_{51}	9.3×10^{-9}	6.5×10^{-8}	2.9×10^{-4}	3.2×10^{-3}	0.13
GRS_{72}	2.23	2.10	1.79	1.78	1.55
p_{72}	2.2×10^{-6}	1.3×10^{-5}	5.3×10^{-4}	5.8×10^{-4}	7.3×10^{-3}
Number of $min\|\alpha\|$	-	10	11	20	32
Panel C: 73 Anomalies, equally weighted, NYSE/AMEX/NASDAQ deciles					
Average $\|\alpha\|$	0.50	0.53	0.35	0.23	0.21
Average $\|t\|$	3.02	3.72	2.41	1.44	1.32
GRS_{51}	5.85	6.31	5.15	4.19	4.06
p_{51}	2.7×10^{-27}	6.2×10^{-30}	4.3×10^{-23}	2.3×10^{-17}	1.4×10^{-16}
GRS_{72}	2.95	3.25	2.64	2.41	2.59
p_{72}	1.4×10^{-10}	3.0×10^{-12}	1.1×10^{-8}	2.3×10^{-7}	2.2×10^{-8}
Number of $min\|\alpha\|$	-	7	10	25	31

This table recomputes Table 5 with the SMB factor constructed by averaging across the three mispricing categories instead of using only stocks in the middle category. The table reports measures that summarize the degree to which anomalies produce alpha under four different factor models: the three-factor model of Fama and French (1993), denoted FF-3; the five-factor model of Fama and French (2015), denoted FF-5; the four-factor model of Hou, Xue, and Zhang (2015a), denoted q-4; and the four-factor mispricing-factor model introduced in this study, denoted M-4. Also reported are measures based on the average unadjusted return spreads (the alphas in a model with no factors). For each model, the table reports the average absolute alpha, average absolute t-statistic, the F-statistic and associated p-value for the "GRS" test of Gibbons, Ross, and Shanken (1989), and the number of anomalies for which the model produces the smallest absolute alpha among the models being compared in the table. In Panel A, two versions of the GRS test are reported. Data for the distress anomaly begin in October 1974, and data for the return-on-assets anomaly begin in November 1971. The other ten anomalies begin at the start of the sample period, which is from January 1967 through December 2013 (564 months). GRS_{10} tests whether the ten alphas for the full-sample anomalies equal zero, while GRS_{12} uses all 12 anomalies and begins the sample in October 1974. Panels B and C also report two versions of the GRS test: GRS_{51} uses 51 anomalies whose data are available by January 1967, while GRS_{72} use those anomalies plus 21 others whose data are available by February 1986.

Measure	Unadjusted	FF-3	FF-5	q-4	M-4
Panel A: 12 Anomalies, value-weighted, NYSE deciles					
Average $|\alpha|$	0.52	0.67	0.45	0.34	0.17
Average $|t|$	3.14	4.44	2.93	2.34	1.23
GRS_{10}	6.89	10.10	6.71	5.99	1.76
p_{10}	$3.4{\times}10^{-10}$	$1.1{\times}10^{-15}$	$6.9{\times}10^{-10}$	$1.2{\times}10^{-8}$	0.06
GRS_{12}	6.16	7.71	4.17	3.95	1.71
p_{12}	$4.5{\times}10^{-10}$	$4.2{\times}10^{-13}$	$3.1{\times}10^{-6}$	$8.3{\times}10^{-6}$	0.06
Number of $min|\alpha|$	-	0	1	2	9
Panel B: 73 Anomalies, value-weighted, NYSE deciles					
Average $|\alpha|$	0.39	0.44	0.30	0.20	0.19
Average $|t|$	2.14	2.74	1.77	1.15	1.06
GRS_{51}	2.74	2.60	1.91	1.68	1.35
p_{51}	$9.3{\times}10^{-9}$	$6.5{\times}10^{-8}$	$2.9{\times}10^{-4}$	$3.2{\times}10^{-3}$	0.06
GRS_{72}	2.23	2.10	1.79	1.78	1.61
p_{72}	$2.2{\times}10^{-6}$	$1.3{\times}10^{-5}$	$5.3{\times}10^{-4}$	$5.8{\times}10^{-4}$	$3.9{\times}10^{-3}$
Number of $min|\alpha|$	-	9	12	24	28
Panel C: 73 Anomalies, equally weighted, NYSE/AMEX/NASDAQ deciles					
Average $|\alpha|$	0.50	0.53	0.35	0.23	0.21
Average $|t|$	3.02	3.72	2.41	1.44	1.34
GRS_{51}	5.85	6.31	5.15	4.19	4.04
p_{51}	$2.7{\times}10^{-27}$	$6.2{\times}10^{-30}$	$4.3{\times}10^{-23}$	$2.3{\times}10^{-17}$	$1.8{\times}10^{-16}$
GRS_{72}	2.95	3.25	2.64	2.41	2.38
p_{72}	$1.4{\times}10^{-10}$	$3.0{\times}10^{-12}$	$1.1{\times}10^{-8}$	$2.3{\times}10^{-7}$	$3.2{\times}10^{-7}$
Number of $min|\alpha|$	-	8	13	21	31

4

Summary Measures of Models' Abilities to Explain Anomalies,
Factors Constructed Using All Modifications in Tables B1–B3

This table recomputes Table 5 with the factors constructed using mispricing-measure breakpoints of 30% and 70% for the NYSE instead of 20% and 80% for NYSE/AMEX/NASDAQ. Also, the SMB factor is constructed by averaging across the three mispricing categories instead of using only stocks in the middle category. The table reports measures that summarize the degree to which anomalies produce alpha under four different factor models: the three-factor model of Fama and French (1993), denoted FF-3; the five-factor model of Fama and French (2015), denoted FF-5; the four-factor model of Hou, Xue, and Zhang (2015a), denoted q-4; and the four-factor mispricing-factor model introduced in this study, denoted M-4. Also reported are measures based on the average unadjusted return spreads (the alphas in a model with no factors). For each model, the table reports the average absolute alpha, average absolute t-statistic, the F-statistic and associated p-value for the "GRS" test of Gibbons, Ross, and Shanken (1989), and the number of anomalies for which the model produces the smallest absolute alpha among the models being compared in the table. In Panel A, two versions of the GRS test are reported. Data for the distress anomaly begin in October 1974, and data for the return-on-assets anomaly begin in November 1971. The other ten anomalies begin at the start of the sample period, which is from January 1967 through December 2013 (564 months). GRS_{10} tests whether the ten alphas for the full-sample anomalies equal zero, while GRS_{12} uses all 12 anomalies and begins the sample in October 1974. Panels B and C also report two versions of the GRS test: GRS_{51} uses 51 anomalies whose data are available by January 1967, while GRS_{72} use those anomalies plus 21 others whose data are available by February 1986.

Measure	Unadjusted	FF-3	FF-5	q-4	M-4
Panel A: 12 Anomalies, value-weighted, NYSE deciles					
Average $|\alpha|$	0.52	0.67	0.45	0.34	0.16
Average $|t|$	3.14	4.44	2.93	2.34	1.22
GRS_{10}	6.89	10.10	6.71	5.99	2.26
p_{10}	3.4×10^{-10}	1.1×10^{-15}	6.9×10^{-10}	1.2×10^{-8}	0.01
GRS_{12}	6.16	7.71	4.17	3.95	2.14
p_{12}	4.5×10^{-10}	4.2×10^{-13}	3.1×10^{-6}	8.3×10^{-6}	0.01
Number of $min|\alpha|$	-	0	1	1	10
Panel B: 73 Anomalies, value-weighted, NYSE deciles					
Average $|\alpha|$	0.39	0.44	0.30	0.20	0.18
Average $|t|$	2.14	2.74	1.77	1.15	1.02
GRS_{51}	2.74	2.60	1.91	1.68	1.34
p_{51}	9.3×10^{-9}	6.5×10^{-8}	2.9×10^{-4}	3.2×10^{-3}	0.07
GRS_{72}	2.23	2.10	1.79	1.78	1.59
p_{72}	2.2×10^{-6}	1.3×10^{-5}	5.3×10^{-4}	5.8×10^{-4}	4.8×10^{-3}
Number of $min|\alpha|$	-	10	9	21	33
Panel C: 73 Anomalies, equally weighted, NYSE/AMEX/NASDAQ deciles					
Average $|\alpha|$	0.50	0.53	0.35	0.23	0.22
Average $|t|$	3.02	3.72	2.41	1.44	1.38
GRS_{51}	5.85	6.31	5.15	4.19	4.29
p_{51}	2.7×10^{-27}	6.2×10^{-30}	4.3×10^{-23}	2.3×10^{-17}	6.2×10^{-18}
GRS_{72}	2.95	3.25	2.64	2.41	2.47
p_{72}	1.4×10^{-10}	3.0×10^{-12}	1.1×10^{-8}	2.3×10^{-7}	1.0×10^{-7}
Number of $min|\alpha|$	-	5	13	26	29

Table B5
Summary Measures of Models' Abilities to Explain Anomalies
Less Correlated with Factors in Models q-4 and M-4,
Factors Constructed Using 30% and 70% Breakpoints

This table recomputes Table 6 with the factors constructed using mispricing-measure breakpoints of 30% and 70% instead of 20% and 80%. The table reports measures that summarize the degree to which anomalies produce alpha under four different factor models: the three-factor model of Fama and French (1993), denoted FF-3; the five-factor model of Fama and French (2015), denoted FF-5; the four-factor model of Hou, Xue, and Zhang (2015a), denoted q-4; and the four-factor mispricing-factor model introduced in this study, denoted M-4. Also reported are measures based on the average unadjusted return spreads (the alphas in a model with no factors). For each model, the table reports the average absolute alpha, average absolute t-statistic, the F-statistic and associated p-value for the "GRS" test of Gibbons, Ross, and Shanken (1989), and the number of anomalies for which the model produces the smallest absolute alpha among the models being compared in the table. Panels A and B correspond to Panels B and C of Table 5, except the set of anomalies is reduced: For each of four factors—those in models q-4 and M-4 other than the market and size factors—the five anomalies (of the 73) whose long-short returns are most highly correlated with the factor are eliminated. This procedure leaves 56 anomalies in Panel A and 54 in Panel B. The sample period is from January 1967 through December 2013 (564 months). Two versions of the GRS test are reported: GRS_{41} in Panel A (or GRS_{40} in Panel B) uses the 41 (or 40) anomalies whose data are available by January 1967, while GRS_{55} in Panel A (or GRS_{53} in Panel B) uses the remaining anomalies whose data are available by February 1986.

Measure	Unadjusted	FF-3	FF-5	q-4	M-4
Panel A: 56 Anomalies, value-weighted, NYSE deciles					
Average $|\alpha|$	0.37	0.38	0.26	0.20	0.18
Average $|t|$	2.13	2.57	1.69	1.17	1.00
GRS_{41}	3.01	2.81	1.95	1.63	1.35
p_{41}	6.7×10^{-9}	6.1×10^{-8}	5.8×10^{-4}	9.3×10^{-3}	0.08
GRS_{55}	2.28	2.17	1.75	1.66	1.58
p_{55}	7.2×10^{-6}	2.5×10^{-5}	2.0×10^{-3}	4.5×10^{-3}	9.8×10^{-3}
Number of $min|\alpha|$	-	7	7	19	23
Number of $min|\alpha|$, q-4 vs. M-4	-	-	-	27	29
Panel B: 54 Anomalies, equally weighted, NYSE/AMEX/NASDAQ deciles					
Average $|\alpha|$	0.45	0.46	0.32	0.25	0.22
Average $|t|$	2.94	3.53	2.41	1.61	1.42
GRS_{40}	6.38	7.03	5.78	4.69	4.54
p_{40}	7.3×10^{-26}	5.0×10^{-29}	8.3×10^{-23}	2.9×10^{-17}	1.8×10^{-16}
GRS_{53}	3.36	4.07	3.12	2.82	2.87
p_{53}	3.6×10^{-11}	1.0×10^{-14}	6.7×10^{-10}	2.0×10^{-8}	1.2×10^{-8}
Number of $min|\alpha|$	-	6	11	16	21
Number of $min|\alpha|$, q-4 vs. M-4	-	-	-	21	33

Table B6
Summary Measures of Models' Abilities to Explain Anomalies
Less Correlated with Factors in Models q-4 and M-4,
Factors Constructed Using NYSE Breakpoints

This table recomputes Table 6 with the factors constructed using mispricing-measure breakpoints for the NYSE instead of NYSE/AMEX/NASDAQ. The table reports measures that summarize the degree to which anomalies produce alpha under four different factor models: the three-factor model of Fama and French (1993), denoted FF-3; the five-factor model of Fama and French (2015), denoted FF-5; the four-factor model of Hou, Xue, and Zhang (2015a), denoted q-4; and the four-factor mispricing-factor model introduced in this study, denoted M-4. Also reported are measures based on the average unadjusted return spreads (the alphas in a model with no factors). For each model, the table reports the average absolute alpha, average absolute t-statistic, the F-statistic and associated p-value for the "GRS" test of Gibbons, Ross, and Shanken (1989), and the number of anomalies for which the model produces the smallest absolute alpha among the models being compared in the table. Panels A and B correspond to Panels B and C of Table 5, except the set of anomalies is reduced: For each of four factors—those in models q-4 and M-4 other than the market and size factors—the five anomalies (of the 73) whose long-short returns are most highly correlated with the factor are eliminated. This procedure leaves 56 anomalies in Panel A and 54 in Panel B. The sample period is from January 1967 through December 2013 (564 months). Two versions of the GRS test are reported: GRS_{40} uses the 40 anomalies whose data are available by January 1967, while GRS_{55} in Panel A (or GRS_{53} in Panel B) uses the remaining anomalies whose data are available by February 1986.

Measure	Unadjusted	FF-3	FF-5	q-4	M-4
Panel A: 56 Anomalies, value-weighted, NYSE deciles					
Average $|\alpha|$	0.36	0.38	0.26	0.20	0.17
Average $|t|$	2.13	2.58	1.70	1.17	0.95
GRS_{40}	3.08	2.89	2.00	1.66	1.15
p_{40}	3.9×10^{-9}	3.6×10^{-8}	4.0×10^{-4}	8.0×10^{-3}	0.25
GRS_{55}	2.28	2.14	1.70	1.61	1.42
p_{55}	7.3×10^{-6}	3.3×10^{-5}	3.3×10^{-3}	7.1×10^{-3}	0.04
Number of $min|\alpha|$	-	7	8	15	26
Number of $min|\alpha|$, q-4 vs. M-4	-	-	-	22	34
Panel B: 54 Anomalies, equally weighted, NYSE/AMEX/NASDAQ deciles					
Average $|\alpha|$	0.45	0.46	0.32	0.25	0.22
Average $|t|$	2.94	3.53	2.41	1.61	1.42
GRS_{40}	6.38	7.03	5.78	4.69	4.40
p_{40}	7.3×10^{-26}	5.0×10^{-29}	8.3×10^{-23}	2.9×10^{-17}	8.8×10^{-16}
GRS_{53}	3.36	4.07	3.12	2.82	3.00
p_{53}	3.6×10^{-11}	1.0×10^{-14}	6.7×10^{-10}	2.0×10^{-8}	2.5×10^{-9}
Number of $min|\alpha|$	-	6	8	18	22
Number of $min|\alpha|$, q-4 vs. M-4	-	-	-	25	29

Summary Measures of Models' Abilities to Explain Anomalies
Less Correlated with Factors in Models q-4 and M-4, SMB
Constructed by Averaging Across Mispricing Categories

This table recomputes Table 6 with the SMB factor constructed by averaging across the three mispricing categories instead of using only stocks in the middle category. The table reports measures that summarize the degree to which anomalies produce alpha under four different factor models: the three-factor model of Fama and French (1993), denoted FF-3; the five-factor model of Fama and French (2015), denoted FF-5; the four-factor model of Hou, Xue, and Zhang (2015a), denoted q-4; and the four-factor mispricing-factor model introduced in this study, denoted M-4. Also reported are measures based on the average unadjusted return spreads (the alphas in a model with no factors). For each model, the table reports the average absolute alpha, average absolute t-statistic, the F-statistic and associated p-value for the "GRS" test of Gibbons, Ross, and Shanken (1989), and the number of anomalies for which the model produces the smallest absolute alpha among the models being compared in the table. Panels A and B correspond to Panels B and C of Table 5, except the set of anomalies is reduced: For each of four factors—those in models q-4 and M-4 other than the market and size factors—the five anomalies (of the 73) whose long-short returns are most highly correlated with the factor are eliminated. This procedure leaves 57 anomalies in Panel A and 54 in Panel B. The sample period is from January 1967 through December 2013 (564 months). Two versions of the GRS test are reported: GRS_{41} in Panel A (or GRS_{40} in Panel B) uses the 41 (or 40) anomalies whose data are available by January 1967, while GRS_{56} in Panel A (or GRS_{53} in Panel B) uses the remaining anomalies whose data are available by February 1986.

Measure	Unadjusted	FF-3	FF-5	q-4	M-4
Panel A: 57 Anomalies, value-weighted, NYSE deciles					
Average $\lvert\alpha\rvert$	0.37	0.39	0.27	0.19	0.18
Average $\lvert t\rvert$	2.14	2.60	1.71	1.15	1.05
GRS_{41}	3.01	2.81	1.95	1.63	1.36
p_{41}	6.7×10^{-9}	6.1×10^{-8}	5.8×10^{-4}	9.3×10^{-3}	0.07
GRS_{56}	2.29	2.16	1.73	1.64	1.59
p_{56}	5.5×10^{-6}	2.5×10^{-5}	2.3×10^{-3}	5.2×10^{-3}	8.3×10^{-3}
Number of $min\lvert\alpha\rvert$	-	6	9	20	22
Number of $min\lvert\alpha\rvert$, q-4 vs. M-4	-	-	-	27	30
Panel B: 54 Anomalies, equally weighted, NYSE/AMEX/NASDAQ deciles					
Average $\lvert\alpha\rvert$	0.45	0.46	0.32	0.25	0.23
Average $\lvert t\rvert$	2.94	3.53	2.41	1.61	1.47
GRS_{40}	6.38	7.03	5.78	4.69	4.55
p_{40}	7.3×10^{-26}	5.0×10^{-29}	8.3×10^{-23}	2.9×10^{-17}	1.6×10^{-16}
GRS_{53}	3.36	4.07	3.12	2.82	3.03
p_{53}	3.6×10^{-11}	1.0×10^{-14}	6.7×10^{-10}	2.0×10^{-8}	1.8×10^{-9}
Number of $min\lvert\alpha\rvert$	-	6	9	16	23
Number of $min\lvert\alpha\rvert$, q-4 vs. M-4	-	-	-	21	33

Summary Measures of Models' Abilities to Explain Anomalies
Less Correlated with Factors in Models q-4 and M-4, Factors
Constructed Using All Modifications in Tables B5–B7

This table recomputes Table 6 with the factors constructed using mispricing-measure breakpoints of 30% and 70% for the NYSE instead of 20% and 80% for NYSE/AMEX/NASDAQ. Also, the SMB factor is constructed by averaging across the three mispricing categories instead of using only stocks in the middle category. The table reports measures that summarize the degree to which anomalies produce alpha under four different factor models: the three-factor model of Fama and French (1993), denoted FF-3; the five-factor model of Fama and French (2015), denoted FF-5; the four-factor model of Hou, Xue, and Zhang (2015a), denoted q-4; and the four-factor mispricing-factor model introduced in this study, denoted M-4. Also reported are measures based on the average unadjusted return spreads (the alphas in a model with no factors). For each model, the table reports the average absolute alpha, average absolute t-statistic, the F-statistic and associated p-value for the "GRS" test of Gibbons, Ross, and Shanken (1989), and the number of anomalies for which the model produces the smallest absolute alpha among the models being compared in the table. Panels A and B correspond to Panels B and C of Table 5, except the set of anomalies is reduced: For each of four factors—those in models q-4 and M-4 other than the market and size factors—the five anomalies (of the 73) whose long-short returns are most highly correlated with the factor are eliminated. This procedure leaves 56 anomalies in Panel A and 55 in Panel B. The sample period is from January 1967 through December 2013 (564 months). Two versions of the GRS test are reported: GRS_{41} in Panel A (or GRS_{40} in Panel B) uses the 41 (or 40) anomalies whose data are available by January 1967, while GRS_{55} in Panel A (or GRS_{54} in Panel B) uses the remaining anomalies whose data are available by February 1986.

Measure	Unadjusted	FF-3	FF-5	q-4	M-4
Panel A: 56 Anomalies, value-weighted, NYSE deciles					
Average $|\alpha|$	0.37	0.38	0.26	0.20	0.18
Average $|t|$	2.13	2.57	1.69	1.17	1.03
GRS_{41}	3.01	2.81	1.95	1.63	1.34
p_{41}	6.7×10^{-9}	6.1×10^{-8}	5.8×10^{-4}	9.3×10^{-3}	0.08
GRS_{55}	2.28	2.17	1.75	1.66	1.58
p_{55}	7.2×10^{-6}	2.5×10^{-5}	2.0×10^{-3}	4.5×10^{-3}	0.01
Number of $min|\alpha|$	-	7	6	17	26
Number of $min|\alpha|$, q-4 vs. M-4	-	-	-	27	29
Panel B: 55 Anomalies, equally weighted, NYSE/AMEX/NASDAQ deciles					
Average $|\alpha|$	0.45	0.47	0.32	0.25	0.23
Average $|t|$	2.96	3.57	2.42	1.61	1.49
GRS_{40}	6.38	7.03	5.78	4.69	4.69
p_{40}	7.3×10^{-26}	5.0×10^{-29}	8.3×10^{-23}	2.9×10^{-17}	3.1×10^{-17}
GRS_{54}	3.33	3.99	3.07	2.78	2.88
p_{54}	3.8×10^{-11}	1.9×10^{-14}	9.5×10^{-10}	2.6×10^{-8}	8.1×10^{-9}
Number of $min|\alpha|$	-	4	10	18	23
Number of $min|\alpha|$, q-4 vs. M-4	-	-	-	25	30

**Models' Abilities to Explain Each Other's Factors, Factors
Constructed Using 30% and 70% Breakpoints**

This table recomputes Table 7 with the factors constructed using mispricing-measure breakpoints of 30% and 70% instead of 20% and 80%. Panel A reports a factor's estimated monthly alpha (in percent) with respect to each of the other models (with White (1980) heteroskedasticity-consistent t-statistics in parentheses). Panel B computes the Gibbons-Ross-Shanken (1989) F-test of whether a given model produces zero alphas for the factors of an alternative model(with p-values in parentheses). The models considered are the five-factor model of Fama and French (2015), denoted FF-5, which includes the factors HML, RMW, and CMA; the four-factor model of Hou, Xue, and Zhang (2015a), denoted q-4, which includes the factors I/A and ROE; and the four-factor mispricing-factor model, denoted M-4, which includes the factors $MGMT$ and $PERF$. The sample period is from January 1967 through December 2013 (564 months).

	Alpha computed with respect to model		
Factors	FF-5	q-4	M-4
Panel A: Alpha (t-statistic)			
Factors in FF-5			
HML	-	0.04	0.05
		(0.43)	(0.50)
RMW	-	0.04	0.11
		(0.55)	(1.26)
CMA	-	0.02	-0.02
		(0.47)	(-0.43)
Factors in q-4			
I/A	0.12	-	0.09
	(3.48)		(1.57)
ROE	0.45	-	0.36
	(5.53)		(4.00)
Factors in M-4			
$MGMT$	0.22	0.22	-
	(4.23)	(3.55)	
$PERF$	0.48	0.26	-
	(4.15)	(2.15)	
Panel B: GRS F-statistic (p-value)			
HML, RMW, CMA	-	0.23	0.53
		(0.87)	(0.66)
I/A, ROE	19.06	-	6.67
	(9.8×10^{-9})		(1.4×10^{-3})
$MGMT$, $PERF$	22.25	11.39	-
	(5.1×10^{-10})	(1.4×10^{-5})	

Models' Abilities to Explain Each Other's Factors,
Factors Constructed Using NYSE Breakpoints

This table recomputes Table 7 with the factors constructed using mispricing-measure breakpoints for the NYSE instead of NYSE/AMEX/NASDAQ. Panel A reports a factor's estimated monthly alpha (in percent) with respect to each of the other models (with White (1980) heteroskedasticity-consistent t-statistics in parentheses). Panel B computes the Gibbons-Ross-Shanken (1989) F-test of whether a given model produces zero alphas for the factors of an alternative model(with p-values in parentheses). The models considered are the five-factor model of Fama and French (2015), denoted FF-5, which includes the factors HML, RMW, and CMA; the four-factor model of Hou, Xue, and Zhang (2015a), denoted q-4, which includes the factors I/A and ROE; and the four-factor mispricing-factor model, denoted M-4, which includes the factors $MGMT$ and $PERF$. The sample period is from January 1967 through December 2013 (564 months).

	Alpha computed with respect to model		
Factors	FF-5	q-4	M-4
Panel A: Alpha (t-statistic)			
Factors in FF-5			
HML	-	0.04	0.03
		(0.43)	(0.29)
RMW	-	0.04	0.13
		(0.55)	(1.50)
CMA	-	0.02	-0.04
		(0.47)	(-0.63)
Factors in q-4			
I/A	0.12	-	0.09
	(3.48)		(1.57)
ROE	0.45	-	0.36
	(5.53)		(4.00)
Factors in M-4			
$MGMT$	0.29	0.31	-
	(4.72)	(4.38)	
$PERF$	0.60	0.33	-
	(4.46)	(2.26)	
Panel B: GRS F-statistic (p-value)			
HML, RMW, CMA	-	0.23	0.70
		(0.87)	(0.55)
$I/A, ROE$	19.06	-	7.56
	(9.8×10^{-9})		(5.7×10^{-4})
$MGMT, PERF$	26.84	16.02	-
	(7.5×10^{-12})	(1.7×10^{-7})	

Models' Abilities to Explain Each Other's Factors, SMB
Constructed by Averaging Across Mispricing Categories

This table recomputes Table 7 with the SMB factor constructed by averaging across the three mispricing categories instead of using only stocks in the middle category. Panel A reports a factor's estimated monthly alpha (in percent) with respect to each of the other models (with White (1980) heteroskedasticity-consistent t-statistics in parentheses). Panel B computes the Gibbons-Ross-Shanken (1989) F-test of whether a given model produces zero alphas for the factors of an alternative model(with p-values in parentheses). The models considered are the five-factor model of Fama and French (2015), denoted FF-5, which includes the factors HML, RMW, and CMA; the four-factor model of Hou, Xue, and Zhang (2015a), denoted q-4, which includes the factors I/A and ROE; and the four-factor mispricing-factor model, denoted M-4, which includes the factors $MGMT$ and $PERF$. The sample period is from January 1967 through December 2013 (564 months).

	Alpha computed with respect to model		
Factors	FF-5	q-4	M-4
Panel A: Alpha (t-statistic)			
Factors in FF-5			
HML	-	0.04	-0.02
		(0.43)	(-0.19)
RMW	-	0.04	0.10
		(0.55)	(1.16)
CMA	-	0.02	-0.02
		(0.47)	(-0.40)
Factors in q-4			
I/A	0.12	-	0.09
	(3.48)		(1.57)
ROE	0.45	-	0.36
	(5.53)		(4.00)
Factors in M-4			
$MGMT$	0.33	0.36	-
	(4.93)	(4.54)	
$PERF$	0.64	0.35	-
	(4.17)	(2.24)	
Panel B: GRS F-statistic (p-value)			
HML, RMW, CMA	-	0.23	0.49
		(0.87)	(0.69)
I/A, ROE	19.06	-	8.89
	$(9.8{\times}10^{-9})$		$(1.6{\times}10^{-4})$
$MGMT$, $PERF$	25.35	15.66	-
	$(2.9{\times}10^{-11})$	$(2.4{\times}10^{-7})$	

**Models' Abilities to Explain Each Other's Factors, Factors
Constructed Using All Modifications in Tables B9–B11**

This table recomputes Table 7 with the factors constructed using mispricing-measure breakpoints of 30% and 70% for the NYSE instead of 20% and 80% for NYSE/AMEX/NASDAQ. Also, the SMB factor is constructed by averaging across the three mispricing categories instead of using only stocks in the middle category. Panel A reports a factor's estimated monthly alpha (in percent) with respect to each of the other models (with White (1980) heteroskedasticity-consistent t-statistics in parentheses). Panel B computes the Gibbons-Ross-Shanken (1989) F-test of whether a given model produces zero alphas for the factors of an alternative model(with p-values in parentheses). The models considered are the five-factor model of Fama and French (2015), denoted FF-5, which includes the factors HML, RMW, and CMA; the four-factor model of Hou, Xue, and Zhang (2015a), denoted q-4, which includes the factors I/A and ROE; and the four-factor mispricing-factor model, denoted M-4, which includes the factors $MGMT$ and $PERF$. The sample period is from January 1967 through December 2013 (564 months).

	Alpha computed with respect to model		
Factors	FF-5	q-4	M-4
Panel A: Alpha (t-statistic)			
Factors in FF-5			
HML	-	0.04	0.09
		(0.43)	(1.02)
RMW	-	0.04	0.13
		(0.55)	(1.47)
CMA	-	0.02	-0.02
		(0.47)	(-0.30)
Factors in q-4			
I/A	0.12	-	0.09
	(3.48)		(1.57)
ROE	0.45	-	0.36
	(5.53)		(4.00)
Factors in M-4			
$MGMT$	0.19	0.19	-
	(3.90)	(3.32)	
$PERF$	0.49	0.28	-
	(4.65)	(2.48)	
Panel B: GRS F-statistic (p-value)			
HML, RMW, CMA	-	0.23	0.91
		(0.87)	(0.44)
I/A, ROE	19.06	-	7.79
	(9.8×10^{-9})		(4.6×10^{-4})
$MGMT$, $PERF$	23.74	12.12	-
	(1.3×10^{-10})	(7.1×10^{-6})	

Summary Measures of Models' Abilities to Explain Anomalies, Annual-Data Factors
for Models FF-3, FF-5, and q-4 Reconstructed with Minimum Four-Month Gap

This table recomputes Table 5 after reconstructing factors in models FF-3, FF-5, and q-4 that use annual data. A minimum gap of 4 months after the end of the fiscal year (instead of a gap of 6 to 18 months) is required before using an annual variable in the factor construction, while using immediate prior-month values of market capitalization. We apply this change to SMB and HML in model FF-3, to SMB, HML, RMW, and CMA in model FF-5, and to I/A in model q-4. The table reports measures that summarize the degree to which anomalies produce alpha under four different factor models: the three-factor model of Fama and French (1993), denoted FF-3; the five-factor model of Fama and French (2015), denoted FF-5; the four-factor model of Hou, Xue, and Zhang (2015a), denoted q-4; and the four-factor mispricing-factor model introduced in this study, denoted M-4. Also reported are measures based on the average unadjusted return spreads (the alphas in a model with no factors). For each model, the table reports the average absolute alpha, average absolute t-statistic, the F-statistic and associated p-value for the "GRS" test of Gibbons, Ross, and Shanken (1989), and the number of anomalies for which the model produces the smallest absolute alpha among the models being compared in the table. In Panel A, two versions of the GRS test are reported. Data for the distress anomaly begin in October 1974, and data for the return-on-assets anomaly begin in November 1971. The other ten anomalies begin at the start of the sample period, which is from January 1967 through December 2013 (564 months). GRS_{10} tests whether the ten alphas for the full-sample anomalies equal zero, while GRS_{12} uses all 12 anomalies and begins the sample in October 1974. Panels B and C also report two versions of the GRS test: GRS_{51} uses 51 anomalies whose data are available by January 1967, while GRS_{72} use those anomalies plus 21 others whose data are available by February 1986.

Measure	Unadjusted	FF-3	FF-5	q-4	M-4
Panel A: 12 Anomalies, value-weighted, NYSE deciles					
Average $\|\alpha\|$	0.52	0.68	0.45	0.30	0.18
Average $\|t\|$	3.14	4.55	2.85	2.17	1.29
GRS_{10}	6.89	11.47	7.59	5.83	1.84
p_{10}	$3.4{\times}10^{-10}$	$5.5{\times}10^{-10}$	$2.2{\times}10^{-11}$	$2.2{\times}10^{-8}$	0.05
GRS_{12}	6.16	8.38	4.56	3.73	1.88
p_{12}	$4.5{\times}10^{-10}$	$2.2{\times}10^{-14}$	$5.8{\times}10^{-7}$	$2.1{\times}10^{-5}$	0.03
Number of $min\|\alpha\|$	-	0	1	3	8
Panel B: 73 Anomalies, value-weighted, NYSE deciles					
Average $\|\alpha\|$	0.39	0.44	0.32	0.21	0.18
Average $\|t\|$	2.14	2.73	1.90	1.20	0.99
GRS_{51}	2.74	2.66	2.10	1.75	1.28
p_{51}	$9.3{\times}10^{-9}$	$3.0{\times}10^{-8}$	$3.3{\times}10^{-5}$	$1.6{\times}10^{-3}$	0.10
GRS_{72}	2.23	2.14	1.67	1.83	1.54
p_{72}	$2.2{\times}10^{-6}$	$7.2{\times}10^{-6}$	$2.0{\times}10^{-3}$	$3.3{\times}10^{-4}$	$8.1{\times}10^{-3}$
Number of $min\|\alpha\|$	-	8	6	22	37
Panel C: 73 Anomalies, equally weighted, NYSE/AMEX/NASDAQ deciles					
Average $\|\alpha\|$	0.50	0.53	0.38	0.24	0.22
Average $\|t\|$	3.02	3.76	2.66	1.45	1.38
GRS_{51}	5.85	6.43	5.43	4.28	4.17
p_{51}	$2.7{\times}10^{-27}$	$1.2{\times}10^{-30}$	$8.9{\times}10^{-25}$	$6.5{\times}10^{-18}$	$3.1{\times}10^{-17}$
GRS_{72}	2.95	3.41	2.75	2.46	2.68
p_{72}	$1.4{\times}10^{-10}$	$3.3{\times}10^{-13}$	$2.6{\times}10^{-9}$	$1.1{\times}10^{-7}$	$6.2{\times}10^{-9}$
Number of $min\|\alpha\|$	-	7	6	27	33

**Summary Measures of Models' Abilities to Explain Anomalies Less Correlated
with Factors in Models q-4 and M-4, Annual-Data Factors for Models
FF-3, FF-5, and q-4 Reconstructed with Minimum Four-Month Gap**

This table recomputes Table 6 after reconstructing factors in models FF-3, FF-5, and q-4 that use annual data. A minimum gap of 4 months after the end of the fiscal year (instead of a gap of 6 to 18 months) is required before using an annual variable in the factor construction, while using immediate prior-month values of market capitalization. We apply this change to SMB and HML in model FF-3, to SMB, HML, RMW, and CMA in model FF-5, and to I/A in model q-4. The table reports measures that summarize the degree to which anomalies produce alpha under four different factor models: the three-factor model of Fama and French (1993), denoted FF-3; the five-factor model of Fama and French (2015), denoted FF-5; the four-factor model of Hou, Xue, and Zhang (2015a), denoted q-4; and the four-factor mispricing-factor model introduced in this study, denoted M-4. Also reported are measures based on the average unadjusted return spreads (the alphas in a model with no factors). For each model, the table reports the average absolute alpha, average absolute t-statistic, the F-statistic and associated p-value for the "GRS" test of Gibbons, Ross, and Shanken (1989), and the number of anomalies for which the model produces the smallest absolute alpha among the models being compared in the table. Panels A and B correspond to Panels B and C of Table 5, except the set of anomalies is reduced: For each of four factors—those in models q-4 and M-4 other than the market and size factors—the five anomalies (of the 73) whose long-short returns are most highly correlated with the factor are eliminated. This procedure leaves 57 anomalies in Panel A and 54 in Panel B. The sample period is from January 1967 through December 2013 (564 months). Two versions of the GRS test are reported: GRS_{41} in Panel A (or GRS_{40} in Panel B) uses the 41 (or 40) anomalies whose data are available by January 1967, while GRS_{56} in Panel A (or GRS_{53} in Panel B) uses the remaining anomalies whose data are available by February 1986.

Measure	Unadjusted	FF-3	FF-5	q-4	M-4
Panel A: 57 Anomalies, value-weighted, NYSE deciles					
Average $|\alpha|$	0.37	0.39	0.28	0.21	0.17
Average $|t|$	2.14	2.56	1.83	1.23	0.97
GRS_{41}	3.01	2.87	2.15	1.71	1.26
p_{41}	6.7×10^{-9}	3.2×10^{-8}	7.7×10^{-5}	4.7×10^{-3}	0.13
GRS_{56}	2.29	2.22	1.67	1.68	1.53
p_{56}	5.5×10^{-6}	1.2×10^{-5}	4.0×10^{-3}	3.5×10^{-3}	0.01
Number of $min|\alpha|$	-	6	4	17	30
Number of $min|\alpha|$, q-4 vs. M-4	-	-	-	22	35
Panel B: 54 Anomalies, equally weighted, NYSE/AMEX/NASDAQ deciles					
Average $|\alpha|$	0.45	0.47	0.34	0.26	0.23
Average $|t|$	2.94	3.54	2.63	1.62	1.47
GRS_{40}	6.38	7.19	6.25	4.80	4.55
p_{40}	7.3×10^{-26}	8.7×10^{-30}	3.9×10^{-25}	7.8×10^{-18}	1.6×10^{-16}
GRS_{54}	3.36	4.29	3.31	2.85	3.03
p_{54}	3.6×10^{-11}	7.8×10^{-16}	6.9×10^{-11}	1.4×10^{-8}	1.8×10^{-9}
Number of $min|\alpha|$	-	6	4	21	23
Number of $min|\alpha|$, q-4 vs. M-4	-	-	-	24	30

Table B15

Models' Abilities to Explain Each Other's Factors, Annual-Data Factors for Models FF-3, FF-5, and q-4 Reconstructed with Minimum Four-Month Gap

This table recomputes Table 7 after reconstructing factors in models FF-3, FF-5, and q-4 that use annual data. A minimum gap of 4 months after the end of the fiscal year (instead of a gap of 6 to 18 months) is required before using an annual variable in the factor construction, while using immediate prior-month values of market capitalization. We apply this change to SMB and HML in model FF-3, to SMB, HML, RMW, and CMA in model FF-5, and to I/A in model q-4. Because the FF-5 factor CMA then becomes identical to the q-4 factor I/A, we omit the comparisons involving those factors. Panel A reports a factor's estimated monthly alpha (in percent) with respect to each of the other models (with White (1980) heteroskedasticity-consistent t-statistics in parentheses). Panel B computes the Gibbons-Ross-Shanken (1989) F-test of whether a given model produces zero alphas for the factors of an alternative model(with p-values in parentheses). The models considered are the five-factor model of Fama and French (2015), denoted FF-5, which includes the factors HML, RMW, and CMA; the four-factor model of Hou, Xue, and Zhang (2015a), denoted q-4, which includes the factors I/A and ROE; and the four-factor mispricing-factor model, denoted M-4, which includes the factors $MGMT$ and $PERF$. The sample period is from January 1967 through December 2013 (564 months).

	Alpha computed with respect to model		
Factors	FF-5	q-4	M-4
Panel A: Alpha (t-statistic)			
Factors in FF-5			
HML	-	-0.01	-0.07
		(-0.10)	(-0.69)
RMW	-	0.10	0.15
		(1.07)	(1.61)
CMA	-	-	-0.08
			(-1.38)
Factors in q-4			
I/A	-	-	-0.08
			(-1.38)
ROE	0.49	-	0.36
	(5.92)		(4.00)
Factors in M-4			
$MGMT$	0.32	0.33	-
	(5.53)	(4.67)	
$PERF$	0.64	0.24	-
	(4.18)	(1.55)	
Panel B: GRS F-statistic (p-value)			
HML, RMW, CMA	-	-	1.45
			(0.23)
I/A, ROE	-	-	8.29
			(2.8×10^{-4})
$MGMT$, $PERF$	28.84	16.88	-
	(1.2×10^{-12})	(7.7×10^{-8})	

Summary Measures of Models' Abilities to Explain Anomalies
Subsample 1: 1967–1990

This table recomputes Table 5 for the first subperiod. An anomaly is included in the subperiod analysis if it satisfies the following three conditions: (i) the available data for the anomaly cover the entire sample period from 1967 to 2013, (ii) the estimated FF-3 long-short alpha for the anomaly has the same sign in the early and late subperiods, and (iii) the FF-3 alpha is significant at the one-sided 10% significance level in both subperiods.

Measure	Unadjusted	FF-3	FF-5	q-4	M-4
Panel A: 10 Anomalies, value-weighted, NYSE deciles					
Average $|\alpha|$	0.59	0.57	0.44	0.34	0.08
Average $|t|$	2.81	3.17	2.97	1.90	0.44
GRS_{10}	5.45	6.04	2.94	2.28	0.32
p_{10}	2.2×10^{-7}	2.7×10^{-8}	1.6×10^{-3}	0.01	0.98
Number of $min|\alpha|$	-	0	2	0	8
Panel B: 29 Anomalies, value-weighted, NYSE deciles					
Average $|\alpha|$	0.53	0.66	0.36	0.28	0.31
Average $|t|$	2.38	3.52	2.19	1.48	1.46
GRS_{29}	4.56	5.06	3.70	2.52	2.41
p_{29}	1.0×10^{-11}	2.4×10^{-13}	8.4×10^{-9}	6.7×10^{-5}	1.5×10^{-4}
Number of $min|\alpha|$	-	0	8	10	11
Panel C: 38 Anomalies, equally weighted, NYSE/AMEX/NASDAQ deciles					
Average $|\alpha|$	0.67	0.67	0.36	0.37	0.30
Average $|t|$	3.47	4.39	2.62	2.17	1.68
GRS_{38}	9.23	9.16	7.28	5.48	5.85
p_{38}	5.6×10^{-30}	1.3×10^{-29}	1.8×10^{-23}	4.1×10^{-17}	1.8×10^{-18}
Number of $min|\alpha|$	-	0	11	10	17

Table B17
Summary Measures of Models' Abilities to Explain Anomalies
Subsample 2: 1991–2013

This table recomputes Table 5 for the second subperiod. An anomaly is included in the subperiod analysis if it satisfies the following three conditions: (i) the available data for the anomaly cover the entire sample period from 1967 to 2013, (ii) the estimated FF-3 long-short alpha for the anomaly has the same sign in the early and late subperiods, and (iii) the FF-3 alpha is significant at the one-sided 10% significance level in both subperiods.

Measure	Unadjusted	FF-3	FF-5	q-4	M-4
Panel A: 10 Anomalies, value-weighted, NYSE deciles					
Average $|\alpha|$	0.46	0.60	0.44	0.38	0.20
Average $|t|$	2.01	2.94	2.97	1.94	1.08
GRS_{10}	3.54	5.38	2.99	3.44	1.43
p_{10}	$2.1{\times}10^{-4}$	$3.0{\times}10^{-7}$	$1.4{\times}10^{-3}$	$3.0{\times}10^{-4}$	0.17
Number of $min|\alpha|$	-	0	0	1	9
Panel B: 29 Anomalies, value-weighted, NYSE deciles					
Average $|\alpha|$	0.35	0.60	0.36	0.22	0.24
Average $|t|$	1.40	2.57	2.19	0.91	0.91
GRS_{29}	2.09	2.24	1.47	1.62	1.28
p_{29}	$1.4{\times}10^{-3}$	$5.0{\times}10^{-4}$	0.06	0.03	0.16
Number of $min|\alpha|$	-	1	5	11	12
Panel C: 38 Anomalies, equally weighted, NYSE/AMEX/NASDAQ deciles					
Average $|\alpha|$	0.48	0.64	0.36	0.20	0.19
Average $|t|$	1.85	2.98	2.62	0.96	0.80
GRS_{38}	3.34	3.97	2.91	2.94	2.66
p_{38}	$8.4{\times}10^{-9}$	$3.2{\times}10^{-11}$	$4.6{\times}10^{-7}$	$3.4{\times}10^{-7}$	$3.9{\times}10^{-6}$
Number of $min|\alpha|$	-	0	6	13	19

Abilities of Models FF-5, q-4, and M-4 to Explain Each Other's Factors
Subsample 1: 1967–1990

This table recomputes Table 7 for the first subperiod. Panel A reports a factor's estimated monthly alpha (in percent) with respect to each of the other models (with White (1980) heteroskedasticity-consistent t-statistics in parentheses). Panel B computes the Gibbons-Ross-Shanken (1989) F-test of whether a given model produces zero alphas for the factors of an alternative model (with p-values in parentheses). The factors whose alphas are tested are those other than a model's market and size factors. The models considered are the five-factor model of Fama and French (2015), denoted FF-5, which includes the factors HML, RMW, and CMA; the four-factor model of Hou, Xue, and Zhang (2015a), denoted q-4, which includes the factors I/A and ROE; and the four-factor mispricing-factor model, denoted M-4, which includes the factors $MGMT$ and $PERF$.

| | Alpha computed with respect to model | | |
Factors	FF-5	q-4	M-4
Panel A: Alpha (t-statistic)			
Factors in FF-5			
HML	-	0.38	0.01
		(3.45)	(0.07)
RMW	-	0.21	0.24
		(2.51)	(2.84)
CMA	-	0.01	-0.09
		(0.22)	(-1.37)
Factors in q-4			
I/A	0.22	-	0.12
	(4.98)		(1.71)
ROE	0.78	-	0.69
	(7.27)		(6.33)
Factors in M-4			
$MGMT$	0.28	0.47	-
	(3.93)	(5.01)	
$PERF$	0.69	0.27	-
	(3.81)	(1.47)	
Panel B: GRS F-statistic (p-value)			
HML, RMW, CMA	-	7.01	2.67
		(1.5×10^{-4})	(0.05)
I/A, ROE	29.90	-	19.34
	(1.7×10^{-12})		(1.4×10^{-8})
$MGMT$, $PERF$	17.49	14.87	-
	(6.9×10^{-8})	(7.3×10^{-7})	

19

Table B19
Abilities of Models FF-5, q-4, and M-4 to Explain Each Other's Factors
Subsample 2: 1991–2013

This table recomputes Table 7 for the second subperiod. Panel A reports a factor's estimated monthly alpha (in percent) with respect to each of the other models (with White (1980) heteroskedasticity-consistent t-statistics in parentheses). Panel B computes the Gibbons-Ross-Shanken (1989) F-test of whether a given model produces zero alphas for the factors of an alternative model (with p-values in parentheses). The factors whose alphas are tested are those other than a model's market and size factors. The models considered are the five-factor model of Fama and French (2015), denoted FF-5, which includes the factors HML, RMW, and CMA; the four-factor model of Hou, Xue, and Zhang (2015a), denoted q-4, which includes the factors I/A and ROE; and the four-factor mispricing-factor model, denoted M-4, which includes the factors $MGMT$ and $PERF$.

| | Alpha computed with respect to model | | |
Factors	FF-5	q-4	M-4
Panel A: Alpha (t-statistic)			
Factors in FF-5			
HML	-	-0.10	0.00
		(-0.59)	(-0.01)
RMW	-	0.15	0.25
		(1.24)	(1.95)
CMA	-	0.07	0.01
		(1.15)	(0.07)
Factors in q-4			
I/A	0.03	-	0.05
	(0.53)		(0.50)
ROE	0.30	-	0.25
	(2.62)		(1.99)
Factors in M-4			
$MGMT$	0.29	0.34	-
	(2.51)	(2.65)	
$PERF$	0.61	0.48	-
	(2.62)	(2.05)	
Panel B: GRS F-statistic (p-value)			
HML, RMW, CMA	-	2.18	1.62
		(0.09)	(0.18)
I/A, ROE	3.25	-	2.07
	(0.04)		(0.13)
$MGMT$, $PERF$	7.38	7.24	-
	(7.6×10^{-4})	(8.6×10^{-4})	

20

Table B20
Summary Measures of the Abilities of a Model with a Single
Mispricing Factor to Explain Anomalies
Subsample 1: 1967–1990

This table recomputes Table 9 for the first subperiod. An anomaly is included in the subperiod analysis if it satisfies the following three conditions: (i) the available data for the anomaly cover the entire sample period from 1967 to 2013, (ii) the estimated FF-3 long-short alpha for the anomaly has the same sign in the early and late subperiods, and (iii) the FF-3 alpha is significant at the one-sided 10% significance level in both subperiods.

Panel A: 10 Anomalies, value-weighted, NYSE deciles

Average $|\alpha|$	0.15
Average $|t|$	0.72
GRS_{10}	0.58
p_{10}	0.83

Panel B: 29 Anomalies, value-weighted, NYSE deciles

Average $|\alpha|$	0.29
Average $|t|$	1.35
GRS_{29}	2.56
p_{29}	4.7×10^{-5}

Panel C: 38 Anomalies, equally weighted, NYSE/AMEX/NASDAQ deciles

Average $|\alpha|$	0.37
Average $|t|$	1.90
GRS_{38}	5.88
p_{38}	1.3×10^{-18}

21

Summary Measures of the Abilities of a Model with a Single
Mispricing Factor to Explain Anomalies
Subsample 2: 1991–2013

This table recomputes Table 9 for the second subperiod. An anomaly is included in the subperiod analysis if it satisfies the following three conditions: (i) the available data for the anomaly cover the entire sample period from 1967 to 2013, (ii) the estimated FF-3 long-short alpha for the anomaly has the same sign in the early and late subperiods, and (iii) the FF-3 alpha is significant at the one-sided 10% significance level in both subperiods.

Panel A: 10 Anomalies, value-weighted, NYSE deciles	
Average $|\alpha|$	0.30
Average $|t|$	1.24
GRS_{10}	1.53
p_{10}	0.13

Panel B: 29 Anomalies, value-weighted, NYSE deciles	
Average $|\alpha|$	0.35
Average $|t|$	1.12
GRS_{29}	1.44
p_{29}	0.07

Panel C: 38 Anomalies, equally weighted, NYSE/AMEX/NASDAQ deciles	
Average $|\alpha|$	0.32
Average $|t|$	1.13
GRS_{38}	2.51
p_{38}	1.4×10^{-5}

Abilities of Models FF-5, q-4, and M-3 to Explain Each Other's Factors
Subsample 1: 1967–1990

This table recomputes Table 10 for the first subperiod. Panel A reports a factor's estimated monthly alpha (in percent) with respect to each of the other models (with White (1980) heteroskedasticity-consistent t-statistics in parentheses). Panel B computes the Gibbons-Ross-Shanken (1989) F-test of whether a given model produces zero alphas for the factors of an alternative model (with p-values in parentheses). The factors whose alphas are tested are those other than a model's market and size factors. The models considered are the five-factor model of Fama and French (2015), denoted FF-5, which includes the factors HML, RMW, and CMA; the four-factor model of Hou, Xue, and Zhang (2015a), denoted q-4, which includes the factors I/A and ROE; and the three-factor mispricing factor model, denoted M-3, which includes the factor UMO.

	Alpha computed with respect to model		
Factors	FF-5	q-4	M-3
Panel A: Alpha (t-statistic)			
Factors in FF-5			
HML	-	0.38	0.27
		(3.45)	(1.60)
RMW	-	0.21	0.18
		(2.51)	(1.75)
CMA	-	0.01	0.01
		(0.22)	(0.09)
Factors in q-4			
I/A	0.22	-	0.20
	(4.98)		(1.94)
ROE	0.78	-	0.54
	(7.27)		(3.82)
Factor in M-3			
UMO	0.71	0.53	-
	(5.57)	(3.87)	
Panel B: GRS F-statistic (p-value)			
HML, RMW, CMA	-	7.01	4.20
		(1.5×10^{-4})	(6.3×10^{-3})
I/A, ROE	29.90	-	13.77
	(1.7×10^{-12})		(2.0×10^{-6})
UMO	40.60	18.69	-
	(7.6×10^{-10})	(2.1×10^{-5})	

Table B23
Abilities of Models FF-5, q-4, and M-3 to Explain Each Other's Factors
Subsample 2: 1991–2013

This table recomputes Table 10 for the second subperiod. Panel A reports a factor's estimated monthly alpha (in percent) with respect to each of the other models (with White (1980) heteroskedasticity-consistent t-statistics in parentheses). Panel B computes the Gibbons-Ross-Shanken (1989) F-test of whether a given model produces zero alphas for the factors of an alternative model (with p-values in parentheses). The factors whose alphas are tested are those other than a model's market and size factors. The models considered are the five-factor model of Fama and French (2015), denoted FF-5, which includes the factors HML, RMW, and CMA; the four-factor model of Hou, Xue, and Zhang (2015a), denoted q-4, which includes the factors I/A and ROE; and the three-factor mispricing factor model, denoted M-3, which includes the factor UMO.

Factors	Alpha computed with respect to model		
	FF-5	q-4	M-3
Panel A: Alpha (t-statistic)			
Factors in FF-5			
HML	-	-0.10	0.32
		(-0.59)	(1.47)
RMW	-	0.15	0.25
		(1.24)	(1.88)
CMA	-	0.07	0.15
		(1.15)	(1.16)
Factors in q-4			
I/A	0.03	-	0.19
	(0.53)		(1.39)
ROE	0.30	-	0.10
	(2.62)		(0.73)
Factor in M-3			
UMO	0.74	0.69	-
	(4.54)	(4.25)	
Panel B: GRS F-statistic (p-value)			
HML, RMW, CMA	-	2.18	1.48
		(0.09)	(0.22)
$I/A, ROE$	3.25	-	1.49
	(0.04)		(0.23)
UMO	23.68	22.48	-
	(1.9×10^{-6})	(3.4×10^{-6})	

24

Abilities of Models M-3 and M-4 to Explain Each Other's Factors
Subsample 1: 1967–1990

This table recomputes Table 11 for the first subperiod. Panel A reports a factor's estimated monthly alpha (in percent) with respect to each of the other models (with White (1980) heteroskedasticity-consistent t-statistics in parentheses). Panel B computes the Gibbons-Ross-Shanken (1989) F-test of whether a given model produces zero alphas for the factors of an alternative model (with p-values in parentheses). The factors whose alphas are tested are those other than a model's market and size factors. The models considered are the three-factor mispricing factor model, denoted M-3, which includes the factor UMO, and the four-factor mispricing-factor model, denoted M-4, which includes the factors $MGMT$ and $PERF$.

	Alpha computed with respect to model	
Factors	M-3	M-4
Panel A: Alpha (t-statistic)		
Factor in M-3		
UMO	-	0.15
		(2.46)
Factors in M-4		
$MGMT$	0.24	-
	(1.80)	
$PERF$	-0.15	-
	(-0.90)	
Panel B: GRS F-statistic (p-value)		
UMO	-	5.20
		(0.02)
$MGMT$, $PERF$	2.26	-
	(0.11)	

Table B25
Abilities of Models M-3 and M-4 to Explain Each Other's Factors
**Subsample 2: 1991–2013**

This table recomputes Table 11 for the second subperiod. Panel A reports a factor's estimated monthly alpha (in percent) with respect to each of the other models (with White (1980) heteroskedasticity-consistent t-statistics in parentheses). Panel B computes the Gibbons-Ross-Shanken (1989) F-test of whether a given model produces zero alphas for the factors of an alternative model (with p-values in parentheses). The factors whose alphas are tested are those other than a model's market and size factors. The models considered are the three-factor mispricing factor model, denoted M-3, which includes the factor UMO, and the four-factor mispricing-factor model, denoted M-4, which includes the factors $MGMT$ and $PERF$.

| | Alpha computed with respect to model | |
Factors	M-3	M-4
Panel A: Alpha (t-statistic)		
Factor in M-3		
UMO	-	0.23
		(3.76)
Factors in M-4		
$MGMT$	0.29	-
	(1.67)	
$PERF$	-0.45	-
	(-2.57)	
Panel B: GRS F-statistic (p-value)		
UMO	-	15.49
		(1.1×10^{-4})
$MGMT$, $PERF$	4.08	-
	(0.02)	

26

The table reports summary statistics for alphas computed under six different factor models: the three-factor model of Fama and French (1993), denoted FF-3; the four-factor momentum model of Carhart (1997), denoted MOM-4; the four-factor liquidity-risk model of Pástor and Stambaugh (2003), denoted LIQ-4; the five-factor model of Fama and French (2015), denoted FF-5; the four-factor model of Hou, Xue, and Zhang (2015a), denoted q-4; and the four-factor mispricing-factor model introduced in this study, denoted M-4. The table reports each model's average absolute monthly (percent) alpha, average absolute t-statistic, the F-statistic and associated p-value for the "GRS" test of Gibbons, Ross, and Shanken (1989), and the number of anomalies for which the model produces the smallest absolute alpha among the models being compared in the table. In Panel A, two versions of the GRS test are reported. Data for the distress anomaly begin in October 1974, and data for the return-on-assets anomaly begin in November 1971. The other ten anomalies begin at the start of the sample period, which is from January 1967 through December 2013 (564 months). GRS_{10} tests whether the ten alphas for the full-sample anomalies equal zero, while GRS_{12} uses all 12 anomalies and begins the sample in October 1974. Panels B and C also report two versions of the GRS test: GRS_{51} uses 51 anomalies whose data are available by January 1967, while GRS_{72} use those anomalies plus 21 others whose data are available by February 1986.

	FF-3	MOM-4	LIQ-4	FF-5	q-4	M-4
Panel A: 12 Anomalies (value-weighted, NYSE deciles)						
Average $|\alpha|$	0.67	0.46	0.66	0.45	0.34	0.18
Average $|t|$	4.40	3.52	4.34	2.84	2.31	1.29
GRS_{10}	9.67	7.21	9.34	6.23	5.56	1.62
p_{10}	6.5×10^{-15}	9.9×10^{-11}	2.4×10^{-14}	4.8×10^{-9}	6.6×10^{-8}	0.10
GRS_{12}	7.71	6.09	7.77	4.17	3.95	1.88
p_{12}	4.2×10^{-13}	6.2×10^{-10}	3.2×10^{-13}	3.1×10^{-6}	8.3×10^{-6}	0.03
Number of $min|\alpha|$	0	0	0	1	2	9
Panel B: 73 Anomalies, value-weighted, NYSE deciles						
Average $|\alpha|$	0.44	0.30	0.44	0.29	0.20	0.18
Average $|t|$	2.69	1.92	2.72	1.74	1.13	1.00
GRS_{51}	2.60	2.11	2.64	1.95	1.72	1.36
p_{51}	6.7×10^{-8}	2.7×10^{-5}	3.9×10^{-8}	1.8×10^{-4}	2.0×10^{-3}	0.05
GRS_{72}	2.10	1.96	2.15	1.79	1.78	1.54
p_{72}	1.3×10^{-5}	7.0×10^{-5}	6.8×10^{-6}	5.3×10^{-4}	5.8×10^{-4}	8.1×10^{-3}
Number of $min|\alpha|$	4	6	5	5	21	32
Panel C: 73 Anomalies, equally weighted, NYSE/AMEX/NASDAQ deciles						
Average $|\alpha|$	0.52	0.35	0.52	0.34	0.23	0.22
Average $|t|$	3.63	2.52	3.68	2.37	1.41	1.38
GRS_{51}	6.11	5.46	6.09	4.97	4.12	4.14
p_{51}	1.2×10^{-28}	7.5×10^{-25}	1.8×10^{-28}	5.9×10^{-22}	7.1×10^{-17}	5.9×10^{-17}
GRS_{72}	3.25	3.06	3.18	2.64	2.41	2.68
p_{72}	3.0×10^{-12}	4.1×10^{-11}	7.9×10^{-12}	1.1×10^{-8}	2.3×10^{-07}	6.2×10^{-9}
Number of $min|\alpha|$	3	7	4	7	24	28
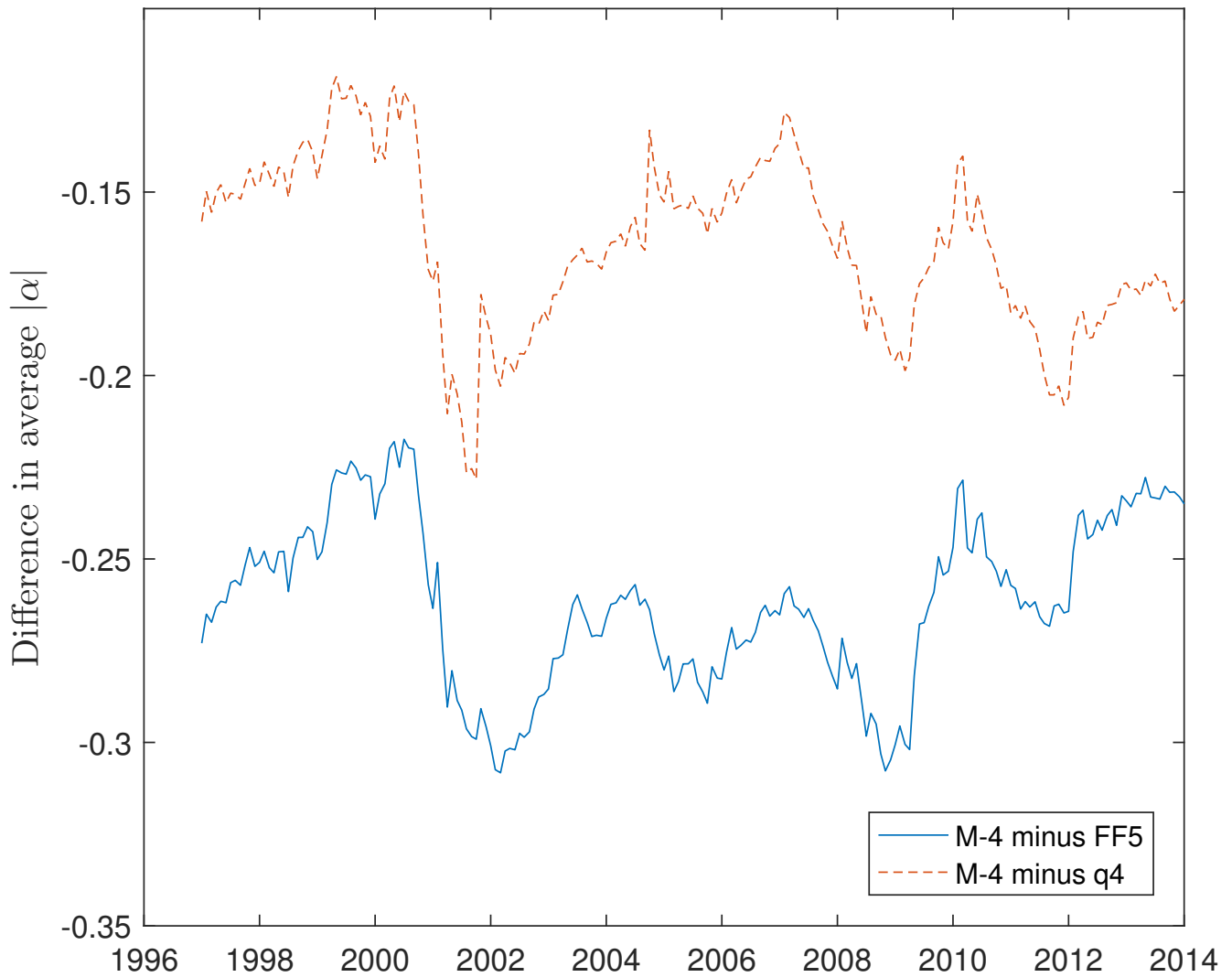
Figure B1.   Rolling comparisons of average absolute alpha for 12 value-weighted anomalies. The 12 anomalies are the same as those in Panel A of Table 5. The plot displays the difference between the anomalies' average absolute alpha under model M-4 and the average under model FF-5 (solid line) and model q-4 (dashed line). Alphas are computed over 30-year subperiods ending at each month plotted. The overall sample period is from January 1967 through December 2013 (564 months)
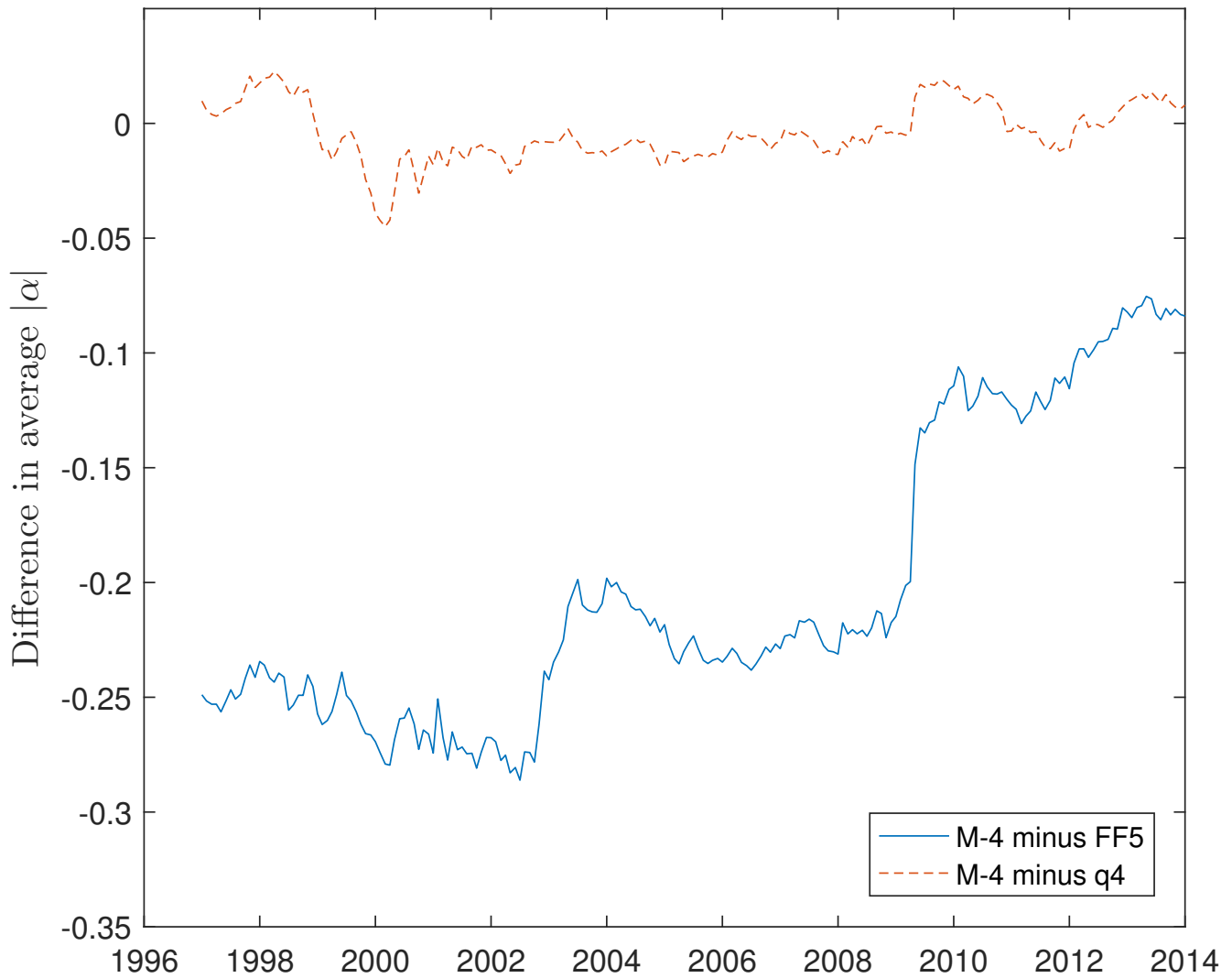
Figure B2. Rolling comparisons of average absolute alpha for 29 value-weighted anomalies. The 29 anomalies consist of the subset of the 73 anomalies in Panel B of Table 5 that (i) have complete data for the overall sample period and (ii) in each half of that period produce a FF-3 alpha that is significant at a one-tailed 10% significance level. (The objective is to analyze only anomalies that are robust over the entire sample period.) The plot displays the difference between the anomalies' average absolute alpha under model M-4 and the average under model FF-5 (solid line) and model q-4 (dashed line). Alphas are computed over 30-year subperiods ending at each month plotted. The overall sample period is from January 1967 through December 2013 (564 months)
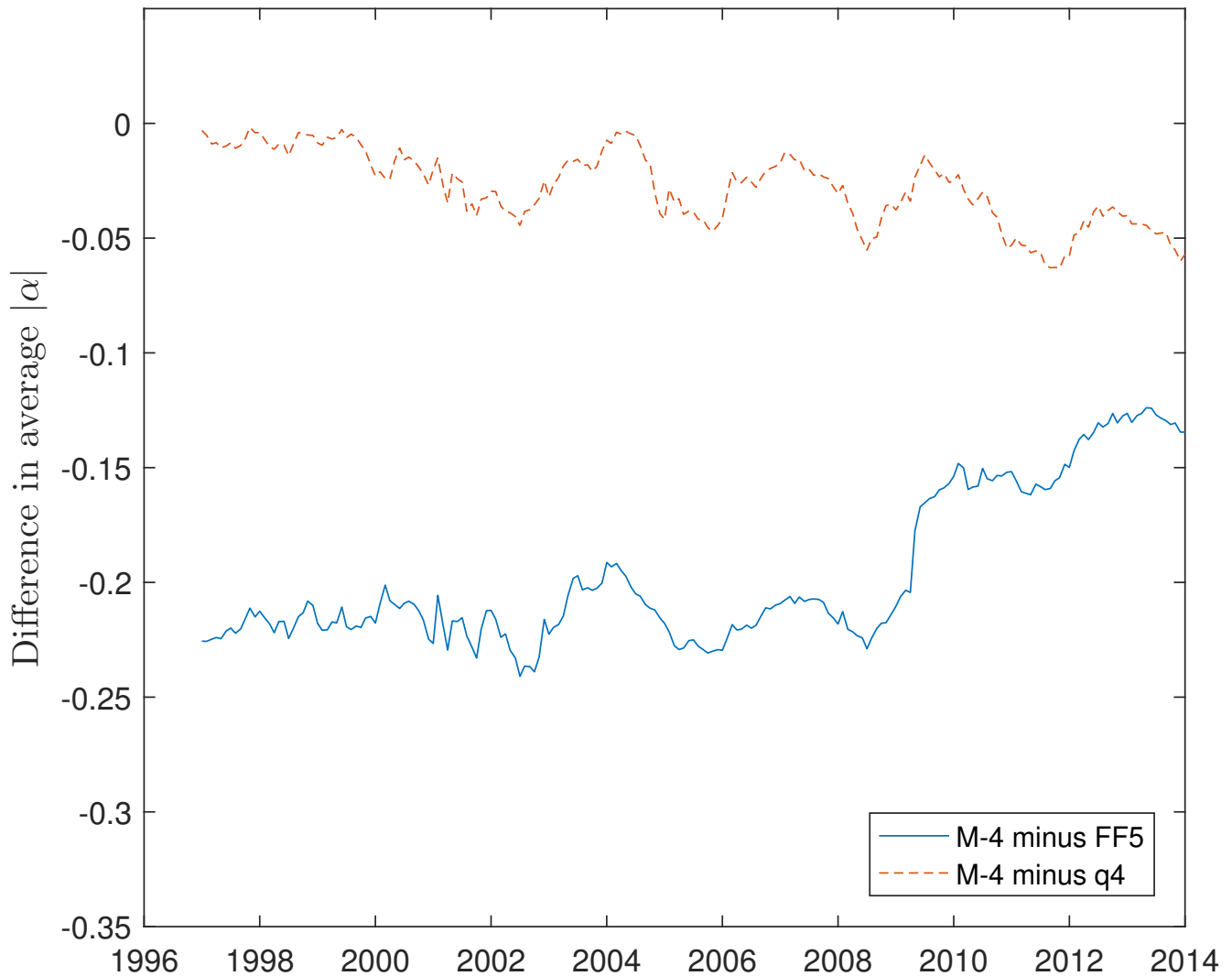
Figure B3. Rolling comparisons of average absolute alpha for 38 equally weighted anomalies. The 38 anomalies consist of the subset of the 73 anomalies in Panel C of Table 5 that (i) have complete data for the overall sample period and (ii) in each half of that period produce a FF-3 alpha that is significant at a one-tailed 10% significance level. (The objective is to analyze only anomalies that are robust over the entire sample period.) The plot displays the difference between the anomalies' average absolute alpha under model M-4 and the average under model FF-5 (solid line) and model q-4 (dashed line). Alphas are computed over 30-year subperiods ending at each month plotted. The overall sample period is from January 1967 through December 2013 (564 months)