Distance-Based Metrics: A Bayesian Solution for Asset-Pricing Tests

Amit Goyal **Zhongzhi (Lawrence) He** **Sahn-Wook Huh***

October 2019

Abstract

We propose a unified set of distance-based performance metrics that address the power problems inherent in traditional measures for asset-pricing tests. From a Bayesian perspective, the distance metrics coherently incorporate both pricing errors and their standard errors. Measured in units of return, the metrics have an economic interpretation as the minimum cost of holding a dogmatic belief in a model. Our metrics identify the six-factor model of Fama and French (2018), the q^5 model of Hou, Mo, Xue, and Zhang (2018), and the Stambaugh and Yuan (2017) model as the top performers whose performance is economically indistinguishable. By contrast, the *GRS* and average-alpha-based statistics often lead to counter-intuitive rankings.

***JEL Classification*:** C11; G11; G12

***Keywords*:** Asset-Pricing Tests; Power Problem; Distance-Based Metrics; Distance Equivalence; Optimal Transport Theory; Bayesian Interpretations; Model Comparisons and Rankings; Marginal Values of Individual Factors

---

*Amit Goyal is from the Swiss Finance Institute at the University of Lausanne, CH-1015 Lausanne, Switzerland (E-mail: amit.goyal@unil.ch; T: +41-21-692-3676), Lawrence He is from the Goodman School of Business, Brock University, St. Catharines, ON L2S 3A1, Canada (E-mail: zhe@brocku.ca; T: 905-688-5550 x.4540), and Sahn-Wook Huh is from the School of Management, University (SUNY) at Buffalo, Buffalo, NY 14260, USA (E-mail: swhuh@buffalo.edu; T: 716-645-5435). We thank Shaun Bond, Kent Daniel, Xuezhong He, Mi-Ae Michelle Kim, Ľuboš Pástor, Jiaping Qiu, Avanidhar Subrahmanyam, Jason Wei, Chen Xue, Tong Yu, and seminar participants at the University of Cyprus for valuable discussions and comments. Our special thanks go to Ken French for sharing the data on the Fama-French factors and portfolios, Lu Zhang for providing the data on the factors used in the q-factor models and on the Hou-Mo-Xue-Zhang portfolios, Ilias Filippou for help with the Green-Hand-Zhang portfolios, and AQR Capital Management for providing the data on the monthly updated value factor. Research assistance was ably provided by Seung-Oh Han. All errors are our own.

1. Introduction

Asset-pricing models are designed to explain excess returns of a universe of left-hand-side (LHS) assets using a small set of right-hand-side (RHS) factors. A variety of multifactor models have been proposed, and models that produce low pricing errors (alphas) with high estimation precision are deemed successful. However, the low-alpha criterion does not always go hand in hand with the high-precision standard in the existing performance metrics, giving rise to what is known as the "power problem" that have long plagued asset-pricing tests.

Barillas and Shanken (2018) point out that a relatively large p-value for a ratio-based statistic may tell us more about the imprecision in estimating a particular model's alphas than about the pricing ability of the model. Therefore, the lack-of-power problem (under-rejection) afflicts asset-pricing tests of the null hypothesis (of no pricing error), when the performance of asset-pricing models is evaluated based on the t-statistic for an individual asset or the F-statistic of Gibbons, Ross, and Shanken (1989) (*GRS*-statistic or *GRS* henceforth) for a group of assets. For example, Fama and French (2012, 2016a) report that *GRS* cannot reject global models in pricing regional stock returns such as in Japan. Similarly, Cochrane (2005) and De Moore, Dhaene, and Sercu (2015) caution against blowing up the residual covariance matrix, which causes a poor model to pass the statistical test.

The power problem also occurs when a ratio-based metric such as *GRS* rejects a model although its pricing errors are economically insignificant but their standard errors are small (too much power or over-rejection). For example, the three- and five-factor models of Fama and French (1993, 2015) produce an average pricing error of less than 0.10% per month for a wide range of test assets, but they are still rejected by *GRS*.

Consequently, most empirical studies rely on summary statistics of alphas jointly with the *GRS*-statistic to judge model performance. Alpha-based statistics, especially the mean absolute pricing error (or alpha) (*MAE*), are routinely reported as the main comparative results (e.g., Fama and French (2015, 2016a, 2016b), Hou, Xue, and Zhang (2015), Hou, Mo, Xue, and Zhang (2018), and Stambaugh and Yu (2017)), and the model with the lowest MAE is considered the best one. However, comparing models based on MAE is neither theoretically founded, nor does it consider the degree of estimation precision for model choices.[1] In particular, when GRS and MAE produce contradicting rankings (as commonly observed in the literature), which criterion should we use and what are the causes of the ranking inconsistency? To what extent does the power problem or ignoring estimation precision lead to sub-optimal model choices?

Given the above questions, the main objective of this study is to develop a unified set of performance metrics that are theoretically motivated and provide Bayesian interpretations for comparing the performance of asset-pricing models as well as assessing the value of specific factors in the models. Our Bayesian metrics are in contrast to the frequentist ratio-based (t- and F-) statistics. Specifically, a frequentist investor views the pricing error as a true (deterministic) but unknown parameter (α), and the measurement error (ϵ) prevents him from observing the true value of it, so that the estimated alpha is $\hat{\alpha} = \alpha + \epsilon$ and its estimation error is $\hat{\sigma}_\alpha$. The lack-of-power problem, thus, arises because of large $\hat{\alpha}$ but even larger $\hat{\sigma}_\alpha$. The too-much-power problem arises because of small (economically insignificant) $\hat{\alpha}$ but even smaller $\hat{\sigma}_\alpha$. In contrast, a Bayesian investor views the sample of data as given for updating her subjective beliefs about a model's mispricing, characterized by a posterior distribution of the alpha, $\alpha^{\text{post}} \sim N(\tilde{\alpha}, \tilde{\sigma}_\alpha^2)$, given the prior distribution, $\alpha^{\text{pre}} \sim N(0, \sigma_\alpha^2)$, characterized by prior estimation uncertainty, σ_α. The posterior estimates of the mean, $\tilde{\alpha}$, and the stadard deviation, $\tilde{\sigma}_\alpha$, measure the expected value of mispricing and the estimation uncertainty (or imprecision), respectively.

This probabilistic view of the mispricing parameters for a Bayesian investor makes it possible to define performance metrics that measure the "distance" between two posterior distributions with different degrees of prior estimation uncertainty, σ_α. Specifically, when a Bayesian investor holds a dogmatic belief in a model a priori (i.e., $\sigma_\alpha = 0$), her posterior

[1] Barillas and Shanken (2017) show examples in which evaluating the performance of models based on alpha-based statistics leads to inconsistent model rankings. However, they do not provide any guidance for the ad hoc practice or a solution to the problem.

estimate of the alpha shrinks to the model-implied value of zero with no uncertainty ($\alpha^{\text{post}} \sim N(0, 0^2)$) . At the other extreme, when the investor is completely skeptical about the model (i.e., $\sigma_\alpha = \infty$), the posterior estimates of the alpha and its estimation uncertainty conform to their OLS sample estimates ($\alpha^{\text{post}} \sim N(\hat{\alpha}, \hat{\sigma}_\alpha^2)$). We define a distance-based metric, average distance (AD), that measures the cost of moving the mass of $N(0, 0^2)$ to $N(\hat{\alpha}, \hat{\sigma}_\alpha^2)$ or vice versa.[2] Intuitively, a Bayesian investor views model performance as the gap between her subjective belief (i.e., model-implied distribution) and the objective reality (i.e., data-based distribution). Differently put, the distance metrics can be viewed as the minimum cost of holding a dogmatic belief in the model. To evaluate the performance of models, therefore, a Bayesian investor simply ranks models based on the size of the distance and then identifies the one with the shortest distance as the best model.

More specifically, the distance (AD) between the two distributions is defined using an $\mathcal{L}_2$ norm. For normal distributions, this leads to AD being defined to be the square root of the sum of the average of squared alphas ($\hat{\alpha}_i^2$) and the average of squared standard errors ($\hat{\sigma}_{\alpha_i}^2$). AD is analogous to GRS in that both summarize the overall performance of the model in a single measure. However, unlike the *ratio*-based GRS-statistic, AD (measured in units of return) is derived as a *sum*. In contrast to a frequentist investor who prefers a low ratio of the alpha estimate to its sampling error, a Bayesian investor views both large dispersion of the alpha and high estimation uncertainty as bad news.

AD is also akin to MAE in that it measures the average performance of a model. However, MAE is problematic in two aspects: (i) it weighs pricing errors with different magnitudes equally; and (ii) it also ignores the degree of estimation precision. In contrast to MAE, the AD metric heavily penalizes extreme pricing errors, choosing models that produce smaller dispersion of alpha estimates and higher estimation precision.

We also define a metric, marginal distance (d_i), to be the square root of the sum of

[2] The distance-based metrics that we propose are derived from the concept of the Wasserstein distance in the optimal transport theory (see Villani (2003, 2009)), and applied in economics and finance by Galichon (2016, 2017).

squared alpha ($\hat{\alpha}_i^2$) and squared standard error ($\hat{\sigma}_{\alpha_i}^2$) for asset i. Thus, d_i measures asset i's contribution to AD and can be used to identify 'troublesome' assets that most contribute to AD. In that sense, d_i for a Bayesian investor is analogous to the t_i-statistic for a frequentist investor. As before, t_i is a ratio-based statistic while d_i is based on a sum.

We note that AD can be used not only for pair-wise comparisons but also for simultaneously comparing a set of different models. These models may be nested or non-nested. By contrast, GRS can be used to compare only nested models in a frequentist setting. Moreover, it is not obvious how to account for estimation errors in inputs when comparing non-nested models.[3] Finally, as explained in more detail below, GRS-implied Sharpe ratio comparisons are suitable for the RHS approach to testing asset-pricing models, but the problem of the LHS approach for non-nested models has not been explored in the literature. These issues can be easily addressed in our Bayesian framework of the distance-based metrics, which can handle the LHS approach to testing nested or non-nested models.

In the empirical part of the paper, we evaluate asset-pricing models using AD as well as other performance measures often used in the literature on comparing models and choosing factors. We rank ten prominent asset-pricing models, which we classify into the following four categories: (1) A single-factor model: CAPM; (2) The Fama-French (FF) models: the three-factor FF3 model of Fama and French (1993), the five-factor FF5 model of Fama and French (2015), the six-factor FF6 model that uses FF5 along with the momentum factor (UMD), and the six-factor BKRS model of Barillas, Kan, Robotti, and Shanken (2019); (3) The q models: the four-factor q^4 model of Hou, Xue, and Zhang (2015) and the five-factor q^5 model of Hou, Mo, Xue, and Zhang (2018); and (4) Other models: the Bayes-factor BS model of Barillas and Shanken (2018), the mispricing factor SY model of Stambaugh and Yu (2017), and the behavioral DHS model of Daniel, Hirshleifer, and Sun (2018).

We perform our analyses for a variety of test assets. In particular, we use returns on decile and high-minus-low (H−L) portfolios formed by sorting on (1) 150 pooled portfolios

[3] Fama and French (2018) use a bootstrap analysis to account for estimation errors. Barillas, Kan, Robotti, and Shanken (2017) derive asymptotic results for comparing Sharpe ratios from non-nested models.

sorted on 15 different anomaly variables, constructed by Fama and French (2015); and (2) 300 pooled portfolios sorted on 30 anomaly variables, formed by Hou, Mo, Xue, and Zhang (2018). Lastly, to get around the potential biases associated with choosing test assets, we also use 85 H−L portfolio returns as test assets following Green, Hand, and Zhang (2017). Our sample period is from January 1972 to December 2015 (528 months) for the most part.

Leaving the details to the main text, we find that, with this universe of test assets, AD identifies FF6, q^5, and SY as the top three models. This ranking is in contrast to that generated by GRS or MAE, which often leads to counter-intuitive results. In particular, AD ranks FF6 highest among the ten models across different sets of test assets (except for Hou, Mo, Xue, and Zhang's (2018) H−L portfolios). However, using the GRS-statistic, a frequentist investor typically picks q^5 as the top model, since it produces the lowest ratio. The q^5 model is ranked highest not because of its best pricing ability but because of its higher estimation imprecision (causing lower power in GRS). Intuitively, this means that in cases where the two models produce the same level of pricing errors, FF6 estimates alphas more precisely than q^5. In general, however, the top three models produce the AD-statistics relatively close to each other.

Given that investors may not use any model as a dogma, nor do they consider the model as completely useless, we next allow investors to have positive values in prior estimation uncertainty (i.e., $0 < \sigma_\alpha < \infty$). For a given model, AD monotonically decreases as σ_α increases, reflecting the shrinkage effect of the posterior estimates. Using this property, we define a concept of 'distance equivalence,' which states that a Bayesian investor is indifferent between two models if their respective AD values (potentially calculated at different levels of prior uncertainty) are the same. We examine the distance-equivalent relations among the ten models using the 85 H−L portfolios. We find that q^5 and SY at $\sigma_\alpha = 2\%$ are distance-equivalent to FF6 at $\sigma_\alpha = 0$. This means that if a Bayesian investor chooses q^5 or SY as an alternative model over FF6 as the benchmark model, she is willing to accept prior estimation uncertainty in mispricing of $\pm 4\%$ per year. $\sigma_\alpha = 2\%$ is a modest level of prior uncertainty, according to Pástor and Stambaugh (2000). We thus infer that the performance difference among FF6, q^5, and SY is economically insignificant.

We also find that FF5 at $\sigma_\alpha = 4\%$ is distance-equivalent to FF6 at $\sigma_\alpha = 0$, implying that a Bayesian investor choosing FF5 over FF6 must accept prior mispricing of as much as $\pm 8\%$. This suggests that *UMD* is an important factor for the FF models. Furthermore, q^4 at $\sigma_\alpha = 4\%$ is distance-equivalent to q^5 at $\sigma_\alpha = 0$, indicating that adding the expected growth factor (*EG*) to q^4 is economically warranted. The key advantage of adding *EG* is to make q^5 generate lower pricing errors; however, the downside is that q^5 estimates alphas less precisely. Nonetheless, the benefit of smaller pricing errors appears to dominate the adverse effect of lower estimation precision for the q^5 model.

The concept of distance equivalence also allows us to evaluate the economic value of individual factors and their combinations for any model. We do so by excluding some factors from a model and computing the *AD*-statistics of the parsimonious model (that deletes the factor(s)). We do this exercise for the top three benchmark models (FF6, q^5, SY) using the 85 H$-$L portfolios as test assets. We find that *MKT* is the single most important factor for both FF6 and q^5, consistent with one of the key findings in Harvey and Liu (2018). Interestingly, we also find that the *EG* factor exerts a strong effect, making the role of other factors in the q^5 model useless or redundant: i.e., the model without *IA* and/or *ROE* produces an even lower *AD*-statistic than the benchmark q^5 model does. Thus, the new model excluding both *IA* and *ROE* from q^5 (with only the *MKT*, *ME*, and *EG* factors) performs best among the set of q models (although the advantage of the parsimonious model measured by *AD* does not always extend to the cases where different sets of portfolios are used as test assets).

Our paper builds upon the Bayesian setting of Pástor and Stambaugh (2000), and shares the same prior specifications as their study. However, our paper differs from theirs in its objectives. The utility-based metrics of their study are designed to examine the impact of varying degree of prior beliefs on portfolio choices, but not to choose one asset-pricing model over another. Our distance-based metrics can be used not only to measure the performance of an asset-pricing model, but also to compare and rank different models (both nested and non-nested) or to evaluate the value of factors in them.

In this regard, our paper shares the same purpose as Barillas and Shanken (2018). But it is different in the methodology. First, Barillas and Shanken use the Bayes factor to compute the posterior model probabilities and then choose the best set of factors. However, our distance-based metrics are cost measures (in units of return), carrying intuitive economic interpretations that are easy to communicate. Second, Barillas and Shanken (2018) belongs to the RHS approach based on spanning regressions of the candidate factors, whereas ours is the LHS approach designed to identify the model that best prices the universes of test assets. Motivated by Fama (1998), the RHS approach runs spanning regressions to evaluate the marginal effects of factors or their potential redundancy, and it often uses the *GRS*-statistic for inferences. This approach has been advocated by Barillas and Shanken (2017, 2018) and recently used in Fama and French (2018). However, Harvey and Liu (2018) argue that the RHS approach leads to poor results when the number of assets is large, in which case the power problem of *GRS* is severe. We stay agnostic on which approach is better for testing asset-pricing models and contribute only to the literature that uses the LHS approach.[4]

In terms of choosing the best models, our paper shares the same purpose as Ahmed, Bu, and Tsvetanov (2019) who rely on the traditional statistical methods. Our paper differs from theirs in that we propose a set of novel metrics to address the power problems inherent in traditional measures. Our distinct finding is that, despite their motivational and statistical differences, FF6 (empirical-based), q^5 (theory-based), and SY (behavioral-based) are economically indistinguishable as the top three models.

Our Bayesian solution to the power problems inherent in the existing metrics is surprisingly simple. In its frequentist-equivalent form, the OLS estimates of pricing errors and their standard errors from time-series regressions are the only required inputs to construct

[4]The LHS approach is still dominant in testing asset pricing models. For examplee, see Daniel, Hirshleifer, and Sun (2018), Fama and French (2015, 2016a, 2016b), Harvey and Liu (2018), Hou, Xue, and Zhang (2015), Hou, Mo, Xue, and Zhang (2018), and Stambaugh and Yuan (2017), among others.

the distance-based metrics: that is, at no more cost than constructing the GRS-statistic. The only difference is how to view pricing errors and their standard errors. Instead of taking a ratio, a Bayesian investor uses the square-root of the sum of the two.

2. Analytical Framework

We start with the following thought experiment to distinguish the Bayesian view from the frequentist view on asset-pricing tests, motivating the rationale for the distance metrics.

Consider a frequentist investor evaluating two models with a test asset. From OLS regressions, Model A generates $\hat{\alpha}_i^A = 5\%$ and $\hat{\sigma}_{\alpha_i}^A = 0\%$ per year, and Model B generates $\hat{\alpha}_i^B = 5\%$ and $\hat{\sigma}_{\alpha_i}^B = 5\%$ per year. Assuming both alpha estimates follow a normal distribution, the sampling theory suggests that Model A estimates the alpha precisely at 5% with no sampling error, while Model B does so within the sampling error ranging from -10% to 10% at the 95% confidence level. Using the ratio-based metric (t_i), the frequentist investor would reject Model A (i.e., $t = \infty$) and choose Model B (i.e., $t = 1$) unequivocally.[5]

However, a rational investor may not consider Model B universally better, because there is roughly a one-third chance that $\hat{\alpha}_i^B$ is outside a one-standard-error range ($\hat{\alpha}_i^B < 0\%$ or $\hat{\alpha}_i^B > 10\%$), a 5% chance that $\hat{\alpha}_i^B$ is outside a two-standard-error range ($\hat{\alpha}_i^B < -5\%$ or $\hat{\alpha}_i^B > 15\%$), and the sampling error can be arbitrarily large with positive probabilities. Thus, a natural question is: how do we compare Model A that is precisely wrong with Model B that is imprecisely wrong? The existing ratio-based statistics (t_i for an individual asset and GRS for a group of assets) do not provide a satisfactory answer to this question.

Now consider a Bayesian investor viewing model performance as the shortest 'distance' between the model-implied posterior distribution $N(0, 0^2)$ (when she holds dogmatic prior belief in the model) and the data-based posterior distribution (when she holds complete skepticism in the model). Economically, the distance is the minimum cost (price to pay)

[5] A frequentist would choose Model B over Model A, even if $\hat{\alpha}_i^A = 0.1\%$ per year (economically insignificant, but still $t = \infty$ with $\hat{\sigma}_{\alpha_i}^A = 0\%$) and $\hat{\alpha}_i^B = 9\%$ per year (economically significant, but still $t < 2$ with $\hat{\sigma}_{\alpha_i}^A = 5\%$). This represents over-rejection of Model A and under-rejection of Model B.

to change one's dogmatic belief in a model to a complete disbelief in it; or intuitively, it is the minimum cost of holding a dogmatic belief in the model. For Model A, getting the distance between $N(0, 0^2)$ and $N(5\%, 0^2)$ requires only a mean shift of the mass of 5%, so that the transport cost is exactly 5%. For Model B, the distance is the minimum cost of moving the mass of $N(5\%, 5\%^2)$ to $N(0, 0^2)$, which involves both a mean shift from 5% to zero and a variance reduction to zero. The minimum cost for this transport is simply $\sqrt{5\%^2 + 5\%^2} \approx 7.07\%$ by our $\mathcal{L}_2$-norm definition of distance (to be derived below). Because she incurs a higher cost for moving the mass in Model B than in Model A, the Bayesian investor chooses Model A over Model B.

As such, our distance-based metrics summarize both the magnitude of alpha estimates and the estimation imprecision into an intuitive cost measure, favoring models that produce small alphas that are estimated more precisely. The rest of this section formalizes this intuition.

2.1. The Bayesian Setup

Following Pástor (2000) and Pástor and Stambaugh (2000), we assume that an investor has T observations on n assets. Let R is a $T \times n$ matrix of asset returns in excess of risk free rate, and $X = [\iota_T \, F]$ is a $T \times (k+1)$ matrix, where the first column (ι_T) is a $T \times 1$ vector of ones and the remaining k columns (F) contain a $T \times k$ matrix of factor returns. Consider a multivariate regression of R on X:

$$R = XB + U, \quad \text{vec}(U) \sim N\left(0, \Sigma \otimes I_T\right), \tag{1}$$

where $B = \begin{bmatrix} \alpha' \\ \beta' \end{bmatrix}$ is a $(k+1) \times n$ matrix in which the first row is a $1 \times n$ vector of alphas, and the second row contains a $k \times n$ matrix of factor loadings, and I_T is an identity matrix of rank T. Also, $\text{vec}(\cdot)$ denotes an operator that stacks the columns of a matrix into a vector, and $\otimes$ is the Kronecker product. The rows of the disturbance matrix U are assumed to be serially uncorrelated and homoscedastic with an $n \times n$ covariance matrix Σ. The setup of equation (1) is the classic multivariate regression model considered by Zellner (1971) and applied by Pástor and Stambaugh in traditional portfolio problems.

An investor's prior belief about α and β is given by the following multivariate normal distribution:

$$p(B|\Sigma) \sim N\left(\begin{bmatrix}\alpha_0' \\ \beta_0'\end{bmatrix}, \begin{bmatrix}\sigma_\alpha^2(\Sigma/s^2) & 0 \\ 0 & \Psi\end{bmatrix}\right). \tag{2}$$

In general, α_0, which is the $n \times 1$ prior mean of α, can take any non-zero values to reflect the investor's proprietary views on the level of mispricing for the set of returns on the LHS assets used for the regression in equation (1). If the prior mean of alphas is centered at zero ($\alpha_0 = 0$: no mispricing in the model), then σ_α represents the investor's prior beliefs about the degree of estimation uncertainty. When $\sigma_\alpha = 0$, the investor has dogmatic beliefs in the model, so mispricing is completely ruled out. When $\sigma_\alpha = \infty$, the investor regards the model as useless and relies solely on the data to detect the level of mispricing of a given model. Between these two extreme views, σ_α can take a range of values to express the investor's prior beliefs in the model.

Note that in equation (2) the prior uncertainty of α_0', $\sigma_\alpha^2(\Sigma/s^2)$, is proportional to the residual covariance matrix, Σ, to reflect the fact that very large mispricing opportunities are improbable.[6] s^2 is a scalar whose value is set equal to the average of the diagonal elements of the sample estimate of Σ to make Σ/s^2 invariant to scaling. Finally, the prior distribution of factor loadings is also typically centered at zero ($\beta_0 = 0$), with a diagonal covariance matrix, Ψ, whose elements take very large values so that the prior distribution for factor loadings is non-informative.

The prior distribution of Σ is typically specified as an inverted-Wishart distribution with

[6] Pástor (2000) and Pástor and Stambaugh (2000) provide a detailed discussion for this prior specification. Barillas and Shanken (2018) also use this prior and regard $k = \sigma_\alpha^2/s^2$ as the information ratio. He (2007) interprets σ_α as the active risk budget assigned to asset managers based on investment policies.

degree of freedom $\upsilon_0 = n + 2$:

$$p(\Sigma) \sim IW(H_0, \upsilon_0) \propto |\Sigma|^{-\frac{\upsilon_0+n+1}{2}} \exp\left\{-\frac{1}{2}\mathrm{Tr}\left(H_0\Sigma^{-1}\right)\right\}, \tag{3}$$

where the scaling matrix is $H_0 = E[\Sigma] = s^2 I_n$ and $\mathrm{Tr}(\cdot)$ is the trace operator.

The prior distributions in equations (2) and (3) are combined with the likelihood function of equation (1) to derive the posterior estimates of regression parameters. In particular, for our choice of $\alpha_0 = 0$ and $\beta_0 = 0$, the posterior means of α and β have a simple and intuitive form (tildes are used to denote the posterior means of the parameters):

$$\tilde{B} = \begin{bmatrix} \tilde{\alpha}' \\ \tilde{\beta}' \end{bmatrix} = (V_0^{-1} + X'X)^{-1}X'R, \tag{4}$$

where $V_0^{-1} = \begin{bmatrix} s^2/\sigma_\alpha^2 & 0 \\ 0 & 0 \end{bmatrix}$ is a $(k+1) \times (k+1)$ matrix. The posterior variance of alpha, $\tilde{V_\alpha} = \mathrm{Var}\,[\alpha|R, F]$, is taken from the (n, n) upper left block of the $n(k+1) \times n(k+1)$ matrix, $\tilde{V_\alpha} \otimes \tilde{\Sigma}$, defined in Appendix A.

The shrinkage effect for $\tilde{\alpha}$ is readily seen in equation (4). For $\sigma_\alpha = \infty$ (complete disbeliefs), $\tilde{B} = (X'X)^{-1}X'R$ reduces to the OLS estimates of α and β, which are based solely on the sample (data) information. For $\sigma_\alpha = 0$ (dogmatic beliefs), $\tilde{\beta}$ stays as the OLS estimates but $\tilde{\alpha}$ reduces to the theoretical value of zero. For any other value of σ_α between 0 and ∞, $\tilde{\alpha}$ is a weighted average of zero and the OLS estimates, with the respective weights being determined by the relative confidence in the prior beliefs (captured by V_0^{-1}) and in the sample of data (captured by $X'X$).

For an asset-pricing model, its mispricing is characterized by its posterior distribution of the alpha, $p(\alpha|R, F, \sigma_\alpha)$, with the value of σ_α pre-specified from 0 (exact pricing by the model) to ∞ (uselessness of the model) to reflect varying degrees of prior confidence in the model. Thus, model performance is formulated as the problem of comparing two posterior distributions of the alpha. The optimal transport theory presented below provides useful metrics that measure the shortest distance between two probability distributions.

2.2. The Optimal Transport Theory and the Wasserstein Distance

Our distance metrics are based on the optimal transport theory rooted in mathematics (Villani (2003, 2009)) with rich applications in economics (Galichon (2016)) and econometrics (Galichon (2017)). The classic problem (Monge (1781)) is to find the shortest distance or the minimum cost to move the mass of one probability distribution to another. This is accomplished by defining a quadratic Wasserstein distance between two probability distributions. This distance measure has an economic interpretation as the minimum expected cost of transporting the mass of one distribution to another distribution. In general, there exists no closed-form formula for this distance measure for general probability distributions. Fortunately, when the two distributions are Gaussian, a closed-form formula for the Wasserstein distance has an analytical form.

Relegating the details to Appendix B, we present only the definition here. Let P_I and P_{II} be Gaussian measures on $\mathbb{R}^n$ with finite second moments such that $P_I \sim N(\alpha_I, V_I)$ and $P_{II} \sim N(\alpha_{II}, V_{II})$, where α_I and α_{II} are two $n \times 1$ vectors of mean, and V_I and V_{II} are two $n \times n$ symmetric, positive-definite covariance matrices. Then, the $\mathcal{L}_2$-norm quadratic Wasserstein distance (WD_2) between P_I and P_{II} is given by

$$\begin{aligned} WD_2 &= \sqrt{||\alpha_{II} - \alpha_I||^2 + ||V_{II} - V_I||} \\ ||V_{II} - V_I|| &= \mathrm{Tr}\left(V_I + V_{II} - 2(V_I^{1/2} V_{II} V_I^{1/2})^{1/2}\right), \end{aligned} \tag{5}$$

where $||\alpha_{II} - \alpha_I||$ is the Euclidean 2-norm of the mean difference vector, $||V_{II} - V_I||$ is the distance between the two covariance matrices, and $V^{1/2}$ is the square-root of the covariance matrix such that $V = V^{1/2}V^{1/2}$.[7] To use this distance measure in our Bayesian

[7] For symmetric and positive-definite V matrix, $V^{1/2}$ is unique, symmetric, and positive-definite. $V^{1/2}$ is computed using the Schur algorithm (Deadman, Higham, and Ralha (2013)). Python library

setting, the first two moments, (α_I, V_I) of P_I and (α_{II}, V_{II}) of P_{II}, are replaced with their model-generated posterior estimates of the alpha and its variance, $(\tilde{\alpha}_I, \tilde{V}_{\alpha_I})$ and $(\tilde{\alpha}_{II}, \tilde{V}_{\alpha_{II}})$, respectively, where I and II represent two distinct distribution specifications about prior mispricing uncertainty (σ_α) for a given asset-pricing model.

In particular, let prior specification I be set as $\sigma_\alpha = 0$ (complete confidence in the model's pricing ability); under such dogmatic beliefs, there is no mispricing uncertainty and hence the posterior estimate of the alpha shrinks to its theoretical value of zero: i.e., both $\tilde{\alpha}_I$ and $\tilde{V}_I$ are zero. On the other hand, let prior specification II be set as $\sigma_\alpha = \infty$ (complete skepticism about the model's pricing ability), in which case the posterior estimates $(\tilde{\alpha}_{II}, \tilde{V}_{\alpha_{II}})$ shrink to their sample estimates based entirely on the sample of data. Given such prior specifications, the quadratic distance metric reduces to $WD_2 = \sqrt{||\tilde{\alpha}_{II}||^2 + \mathrm{Tr}\left(\tilde{V}_{\alpha_{II}}\right)}$. Note that equation (5) was derived under the assumption of normality. However, under the special case of prior specification I with $\sigma_\alpha = 0$, the expression for WD_2 is valid under an arbitrary posterior distribution for alpha. This is because, under the $\mathcal{L}_2$-norm, the distance measure is simply (square root of) the average of squared alphas, which is equal to the sum of square of the average alpha and the variance of alpha $(\mathrm{E}(\mathrm{X}^2) = \mathrm{E}(\mathrm{X})^2 + \mathrm{V}(\mathrm{X}))$.

2.3. The Distance Metrics

Given the non-informativeness in prior specification II, the posterior estimates $\tilde{\alpha}_{II}$ and $\tilde{V}_{\alpha_{II}}$ are identical to the maximum-likelihood estimates of the alpha, $\hat{\alpha}$, and its covariance matrix, $\hat{V}_\alpha$, respectively. Using these facts and WD_2 above, we define three types of distance-based metrics, Total Distance (TD), Average Distance (AD) and Marginal Distance (d_i), as follows:

$$
\begin{aligned}
TD &= \sqrt{\sum\nolimits_{i=1}^{n} (\hat{\alpha}_i^2 + \hat{\sigma}_{\alpha_i}^2)} \\
AD &= \sqrt{\sum\nolimits_{i=1}^{n} (\hat{\alpha}_i^2 + \hat{\sigma}_{\alpha_i}^2) \Big/ n} \equiv \sqrt{MSE_{\hat{\alpha}} + MSE_{\hat{\sigma}_\alpha}} \\
d_i &= \sqrt{\hat{\alpha}_i^2 + \hat{\sigma}_{\alpha_i}^2}\ , \qquad (6)
\end{aligned}
$$

'`scipy.linalg.sqrtm()`' implements this algorithm.

where n is the number of LHS test assets, $\hat{\sigma}_{\alpha_i} = \hat{V}_{\alpha}^{1/2}(i,i)$ is the posterior estimate of the standard error of the alpha, $\hat{\alpha}_i$, for asset i (same as the standard error of $\hat{\alpha}_i$), $MSE_{\hat{\alpha}}$ is the average squared intercept, and $MSE_{\hat{\sigma}_\alpha}$ is the average squared standard error of the intercept.

The Bayesian interpretations of the above metrics are as follows. TD is the minimum total cost of transporting the mass of the model-implied distribution (complete confidence in the model) to the data-based distribution (complete skepticism about the model), or the shortest total distance between the model-implied distribution and the data-based distribution. Similarly, AD is the shortest average distance between the model-implied distribution and the data-based distribution; differently put, AD is the minimum average cost of holding a dogmatic belief in the model.[8] Finally, d_i is the marginal contribution of asset i to TD or AD; equivalently, d_i is the marginal cost of holding a dogmatic belief in the model.

A Comparison of *AD* with *GRS*

AD is akin to the frequentist GRS F-statistic, $\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}$.[9] Both AD and GRS summarize the overall performance of a given model by a single measure that carries economic interpretations. Specifically, the core of the GRS-statistic is the difference between the maximum squared Sharpe ratio of both factors and assets, $\text{Sh}^2(F,R)$, and that of the factors alone, $\text{Sh}^2(F)$. Somewhat similarly, AD has a Bayesian interpretation as the minimum average cost of holding dogmatic beliefs in the model.

Despite their identical estimates of α and V, however, there is a key difference between the frequentist interpretation of GRS and the Bayesian interpretation of AD. The GRS-statistic is the ratio of the sum of squared alphas to the covariance matrix of alpha estimates. Ignoring the correlations, the GRS statistic can be viewed as $MSE_{\hat{\alpha}}/MSE_{\hat{\sigma}_\alpha}$. The frequentist interpretation of the reciprocal of this ratio is best illustrated by Fama and French (2016b, p. 78) "… high values of $MSE_{\hat{\sigma}_\alpha}/MSE_{\hat{\alpha}}$ are good news: they say that much of the dis-

[8] The Wasserstein distance defined in equation (5) depends on the number of test assets, and thus we may not compare the performance across different models when the universes of test assets are different. The AD metric solves this issue as it describes the same properties in terms of average (rather than total) distance.

[9] In the AD formula in equation (6), the covariance terms do not appear (i.e., all the elements of the covariance matrix are zero) given the assumption of a dogmatic belief ($\sigma_\alpha = 0$) in a model. However, when we assume $\sigma_\alpha > 0$ in Section 6, the covariance matrix has non-zero values.

persion of the intercept estimates is due to sampling error rather than to dispersion of the true intercepts.” Thus, a frequentist views a model that produces a higher ratio, or a lower *GRS*-statistic, as a better model.[10] The Bayesian interpretation views both large alphas and large standard errors as bad news, because they together contribute to enlarging the total (or average) distance. Our distance-based measures effectively resolve this power problem by requiring that a good model have *both* $\hat{\alpha}$ and $\hat{\sigma}_\alpha$ be small.

A Comparison of *AD* with *MAE*

AD is also somewhat similar to $MAE_{\hat{\alpha}} = \sum_{i=1}^{n} |\hat{\alpha}_i| / n$. But the inequality below gives the comparative statics for the performance-related metrics:

$$AD > RMSE_{\hat{\alpha}} \geq MAE_{\hat{\alpha}} \tag{7}$$

The first strict inequality follows from the notion that any model is an incomplete description of cross-sectional expected returns (e.g., Fama and French (2015)), because $RMSE_{\hat{\sigma}_\alpha}$ is strictly positive (i.e., alpha estimates are imprecise). The second inequity is a standard statistical property of *RMSE*, and the equality holds if and only if the assets have identical values of pricing errors. Also, the wider the dispersion of pricing errors, the larger the gap between $RMSE_{\hat{\alpha}}$ and $MAE_{\hat{\alpha}}$.

To illustrate the distinction between *RMSE* and *MAE*, suppose that two different models (A and B) are compared on two test assets: pricing errors produced by Model A are 0.15% and 0.17% per month, and those produced by Model B are 0.05% and 0.25% per month. If ranked by *MAE*, Model B is better since $MAE(\hat{\alpha}^B) = 0.15\% < MAE(\hat{\alpha}^A) = 0.16\%$. However, if ranked by *RMSE*, Model A is better since $RMSE(\hat{\alpha}^A) = 0.16\% < RMSE(\hat{\alpha}^B) = 0.18\%$. This example illustrates a notable property of the *RMSE* criterion: it gives higher weights to large pricing errors. As a result, the *RMSE* criterion views models that produce

[10] The *GRS*-statistic suffers from an additional problem – the inverse of the covariance matrix Σ may cause the statistic to be large, leading to rejections of models that may have small values of $MSE_{\hat{\alpha}}$. Relatedly, the weighting scheme in $\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}$ typically have extreme long and short positions that may not reflect the relative importance of factors (Fama and French (2018) and Harvey and Liu (2018)).

extreme alphas as particularly undesirable, penalizing those models more heavily than the MAE criterion.

As a more realistic example, consider a model that produces low alphas for most of assets but extreme alphas for just a few assets (e.g., the Fama and French (2015) 5-factor model for most anomaly portfolios vs. the momentum portfolio). This model's $RMSE_{\hat{\alpha}}$ is dominated by the few extreme alphas. Therefore, the $RMSE_{\hat{\alpha}}$ criterion would rank this model lower than its competitors that produce a higher $MAE_{\hat{\alpha}}$ but fewer extreme alphas. Note further that the $RMSE_{\hat{\alpha}}$ properties equally apply to $RMSE_{\hat{\sigma}_{\alpha}}$, which penalizes low-precision models, especially those that produce extreme standard errors. By contrast, $MAE_{\hat{\alpha}}$ does not consider estimation precision of the model at all, assigning equal weights to different size of pricing errors.

Given the fact that AD accounts for the number of assets and incorporates both components ($MSE_{\hat{\alpha}}$ and $MSE_{\hat{\sigma}_{\alpha}}$) in assessing performance, we use AD as our primary distance-based metric for ranking different asset-pricing models. As illustrated above, AD tends to favor models that produce: (1) low pricing errors; (2) lower standard errors of them (i.e., high estimation precision); and (3) fewer extreme pricing errors and their standard errors.

<u>A Comparison of d_i with t_i</u>

d_i is equivalent to the t-statistic ($t_i = \hat{\alpha}_i/\hat{\sigma}_{\alpha_i}$) for testing the statistical significance of a pricing error for asset i. While d_i and t_i use the same inputs, they are quite different in the ways of computations and interpretations. The t-statistic is a ratio-based measure, favoring models that produce smaller values: from a frequentist perspective, a small t-statistic means that the estimate of a pricing error (alpha) is insignificantly different from zero relative to its sampling error (standard error of the alpha). However, an insignificant t-statistic may not be attributable to a small estimate of the alpha, but instead to an inflated standard error, resulting in high estimation imprecision for individual assets. By contrast, the Bayesian investor views d_i as asset i's marginal contribution to the total (or average) cost of holding complete confidence in a given model that prices the universe of assets. Both large pricing error and large standard error are bad news for the model. Therefore, d_i is useful for singling out such individual assets that most contribute to the total distance.

In Table 1, we summarize the comparisons of the performance metrics: *AD*, *GRS*, and $MAE_{\hat{\alpha}}$. Based on the comparisons and hypothetical examples, we have identified the power problem inherent in the ratio-based *GRS*- and t-statistics. In addition, the mean absolute alpha ($MAE_{\hat{\alpha}}$) often causes a problem since it gives equal weights to extreme pricing errors and completely ignores the degree of estimation precision in asset-pricing models. To what extent are these problems observed in data, leading to inconsistency in model rankings? How does the model ranking based on the distance-based metrics compare to that based on *GRS* or $MAE_{\hat{\alpha}}$? What are the advantages or problems of asset-pricing models commonly used in the literature? These are the empirical issues that we try to explore in the sections below.

3. Factors, Test Assets, and Performance Metrics

3.1. Models and Factors

We limit our comparative analyses to ten prominent asset-pricing models (and 17 different factors). The ten models are categorized into four groups, with their corresponding factors shown in the parentheses, as follows:

Single-Factor Model	q Models
CAPM (*MKT*)	q^4 (*MKT*, *ME*, *IA*, *ROE*)
FF Models	q^5 (*MKT*, *ME*, *IA*, *ROE*, *EG*)
FF3 (*MKT*, *SMB*, *HML*)	Other Models
FF5 (*MKT*, *SMB*, *HML*, *RMW*, *CMA*)	BS (*MKT*, *SMB*, HML^m, *IA*, *ROE*, *UMD*)
FF6 (*MKT*, *SMB*, *HML*, *RMW*, *CMA*, *UMD*)	SY (*MKT*, *SMB*, *MGMT*, *PERF*)
BKRS (*MKT*, *SMB*, HML^m, *RMWCP*, *CMA*, *UMD*)	DHS (*MKT*, *FIN*, *PEAD*)

We only briefly describes the above models and factors, since they are often used in the asset-pricing literature. To gauge the performance of a basic single-factor model, we include CAPM, which uses the market factor (*MKT*) only. FF3 and FF5 are the Fama and French (1993, 2015) three- and five-factor models, the latter of which employs the profitability (*RMW*) and investment (*CMA*) factors. FF6, used in Fama and French (2018), adds Carhart's (1997) *UMD* factor to the FF5 model. As one of the FF-model group, BKRS is the six-factor model of Barillas, Kan, Robotti, and Shanken (2019), which is a variant of FF6 in two ways. First, it uses a monthly updated value factor (HML^m) proposed by Asness and Frazzini (2013), instead of the usual *HML*. Second, BKRS also uses a cash-based profitability factor (*RMWCP*), replacing the original (accrual-based) operating profitability factor (*RMW*) used in FF6.

There are two models in the q-group. q^4 is Hou, Xue, and Zhang's (2015) q-factor model consisting of four factors, three of which are the size (*ME*), investment-to-asset (*IA*), and profitability (*ROE*) factors. Hou, Mo, Xue, and Zhang (2018) add the expected growth factor (*EG*), which results in the q^5 model.

Among the 'Other' group, BS is the Barillas and Shanken (2018) model, a mixture of FF6 and q^4. SY is the four-factor model of Stambaugh and Yuan (2017) that employs two mispricing factors, management (*MGMT*) and performance (*PERF*), in addition to *MKT* and *SMB*. Finally, DHS is a three-factor model that contains two behavioral factors related to financing (*FIN*) and post-earnings-announcement drift (*PEAD*), suggested by Daniel, Hirshleifer, and Sun (2018).

Data on the above factors are obtained from the authors' websites (or, in some cases, directly from the authors). The sample period is from January 1972 to December 2015 spanning 528 months, during which period most of the 17 factors are available, except for the DHS model whose factors are available from January 1972 to December 2014.

3.2. Test Assets

We first use the returns on the 25 (5×5) portfolios sorted on firm size (Size) and momentum (MOM) to illustrate the power problems in the two traditional performance metrics. Next, we choose two broader sets of decile portfolios used in recent asset-pricing studies: (1) 150 pooled portfolios sorted on 15 different anomaly variables, constructed by Fama and French (2015); and (2) 300 pooled portfolios sorted on 30 anomaly variables formed by Hou, Mo, Xue, and Zhang (2018). The returns on the 25 Size-MOM portfolios and the 150 pooled decile portfolios are available from the Kenneth French's website. The returns on the 300 portfolios are obtained from Lu Zhang.

Studies often use all sets of returns (in excess of the risk-free rate) on the portfolios formed by sorting on various anomaly variables, or only the H−L returns computed using the above sets of anomaly-sorted portfolios. Accordingly, we also test the performance of the ten models using only the 15 and 30 H−L returns computed using the above two different universes of pooled decile portfolios.

Lastly, to get around the potential biases associated with choosing test assets, we also use 85 H−L portfolio returns as test assets following Green, Hand, and Zhang (2017). These portfolios are constructed by a third party independent of the authors of the FF- and q-models. These data are also used in Abhyankar, Filippou, Garcia-Ares, and Haykir (2018). We obtain these data from Ilias Filippou; these data are available from January 1980 to December 2015. The definitions of the 15, 30, and 85 sorting variables used to construct the test assets are briefly summarized in Appendix C. For further details on the construction of these portfolios, refer to Fama and French (2015), Hou, Mo, Xue, and Zhang (2018), and Green, Hand, and Zhang (2017).

3.3. Performance Metrics

For each LHS portfolio, we try to explain its average excess return (or, in the case of long-short portfolio, just the return), $\bar{r}_i$, by the factors (F) of each model by running a time-series regression:

$$r_{it} = \alpha_i + \beta_i' F_t + \epsilon_{it}. \tag{8}$$

Using estimates of alpha ($\hat{\alpha}_i$) and its standard error ($\hat{\sigma}_{\alpha_i}$), we compute the distance-based metrics (AD and d_i), as well as the following metrics:

$$\begin{aligned}
RMSE_{\hat{\alpha}} &= \sqrt{\sum_{i=1}^{n} \hat{\alpha}_i^2 \Big/ n} \\
RMSE_{\hat{\sigma}_\alpha} &= \sqrt{\sum_{i=1}^{n} \hat{\sigma}_{\alpha_i}^2 \Big/ n} \\
MAE_{\hat{\alpha}} &= \sum_{i=1}^{n} |\hat{\alpha}_i| \Big/ n \\
MAR_{\bar{r}} &= \sum_{i=1}^{n} |\bar{r}_i| \Big/ n \\
GRS &= ((T-n-k)/n) \times \left(1 + \mathrm{Sh}^2(F)\right)^{-1} \times \hat{\alpha}' \hat{\Sigma}^{-1} \hat{\alpha} \ , \qquad (9)
\end{aligned}$$

where T is the number of observations (months), n is the number of assets (i.e., portfolios), k is the number of factors, $\mathrm{Sh}^2(F)$ is the squared Sharpe ratio for the factors, ($\mathrm{Sh}(F)$ is calculated as $\bar{F}' \hat{\Sigma}_F^{-1} \bar{F}$, where $\bar{F}$ is a $k \times 1$ vector of the mean returns on the factors and $\hat{\Sigma}_F$ is the $k \times k$ variance-covariance matrix of the factor returns), and $\hat{\Sigma}$ is the maximum likelihood estimate of the residual covariance matrix.

$RMSE_{\hat{\alpha}}$ and $RMSE_{\hat{\sigma}_\alpha}$ are the two components in $AD \equiv \sqrt{MSE_{\hat{\alpha}} + MSE_{\hat{\sigma}_\alpha}}$, and help us understand the contribution of pricing error and its estimation uncertainty in the overall distance score.[11] The remaining metrics are used often in the literature (e.g., in Fama and French (2015, 2016a, 2016b), Hou, Mo, Xue, and Zhang (2018), and Stambaugh and Yu (2017)). $MSE_{\hat{\sigma}_\alpha}/MSE_{\hat{\alpha}}$ measures the contribution of mispricing uncertainty relative to the dispersion of pricing errors. $MAE_{\hat{\alpha}}$ is the mean absolute pricing error. $MAE_{\hat{\alpha}}/MAR_{\bar{r}}$ measures the proportion of unexplained average returns. AR^2 measures the cross-sectional average of R^2s from time-series regressions with individual assets (portfolios).[12]

[11] $RMSE_{\hat{\alpha}}$ gives the same model rankings as the proportion of variances in the LHS expected returns not explained by a model, i.e., $A\alpha_i^2/A\bar{r}_i^2$ in Fama and French (2016).

[12] Other reported statistics include: the number of significant alphas based on t-statistics (Hou, Xue, and Zhang (2015), and Hou, Mo, Xue, and Zhang (2018)), the number of rejections by the GRS test (Hou, Xue, and Zhang (2015), and Hou, Mo, Xue, and Zhang (2018)), the number of anomalies for which a model produces the smallest absolute alpha (Stambaugh and Yu (2017)), the average absolute t-statistic (Stambaugh and Yu (2017)), and the maximum squared Sharpe ratio for factors and alphas (Fama and French (2018)).

4. An Illustrative Example Using Size-MOM Portfolios

We use a set of 25 (5×5) portfolios formed by sorting on firm size (Size) and momentum (MOM) as test assets to illustrate how the power problem afflicts traditional performance metrics and how the distance metrics effectively address the problem.

4.1. Performance Metrics for Individual Portfolios

Table 2 reports the posterior estimates of alphas ($\hat{\alpha}_i$), their standard errors ($\hat{\sigma}_{\alpha_i}$), as well as t-statistics (t_i), and the marginal distances (d_i) computed with non-informative priors, generated using the 25 Size-MOM portfolios as test assets. In the table, we compare two models: FF6 (in Panel A) vs. q^5 (in Panel B). Reported in the lower part of each panel are various performance metrics for joint tests, which allow us to compare across the two models. We find that GRS ranks q^5 higher than FF6. However, $MAE_{\hat{\alpha}}$ reverses this ranking. The distance-based metric AD produces a ranking similar to that by $MAE_{\hat{\alpha}}$: i.e., both metrics rank FF6 higher than q^5. We examine in more detail why the power problem of GRS and the ignorance of the power problem by $MAE_{\hat{\alpha}}$ lead to counter-intuitive rankings and how the distance-based metrics address the problems.

We start by looking at estimates of alphas ($\hat{\alpha}_i$) reported in the first five columns in Table 2. Panel A shows that FF6 produces economically significant pricing errors in (1,1) (small/loser), (1,5) (small/winner), and (2,5) portfolios with respective alpha estimates of −0.24%, 0.39%, and 0.24%. $MAE_{\hat{\alpha}}$ is 0.110% and $RMSE_{\hat{\alpha}}$ is 0.143% ($RMSE_{\hat{\alpha}} \geq MAE_{\hat{\alpha}}$ in inequality (7)). Pricing errors for the q^5 model are economically significant for (1,4), (1,5), and (2,5) portfolios with magnitudes of 0.24%, 0.44%, and 0.31%, respectively. As a result, both $MAE_{\hat{\alpha}}$ and $RMSE_{\hat{\alpha}}$ are higher for q^5 model than those for FF6 model. This is not surprising as the q^5 model does not explicitly include a momentum factor.

The second block of columns in Table 2 reports the standard errors of alphas ($\hat{\sigma}_{\alpha_i}$), whose dispersion is summarized by $RMSE_{\hat{\sigma}_\alpha}$. This statistic is 0.079% and 0.122% for the FF6 and the q^5 model, respectively, indicating that the q^5 model estimates alphas more imprecisely than the FF6 model. The t-statistics (t_i) reported in the next five columns of Table 2 show that the null hypothesis of no pricing error is rejected at 5% in seven (four) cases out of 25 for the FF6 (q^5) model. More rejections for q^5 than those for FF6 potentially indicate a 'lack of power' or 'under-rejection' in the t-test for individual assets for the q^5 model.

We further investigate the issue of low power by using the $(1,5)$ portfolio to illustrate the problem of the t-statistic. We find that its t-statistic at 4.46 in Panel A is much larger than 3.65 in Panel B (the t-statistic rejects FF6 more strongly than it does q^5). This is not because its pricing error for the test asset in Panel A is larger than that in Panel B ($\hat{\alpha}$ is 0.39% for FF6 versus 0.44% for q^5), but because the estimation error of the alpha estimate is smaller for FF6 than that for q^5 ($\hat{\sigma}_\alpha$ is 0.09% for FF6 and 0.12% for q^5). Given that q^5 produces a larger alpha that is estimated more imprecisely, a Bayesian performance metric ranks q^5 lower than FF6 for the (1,5) portfolio as a test asset, but the frequentist t-statistic does the opposite. As another example, $\hat{\alpha}$ is 0.16% for both models for the (3,5) portfolio. However, greater estimation uncertainty in the q^5 model than that in the FF6 model makes the q^5 model not reject the null hypothesis of a zero alpha while the FF6 model rejects it.

As for the joint tests, the GRS-statistic of 2.99 (2.64) for FF6 (q^5) suggest that q^5 is rejected less strongly than FF6. The stronger rejection of FF6 is a potential symptom of 'too much power' or 'over-rejection,' which is part of the power problem inherent in the GRS-statistic. The stronger rejection of the FF6 model is also counter-intuitive, in the sense that FF6 produces, on average, smaller alphas with higher estimation precision than does q^5.

The distance-based metrics (d_i for individual assets and AD for a universe of assets) effectively address the power problem (both too much power and lack of power) inherent in the t- and GRS-statistics by coherently incorporating the effects of pricing errors and their standard errors. Among the 25 individual assets, the three that marginally contribute the most to AD for FF6 in Panel A are $d_{(1,5)}= 0.40\%$, $d_{(1,1)}= 0.26\%$, and $d_{(2,5)}= 0.25\%$. The three equivalents for q^5 in Panel B are $d_{(1,5)}= 0.45\%$, $d_{(3,1)}= 0.37\%$, and $d_{(4,1)}= 0.34\%$.

To gauge the overall performance, we find that the average cost of moving the mass of the model-implied alpha distribution to the data-based distribution (measured by AD) for FF6 at 0.164% is smaller than that for q^5 at 0.211%. Therefore, FF6 is ranked higher than q^5, when assessed by the AD metric, which makes a Bayesian investor prefer FF6 to q^5. To sum up, the distance-based metrics consider the dispersion of pricing errors ($RMSE_{\hat{\alpha}}$) and their estimation precision ($RMSE_{\hat{\sigma}_\alpha}$) together, effectively solving the power problem of GRS and the ignorance of the power problem by $MAE_{\hat{\alpha}}$.

4.2. A Geometric View of the Average Distance (AD)

The above situation can be presented by a geometric view of the average distance (AD) in Figure 1, which plots the pricing errors ($\hat{\alpha}_i$) on the X-axis and their standard errors ($\hat{\sigma}_{\alpha_i}$) on the Y-axis generated by the above two models (FF6 in Panel A and q^5 in Panel B) with the 25 Size-MOM portfolios as test assets. The 25 diamond-shaped dots in each panel of are the ($\hat{\alpha}_i, \hat{\sigma}_{\alpha_i}$) pairs obtained from the two models. In each panel, AD is measured by the length of the arrow connecting from the origin (a round dot) to the other round dot. The half circle has a radius of 0.2% (per month) as a benchmark, which translates into 2.4% per annum, considered economically significant (Fama and French (1993, 1996)). The geometric view allows us to visualize how the dispersion of both alphas and their standard errors is summarized in the AD metric.

The plot for FF6 in Panel A shows that the 25 dots are horizontally more dispersed than vertically. The dots located outside of the half circle are the most troublesome portfolios that contribute the most to AD. Reading the plot for q^5 in Panel B, we see that its dots are more dispersed horizontally than those for FF6 in Panel A. Moreover, the 25 dots are conspicuously more dispersed along the Y-axis than those for FF6, with many of them hovering above 0.1% and some reaching up to 0.2%. This confirms that the q^5 model estimates alphas more imprecisely than FF6, potentially causing the lack of power in GRS for q^5. As a result, the arrow for q^5 in Panel B reaches beyond the half circle, whereas that for FF6 in Panel

A is within it, suggesting that the AD value (0.164% from Panel A in Table 2) for FF6 is economically insignificant. In essence, the AD metric ranks FF6 higher than q^5.

4.3. Performance of All the Ten Models

Table 3 reports the performance-related statistics (for joint-tests) generated by all of the ten asset-pricing models described in Subsection 3.1 for the 25 Size-MOM portfolios used in Table 2. The first three columns contain AD and its components ($RMSE_{\hat{\alpha}}$ and $RMSE_{\hat{\sigma}_\alpha}$). The last five columns show the traditional performance metrics described in Subsection 3.3.

The average distance (AD) can be used to compare asset-pricing performance across different models and rank them. In the second column, we find that AD identifies FF6, BS, BKRS, and SY as the top four models in that order (AD = 0.164%, 0.184%, 0.196%, and 0.197%, respectively), with the minimum average cost of holding a dogmatic belief in these models being economically insignificant, given the benchmark of 0.2% per month. We find poor performance of FF5 and especially FF3, which is even worse than CAPM. This strongly suggests that the UMD factor is important for the FF models. AR^2 also ranks FF6, BS, and BKRS highly. The DHS model performs poorly, regardless of whichever metric is used, most likely because it fails to explain the returns on the momentum portfolios.

Of more interest is how GRS and $MAE_{\hat{\alpha}}$ assess the models. We find again that the two metrics rank the ten models differently from what AD does, which is often inconsistent and counter-intuitive. Specifically, the GRS statistic ranks SY (2.58), q^5 (2.64), and q^4 (2.84) as the top three models, all of which are strongly rejected at the 1% level. On the other hand, $MAE_{\hat{\alpha}}$ ranks FF6 (0.110%), q^4 (0.118%), and SY (0.122%) as the top three models, similarly to what the $MAE_{\hat{\alpha}}/MAR_{\bar{r}}$ criterion does (0.39 for FF6, 0.42 for q^4, and 0.44 for SY). We attribute the contradictory rankings produced by GRS and $MAE_{\hat{\alpha}}$ to the power problem (in GRS) as well as the way of weighting pricing errors and treating the extent of estimation precision (in $MAE_{\hat{\alpha}}$).

To examine further the above two issues in the GRS and $MAE_{\hat{\alpha}}$ metrics, we notice that the top three models ranked by GRS tend to have larger estimation uncertainty (or imprecision) measured by $RMSE_{\hat{\sigma}_\alpha}$: 0.106% for SY, 0.122% for q^5, and 0.112% for q^4. This causes the ratio-based GRS statistic to be smaller for these models, inducing the lack of power. The lack of power in GRS for these models can also be seen in $MSE_{\hat{\sigma}_\alpha}/MSE_{\hat{\alpha}}$ in the fourth column. Indeed, these three models have the largest values among all the ten models (0.41 for SY, 0.44 for q^4, and 0.50 for q^5), which suggests that much of the dispersion of alpha estimates is due to estimation imprecision, resulting in lower GRS statistics. By contrast, the values of $MSE_{\hat{\sigma}_\alpha}/MSE_{\hat{\alpha}}$ for the other seven models are far below 0.35.

As for the ranking produced by $MAE_{\hat{\alpha}}$, recall that this criterion ignores the information on the extent of estimation precision. This makes BS ranked lower than q^4 and SY, despite the fact that the BS model estimates alphas more precisely than the two models ($RMSE_{\hat{\sigma}_\alpha} =$ 0.082% for BS vs. 0.112% for q^4 and 0.106% for SY). Second, as discussed before, $MAE_{\hat{\alpha}}$ treats all pricing errors equally whereas $RMSE_{\hat{\alpha}}$ places higher weights on extreme values. The consequence is apparent from the fact that although the values of $MAE_{\hat{\alpha}}$ for q^5 and q^4 are smaller, the levels of $RMSE_{\hat{\alpha}}$ for them are larger than FF6 or BS, because q^5 and q^4 produce more extreme alphas than FF6 or BS.

5. Performance with Broader Sets of Test Assets

We now compare the performance of all the ten models by using a much broader sets of LHS assets formed by sorting on a wide range of anomaly variables. We rank the models by the performance metrics, thereby identifying the top performers, and interpret the implications.

5.1. Performance with Decile Portfolios

5.1.1. Fama and French's (2015) 150 Decile Portfolios

Panel A in Table 4 reports the results with a pooled set of decile portfolios formed by sorting on 15 different variables (150 portfolios in total). With these 150 portfolios, AD identifies

FF6 (0.138%), BKRS (0.146%), and SY (0.152%) as the top three models, while *GRS* picks SY (1.55), FF6 (1.87), and FF5 (1.88).

From a Bayesian perspective, it is counter-intuitive that SY is chosen as the best model by *GRS*, given that SY has higher dispersion of alphas ($RMSE_{\hat{\alpha}}$): 0.123% for SY compared to 0.114% for FF6. SY also estimates alphas less precisely ($RMSE_{\hat{\sigma}_\alpha} = 0.089\%$ for SY vs. 0.078% for FF6). It is even more surprising that *GRS* identifies FF5 as one of the top three, despite its high dispersion of alphas ($RMSE_{\hat{\alpha}} = 0.150\%$). *AD* ranks FF5 (0.171%) only sixth behind the two q models. Among the 15 sets of decile portfolios, we find that the only troublesome set of assets for FF5 is the one sorted on momentum (MOM). In unreported results, we find that without the MOM-sorted decile portfolios, *AD* ranks FF5 third among the ten models, close to BKRS.

The above experiments have two implications for asset-pricing studies. First, given the failure of the FF3 and FF5 in explaining momentum as well as the *RMSE* property of heavily penalizing models that produce extreme alphas, it is necessary to add *UMD*. This dramatically improves the performance of the FF models by reducing the momentum-induced pricing errors. Second, adding *UMD* also helps the FF models better explain other anomalies than momentum. Therefore, *UMD* is an essential factor for the FF models.

5.1.2. Hou, Mo, Xue, and Zhang's (2018) 300 Decile Portfolios

In Panel B of Table 4, we report the results with the pooled set of decile portfolios formed by sorting on 30 anomaly variables (300 portfolios in total) used in Hou, Mo, Xue, and Zhang (2018). With this set of LHS assets, *AD* chooses SY (0.154%), FF6 (0.155%), and q^5 (0.156%) as the top three models, with their statistics being very close to each other. $MAE_{\hat{\alpha}}$ also picks the same set as the top three performers in slightly different order: SY (0.093%), q^5 (0.095%), and FF6 (0.099%).

GRS identifies SY (1.57) and q^5 (1.59) as the top two performers, similarly to the picks by $MAE_{\hat{\alpha}}$. However, its next pick is BKRS (1.60), which is not reasonable from a Bayesian perspective. Given the larger mispricing dispersion and higher estimation uncertainty for BKRS ($RMSE_{\hat{\alpha}} = 0.141\%$ and $RMSE_{\hat{\sigma}_\alpha} = 0.095\%$) relative to those for FF6 ($RMSE_{\hat{\alpha}} = 0.129\%$ and $RMSE_{\hat{\sigma}_\alpha} = 0.086\%$), a Bayesian investor should prefer FF6 to BKRS.

A similar logic applies to the DHS model: *GRS* ranks DHS (1.62) higher than FF6 (1.80), although DHS produces both higher $RMSE_{\hat{\alpha}}$ and $RMSE_{\hat{\sigma}_\alpha}$ than FF6. Overall, SY, FF6, and q^5 are the top three performers, producing quite close statistics in AD and $MAE_{\hat{\alpha}}$, when tested with this set of 300 decile portfolios.

5.2. Performance with H−L Portfolios

Although asset-pricing models should leave no pricing errors for all portfolios, the literature often focuses only on the long-short H−L portfolios computed using the highest and the lowest anomaly portfolios (e.g., Hou, Xue, and Zhang (2015); Hou, Mo, Xue, and Zhang (2018)). One advantage of using this subset of portfolios is that they generate a large cross-sectional variation in returns. In contrast, the average returns across different portfolios (e.g., decile four vs. decile five) in the universe of assets may be very similar. Second, the extreme portfolios (and the long-short portfolio) are more interesting because these are the portfolios where the asset-pricing models may have the biggest problem. In frequentist terms, long-short portfolios are more powerful tests than the decile portfolios of asset pricing models. For a Bayesian, H−L portfolios might be the ones that move the priors the most. Accordingly, in this subsection, we use three sets of H−L portfolios.

5.2.1. Fama and French's (2015) 15 H−L Portfolios

Panel A in Table 5 contains the performance-related statistics with 15 H−L portfolio returns computed from the 150 decile portfolios used in Panel A of Table 4. As expected, we find that the averages of estimated alphas are much larger than the corresponding values reported in Panel A of Table 4. This is turn causes the key metrics of interest (e.g., AD, *GRS*, and $MAE_{\hat{\alpha}}$) to have larger variation in their statistics across different models. The statistics in

Panel A show that the metrics identify three different sets as top three in different order. GRS's top picks are (i) q^5 (1.58), (ii) SY (1.75), and (iii) q^4 (1.92). $MAE_{\hat{\alpha}}$ chooses (i) q^5 (0.149%), (ii) SY (0.157%), and (iii) FF6 (0.160%). However, the AD metric ranks differently: (i) FF6 (0.260%), (ii) SY (0.285%), and (iii) q^5 (0.288%).

q^5 has the lowest value of $MAE_{\hat{\alpha}}$ (0.149%) and $RMSE_{\hat{\alpha}}$ (0.211%). However, it has one of the highest estimation imprecision that is 6 bps higher than that for FF6 ($RMSE_{\hat{\sigma}_\alpha} =$ 0.196% for q^5 vs. 0.144% for FF6). As a result, it has the highest value of $MSE_{\hat{\sigma}_\alpha}/MSE_{\hat{\alpha}}$ among the ten models at 0.86. From a Bayesian view, this close-to-one ratio is a typical symptom of low power in GRS for q^5. The low power of q^5 is shown not just by the high level in $RMSE_{\hat{\sigma}_\alpha}$ but also by the low level (0.44) in its AR^2: 16% lower than that of FF6 (0.60). As a result, the GRS metric (which favors high sampling errors) ranks q^5 higher than AD does. The $MAE_{\hat{\alpha}}$ metric also ranks q^5 as the highest as this metric does not utilize the information on estimation precision at all.

For the remaining models, AD puts q^4 as a close fourth to q^5, given its higher dispersion of alphas ($RMSE_{\hat{\alpha}} = 0.227\%$) but lower estimation uncertainty ($RMSE_{\hat{\sigma}_\alpha} = 0.179\%$). The next two models in AD's ranking list are BS and BKRS, whose level in $RMSE_{\hat{\alpha}}$ is high (above 0.26%). As one would expect, the poor performance of FF5 stems from its failure to explain the H−L returns on momentum portfolio. Unreported results show that the H−L momentum portfolio produces the largest marginal distance for FF5 (its d_i is as large as 1.39%), which is almost entirely due to its sheer size of alpha (1.35%). We again observe that DHS performs better than only CAPM and FF3.

5.2.2. Hou, Mo, Xue, and Zhang's (2018) 30 H−L Portfolios

In Panel B of Table 5, we report the performance statistics with the 30 H−L portfolio returns computed using the 300 decile portfolios used in Panel B of Table 4. For these portfolios, q^5 beats all the other nine models by a wide margin when assessed by AD as well as the other four metrics ($RMSE_{\hat{\alpha}}$, GRS, $MAE_{\hat{\alpha}}$, and $MAE_{\hat{\alpha}}/MAE_{\bar{r}}$). In particular, its $MAE_{\hat{\alpha}}$ of

0.137% is almost 10 bps lower than that for its closest competitor, SY (0.227%). Moreover, the *GRS*-statistic of 1.48 suggests that q^5 is the only model that cannot be rejected even at the 10% level, whereas almost all the others are rejected at 1% level. In terms of the distance measure, AD of 0.279% for q^5 is lower by 7 bps than the value for its closest competitor (0.352% for SY) and by 9 bps lower than that for the third best model (0.366% for FF6).

Despite the top performance of q^5 when assessed by the six metrics with these 30 H−L portfolios, it should be noted that the q^5 model has high estimation uncertainty ($RMSE_{\hat{\sigma}_\alpha} =$ 0.184%) (which causes lower power for GRS) than its two closest competitors ($RMSE_{\hat{\sigma}_\alpha} =$ 0.174% for SY and 0.150% for FF6). However, q^5's lower pricing errors appear to more than offset the adverse effect of its high estimation uncertainty.

For a visual inspection, in Figure 2, we plot the $(\hat{\alpha}_i, \hat{\sigma}_{\alpha_i})$ pairs generated by the q^5, FF6, and SY models for the 30 H−L portfolio. In the figure, the half-circle is drawn with a radius of 0.4% (vis-à-vis 0.2% in Figure 1), given that the H−L returns are harder to explain by the models (hence generating much larger alphas on average). We first observe that alphas are positively skewed for the three models, especially FF6 whose alphas are all above −0.1% in Panel A. The plot for SY in Panel C is similar, although more dispersed horizontally. Second, the vertical distribution of the dots exhibits the degree of estimation precision, which affects the power of *GRS*. As seen in Panel B, the dots for q^5 are vertically more dispersed than those for FF6. Third, q^5 has the shortest arrow: a geometric presentation of its outperformance. Finally, the slope of the arrow is steeper for q^5 than for FF6 or SY, suggesting that estimation imprecision contributes more to AD for q^5 than for the other two models.

5.2.3. Green, Hand, and Zhang's (2017) 85 H−L Portfolios

There is an interesting 'home-bias' in the results so far. Using Fama and French's (2015) portfolios, FF6 performs the best (Panels A in Tables 4 and 5). Using Hou, Mo, Xue, and Zhang's (2018) portfolios, especially their H−L portfolios, q^5 performs the best (Panel B in Table 5). Although FF6, q^5, and SY are the close competitors, the above finding makes us wonder how the three models perform when we use a different universe of portfolios constructed by a third party. Therefore, we repeat our tests with a set of 85 H−L returns computed using the 776 portfolios formed by Green, Hand, and Zhang (2017).[13] Using this independent set of LHS assets may alleviate potential biases associated with the issues of selecting test assets.

Panel C in Table 5 contains the statistics for these 85 H−L portfolios. The top three models chosen by *GRS* are q^5 (1.47), BKRS (1.54), and SY (1.62), whereas those by $MAE_{\hat{\alpha}}$ are q^5 (0.220%), FF6 (0.225%), and SY (0.227%). By contrast, the *AD* metric ranks differently: (i) FF6 (0.339%), (ii) q^5 (0.348%), and (iii) SY (0.354%). Again, the reasons for *GRS* and $MAE_{\hat{\alpha}}$ picking q^5 as the top performer are: (a) the q^5 model produces lower alphas but estimates them most imprecisely (note the highest $RMSE_{\hat{\sigma}_\alpha}$ of 0.207%), making *GRS* reject the model less often; and (b) $MAE_{\hat{\alpha}}$ does not take estimation uncertainty into account.

To get a sense of how the three models perform in the individual portfolios, we present a graphic view of the 85 $(\hat{\alpha}_i, \hat{\sigma}_{\alpha_i})$ pairs in Figure 3. Looking horizontally, we find in Panel A that alphas generated by FF6 are more widely spread (from −0.73% to 0.80%) than those generated by q^5 (from −0.77% to 0.70%) as well as those generated by SY (from −0.75% to 0.62%). Vertically, however, FF6 has the smallest dispersion of standard errors (most of the time below 0.3%). This is distinguishable from the plot for q^5, whose dots in Panel B are vertically more scattered at a higher level on average than those for FF6. The vertical dispersion for SY in Panel C is somewhere in between FF6 and q^5. Consequently, we again observe a typical feature of q^5: it has the steepest arrow, meaning that its estimation uncertainty contributes the most to *AD* among the three models.

To summarize, *AD* consistently identifies FF6, q^5, and SY as the top three models when tested with different universes of LHS assets. We reemphasize the importance of the *UMD* factor for the FF-model group; adding *UMD* not only alleviates the momentum-induced pricing errors but also improves the explanatory power for the FF models. One advantage

[13] Some portfolios formed based on their sorting variables are excluded, because they are not available for the full sample period (1980:01-2015:12).

of the q^5 model is that it produces lower pricing errors. However, the main issue of this model lies in its highest estimation uncertainty. Small alphas estimates, coupled with low estimation precision, often make q^5 identified as one of the top three by the traditional metrics such as *GRS* and *MAE*. From a Bayesian perspective, however, the performance of q^5 is compromised by its high estimation imprecision. Among the 'Other' group, SY stands out consistently as one of the top performers across different sets of test assets.

6. Bayesian Assessment of Choosing Alternative Models

In the previous section, we have compared models using our proposed metrics (AD and d_i), which measure the distances between the purely model-based distribution (complete confidence in a model: $\sigma_\alpha = 0$) and the purely data-based distribution (complete skepticism about a model: $\sigma_\alpha = \infty$). This Bayesian approach is akin to the frequentist approach of testing the null hypothesis of zero alphas. However, it is reasonable to assume that investors generally have some degree of skepticism ($\sigma_\alpha > 0$) about any model a priori. As Pástor (2000) and Barillas and Shanken (2018) note, investors neither use a model as a dogma, nor do they regard the model as completely useless. Therefore, we now allow investors to have varying degrees of prior estimation uncertainty (i.e., $0 < \sigma_\alpha < \infty$) about the models.

6.1. Prior Estimation Uncertainty

For $\sigma_\alpha \in (0, \infty)$, the posterior distribution of mispricing is centered around a non-zero mean with a non-zero covariance matrix: i.e., $p\,(\alpha|R, \sigma_\alpha) \sim N(\tilde{\alpha}, \tilde{V})$, where $\tilde{\alpha}$ and $\tilde{V}$ are the posterior estimates of the mean and variance for the distribution with prior specification $\sigma_\alpha > 0$. The exact expressions for these posterior estimates are given in equation (4) and Appendix A. Using equation (5) the distance metric between the posterior distribution $N(\tilde{\alpha}, \tilde{V})$ (implied by prior with $0 < \sigma_\alpha < \infty$) and the posterior distribution $N(\hat{\alpha}, \hat{\Sigma})$ (implied by prior with $\sigma_\alpha = \infty$) is then given by:

$$
\begin{aligned}
MSE_{\tilde{\alpha}-\hat{\alpha}} &= \sum\nolimits_{i=1}^{n} \left(\tilde{\alpha}_i - \hat{\alpha}_i\right)^2 \Big/ n \\
MSE_{\tilde{V}-\hat{\Sigma}} &= ||\tilde{V} - \hat{\Sigma}|| \Big/ n = \mathrm{Tr}\left(\tilde{V} + \hat{\Sigma} - 2(\tilde{V}^{1/2}\hat{\Sigma}\tilde{V}^{1/2})^{1/2}\right) \Big/ n \\
AD &= \sqrt{MSE_{\tilde{\alpha}-\hat{\alpha}} + MSE_{\tilde{V}-\hat{\Sigma}}} \qquad (10)
\end{aligned}
$$

From this Bayesian perspective, by fixing σ_α at a positive value for the prior distribution, AD is now interpreted as the minimum average cost of holding some degree of skepticism (as opposed to a dogmatic belief) about a model; intuitively, this is the price to pay for changing one's somewhat skeptical view about a model to a complete disbelief in the model. As an investor departs from her dogmatic belief and increases the degree of doubt on the validity of the model, the mispricing distribution under her skeptical view becomes closer to the data-based distribution ($N(\tilde{\alpha}, \tilde{V}) \to N(\hat{\alpha}, \hat{\Sigma})$). For example, when σ_α increases from a small value to a very large one, the posterior estimate of the alpha ($\tilde{\alpha}$) also moves from a value close to zero toward its OLS estimate ($\hat{\alpha}$). Thus, AD decreases monotonically as σ_α increases. Intuitively, this means that the higher the prior degree of skepticism in the model, the lower the cost to change one's belief to accept the data-based distribution.

For each model, we gradually increase σ_α starting from 0% up to 10%. In specifying the range and assessing the economic magnitude of σ_α, we use the average return volatility of the LHS assets as a guideline, following Pástor and Stambaugh (2000, p. 353). In this setup, $\sigma_\alpha = 2\%$ is a modest degree of prior estimation uncertainty, and $\sigma_\alpha > 2\%$ represents a significant degree of it.

Figure 4 plots the AD values (the Y-axis) generated by the 10 models with the set of the 85 H−L portfolios used in Panel C of Table 5, as σ_α (the X-axis) increases from 0% to 10%. In the figure, each line represents the changes in AD for a given model as σ_α increases. From the top, the lines are aligned in order of CAPM, FF3, BS, DHS, q^4, FF5, BKRS, SY, q^5, and FF6, suggesting that FF6 is the best model. The lines are monotonically decreasing in a non-linear and non-parallel fashion. The spread between the top and bottom values on a given line is the gap between AD at $\sigma_\alpha = 0\%$ (a dogmatic belief) and that at $\sigma_\alpha = 10\%$ (a high level of disbelief), reflecting the shrinkage effect of the posterior estimates. As the skepticism about a model continues to grow to a complete disbelief ($\sigma_\alpha = \infty$), each model's AD is zero, and all the lines eventually converge to the same point on the X-axis. Among the 10 models, FF5 and FF6 tend to shrink faster than the other models: the FF5 line intersects the BKRS line at around $\sigma_\alpha = 7\%$, and the FF6 line exhibits a discernibly wider gap from the q^5 line after $\sigma_\alpha = 4\%$.

Table 6 reports the values in AD and its two $RMSE$ components using the same 85 H$-$L portfolios as test assets, as σ_α increases from 2% to 10% in Panels B to F with a 2% increment across the panels. Panel A reproduces the benchmark values from Table 5. Table 6 and Figure 4 jointly provide a comprehensive picture about the shrinkage effect of AD, allowing our Bayesian assessment for model comparisons as detailed below.

6.2. The Distance-Equivalence Measure For Choosing Models

A question that we like to answer in this and next subsections is: if an investor chooses an alternative model 'A' over the benchmark 'B', how much prior estimation uncertainty in mispricing should she accept? Or conversely, by choosing the benchmark model over an alternative, by how much does her prior concern about mispricing alleviate? To answer the above questions, we introduce a concept of 'distance equivalence' as follows.

Distance Equivalence: *By the monotonically decreasing property of the AD metric, there exists a unique and positive* σ^*_α *for Model A such that AD (Model A, prior* $\sigma^*_\alpha > 0$) $=$ *AD (Model B, prior* $\sigma_\alpha = 0$). *With such a value of* σ^*_α, *it is defined that "Model A with* $\sigma_\alpha = \sigma^*_\alpha$ *is distance-equivalent to Model B with* $\sigma_\alpha = 0$."

Being 'distance-equivalent' above means that a Bayesian investor is indifferent between holding a skeptical view about an alternative model (Model A) and holding a dogmatic belief in the benchmark model (Model B). Thus, if the investor chooses the alternative (A) over the benchmark (B), σ_α^* for Model A is the level of prior estimation uncertainty she must accept. The concept of distance equivalence is easy to understand graphically with Figure 4. For example, at $\sigma_\alpha = 6\%$ the vertical value (AD) of the BS model (the solid line with triangles) is 0.385% in the figure. This is very close to the vertical value ($AD = 0.380\%$) of the DHS model (the dotted line with circles) at $\sigma_\alpha = 0\%$. Therefore, BS with $\sigma_\alpha = 6\%$ ($= \sigma_\alpha^*$) is (approximately) distance-equivalent to DHS at $\sigma_\alpha = 0\%$. Here, $\sigma_\alpha^* = 6\%$ is the level of prior estimation uncertainty that makes the two models (BS and DHS) distance-equivalent.[14]

Panel B of Table 6 reports the results with $\sigma_\alpha = 2\%$, for which the posterior distribution of the alpha is only moderately closer to the data-based distribution ($\sigma_\alpha = \infty$). To apply the distance-equivalence concept, we identify models whose AD values shown in Panel B ($\sigma_\alpha = 2\%$) are close (within a 0.5 bps range) to the AD value of any of the ten benchmark models shown in Panel A ($\sigma_\alpha = 0\%$). For instance, q^5's AD of 0.334% in Panel B is close to FF6's AD of 0.339% in Panel A. Graphically in Figure 4, the line connecting $AD = 0.334\%$ of q^5 (the dashed line with squares) at $\sigma_\alpha = 2\%$ to $AD = 0.339\%$ of FF6 (the dashed line) at $\sigma_\alpha = 0\%$ is approximately horizontal. This means that a Bayesian investor is indifferent between holding a modestly skeptical view ($\sigma_\alpha = 2\%$) about the q^5 model and holding a dogmatic belief in the FF6 model. That is, if the investor chooses q^5 (the alternative) over FF6 (the benchmark), her prior belief (at the 95% confidence interval) is that an annualized pricing error of about $\pm 4\%$ ($= \pm 2 \times \sigma_\alpha$) is acceptable to her. Or conversely, her choice of FF6 over q^5 alleviates her prior concern about mispricing by $\pm 4\%$ per year.

Similarly, the SY model at $\sigma_\alpha = 2\%$ ($AD = 0.341\%$ in Panel B of Table 6) is also distance-equivalent to holding a dogmatic belief in FF6. By choosing SY over FF6, a Bayesian investor is willing to accept prior mispricing of about $\pm 4\%$ per year. Since Pástor and Stambaugh (2000) regard $\sigma_\alpha = 2\%$ as a modest level of prior uncertainty, we infer that

[14] The concept of distance equivalence is defined using an equality in the level of AD. However, when σ_α is discretized with a 2% increment, we consider Models A and B are approximately distance-equivalent if $|AD$ (Model A, prior $\sigma_\alpha^A > 0) - AD$ (Model B, prior $\sigma_\alpha^B = 0)| \leq 0.5$ bps.

the performance difference among the top three models (FF6, q^5, and SY) identified by the distance-equivalence concept is economically insignificant.

Panel C of Table 6 shows the results with $\sigma_\alpha = 4\%$. At this (larger) level of prior estimation uncertainty, BKRS ($AD = 0.336\%$) and FF5 ($AD = 0.340\%$) are distance-equivalent to FF6 ($AD = 0.339\%$ in Panel A), meaning that a Bayesian investor choosing BKRS or FF5 over FF6 must accept prior uncertainty in mispricing of as much as $\pm 8\%$ per year. This in turn has two important implications: (1) *UMD* is a crucial factor for the FF models; (2) changing the factor composition (*HML* to HML^m; and *RMW* to *RMWCP*) in BKRS does not help much to improve the performance of the FF models.

At $\sigma_\alpha = 4\%$, q^4 ($AD = 0.345\%$) and DHS ($AD = 0.350\%$) are also distance-equivalent to q^5 ($AD = 0.348\%$ in Panel A), suggesting that adding the *EG* factor to q^4 is important for q^5. In addition, at $\sigma_\alpha = 4\%$, and even at $\sigma_\alpha = 6\%$ in Panel D of Table 6, the *AD* levels for BS, FF3 and CAPM are well above the *AD* levels for the other seven models at $\sigma_\alpha = 0\%$ (i.e., the seven points lying on the Y-axis) in Figure 4. This indicates that much higher levels of prior estimation uncertainty in mispricing (σ_α) are required for these three models to be distance-equivalent to the seven models.

We see in Panel E of Table 6 that at $\sigma_\alpha = 8\%$, BS ($AD = 0.355\%$) is distance-equivalent to SY ($AD = 0.354\%$ in Panel A), and so is FF3 ($AD = 0.375\%$) to FF5 ($AD = 0.372\%$ in Panel A). For a Bayesian investor, choosing SY over BS, or FF5 over FF3, relieves her prior concern about mispricing of as much as $\pm 16\%$ per year. Such a magnitude is economically too large to be ignored, and it signifies the failure of BS and FF3.[15] Finally, we find in Panel F of Table 6 that at $\sigma_\alpha = 10\%$, CAPM ($AD = 0.458\%$) is roughly distance-equivalent to FF3 ($AD = 0.465\%$ in Panel A), implying an outright failure of CAPM.

[15] It is surprising that the BS model, which has the highest posterior probabilities by the Bayes-factor measure, does not perform well in pricing the test assets. It is apparent in Figure 4 that the BS line is positioned below the lines of only CAPM and FF3, but far above the lines of the other seven models.

6.3. Values of Individual Factors in the Benchmark Models

Given the AD metric and the concept of distance equivalence, we have shown above that we can easily assess the economic costs of choosing alternative models across both nested (e.g., FF5 vs. FF6) and non-nested (e.g., q^5 vs. FF6) models. In this subsection, we use the same methodology to evaluate the economic values of individual factors and their combinations in each of the top three benchmark models (FF6, q^5, SY) using the 85 H−L portfolios as test assets. We do so by excluding some factors from a model and computing the AD-statistics of the parsimonious model (that deletes the factor(s)). If an excluded factor(s) is indeed important, the AD of the simpler model should be significantly larger than that of the benchmark model. To assess the economic values of the excluded factors, we compute the level of prior estimation uncertainty, $\sigma_\alpha = \sigma_\alpha^* > 0$, that makes the parsimonious model and the (full) benchmark model distance-equivalent. By focusing on the marginal effects of specific factors, our analyses in the nested-model setting can offer further insights into the relative importance of those factors in each of the top three benchmark models.

Table 7 reports the values in AD and the levels of prior estimation uncertainty (σ_α^* in square brackets) that makes the two (parsimonious and benchmark) models distance-equivalent for each combination of factors excluded from the three benchmark models: FF6 in Panel A, q^5 in Panel B, and SY in Panel C. The diagonal entries in each panel contain the values of AD and σ_α^* when the corresponding (one) factor is excluded from the full model. The set of excluded factors below the diagonal consists of the diagonal factor plus all the factors up to that row. For example, the (6, 2) entry in Panel A, the (5, 2) entry in Panel B, and the (4, 2) entry in Panel C exclude all factors except for the MKT factor (meaning that these entries correspond to the CAPM model). Similarly, the model in the (6, 4) entry in Panel A is FF3, and that in the (6, 6) entry in Panel A is FF5.

The (1, 1) entry in Panel A shows that when MKT is excluded from FF6, the simpler model (consisting of *SMB*, *HML*, *CMA*, *RMW*, and *UMD*) produces $AD = 0.415\%$, which is higher than that of FF6 whose $AD = 0.339\%$ (see Panel A of Table 6). This implies that the marginal value of MKT measured by AD is $0.415\% - 0.339\% = 0.076\%$. It also shows that an investor needs to hold a strong disbelief (at $\sigma_\alpha = 7.1\%$) in the parsimonious model (that excludes MKT from FF6) for her to be indifferent between the two models: differently put, adding MKT to FF5 mitigates her prior mispricing concern by $\pm 14.2\%$ annually. In addition, the values in the diagonal grids in Panel A show that prior uncertainties (σ^*_α) of most factors are larger than 2%, except for HML whose σ^*_α is economically insignificant at 1.5%. This in turn suggests that from a Bayesian view HML is a redundant factor (Fama and French (2015)). Among the other factors, we emphasize our earlier finding that the UMD factor is an important addition to the FF models. Excluding UMD from the FF6 model as shown in the (6, 6) entry in Panel A is equivalent to accepting a prior mispricing uncertainty of $\pm 8.2\%$ per year, which is too large for a Bayesian investor to ignore.

By examining the diagonal values in each of the first two panels of Table 7, we find that MKT is the single most important factor for both FF6 and q^5, which is consistent with one of the key findings of Harvey and Liu (2018). Thus, even though the results in the previous subsection imply that the CAPM model is resoundingly dominated when assessed by the AD metric, MKT remains as the most important factor in the two benchmark asset-pricing models (FF6 and q^5).

The result for q^5 in Panel B of Table 7 shows that adding the expected growth factor (EG) to q^4 is economically warranted. In the (5, 5) entry, we find that the AD of the parsimonious model (that excludes EG) is 0.376%, which is 0.028% higher than the AD of the full q^5 model (0.348% in Panel A of Table 6). The prior uncertainty of $\sigma^*_\alpha = 3.7\%$ that makes the two models distance-equivalent is economically significant. As shown in Table 5, the key advantage of adding EG is that it further reduces the size of pricing errors, making q^5 as one of the top performers, given its lowest levels in $MAE_{\hat{\alpha}}$ and $RMSE_{\hat{\alpha}}$ when tested with all the three sets of (H$-$L) portfolios. However, it should be cautioned that adding EG to the q^4 model is not without a cost. We observe in Table 5 that q^5 generates higher values in $RMSE_{\hat{\sigma}_\alpha}$ than q^4 across the three sets of test assets. While this is good news for the GRS

and MAE metrics, the distance-based metrics inherently recognize the trade-off between lower alpha estimates and higher estimation uncertainty for the q^5 model. Nonetheless, we find that by adding the EG factor to the q^4 model, the smaller pricing errors for q^5 more than offset the unfavorable effect of its higher estimation uncertainty.

Furthermore, we find that adding EG makes the role of other factors in q^5 useless or redundant. For instance, the (3, 3) entry in Panel B shows that the AD of a model that excludes IA is 0.324%. In the (4, 3) entry, the AD of a model that excludes both IA and ROE is 0.322%. Interestingly, the ADs of the above two parsimonious models are even smaller than the AD of the benchmark q^5 model (0.348%). This suggests that IA and/or ROE may be redundant in the q^5 model. The negative sign of σ_α in the two diagonal grids in Panel B also indicates that excluding IA (ROE) from the q^5 model rather *reduces* the prior concern of mispricing by ±6.6% (±4.2%). In addition, $AD = 0.324\%$ in the (3, 3) entry suggests that excluding IA from q^5 makes the parsimonious model (with the MKT, ME, ROE, and EG factors) outperform the FF6 model (whose $AD = 0.339\%$).

The (4, 3) entry in Panel B also shows that excluding both IA and ROE makes the parsimonious three-factor model (with MKT, ME, and EG) perform best among the models with different factor combinations in Panel B, with its AD being at only 0.322%.[16] In unreported results, we find that the lowest value of its AD is a result of its much lower $RMSE_{\hat{\alpha}}$ (0.227% vs. 0.279% for the q^5 model), although its $RMSE_{\hat{\sigma}_\alpha}$ is slightly higher (0.228% vs. 0.207% for the q^5 model).[17] These findings suggest that more efforts are warranted in designing multi-factor models that most improve the pricing performance.

Finally, as to the SY model in Panel C, the values in the diagonal grids implies that all the four factors in the model play economically significant roles, with the most important factor being $MGMT$.

[16] Adding more factors does not necessarily improve model performance. In unreported results, we find that a model that has ten factors (MKT, SMB, HML, UMD, RMW, CMA, HLM^m, ME, IA, and ROE) does not outperform FF6. As another example, removing HML^m from the BS model improves the performance of that model to be closer to that of FF6.

[17] We note however that the improvements in AD of the parsimonious models (relative to the full q^5 model) do not always extend to the cases where we analyze with other test assets used in this paper.

Note in each panel that the entry in the last row and the first column corresponds to a model that excludes all the factors in each of the three models; in this case, data-based alpha estimates are simply the mean returns of the 85 H−L portfolios, and the residual covariance matrix is simply the covariance matrix of the H−L portfolio returns. We find in the (6, 1) entry of Panel A that, with $AD = 0.423\%$ and $\sigma_\alpha = 9.2\%$, such a naïve zero-factor model outperforms prominent models shown in the same panel such as FF3 ($AD = 0.465\%$) in the (6, 4) entry, CAPM ($AD = 0.576\%$) in the (6, 2) entry, as well as the model with only the MKT and SMB factors ($AD = 0.581\%$) in the (6, 3) entry.

7. Conclusion

In evaluating asset-pricing performance via ratio-based statistics, a model with high estimation precision may have small p-values, and is thus statistically rejected even if it produces economically insignificant pricing errors (alphas). Conversely, a model with a big covariance matrix of residual returns may produce large p-values, and thus pass the statistical test even with economically significant alphas. This nature of the power problem in statistical tests has long been recognized since Fama and French (1993). The power problem makes comparing p-values across different models problematic and the results difficult to interpret (Harvey (2017)). As a compromise, most empirical studies rely on various alpha-based statistics (e.g., MAE, which ignores the extent of estimation precision), jointly with the p-values of the GRS test. However, the GRS and average-alpha-based statistics often lead to contradicting and counter-intuitive model choices.

We adopt a Bayesian approach to address the above challenges in asset-pricing studies. A Bayesian investor views the mispricing parameter as a posterior distribution, updating her subjective belief with the data. From this probabilistic view, we define the distance between two posterior distributions, and propose distance-based metrics derived from the optimal transport theory. The distance is interpreted as the minimum cost of holding a dogmatic belief in a specific model (as a benchmark). Furthermore, when an investor chooses an alternative model over the benchmark, we can also quantify her prior uncertainty in mispricing (or economic price) using the concept of distance equivalence.

Our results show that the AD metric consistently identifies FF6, q^5, and SY as the top three models, whose performance differences are economically insignificant. We also find that a Bayesian investor choosing FF5 over FF6, or choosing q^4 over q^5, must accept prior uncertainty in mispricing that is economically too large to ignore. This in turn suggests that the momentum factor (UMD) is essential for the success of FF6 (and the other FF models). Similarly, adding the expected growth factor (EG) to the q^4 model is economically warranted. The main advantage of adding EG is that it helps the q^5 model significantly reduce its pricing errors, although the q^5 model estimates them less precisely than does the q^4 model. Overall, however, the benefit of smaller pricing errors for the q^5 model dominates the adverse effect of its lower estimation precision.

Using the AD metric together with the notion of distance equivalence, we can assess the values of individual factors or their combinations in the three best-performing models. The results show that HML tends to be a redundant factor in FF6, whereas MKT is the single most important factor in the FF6 and q^5 models. We also find that $MGMT$ plays the most important role in the SY model. In addition, the EG factor exerts strong influence in q^5, making some of the other four factors become useless in the model. This suggests that the model could be more parsimonious and thus there is room for improving upon q^5.

Our distance-based metrics along with Bayesian interpretations complement the frequentist approach, which is dominant in asset-pricing tests. The AD metric (as well as TD when the number of test assets are the same) provides a useful diagnostic tool, and d_i can further help identify troublesome assets in the investment universe. Our distance-based metrics may be jointly used in asset-pricing tests and model comparisons.

Appendices

A. Posterior Estimates of the Model Parameters

To derive the posterior estimates of the model parameters for the multivariate regression in equation (1), we can express the likelihood function of the model as:

$$\begin{aligned} p(R|B,\Sigma) &\propto |\Sigma|^{-\frac{T}{2}} \exp\left\{-\frac{1}{2}Tr(R-XB)'(R-XB)\Sigma^{-1}\right\} \\ &\propto |\Sigma|^{-\frac{T}{2}} \exp\left\{-\frac{1}{2}Tr\left(S+(B-\hat{B})'X'X(B-\hat{B})\right)\Sigma^{-1}\right\}, \end{aligned} \tag{A1}$$

where $S = (R - X\hat{B})'(R - X\hat{B})$, $\hat{B} = (X'X)^{-1}X'R$ and $Tr(.)$ is the trace operator. Also, the prior distribution of model parameters is: $p(B, \Sigma) = p(B|\Sigma)p(\Sigma)$, where

$$p(B|\Sigma) \sim N(B_0, \Sigma \otimes V_0) \propto |\Sigma|^{-\frac{k+1}{2}} \exp\left\{-\frac{1}{2}Tr(B-B_0)'V_0^{-1}(B-B_0)\Sigma^{-1}\right\}, \tag{A2}$$

where $V_0^{-1} = \begin{bmatrix} s^2/\sigma_\alpha^2 & 0 \\ 0 & 0 \end{bmatrix}$ is a $(k+1)\times(k+1)$ matrix whose (1,1) element is s^2/σ_α^2 and all other elements are zero; and

$$p(\Sigma) \sim IW(H_0, \upsilon_0) \propto |\Sigma|^{-\frac{\upsilon_0+n+1}{2}} \exp\left\{-\frac{1}{2}Tr(H_0\Sigma^{-1})\right\} \tag{A3}$$

is an inverted-Wishart distribution with degree of freedom $\upsilon_0 = n + 2$, so that the scaling matrix is $H_0 = \mathrm{E}[\Sigma] = \frac{s^2(\upsilon_0-n-1)}{\upsilon_0-n-1}I_n = s^2 I_n$.

Combining the prior distribution (A2) and (A3) with the likelihood function (A1) gives the following posterior distribution:

$$\begin{aligned} p(B,\Sigma|R) &= p(R|B,\Sigma)p(B|\Sigma)p(\Sigma) \propto |\Sigma|^{-\frac{T+k+1+\upsilon_0+n+1}{2}} \exp\left\{-\frac{1}{2}Tr\Big(H_0+S\right. \\ &\left.\quad +(B-\hat{B})'X'X(B-\hat{B})+(B-B_0)'V_0^{-1}(B-B_0)\Big)\Sigma^{-1}\right\}. \end{aligned} \tag{A4}$$

Completing the squares on B and collecting the remaining terms in $(\cdot)$ yields:

$$H_0+S+(B-\hat{B})'X'X(B-\hat{B})+(B-B_0)'V_0^{-1}(B-B_0) = (B-\tilde{B})'\tilde{V}^{-1}(B-\tilde{B})+\tilde{H}, \tag{A5}$$

where $\tilde{B} = (V_0^{-1} + X'X)^{-1}(V_0^{-1}B_0 + X'R)$, $\tilde{V} = (V_0^{-1} + X'X)^{-1}$, and $\tilde{H} = H_0 + B_0'V_0^{-1}B_0 + S + \hat{B}'X'X\hat{B} - \tilde{B}'\tilde{V}^{-1}\tilde{B}$.

The posterior distribution (A4) can be separated into two known distributions:

$$p(B,\Sigma|R) = p(B|\Sigma,R)\times p(\Sigma|R) \propto$$
$$\underbrace{|\Sigma|^{-\frac{k+1}{2}}\exp\left\{-\frac{1}{2}Tr(B-\tilde{B})'\tilde{V}^{-1}(B-\tilde{B})\Sigma^{-1}\right\}}_{N(\tilde{B},\Sigma\otimes\tilde{V})}\times\underbrace{|\Sigma|^{-\frac{T+v_0+n+1}{2}}\exp\left\{-\frac{1}{2}Tr\tilde{H}\Sigma^{-1}\right\}}_{IW(\tilde{H},\tilde{v})}. \quad \text{(A6)}$$

In words, $p(B|\Sigma,R)$ is normally distributed with posterior mean $\tilde{B}$ and posterior variance $\Sigma\otimes\tilde{V}$, and $p(\Sigma|R)$ is inverted-Wishart distributed with degree of freedom $\tilde{v} = T + v_0$ and scaling matrix $\tilde{H}$. Denote the posterior estimate of the residual covariance matrix by $\tilde{\Sigma}$ which, from the properties of the Wishart distribution (Zellner (1971)), is given by:

$$\tilde{\Sigma} = \mathrm{E}[\Sigma|R] = \frac{\tilde{H}}{\tilde{v}-n-1} = \frac{\tilde{H}}{T+1}. \quad \text{(A7)}$$

From (A6) and (A7), the posterior distribution of the alpha is normal with its posterior mean $\tilde{\alpha}' = \mathrm{E}[\alpha|R,F]$ taken from the first row of $\tilde{B}$ and its posterior variance $\tilde{V}_\alpha = \mathrm{Var}[\alpha|R,F]$ taken from the (n,n) upper left block of $\tilde{V}\otimes\tilde{\Sigma}$.

With dogmatic beliefs in an asset-pricing model, mispricing is ruled out by setting the prior alpha uncertainty at zero (i.e., $\sigma_\alpha = 0$), so that both $\tilde{\alpha}$ and $\tilde{V}_\alpha$ are zero. That is, $p(\alpha|R,F,\sigma_\alpha = 0) \sim N(0,0)$. At the other end of the spectrum, $\sigma_\alpha = \infty$, in which case investors are completely skeptical about the model, so that the posterior mean $\tilde{\alpha}$ and $\tilde{V}_\alpha$ of the alpha conform to the sampling theory results, which are:

$$p(\alpha|R,F,\sigma_\alpha = \infty) \sim N\left(\bar{R}-\hat{\beta}\bar{F}, \left(1+\bar{F}'\hat{\Omega}\bar{F}\right)\frac{\hat{\Sigma}}{T}\right), \quad \text{(A8)}$$

where $\bar{R}$ is an $n\times 1$ vector of the sample mean of LHS excess returns, $\bar{F}$ is a $k\times 1$ vector of the sample mean of factor returns, $\hat{\beta} = (F'F)^{-1}F'R$ is the $n\times k$ matrix of OLS estimates of the factor loadings, $\hat{\Omega}$ is the $k\times k$ sample covariance matrix of factor returns, and $\hat{\Sigma} = (H_0+S)/(T+1)$ is the $n\times n$ residual covariance matrix of LHS returns estimated from the sample that dominates its non-informative prior, H_0.

B. Wasserstein Distance

Following Villani (2009, Definition 6.1), the Wasserstein distance between two probability distributions is defined as follows.[18]

Definition: *Let (S, d) be a Polish metric space. Assume that two probability measures P_I and P_{II} on S are continuous and have finite moments of order $p \in [1, \infty]$. The Wasserstein distance between P_I and P_{II} is defined as*

$$\begin{aligned} WD_p(P_I, P_{II}) &= \left[\inf \int_S d^p(x,y) d\pi(x,y) \right]^{1/p} \\ &= \inf\{[\mathrm{E}\left(d(X,Y)^p\right)]^{1/p},\ \mathrm{law}(X) = P_I \text{ and } \mathrm{law}(Y) = P_{II}\}, \end{aligned} \quad \text{(B1)}$$

where the infimum is taken over all $\pi(x, y)$ in $\Pi(P_I, P_{II})$, which is the set of joint probability measures on random variables $X \times Y$ with marginals P_I on X and P_{II} on Y. Especially, with p = 2, the quadratic Wasserstein distance is defined as:

$$WD_2(P_I, P_{II}) = inf(\mathrm{E}_\pi ||X - Y||^2)^{1/2}, \quad \text{(B2)}$$

where the infimum is taken over all the transport plans $\pi(x, y)$ in $\Pi(P_I, P_{II})$, with the marginal distribution of P_I on X and P_{II} on Y.

Given the definition of WD_2, the following theoretical properties are shown in the optimal transport literature:

(1) There exists a unique solution to the optimal transport problem of moving the mass of distribution P_I to distribution P_{II}. The one-to-one mapping is known as the optimal transport plan $y = T(x)$, where $x \sim P_I$ is mapped to $y \sim P_{II}$ via $T(x)$.

(2) Under the optimal transport plan, random vectors $X \sim P_I$ and $Y \sim P_{II}$ are maximally correlated with each other.

The above descriptions are standard results in the optimal transport theory (Villani, 2003, 2009). The quadratic Wasserstein distance, WD_2, has an economic interpretation as the minimum expected cost of transporting the mass of distribution P_I to distribution P_{II}. In general, there exists no closed-form formula for WD_2 or $T(x)$ for general probability distributions. Fortunately, when P_I and P_{II} are Gaussian, closed-form formula for WD_2 and $T(x)$ can be derived, with the key results being summarized in the following theorem.

Theorem: *Let P_I and P_{II} be Gaussian measures on $\mathbb{R}^n$ with finite second moments such that $P_I \sim N(\alpha_I, V_I)$ and $P_{II} \sim N(\alpha_{II}, V_{II})$, where α_I and α_{II} are two $n \times 1$ vectors of mean, and V_I and V_{II} are two $n \times n$ symmetric, positive-definite covariance matrices. Then, the quadratic Wasserstein distance (WD_2) between P_I and P_{II} is given by*

$$\begin{aligned} WD_2 &= \sqrt{||\alpha_{II} - \alpha_I||^2 + ||V_{II} - V_I||} \\ ||V_{II} - V_I|| &= \mathrm{Tr}\left(V_I + V_{II} - 2(V_I^{1/2} V_{II} V_I^{1/2})^{1/2} \right), \end{aligned} \quad \text{(B3)}$$

[18] The Wasserstein distance is also known as the Monge-Kantorovich distance.

where $||\alpha_{II} - \alpha_I||$ is the Euclidean 2-norm of the mean difference vector; $||V_{II} - V_I||$ is *the distance between the two covariance matrices; $Tr(\cdot)$ is the trace operator; and $V^{1/2}$ is the square-root of the covariance matrix V such that $V = V^{1/2}V^{1/2}$.*

Proof: Given the two normally distributed random vectors, $X \sim N(\alpha_I, V_I)$ and $Y \sim N(\alpha_{II}, V_{II})$, in $\mathbb{R}^n$, let the two demeaned random vectors be $\underline{X} = X - \alpha_I$ and $\underline{Y} = Y - \alpha_{II}$. The squared quadratic Wasserstein distance is defined by $WD_2^2 \equiv \mathrm{E}[||Y - X||^2] = ||\alpha_{II} - \alpha_I||^2 + \mathrm{E}[||\underline{Y} - \underline{X}||^2]$. For expositional convenience, denote $||V_{II} - V_I|| \equiv \mathrm{E}[||\underline{Y} - \underline{X}||^2]$. Then we have:

$$WD_2 = \sqrt{||\alpha_{II} - \alpha_I||^2 + ||V_{II} - V_I||}. \tag{B4}$$

For a closed-form expression of the distance under Gaussian measures, we need to show that:

$$||V_{II} - V_I|| = \mathrm{Tr}\left(V_I + V_{II} - 2(V_I^{1/2}V_{II}V_I^{1/2})^{1/2}\right). \tag{B5}$$

For the augmented random vector $(\underline{X}, \underline{Y})$ in $\mathbb{R}^{2n}$, denote its covariance matrix by

$$\Psi = \begin{bmatrix} V_I & C \\ C' & V_{II} \end{bmatrix}. \tag{B6}$$

Then, $||V_{II} - V_I|| = \mathrm{Tr}(V_I + V_{II} - 2C)$, and the infimum of $||V_{II} - V_I||$ is to find $C = \mathrm{E}[\underline{X}\ \underline{Y}']$ so that $\underline{X}$ and $\underline{Y}$ are maximally correlated, subject to the constraint that Ψ is a positive-definite covariance matrix. Thus, the optimization problem becomes:

$$\max_C \ \mathrm{Tr}(C) \tag{B7}$$

$$\text{s.t. } V_I - CV_{II}^{-1}C' > 0, \tag{B8}$$

where (B8) is the Schur complement constraint. The solution of (B7) subject to (B8) leads to (B5). The detailed proof is given by Dawson and Landau (1982) and Givens and Shortt (1984), who also term WD_2 as the Fréchet distance.

Equation (B5) can also be derived from the optimal transport mapping (Knott and Smith (1984) and Olkin and Pukelsheim (1982)). To check that the optimal transport plan maps $N(\alpha_I, V_I)$ to $N(\alpha_{II}, V_{II})$, for the zero-mean random vector $\underline{X} \sim N(0, V_I)$, let $\underline{Y} = T_I\underline{X}$, where the optimal mapping matrix T_I is given by

$$T_I = V_I^{-1/2}(V_I^{1/2}V_{II}V_I^{1/2})^{1/2}V_I^{-1/2}. \tag{B9}$$

Given $\underline{Y} = T_I\underline{X}$ and equation (B9), we have

$$\begin{aligned} \mathrm{E}[\underline{Y}\ \underline{Y}'] &= T_I\mathrm{E}[\underline{X}\ \underline{X}']T_I' = V_I^{-1/2}(V_I^{1/2}V_{II}V_I^{1/2})^{1/2}V_I^{-1/2}V_IV_I^{-1/2}(V_I^{1/2}V_{II}V_I^{1/2})^{1/2}V_I^{-1/2} \\ &= V_I^{-1/2}(V_I^{1/2}V_{II}V_I^{1/2})^{1/2}(V_I^{1/2}V_{II}V_I^{1/2})^{1/2}V_I^{-1/2} = V_I^{-1/2}(V_I^{1/2}V_{II}V_I^{1/2})V_I^{-1/2} \\ &= V_{II}\,. \end{aligned} \tag{B10}$$

In the univariate case where $V_I = \sigma_I^2$ and $V_{II} = \sigma_{II}^2$ are scalers, the optimal mapping matrix simplifies to a scaler $T_I = \sigma_I^{-1}(\sigma_I\sigma_{II}^2\sigma_I)^{1/2}\sigma_I^{-1} = \sigma_{II}/\sigma_I$. Then $\sigma_{II}^2 = (T_I\sigma_I)^2$ is easily verified.

To check that T_I is indeed optimal, we have:

$$\begin{aligned}
\mathrm{E}[||\underline{Y}-\underline{X}||^2] &= \mathrm{E}[||\underline{X}||^2]+\mathrm{E}[||\underline{Y}||^2]-2\mathrm{E}[\langle \underline{X},\underline{Y}\rangle] = \mathrm{Tr}(V_I)+\mathrm{Tr}(V_{II})-2\mathrm{E}[\langle \underline{X},T_I\underline{X}\rangle] \\
&= \mathrm{Tr}(V_I)+\mathrm{Tr}(V_{II})-2\mathrm{Tr}(V_I T_I) \\
&= \mathrm{Tr}(V_I)+\mathrm{Tr}(V_{II})-2\mathrm{Tr}\Big((V_I^{1/2}V_{II}V_I^{1/2})^{1/2}\Big) \\
&= \mathrm{Tr}\Big(V_I+V_{II}-2(V_I^{1/2}V_{II}V_I^{1/2})^{1/2}\Big). \qquad \text{(B11)}
\end{aligned}$$

The second last equality is obtained by the cyclic property of the trace operator. The converse optimal transport mapping can also be derived. Let $\underline{X}=T_{II}\underline{Y}$, where the optimal mapping matrix T_{II} is given by $T_{II}=V_{II}^{-1/2}(V_{II}^{1/2}V_I V_{II}^{1/2})^{1/2}V_{II}^{-1/2}$. It is easy to verify that $T_{II}=T_I^{-1}$.

Note that, given the non-informativeness in prior specification II, the posterior estimates $\tilde{\alpha}_{II}$ and $\tilde{V}_{\alpha_{II}}$ are identical to the maximum-likelihood estimates of the alpha, $\hat{\alpha}$, and its covariance matrix, $\hat{V}_\alpha$, respectively. Hence, $WD_2=\sqrt{||\tilde{\alpha}_{II}||^2+\mathrm{Tr}(\tilde{V}_{\alpha_{II}})}$ shown in Subsection 2.2. has its frequentist-equivalent form as

$$TD=\sqrt{||\hat{\alpha}||^2+\mathrm{Tr}(\hat{V}_\alpha)}\ , \qquad \text{(B12)}$$

which is intuitively interpretable as follows.

TD is the square-root of two sum-of-squared components. The first component, $||\hat{\alpha}||^2$ or $||\tilde{\alpha}||^2$ (the sample and posterior estimates can be used interchangeably, because of the non-informativeness in its prior specification), is the sum of squared alphas of the LHS returns in asset-pricing tests. The second component, $\mathrm{Tr}(\hat{V}_\alpha)$ or $\mathrm{Tr}(\tilde{V}_\alpha)$, is the sum of variances of alpha estimates for individual assets. Note that the two components in equation (B12) is sum-of-squared terms, $||\tilde{\alpha}||^2=\sum_{i=1}^n\tilde{\alpha}_i^2$ and $\mathrm{Tr}(\tilde{V})=\sum_{i=1}^n\tilde{\sigma}_{\alpha_i}^2$, where $\tilde{\sigma}_{\alpha_i}=\tilde{V}_\alpha^{1/2}(i,i)$ is the posterior estimate of the standard error of the alpha for asset i. More intuitively, we can express TD as

$$TD=\sqrt{\sum\nolimits_{i=1}^n\left(\tilde{\alpha}_i^2+\tilde{\sigma}_{\alpha_i}^2\right)}\ . \qquad \text{(B13)}$$

To compare the performance across different models when the universes of assets are different, we divide the two components in equation (B13) by n, resulting in the average distance (AD) defined as follows:

$$AD=\sqrt{\sum\nolimits_{i=1}^n(\tilde{\alpha}_i^2+\tilde{\sigma}_{\alpha_i}^2)/n}\ . \qquad \text{(B14)}$$

With only one LHS asset ($n=1$) (asset i), the marginal distance (d_i) is given by:

$$d_i=\sqrt{\tilde{\alpha}_i^2+\tilde{\sigma}_{\alpha_i}^2}\ . \qquad \text{(B15)}$$

C. Test Assets

Fama and French (2015)		Hou, Xue, and Zhang (2018)	
1.	ME: market capitalization	1.	Sue1: Earnings surprise (1-month holding period)
2.	BM: book-to-market ratio	2.	$R^6 6$: Price momentum (6-month prior returns, 6-month holding period)
3.	OP: operating profitability	3.	Im1: Industry momentum (1-month holding period)
4.	INV: investment ratio	4.	$\epsilon^6 6$: Six-month residual momentum (6-month holding period)
5.	EP: earnings-to-price	5.	Sim1: Supplier industries momentum (1-month holding period)
6.	CFP: Cash flow-to-price	6.	Cim1: Customer industries momentum (1-month holding period)
7.	DP: dividend yield	7.	Bm: Book-to-market equity
8.	MOM: momentum	8.	Em: Enterprise multiple
9.	STR: short-term reversal	9.	Sp: Sales-to-price
10.	LTR: long-term reversal	10.	Ir: Intangible return
11.	AC: accruals	11.	Vhp: Intrinsic value-to-market
12.	NI: net share issues	12.	Dur: Equity duration
13.	Beta: market beta	13.	I/A: Investment-to-assets
14.	VAR: return variance	14.	Ig: Investment growth
15.	RVAR: residual variance	15.	Ivc: Inventory changes
		16.	Nsi: Net stock issues
		17.	Cei: Composite equity issuance
		18.	Oa: Operating accruals
		19.	Roe1: Return on equity (1-month holding period)
		20.	dRoe1: Change in Roe (1-month holding period)
		21.	Gpa: Gross profits-to-assets
		22.	Opa: Operating profits-to-assets
		23.	Cop: Cash-based operating profitability
		24.	Oca: Organizational capital/assets
		25.	Ol: Operating leverage
		26.	R_a^1: 12-month-lagged return
		27.	$R_a^{[2,5]}$: Years 2-5 lagged returns (annual)
		28.	$R_n^{[2,5]}$: Years 2-5 lagged returns (non-annual)
		29.	Dtv12: Dollar trading volume (12-month holding period)
		30.	Isff1: Idiosyncratic skewness per the FF 3-factor model (1-month holding period)

Green, Hand, and Zhang (2017)

1. absacc: Absolute accruals
2. acc: Working capital accruals
3. aeavol: Abnormal earnings announcement volume
4. age: # years since first Compustat coverage
5. agr: Asset growth
6. baspread: Bid-ask spread
7. beta: Beta
8. bm: Book-to-market
9. bm_ia: Industry-adjusted book-to-market
10. cash: Cash holdings
11. cashdebt: Cash flow-to-debt
12. cashpr: Cash productivity
13. cfp: Cash flow-to-price ratio
14. cfp_ia: Industry-adjusted cash flow-to-price ratio
15. chatoia: Industry-adjusted change in asset turnover
16. chcsho: Change in shares outstanding
17. chempia: Industry-adjusted change in employees
18. chinv: Change in inventory
19. chmom: Change in 6-month momentum
20. chpmia: Industry-adjusted change in profit margin
21. chtx: Change in tax expense
22. cinvest: Corporate investment
23. convind: Convertible debt indicator
24. currat: Current ratio
25. depr: Depreciation-to-PP&E
26. divi: Dividend initiation
27. divo: Dividend omission
28. dy: Dividend-to-price
29. ear: Earnings announcement return
30. egr: Growth in common shareholder equity
31. ep: Earnings-to-price
32. gma: Gross profitability
33. grCAPX: Growth in capital expenditures
34. grltnoa: Growth in long term net operating assets
35. herf: Industry sales concentration
36. hire: Employee growth rate
37. idiovol: Idiosyncratic return volatility
38. ill: Illiquidity
39. indmom: Industry momentum
40. invest: Capital expenditures and inventory
41. IPO: New equity issue
42. lev: Leverage
43. mom12m: 12-month momentum
44. mom1m: 1-month momentum
45. mom36m: 36-month momentum
46. ms: Financial statement score
47. mve: Size
48. mve_ia: Industry-adjusted size
49. nincr: Number of earnings increases
50. operprof: Operating profitability
51. orgcap: Organizational capital
52. pchcapx_ia: Industry adjusted % change in capital expenditures
53. pchcurrat: % change in current ratio
54. pchdepr: % change in depreciation
55. pchgm_pchsale: % change in gross margin − % change in sales
56. pchsale_pchinvt: % change in sales − % change in inventory
57. pchsale_pchrect: % change in sales − % change in A/R
58. pchsale_pchxsga: % change in sales − % change in SG&A
59. pchsaleinv: % change sales-to-inventory
60. pctacc: Percent accruals
61. pricedelay: Price delay
62. ps: Financial statements score
63. rd: R&D increase
64. rd_mve: R&D-to-market capitalization
65. rd_sale: R&D-to-sales
66. retvol: Return volatility
67. roaq: Return on assets
68. roavol: Earnings volatility
69. roeq: Return on equity
70. roic: Return on invested capital
71. rsup: Revenue surprise
72. salecash: Sales-to-cash
73. saleinv: Sales-to-inventory
74. salerec: Sales to receivables
75. sgr: Sales growth
76. sin: Sin stocks
77. sp: Sales-to-price
78. std_dolvol: Volatility of liquidity (dollar trading volume)
79. std_turn: Volatility of liquidity (share turnover)
80. stdcf: Cash flow volatility
81. sue: Unexpected quarterly earnings
82. tang: Debt capacity-to-firm tangibility
83. tb: Tax income-to-book income
84. turn: Share turnover
85. zerotrade: Zero trading days

References

Abhyankar, Abhay, Ilias Filippou, Pedro Garcia-Ares, and Ozkan Haykir, 2019, Overcoming Arbitrage Limits: Option Trading and Momentum Returns, Working paper available at SSRN: `https://ssrn.com/abstract=3206873`.

Ahmed, Shamim, Ziwen Bu, and Daniel Tsvetanov, 2019, Best of the Best: A Comparison of Factor Models, *Journal of Financial and Quantitative Analysis* 54, 1713–1758.

Asness, Cliff, and Andrea Frazzini, 2013, The Devil in HML's Details, *Journal of Portfolio Management* 39, 49–68.

Barillas, Francisco, Raymond Kan, Cesare Robotti, and Jay Shanken, 2019, Model Comparison with Sharpe Ratios, forthcoming *Journal of Financial and Quantitative Analysis*.

Barillas, Francisco, and Jay Shanken, 2017, Which Alpha? *Review of Financial Studies* 30, 1316–1338.

Barillas, Francisco, and Jay Shanken, 2018, Comparing Asset Pricing Models, *Journal of Finance* 73, 715–753.

Carhart, Mark, 1997, On Persistence in Mutual Fund Performance, *Journal of Finance* 52, 57-82.

Cochrane, John H., 2005, *Asset Pricing*, Princeton University Press.

Daniel, Kent, David Hirshleifer, and Lin Sun, 2018, Short- and Long-Horizon Behavioral Factors, forthcoming *Review of Financial Studies*.

Deadman, Edvin, Nicholas J. Higham, and Rui Ralha, 2013, Blocked Schur Algorithms for Computing the Matrix Square Root, *Lecture Notes in Computer Science* 7782, 171–182.

Dawson, D. C., and B.V. Landau, 1982, The Fréchet Distance between Multivariate Normal Distributions, *Journal of Multivariate Analysis* 12, 450–455.

De Moore, Lieven, Geert Dhaene, and Piet Sercu, 2015, On Comparing Zero-Alpha Tests Across Multifactor Asset Pricing Models, *Journal of Banking and Finance* 61, 235–240.

Fama, Eugene F., 1998, Determining the Number of Priced State Variables in the ICAPM, *Journal of Financial and Quantitative Analysis* 33, 217–231.

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

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

Fama, Eugene F., and Kenneth R. French, 1996, Multifactor Explanations of Asset Pricing Anomalies, *Journal of Finance* 51, 55–84.

Fama, Eugene F., and Kenneth R. French, 2012, Size, Value, and Momentum in International Stock Returns, *Journal of Financial Economics* 105, 457–472.

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

Fama, Eugene F., and Kenneth R. French, 2016a, International Tests of a Five-Factor Asset Pricing Model, *Journal of Financial Economics* 123, 441–463.

Fama, Eugene F., and Kenneth R. French, 2016b, Dissecting Anomalies With a Five-Factor Model, *Review of Financial Studies* 29, 69–103.

Fama, Eugene F., and Kenneth R. French, 2018, Choosing Factors, *Journal of Financial Economics* 128, 234–252.

Galichon, Alfred, 2016, *Optimal Transport Methods in Economics*, Princeton University Press.

Galichon, Alfred, 2017, A Survey of Some Recent Applications of Optimal Transport Methods to Econometrics, *The Econometrics Journal* 20, C1–C11.

Green, Jeremiah, John R.M. Hand, and X. Frank Zhang, 2017, The Characteristics that Provide Independent Information about Average US Monthly Stock Returns, *Review of Financial Studies* 30, 4389–4436.

Gibbons, Michael R., Steven A. Ross, and Jay Shanken, 1989, A Test of the Efficiency of a Given Portfolio, *Econometrica* 57, 1121–1152.

Givens, Clark R., and Rae Michael Shortt, 1984, A Class of Wasserstein Metrics for Probability Distributions, *Michigan Mathematical Journal* 31, 231–240.

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

Harvey, Campbell R., and Yan Liu, Lucky Factors, 2018, Working paper.

Harvey, Campbell R., Yan Liu, and Heqing Zhu, 2016, ...and the Cross-Section of Expected Returns, *Review of Financial Studies* 29, 5–68.

He, Zhongzhi (Lawrence), 2007, Incorporating Alpha Uncertainty into Portfolio Decisions: A Bayesian Revisit of the Treynor-Black Model, *Journal of Asset Management* 8, 161–175.

Hou, Kewei, Chen Xue, and Lu Zhang, 2015, Digesting Anomalies: An Investment Approach, *Review of Financial Studies* 28, 650–705.

Hou, Kewei, Chen Xue, and Lu Zhang, 2016, A Comparison of New Factor Models, Working paper.

Hou, Kewei, Haitao Mo, Chen Xue, and Lu Zhang, 2018, q^5, Working paper available at SSRN: `https://ssrn.com/abstract=3191167`.

Hou, Kewei, Haitao Mo, Chen Xue, and Lu Zhang, 2019, Which Factors? *Review of Finance* 23, 1–35.

Kan, Raymond, and Cesare Robotti, 2008, Specification Tests of Asset Pricing Models Using Excess Returns, *Journal of Empirical Finance* 15, 816–838.

Kan, Raymond, Cesare Robotti, and Jay Shanken, 2013, Pricing Model Performance and the Two-Pass Cross-Sectional Regression Methodology, *Journal of Finance* 68, 2617–2649.

Kan, Raymond, and Chu Zhang, 1999, Two-Pass Tests of Asset Pricing Models with Useless Factors, *Journal of Finance* 54, 203–235.

Knott, M., and C.S. Smith, 1984, On the Optimal Mapping of Distributions, *Journal of Optimization Theory and its Applications* 43, 39–49.

Monge, Gaspard, 1781, Mémoire sur la Théorie des Déblais et des Remblais, In *Histoire de l'Académie Royale des Sciences de Paris*, 666–704.

Olkin, I., and F. Pukelsheim, 1982, The Distance between Two Random Vectors with Given Dispersion Matrices, *Linear Algebra and its Applications* 48, 257–263.

Pástor, Ľuboš, 2000, Portfolio Selection and Asset Pricing Models, *Journal of Finance* 55, 179–223.

Pástor, Ľuboš, and Robert F. Stambaugh, 2000, Comparing Asset Pricing Models: An Investment Perspective, *Journal of Financial Economics* 56, 335–381.

Shanken, Jay, and Guofu Zhou, 2007, Estimating and Testing Beta Pricing Models: Alternative Methods and Their Performance in Simulations, *Journal of Financial Economics* 84, 40–86.

Stambaugh, Robert F., and Yu Yuan, 2017, Mispricing Factors, *Review of Financial Studies* 30, 1270–1315.

Villani, Cédric, 2003, *Topics in Optimal Transportation*, American Mathematical Society.

Villani, Cédric, 2009, *Optimal Transport: Old and New*, Springer, Berlin.

Zellner, Arnold, 1971, *An Introduction to Bayesian Inference in Econometrics*, Wiley, New York.

Figure 1: A Geometric View of the Average Distance (AD) with 25 Size-MOM Portfolios: FF6 vs. q^5

This figure plots the pricing errors (alpha) ($\hat{\alpha}_i$ on the X-axis) and their standard errors (stderr) ($\hat{\sigma}_{\alpha_i}$ on the Y-axis) generated by two asset-pricing models with the 25 portfolios sorted on firm size (Size) and momentum (MOM) over the sample period of 1972:01-2015:12. Panels A and B show the plots generated by FF6 and q^5, respectively. In each panel, AD is measured by the length of the arrow connecting from the origin (indicated by a round dot) to the other round dot. The half circle has a radius of 0.2% per month as a benchmark.

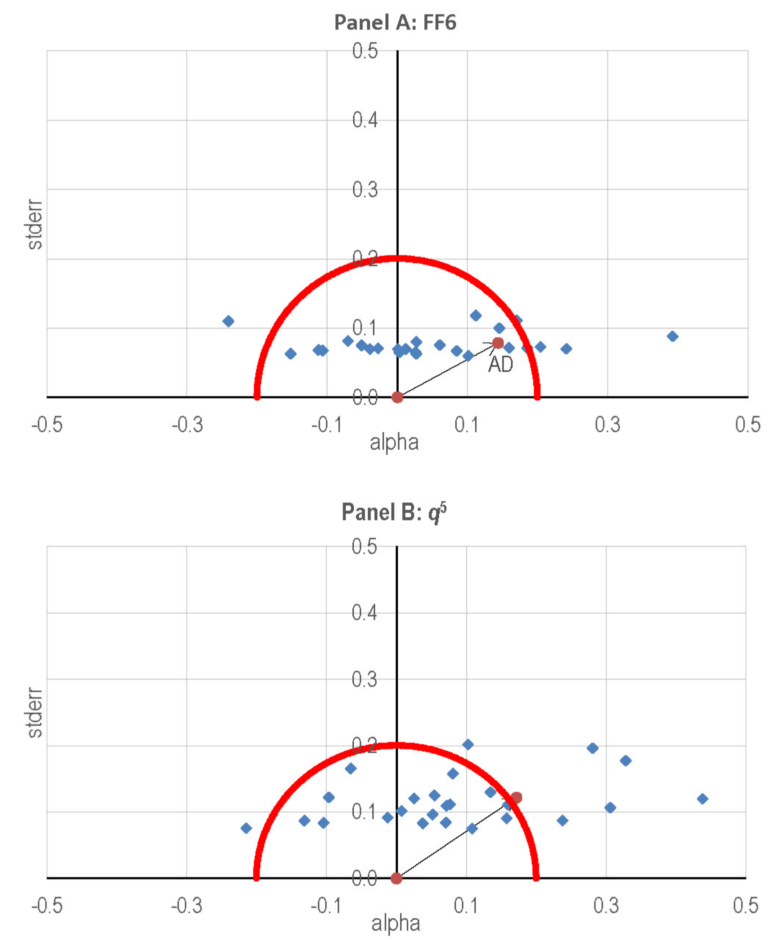
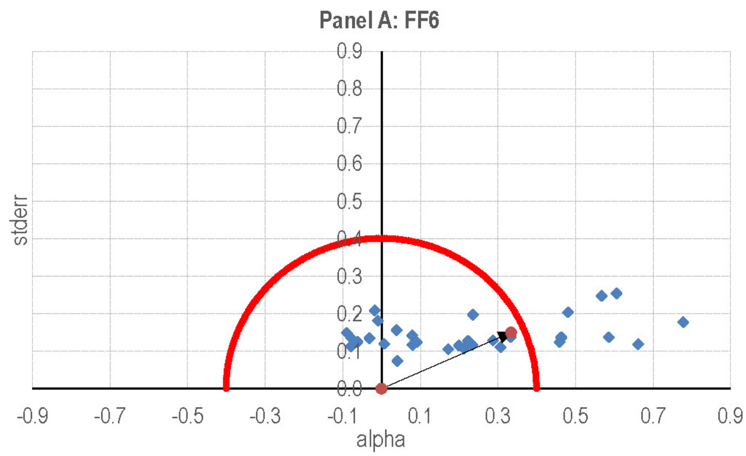
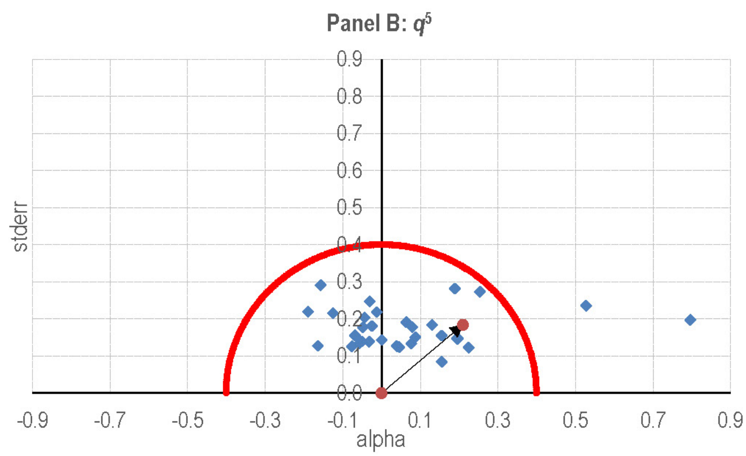


Figure 2: A Geometric View of the Average Distance (AD) with Hou, Mo, Xue, and Zhang's (2018) 30 H−L Portfolios: FF6 vs. q^5 vs. SY

This figure plots the pricing errors (alpha) ($\hat{\alpha}_i$ on the X-axis) and their standard errors (stderr) ($\hat{\sigma}_{\alpha_i}$ on the Y-axis) generated by the three asset-pricing models with 30 H−L portfolios computed from 300 decile portfolios of Hou, Mo, Xue, and Zhang (2018) over the sample period of 1972:01-2015:12. Panels A to C show the plots generated by FF6, q^5, and SY, respectively. In each panel, AD is measured by the length of the arrow connecting from the origin (a round dot) to the other round dot. The half circle has a radius of 0.4% per month as a benchmark.





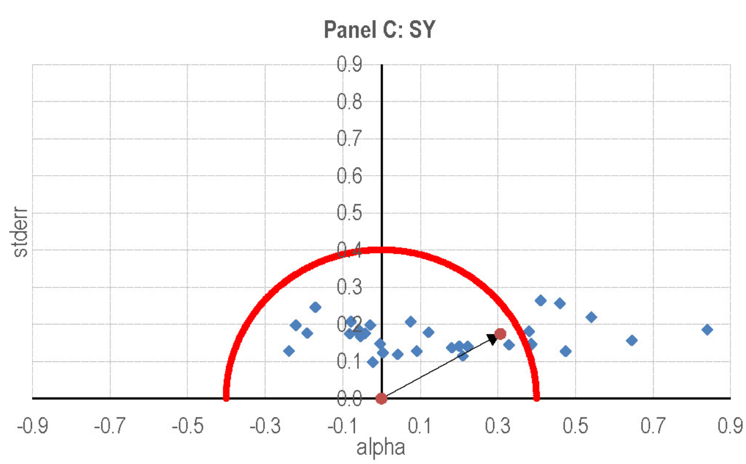


Figure 3: A Geometric View of the Average Distance (AD) with Green, Hand, and Zhang's (2017) 85 H−L Portfolios: FF6 vs. q^5 vs. SY

This figure plots the pricing errors (alpha) ($\hat{\alpha}_i$ on the X-axis) and their standard errors (stderr) ($\hat{\sigma}_{\alpha_i}$ on the Y-axis) generated by the three asset-pricing models with 85 H-L portfolios computed from 776 decile portfolios of Green, Hand, and Zhang (2017) over the sample period of 1972:01-2015:12. Panels A to C show the plots generated by FF6, q^5, and SY, respectively. In each panel, AD is measured by the length of the arrow connecting from the origin (a round dot) to the other round dot. The half circle has a radius of 0.4% per month as a benchmark.

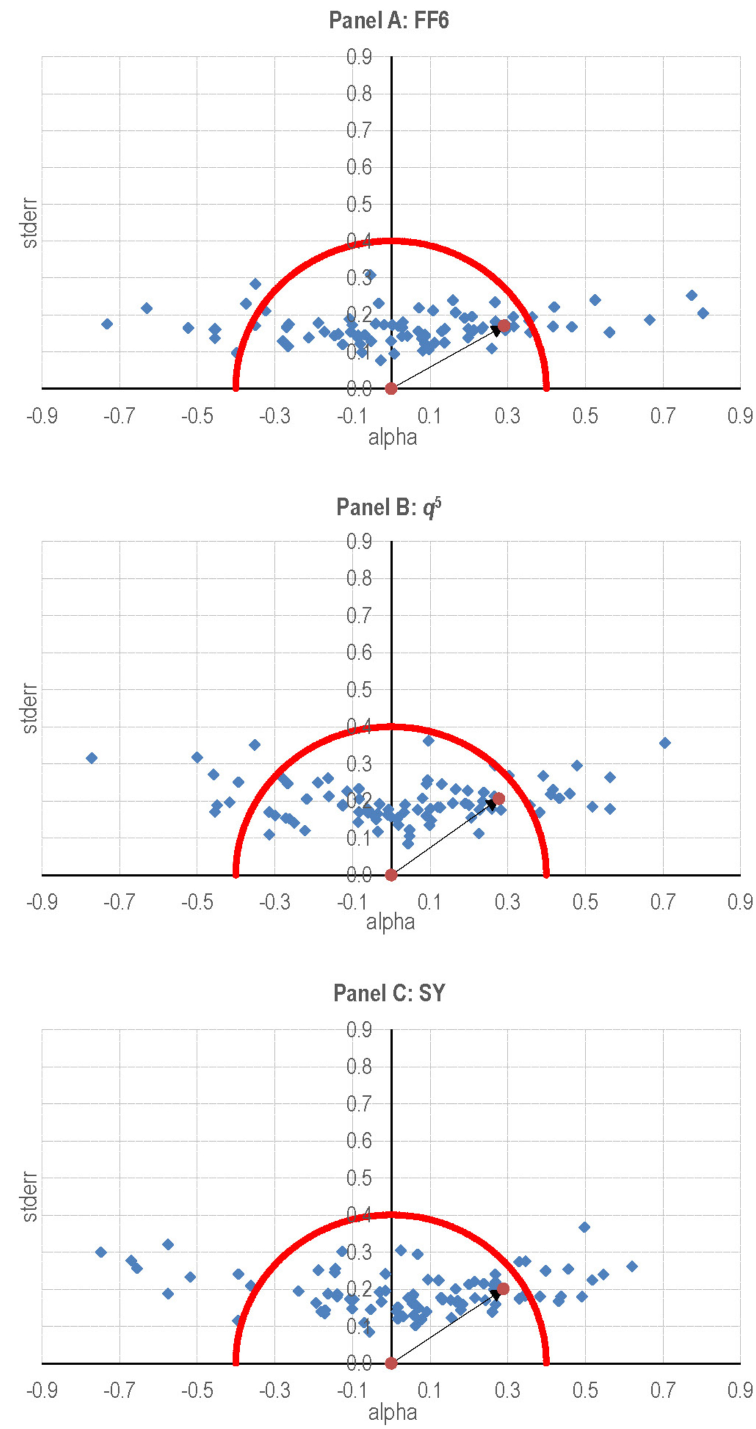


Figure 4: The Average Distance (AD) under Varying Degrees of Prior Estimation Uncertainty

This figure plots the average distance (AD) (on the Y-axis) generated by the 10 asset-pricing models with a set of 85 H−L portfolio returns, when the models have varying degrees of prior estimation uncertainty (i.e., as σ_α increases from 0% to 10% on the X-axis). The 85 H−L returns are computed using the 776 portfolios from by Green, Hand, and Zhang (2017) as in Table 6 as well as Panel C in Table 5. The sample period is 1980:01-2015:12.

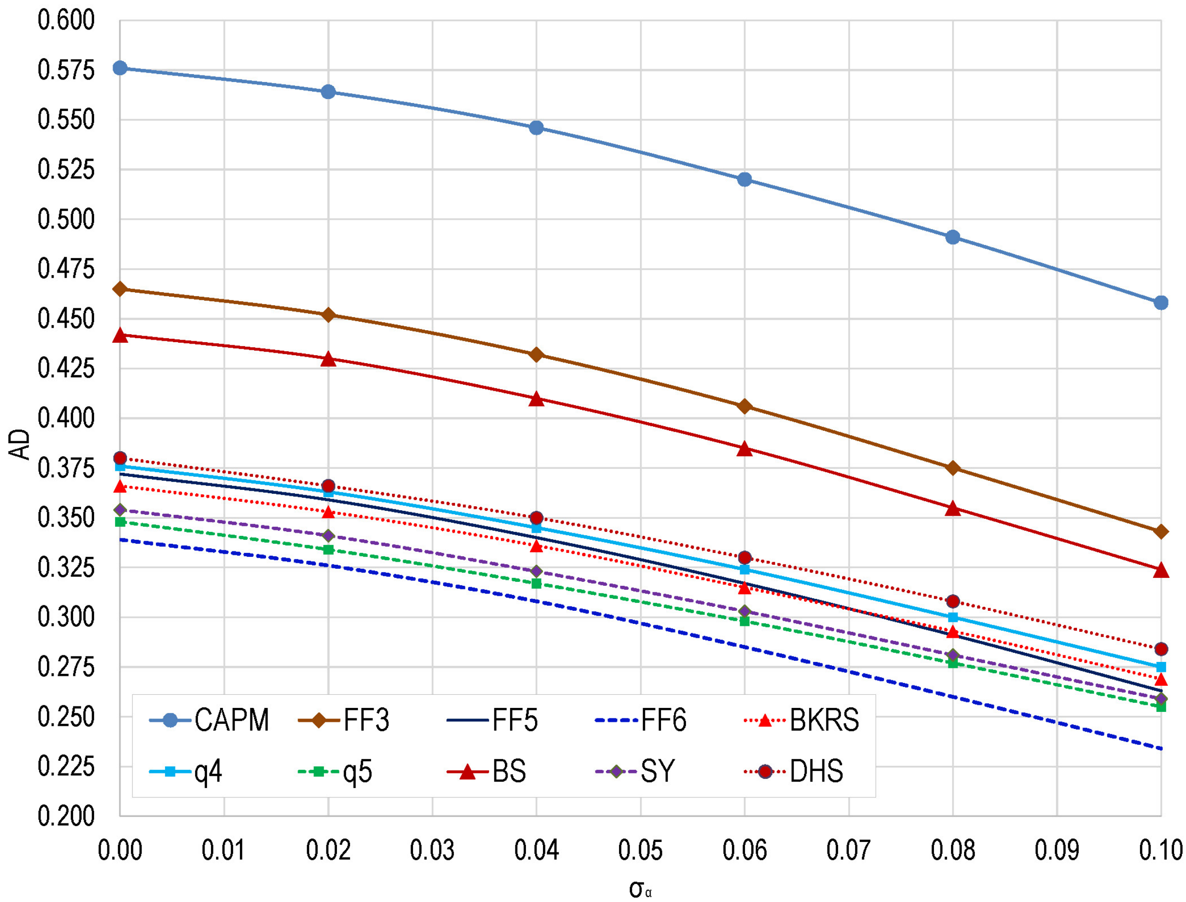


Table 1: Comparison of Performance Metrics: Distance-Based Metrics, *GRS*-Statistic and Mean Absolute Pricing Error (*MAE*)

This table compares the performance metrics for asset-pricing tests. AD and d_i are the average distance and the marginal distance, respectively. GRS is the F-statistic of Gibbons, Ross, and Shanken (1989) for a joint test of the zero-alpha restriction in time-series regressions. $MAE_{\hat{\alpha}}$ is the mean absolute intercept (alpha) from the time-series regressions. $\mathrm{Sh}^2(F)$ is the squared Sharpe ratio for the factors F.

Item / Metric	Distance-Based Metrics	*GRS*	*MAE*
Definition	$AD = \sqrt{\sum_{i=1}^{n}\left(\hat{\alpha}_i^2 + \hat{\sigma}_{\alpha_i}^2\right)/n}$ if $\sigma_\alpha = 0$ in prior specification	$GRS = c \cdot \hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}$ where c is a constant	$MAE_{\hat{\alpha}} = \sum_{i=1}^{n} \lvert\hat{\alpha}_i\rvert/n$
Marginal contribution of asset i	$d_i = \sqrt{\hat{\alpha}_i^2 + \hat{\sigma}_{\alpha_i}^2}$ if $\sigma_\alpha = 0$ in prior specification	$t_i = \hat{\alpha}_i/\hat{\sigma}_{\alpha_i}$	$\lvert\hat{\alpha}_i\rvert$
Measurement unit	Return in %	F- or t-statistic	Return in %
Mispricing parameter	(Bayesian view) The alpha is random, data are given for updating the posterior distribution of alphas; performance is distance-based; small alpha and high estimation precision are preferred	(Frequentist view) The alpha is fixed, data are random; relies on sampling theory to derive test statistics; performance is ratio-based; a lower ratio of alpha estimates to sampling errors is preferred	Simple statistical artifact regardless of the view
Theoretical motivation	Bayesian method (Pástor and Stambaugh (2000)); Optimal transport theory (Villani (2003))	Sampling theory of multivariate statistics (Gibbons, Ross, and Shanken (1989))	Ad-hoc statistical measure
Economic interpretation	The minimum cost of holding dogmatic beliefs in the model	The difference between the $\mathrm{Sh}^2(R, F)$ and $\mathrm{Sh}^2(F)$; tests if factors span the mean-variance-efficient tangency portfolio	No theory or explanation
How to treat pricing errors (Alphas)	$RMSE_{\hat{\alpha}} = \sqrt{\sum_{i=1}^{n}\hat{\alpha}_i^2/n}$ if $\sigma_\alpha = 0$ in prior specification; consider large pricing errors highly undesirable, and heavily penalizes models that produce extreme alphas	Squared alphas are weighted by the inverse of covariance matrix of alphas, so large $\left(\hat{\alpha}_i/\hat{\sigma}_{\alpha_i}\right)^2$ dominate the F-statistic; this creates too much power for the GRS test to reject any pricing model for a large cross-section of assets	Different magnitude of pricing errors is treated equally
How to treat estimation precision	$RMSE_{\hat{\sigma}_\alpha} = \sqrt{\sum_{i=1}^{n}\hat{\sigma}_{\alpha_i}^2/n}$ if $\sigma_\alpha = 0$ in prior specification; considers large standard errors of alphas highly undesirable, and penalizes models of low estimation precision	Models with large sampling errors tend to produce smaller F-statistics and p-values; this causes too small power for *GRS*, rejecting bad models less often	Not considered
Characteristics of good models	Low dispersion of alphas; high estimation precision; i.e., smaller (also fewer extreme) pricing errors and low standard errors	Low dispersion of alphas; large sampling errors (i.e., larger standard errors of alphas)	Small size of alphas

Table 2: Performance of Two Asset-Pricing Models with the 25 Size-MOM Portfolios

This table reports the posterior estimates of alphas ($\hat{\alpha}_i$), their standard errors ($\hat{\sigma}_{\alpha_i}$) and t-statistics ($t_i = \hat{\alpha}_i/\hat{\sigma}_{\alpha_i}$), as well as the marginal distance ($d_i = \sqrt{\hat{\alpha}_i^2 + \hat{\sigma}_{\alpha_i}^2}$) with non-informative priors, generated by two asset-pricing models with the 25 (5×5) portfolios sorted on firm size (Size: Small to Big) and momentum (MOM: loser (Los) to winner (Win)). The details on portfolio construction are available from Kenneth French's website. The two models (and factors) are FF6 (*MKT, SMB, HML, RMW, CMA, UMD*) and q^5 (*MKT, ME, IA, ROE, EG*). Further details on the models are provided in the text. The performance metrics reported at the bottom of each panel are defined in equations (6) and (9). AR^2 is the cross-sectional average of R^2's from the time-series regressions. The sample period is from 1972:01 to 2015:12 (528 months).

	$\hat{\alpha}_i$					$\hat{\sigma}_{\alpha_i}$					t_i					d_i				
	Los	2	3	4	Win	Los	2	3	4	Win	Los	2	3	4	Win	Los	2	3	4	Win
Panel A: FF6 Model																				
Small	−0.24	−0.04	0.08	0.19	0.39	0.11	0.07	0.07	0.07	0.09	−2.19	−0.56	1.26	2.61	4.46	0.26	0.08	0.11	0.20	0.40
2	−0.07	0.03	0.03	0.10	0.24	0.08	0.07	0.06	0.06	0.07	−0.87	0.40	0.43	1.68	3.45	0.11	0.07	0.07	0.12	0.25
3	0.15	0.01	0.00	−0.11	0.16	0.10	0.07	0.06	0.07	0.07	1.46	0.17	0.04	−1.65	2.23	0.18	0.07	0.06	0.13	0.17
4	0.17	0.06	0.00	−0.03	0.03	0.11	0.08	0.07	0.07	0.08	1.53	0.81	0.01	−0.39	0.34	0.20	0.10	0.07	0.08	0.08
Big	0.11	0.20	−0.11	−0.15	−0.05	0.12	0.07	0.07	0.06	0.07	0.94	2.80	−1.58	−2.42	−0.68	0.16	0.22	0.13	0.16	0.09
$AD = 0.164$; $RMSE_{\hat{\alpha}} = 0.143$; $RMSE_{\hat{\sigma}_\alpha} = 0.079$; $\frac{MSE_{\hat{\sigma}_\alpha}}{MSE_{\hat{\alpha}}} = 0.30$; $GRS = 2.99$; $MAE_{\hat{\alpha}} = 0.110$; $\frac{MAE_{\hat{\alpha}}}{MAR_{\tilde{r}}} = 0.39$; $AR2 = 0.92$																				
Panel B: q^5 Model																				
Small	−0.07	0.01	0.16	0.24	0.44	0.17	0.10	0.09	0.09	0.12	−0.40	0.07	1.74	2.71	3.65	0.18	0.10	0.18	0.25	0.45
2	0.08	0.07	0.07	0.11	0.31	0.16	0.11	0.08	0.08	0.11	0.51	0.65	0.84	1.44	2.87	0.18	0.13	0.11	0.13	0.32
3	0.33	0.08	−0.01	−0.10	0.16	0.18	0.11	0.09	0.08	0.11	1.85	0.69	−0.14	−1.24	1.43	0.37	0.14	0.09	0.13	0.19
4	0.28	0.05	0.05	0.04	0.03	0.20	0.12	0.10	0.08	0.12	1.43	0.43	0.54	0.45	0.21	0.34	0.14	0.11	0.09	0.12
Big	0.10	0.13	−0.13	−0.21	−0.10	0.20	0.13	0.09	0.08	0.12	0.51	1.04	−1.52	−2.84	−0.79	0.23	0.19	0.16	0.23	0.16
$AD = 0.211$; $RMSE_{\hat{\alpha}} = 0.172$; $RMSE_{\hat{\sigma}_\alpha} = 0.122$; $\frac{MSE_{\hat{\sigma}_\alpha}}{MSE_{\hat{\alpha}}} = 0.50$; $GRS = 2.64$; $MAE_{\hat{\alpha}} = 0.134$; $\frac{MAE_{\hat{\alpha}}}{MAR_{\tilde{r}}} = 0.48$; $AR2 = 0.85$																				

Table 3: Performance of the Ten Models Using 25 Size-MOM Portfolios

This table reports the performance-related statistics generated by ten asset-pricing models with the 25 portfolios formed by sorting on firm size (Size) and momentum (MOM). The details on portfolio construction are available from Kenneth French's website. $\hat{\alpha}_i$ and $\hat{\sigma}_{\alpha_i}$ are the posterior estimates of a pricing error and its standard error for asset i with non-informative priors. Further details on the models and the definitions of the factors are provided in the text. The performance metrics are defined in equations (6) and (9). The sample period is from 1972:01 to 2015:12 (528 months) (except for the DHS model whose factors are available from 1972:01 to 2014:12).

Model	AD	$RMSE_{\hat{\alpha}}$	$RMSE_{\hat{\sigma}_\alpha}$	$\frac{MSE_{\hat{\sigma}_\alpha}}{MSE_{\hat{\alpha}}}$	GRS	$MAE_{\hat{\alpha}}$	$\frac{MAE_{\hat{\alpha}}}{MAR_{\bar{r}}}$	AR^2
CAPM	0.427	0.402	0.142	0.12	4.60	0.331	1.18	0.73
FF Models								
FF3	0.450	0.436	0.111	0.06	4.49	0.338	1.21	0.84
FF5	0.359	0.342	0.112	0.10	3.63	0.275	0.98	0.85
FF6	0.164	0.143	0.079	0.30	2.99	0.110	0.39	0.92
BKRS	0.196	0.175	0.089	0.26	3.42	0.135	0.48	0.91
q Models								
q^4	0.202	0.168	0.112	0.44	2.84	0.118	0.42	0.85
q^5	0.211	0.172	0.122	0.50	2.64	0.134	0.48	0.85
Other Models								
BS	0.184	0.164	0.082	0.25	3.18	0.143	0.51	0.92
SY	0.197	0.166	0.106	0.41	2.58	0.122	0.44	0.87
DHS	0.434	0.410	0.142	0.12	4.23	0.335	1.23	0.74

Table 4: Performance of the Ten Models Using Decile Portfolios

This table reports the performance-related statistics generated by the ten asset-pricing models using two large sets of decile portfolios as test assets. The test assets in Panel A are a pooled set of decile portfolios sorted on 15 different anomaly variables (150 portfolios in total) from Fama and French (2015). The test assets in Panel B are a pooled set of decile portfolios sorted on 30 anomaly variables (300 portfolios in total) from Hou, Mo, Xue, and Zhang (2018). Further details on the models and the definitions of the factors are provided in the text. The performance metrics are defined in equations (6) and (9). The sample period is from 1972:01 to 2015:12 (528 months), except for the DHS model whose factors are available from 1972:01 to 2014:12.

Model	AD	$RMSE_{\hat{\alpha}}$	$RMSE_{\hat{\sigma}_\alpha}$	$\frac{MSE_{\hat{\sigma}_\alpha}}{MSE_{\hat{\alpha}}}$	GRS	$MAE_{\hat{\alpha}}$	$\frac{MAE_{\hat{\alpha}}}{MAR_{\bar{r}}}$	AR^2
	Panel A: Fama and French's (2015) 150 Portfolios							
CAPM	0.248	0.228	0.097	0.18	2.12	0.167	1.48	0.83
FF Models								
FF3	0.219	0.202	0.083	0.17	2.03	0.119	1.06	0.87
FF5	0.171	0.150	0.082	0.30	1.88	0.103	0.92	0.88
FF6	0.138	0.114	0.078	0.47	1.87	0.087	0.77	0.89
BKRS	0.146	0.115	0.089	0.59	1.89	0.092	0.82	0.88
q Models								
q^4	0.156	0.129	0.088	0.47	2.11	0.101	0.90	0.87
q^5	0.163	0.132	0.096	0.54	1.92	0.110	0.98	0.87
Other Models								
BS	0.184	0.164	0.083	0.26	2.15	0.135	1.20	0.89
SY	0.152	0.123	0.089	0.53	1.55	0.097	0.86	0.87
DHS	0.214	0.193	0.093	0.23	2.14	0.137	1.19	0.85
	Panel B: Hou, Mo, Xue, and Zhang's (2018) 300 Portfolios							
CAPM	0.230	0.210	0.093	0.20	2.00	0.165	1.26	0.83
FF Models								
FF3	0.208	0.189	0.087	0.21	1.91	0.141	1.08	0.85
FF5	0.180	0.158	0.087	0.30	1.85	0.117	0.89	0.86
FF6	0.155	0.129	0.086	0.44	1.80	0.099	0.76	0.86
BKRS	0.170	0.141	0.095	0.46	1.60	0.114	0.87	0.86
q Models								
q^4	0.171	0.145	0.092	0.40	1.80	0.112	0.86	0.85
q^5	0.156	0.119	0.100	0.70	1.59	0.095	0.73	0.85
Other Models								
BS	0.190	0.167	0.091	0.29	1.82	0.130	0.99	0.86
SY	0.154	0.122	0.094	0.60	1.57	0.093	0.71	0.85
DHS	0.184	0.153	0.102	0.45	1.62	0.113	0.84	0.84

Table 5: Performance of the Ten Models Using H−L Portfolios

This table reports the performance-related statistics generated by the ten asset-pricing models using three sets of H−L portfolios as test assets. The test assets used in Panel A are 15 H−L portfolios from 150 decile portfolios of Fama and French (2015), those in Panel B are 30 H−L portfolios from 300 decile portfolios of Hou, Mo, Xue, and Zhang (2018), and those in Panel C are 85 H−L portfolios from 776 portfolios of Green, Hand, and Zhang (2017). Further details on the models and the definitions of the factors are provided in the text. The performance metrics are defined in equations (6) and (9). The sample period for Panels A and B is from 1972:01 to 2015:12 (528 months), except for the DHS model whose factors are available from 1972:01 to 2014:12. The sample period for Panel C is from 1980:01 to 2015:12 (432 months), except for the DHS model whose factors are available from 1980:01 to 2014:12.

Model	AD	$RMSE_{\hat{\alpha}}$	$RMSE_{\hat{\sigma}_\alpha}$	$\frac{MSE_{\hat{\sigma}_\alpha}}{MSE_{\hat{\alpha}}}$	GRS	$MAE_{\hat{\alpha}}$	$\frac{MAE_{\hat{\alpha}}}{MAR_{\tilde{r}}}$	AR^2
	Panel A: Fama and French's (2015) 15 H−L Portfolios							
CAPM	0.741	0.710	0.212	0.09	5.90	0.603	1.37	0.13
FF Models								
FF3	0.681	0.660	0.169	0.07	5.16	0.458	1.04	0.46
FF5	0.462	0.432	0.164	0.14	3.12	0.267	0.61	0.53
FF6	0.260	0.216	0.144	0.44	2.13	0.160	0.36	0.60
BKRS	0.320	0.273	0.167	0.37	3.34	0.227	0.52	0.55
q Models								
q^4	0.289	0.227	0.179	0.63	1.92	0.177	0.40	0.44
q^5	0.288	0.211	0.196	0.86	1.58	0.149	0.34	0.44
Other Models								
BS	0.308	0.265	0.157	0.35	2.85	0.217	0.49	0.55
SY	0.285	0.223	0.177	0.64	1.75	0.157	0.36	0.46
DHS	0.478	0.427	0.214	0.25	2.91	0.335	0.71	0.29
	Panel B: Hou, Mo, Xue, and Zhang's (2018) 30 H−L Portfolios							
CAPM	0.652	0.626	0.184	0.09	5.80	0.600	4.39	0.04
FF Models								
FF3	0.628	0.607	0.162	0.07	5.26	0.501	3.67	0.26
FF5	0.502	0.476	0.159	0.11	4.01	0.374	2.74	0.33
FF6	0.366	0.334	0.150	0.20	3.44	0.250	1.83	0.40
BKRS	0.378	0.338	0.169	0.25	2.53	0.282	2.07	0.36
q Models								
q^4	0.384	0.345	0.170	0.24	3.51	0.264	1.93	0.28
q^5	0.279	0.210	0.184	0.77	1.48	0.137	1.00	0.30
Other Models								
BS	0.409	0.376	0.159	0.18	3.98	0.304	2.23	0.38
SY	0.352	0.306	0.174	0.32	2.86	0.227	1.66	0.27
DHS	0.367	0.310	0.196	0.40	2.92	0.260	2.05	0.15

Model	AD	$RMSE_{\hat{\alpha}}$	$RMSE_{\hat{\sigma}_\alpha}$	$\frac{MSE_{\hat{\sigma}_\alpha}}{MSE_{\hat{\alpha}}}$	GRS	$MAE_{\hat{\alpha}}$	$\frac{MAE_{\hat{\alpha}}}{MAR_{\bar{r}}}$	AR^2
	Panel C: Green, Hand, and Zhang's (2017) 85 H−L portfolios							
CAPM	0.576	0.533	0.217	0.17	2.29	0.424	1.48	0.10
FF Models								
FF3	0.465	0.427	0.184	0.19	2.27	0.331	1.15	0.31
FF5	0.372	0.327	0.177	0.29	1.87	0.243	0.85	0.39
FF6	0.339	0.293	0.172	0.34	1.77	0.225	0.78	0.41
BKRS	0.366	0.308	0.199	0.42	1.54	0.232	0.81	0.36
q Models								
q^4	0.376	0.322	0.194	0.36	1.99	0.242	0.84	0.33
q^5	0.348	0.279	0.207	0.55	1.47	0.220	0.77	0.34
Other Models								
BS	0.442	0.403	0.182	0.20	2.16	0.299	1.04	0.39
SY	0.354	0.290	0.202	0.48	1.62	0.227	0.79	0.31
DHS	0.380	0.315	0.211	0.45	1.93	0.249	0.85	0.26

Table 6: Performance Metrics under Varying Degrees of Prior Estimation Uncertainty Using 85 H−L Portfolios

This table reports the average distance (AD), its components, and a related metric generated by the ten asset-pricing models using a set of 85 H−L portfolio returns under varying degrees of prior estimation uncertainty (σ_α). The 85 H−L returns are computed using the 776 portfolios from Green, Hand, and Zhang (2017). Further details on the models and the definitions of the factors are provided in the text. The performance metrics are defined in equation (10). Panels A to F use the levels of prior uncertainty at $\sigma_\alpha = 0$, 0.02, 0.04, 0.06, 0.08, and 0.10, respectively. The sample period is from 1980:01 to 2015:12, except for the DHS model whose factors are available from 1980:01 to 2014:12.

Model	AD	$RMSE_{\tilde{\alpha}-\hat{\alpha}}$	$RMSE_{\tilde{\sigma}_\alpha-\hat{\sigma}_\alpha}$	$\frac{MSE_{\tilde{\sigma}_\alpha-\hat{\sigma}_\alpha}}{MSE_{\tilde{\alpha}-\hat{\alpha}}}$	AD	$RMSE_{\tilde{\alpha}-\hat{\alpha}}$	$RMSE_{\tilde{V}-\hat{\Sigma}}$	$\frac{MSE_{\tilde{V}-\hat{\Sigma}}}{MSE_{\tilde{\alpha}-\hat{\alpha}}}$
	Panel A: $\sigma_\alpha = 0\%$				Panel B: $\sigma_\alpha = 2\%$			
CAPM	0.576	0.533	0.217	0.17	0.564	0.529	0.197	0.14
FF Models								
FF3	0.465	0.427	0.184	0.19	0.452	0.422	0.164	0.15
FF5	0.372	0.327	0.177	0.29	0.359	0.322	0.157	0.24
FF6	0.339	0.293	0.172	0.34	0.326	0.289	0.152	0.28
BKRS	0.366	0.308	0.199	0.42	0.353	0.305	0.179	0.34
q Models								
q^4	0.376	0.322	0.194	0.36	0.363	0.318	0.174	0.30
q^5	0.348	0.279	0.207	0.55	0.334	0.277	0.188	0.46
Other Models								
BS	0.442	0.403	0.182	0.20	0.430	0.399	0.162	0.17
SY	0.354	0.290	0.202	0.48	0.341	0.287	0.182	0.40
DHS	0.380	0.315	0.211	0.45	0.366	0.313	0.191	0.37

Model	AD	$RMSE_{\tilde{\alpha}-\hat{\alpha}}$	$RMSE_{\tilde{\sigma}_\alpha-\hat{\sigma}_\alpha}$	$\frac{MSE_{\tilde{\sigma}_\alpha-\hat{\sigma}_\alpha}}{MSE_{\tilde{\alpha}-\hat{\alpha}}}$	AD	$RMSE_{\tilde{\alpha}-\hat{\alpha}}$	$RMSE_{\tilde{V}-\hat{\Sigma}}$	$\frac{MSE_{\tilde{V}-\hat{\Sigma}}}{MSE_{\tilde{\alpha}-\hat{\alpha}}}$
	Panel C: $\sigma_\alpha = 4\%$				Panel D: $\sigma_\alpha = 6\%$			
CAPM	0.546	0.516	0.178	0.12	0.520	0.496	0.159	0.10
FF Models								
FF3	0.432	0.407	0.145	0.13	0.406	0.386	0.127	0.11
FF5	0.340	0.311	0.138	0.20	0.317	0.293	0.120	0.17
FF6	0.308	0.278	0.132	0.23	0.285	0.261	0.115	0.19
BKRS	0.336	0.296	0.160	0.29	0.315	0.282	0.141	0.25
q Models								
q^4	0.345	0.309	0.155	0.25	0.324	0.294	0.136	0.22
q^5	0.317	0.269	0.168	0.39	0.298	0.258	0.150	0.34
Other Models								
BS	0.410	0.385	0.143	0.14	0.385	0.364	0.125	0.12
SY	0.323	0.279	0.162	0.33	0.303	0.267	0.144	0.29
DHS	0.350	0.304	0.172	0.32	0.330	0.292	0.153	0.28
	Panel E: $\sigma_\alpha = 8\%$				Panel F: $\sigma_\alpha = 10\%$			
CAPM	0.491	0.470	0.142	0.09	0.458	0.440	0.126	0.08
FF Models								
FF3	0.375	0.359	0.110	0.09	0.343	0.329	0.096	0.08
FF5	0.291	0.271	0.104	0.15	0.263	0.248	0.090	0.13
FF6	0.260	0.240	0.099	0.17	0.234	0.218	0.085	0.15
BKRS	0.293	0.265	0.124	0.22	0.269	0.246	0.109	0.20
q Models								
q^4	0.300	0.275	0.120	0.19	0.275	0.254	0.105	0.17
q^5	0.277	0.243	0.133	0.30	0.255	0.227	0.117	0.27
Other Models								
BS	0.355	0.338	0.108	0.10	0.324	0.310	0.094	0.09
SY	0.281	0.251	0.127	0.26	0.259	0.233	0.112	0.23
DHS	0.308	0.276	0.136	0.24	0.284	0.258	0.121	0.22

Table 7: Values of Individual Factors and their Combinations in the Three Benchmark Models

This table reports the values in AD and the levels of prior estimation uncertainty ($\sigma_\alpha = \sigma^*_\alpha$) (in square brackets) that makes the two (parsimonious and benchmark) models distance-equivalent when each combination of factors are excluded from the three (full) benchmark models (FF6 in Panel A, q^5 in Panel B, and SY in Panel C). The test assets used in the table are the set of 85 H−L portfolio returns, which are computed using the 776 portfolios from Green, Hand, and Zhang (2017). AD is the average distance computed based on equation (6) under a dogmatic prior belief ($\sigma_\alpha = 0$) in the model that excludes some of the factors from the benchmark model. The value in the square bracket represents the prior uncertainty in mispricing that an investor must accept when she chooses the model that excludes some factors from the respective benchmark model (instead of choosing the benchmark model itself). The diagonal grids in each panel contain the values of AD and σ^*_α when the corresponding (one) factor is excluded from the full model. The set of excluded factors below the diagonal grids consists of the diagonal factor plus all the factors up to that row. The sample period is from 1980:01 to 2015:12.

Panel A: FF6						
	MKT	*SMB*	*HML*	*RMW*	*CMA*	*UMD*
MKT	0.415					
	[7.1%]					
SMB	0.481	0.359				
	[10.7%]	[2.8%]				
HML	0.504	0.373	0.349			
	[12.0%]	[4.2%]	[1.5%]			
RMW	0.372	0.404	0.393	0.380		
	[4.3%]	[7.2%]	[6.0%]	[4.8%]		
CMA	0.393	0.571	0.565	0.387	0.355	
	[6.5%]	[16.7%]	[15.3%]	[5.4%]	[2.3%]	
UMD	0.423	0.576	0.581	0.465	0.414	0.372
	[9.2%]	[17.1%]	[16.1%]	[10.2%]	[7.2%]	[4.1%]

Panel B: q^5					
	MKT	*ME*	*IA*	*ROE*	*EG*
MKT	0.555				
	[14.7%]				
ME	0.741	0.481			
	[21.6%]	[11.9%]			
IA	0.683	0.410	0.324		
	[21.5%]	[7.3%]	[−3.3%]		
ROE	0.689	0.412	0.322	0.333	
	[22.3%]	[7.5%]	[−3.6%]	[−2.1%]	
EG	0.423	0.571	0.595	0.427	0.376
	[9.2%]	[16.3%]	[16.4%]	[8.0%]	[3.7%]

	Panel C: SY			
	MKT	*SMB*	*MGMT*	*PERF*
MKT	0.453			
	[9.1%]			
SMB	0.556	0.398		
	[14.3%]	[5.4%]		
MGMT	0.371	0.545	0.617	
	[2.5%]	[14.9%]	[17.1%]	
PERF	0.423	0.576	0.659	0.439
	[9.2%]	[16.2%]	[18.6%]	[8.5%]