Index + Factors + Alpha

This Version: March 13, 2021

We establish, under both theoretical conditions and empirical application, the separate roles of (1) market asset class exposure through index funds; (2) style factor exposure like value, momentum, and quality which have traditionally delivered higher and differentiated returns than market index exposure; and (3) pure alpha-seeking sources of return *in excess* of index and factor returns. A new methodology determines optimal allocations of index, factors, and alpha-seeking funds by imposing priors on the information ratios of factors and alpha strategies. We expect in many cases, prior standard deviations for factor funds should be smaller than alpha strategies, whereas prior means for alpha strategies may be larger than those for factor funds.

MASH0820U-1275625-0/47

Market cap index. Factors. Alpha. These well-known sources of returns each have extensive histories in academic literature and in practice.[1] The seminal role of the market in the first asset pricing model of the CAPM has played a major role in finance since Sharpe (1964) and Lintner (1965). Today's smart beta industry directly targets style factors like value, momentum, quality, size, and minimum volatility, and owes its beginning to the first formal factor model by Ross (1976) and the many empirical studies since then. Perhaps the longest history is that of active management—active managers have attempted to beat benchmarks in comingled vehicles since the 18th century (see Rouwenhorst, 2005). While these distinct sources of return—index, factors, and alpha—are well documented, there is less academic research showing how to combine these distinct sources of return in an optimal portfolio.[2]

We present a new methodology for combining market cap index (or simply "index"), factors, and alpha-seeking strategies. The intuition behind our framework is as follows. If an investor is tracking a market cap benchmark with no active risk, or tracking error, budget, then the portfolio must be 100% index. As the investor is allowed to deviate from the market, then she can hold factor and alpha funds, with higher active risk budgets resulting in lower allocations to index funds. The higher the conviction an investor has on alpha-seeking funds, the higher the prior mean and tighter the prior standard deviation on the active funds' IRs, and more the active risk budget will be allocated to alpha vs. factor strategies. Thus, IRs can be set by investors', beliefs on manager skill, and the observed track records of funds. In one extreme case, an investor with a tracking error target but no conviction in alpha would allocate only to index and factor strategies.

---

[1] See Ang (2014) for a summary of the large empirical literature on the CAPM, factors, and performance measurement of alpha-seeking strategies.

[2] There are only a few studies optimizing index, factors, and alpha-seeking strategies in a formal framework. Homescu (2015) applies a regime switching framework, Carson, Shores, and Nefouse (2017) use all three sources of returns in target date funds, and Bellord et al. (2019) develop an expected short fall methodology with a tracking error limit. Like this paper, Aliaga-Diaz et al. (2020) treat index, factor, and active strategies as distinct sources of returns and propose the investor has three different risk aversion coefficients with respect to each—they acknowledge the difficulty of calibrating risk aversion coefficients, and we use only one as is commonly used in practice. They also do not consider parameter uncertainty. Corum, Malenko, and Malenko (2020) also consider allocations between index and active in a governance setting. None of these papers consider formulating expected returns based on incorporating priors on Sharpe ratios or information ratios, or using prior information in building optimal portfolios of index, factors, and alpha.

Factor strategies have rigorous economic rationales, there are long histories available, and they can be implemented in low-cost and transparent vehicles. Therefore, an investor's degree of confidence in factors is likely to be different than her conviction in alpha-seeking strategies, which are likely to have smaller samples for evaluation. Alpha should be delivering returns in excess of broad and persistently rewarded factor exposures—this requires specialized skills to take advantage of market inefficiencies or dislocations. Ang, Goetzmann and Schaefer (2011) conclude after summarizing a very large literature that these skills do exist, but are scarce. Access to alpha may also be restricted by manager selection capabilities.

We model this intuition with a Bayesian framework where the investor sets priors on Sharpe ratios or information ratios. In typical applications, the factors would have prior distributions on reward-to-risk ratios which would be informed by relatively long data samples. The prior standard deviation on the Sharpe ratios of factor strategies would be relatively tight. In contrast, the prior distribution on the information ratios (IRs) of alpha-seeking strategies would typically have more dispersion, but may have a higher prior mean than for factors. It is important that we model the IR of alpha funds *in excess* of the index and factor strategies. That is, in the procedure, we estimate exposures of the alpha funds to factors and the prior distribution is imposed on the idiosyncratic component of the alpha returns. Finally, we explicitly model management costs as the costs of factor strategies are, on average, significantly lower than alpha funds.

We employ a recent advance in Bayesian computation methods, the No-U-Turn (NUTS) Sampler implementation of Hoffman and Gelman (2014).[3] Standard Bayesian techniques also allow us to infer missing data points, so we can extract information from the longer histories of factor strategies vs. alpha funds, along with samples of unequal lengths of different fund managers. We set our priors directly on Sharpe ratios or IRs, which has the advantage that these are important statistics used in evaluating fund managers and in asset allocation. We derive the posterior parameters and the moments of the predictive distribution.

[3] This is an implementation of Hamilton Monte Carlo, as developed by Duane et al. (1987) and Neal (1994, 2011). These algorithms are much faster methods to construct the posterior distribution of parameters than traditional Markov Chain Monte Carlo (MCMC) methods like Metropolis et al. (1953) or Gibbs sampling (Geman and Geman, 1984).

The procedure has an important advantage in that it complements a traditional optimization-based investment process rather than replacing it. Quantitative-based investment approaches revolve around an optimization, which is typically based on a quadratic objective function. These nest traditional Markowitz (1952) mean-variance and models with ambiguity around certain parameters, as shown by Garlappi, Uppal, and Wang (2007) and Maccheroni, Marinacci, and Rustichini (2006). Those optimizations can be used as per usual—but we introduce a step before the optimization which produces the mean and covariance inputs in a way that trades off the priors, data lengths, and interaction terms between index, factors, and alpha.[4] Recently, Pedersen, Babu, and Levine (2020) show how traditional mean-variance can be adapted to incorporating shrinkage methods for the covariance. Our approach is different in that we place priors on information ratios on index, factors, and alpha to come up with outputs of means, covariances, and possibly whole predictive distributions—which can be used in any optimization.

We apply our methodology to equities, but the same methodology can be used to allocate to other index-factors-alpha in other asset classes and across a multi-asset portfolio. We select alpha-seeking funds from the Morningstar universe of active funds constituting Large, Mid-Cap and Small funds in Value, Growth, and Blend styles, and Technology funds. We use standard long-only factor investment strategies tracking minimum volatility, momentum, value, small size and quality factors, all which are easily accessible in low-cost and transparent ETFs. These factor strategies have longer sample lengths than the alpha funds. We set different prior beliefs on factors and alpha IRs, which we compare with the resulting posterior and predictive means and variances. In a second stage, we feed these updated means and variances into a regular mean-variance utility maximization to determine optimal index, factor, and alpha allocations.

[4] Our procedure also generates the whole predictive distribution, which could be used directly in general convex optimization problems (see, for example, Boyd et al. 2016) or potentially in non-convex utility functions like disappointment aversion or loss aversion (see, for example, Ang, Bekaert and Liu, 2005; Routledge and Zin, 2010).

Our approach is most related to Pastor and Stambaugh (1999, 2000), Tu and Zhou (2004), and Avramov (2004) who formulate asset allocation models in a Bayesian setting.[5] This literature does not directly specify priors directly in terms of Sharpe ratios or information ratios, although Pastor and Stambaugh's prior is proportional to residual variance. Consistent with most asset pricing models which have implications for reward-to-risk ratios (with a large literature starting from Hansen and Jagannathan, 1991), our theoretical framework sets the prior directly with Sharpe or information ratios. Empirically, our advance is to use much faster updating methods than the Monte Carlo Markov Chain (MCMC) methods used by the literature. Practically, we take care to differentiate between returns, which are stochastic, and costs, which are known with certainty, and derive moments of the predictive distribution for optimal allocation.[6]

1. Theoretical Setting

Section 1.1 highlights the different roles of index, factors, and alpha-seeking strategies. We show that investors allocate to alpha only if those strategies are generating returns in excess of index *and* factors. In Section 1.2, we take a special case of uncertainty on only one risky asset (alpha or factors) relative to the market cap benchmark and, show in closed form, the effects of prior uncertainty, sample length, and the data likelihood. The closed form expression allows us to explore the economic intuition. We leave a detailed description of the full model, which allows uncertainty on all parameters, to the Appendix.

[5] An earlier literature on using Bayesian techniques to examine the effects of uncertainty on asset pricing and asset allocation begins with Barry (1974), Brown (1979), and Bawa, Brown, and Klein (1979).

[6] Black and Litterman (1992) also derive posterior means and variances imposing prior beliefs around equilibrium views from the CAPM. However, they do not consider the predictive distribution. Note that the predictive distributions depends on the posterior mean and variance as well as sampling error. Black and Litterman only use posterior means and variances, without accounting for sampling error.

1.1 Alpha is Excess of Index *and* Factors

We work with three assets: index (r_m), factor (f), and pure alpha-seeking (r_i) strategies. (We also refer to the market index as a benchmark portfolio.) We assume that the strategies are stated in excess of the risk-free rate.

We assume that the factor, f, is uncorrelated with the market, r_m:

$$r_m = \mu_m + \varepsilon_m, \tag{1}$$

and

$$f = \mu_f + \varepsilon_f, \tag{2}$$

with orthogonal zero mean shocks ε_m and ε_f with variances σ_m^2 and σ_f^2, respectively.[7]

The alpha-seeking fund, r_i, loads on both the market and factor portfolios:[8]

$$r_i = \alpha_i + r_m + f + \varepsilon_i, \tag{3}$$

which can also be written as

$$r_i = \alpha_i + \mu_m + \mu_f + \varepsilon_m + \varepsilon_f + \varepsilon_i,$$

where the stock-specific shock, ε_i, has variance σ_i^2 and is uncorrelated with shocks ε_m and ε_f to the market and factor funds, respectively. This is consistent with a Ross (1976) factor model with the standard assumption that the residuals are uncorrelated. Importantly, the alpha fund has a premium, α, *in excess* of the market premium, μ_m, *and* the long-term factor return, μ_f. We stack the excess returns of the three funds in the vector $\mu = (\mu_m \; \mu_f \; \alpha_i)'$. We can consider equation (3) to represent a fund with a market beta of one, and the fund has unit factor exposure, f, as well as unique alpha insights in excess of the factor returns, α_i.

---

[7] The factor dynamics in equation (2) relative to the index portfolio are less restrictive than seems. Following Asness (2004), a long-only equity factor fund containing overweight positions in value stocks and short positions in growth stocks relative to a market-cap benchmark can be expressed as $r_m + (V - G) = r_m + (V - r_m) + (r_m - G)$, and if the value, V, and growth, G, legs are appropriately constructed, they can remove a significant part of market exposure. Grinold and Kahn (2000) refer to the long-short portfolio as a "characteristic portfolio."

[8] These results will be the same if equation (3) is changed to allow for factor loadings, $r_i = \alpha + \beta_i r_m + \gamma_i f + \varepsilon_i$. In particular, Treynor and Black (1973) and Roll (1977) show how the optimal weights in a tangency portfolio will adjust to take into account the factor loadings, but the maximum achievable Sharpe ratios by holding index, factors, and alpha funds will be unchanged.

The covariance of $(r_m \; f \; r_i)'$ from equations (1)-(3) is given by:

$$\Sigma = \begin{bmatrix} \sigma_m^2 & 0 & \sigma_m^2 \\ 0 & \sigma_f^2 & \sigma_f^2 \\ \sigma_m^2 & \sigma_f^2 & \sigma_m^2 + \sigma_f^2 + \sigma_i^2 \end{bmatrix}. \tag{4}$$

With standard mean-variance mathematics, the optimal holdings, $w = \left(w_m \; w_f \; w_i\right)'$ to maximize the portfolio's Sharpe ratio (SR) are proportional to $w \propto \Sigma^{-1}\mu$, which is given by:

$$w \propto \begin{bmatrix} {\mu_m}/{\sigma_m^2} - {\alpha_i}/{\sigma_i^2} \\ {\mu_f}/{\sigma_f^2} - {\alpha_i}/{\sigma_i^2} \\ {\alpha_i}/{\sigma_i^2} \end{bmatrix}. \tag{5}$$

From equation (5) we have:

Proposition 1: If there is no alpha present ($\alpha_i = 0$), then the holdings of the active manager is equal to zero. With conviction on alpha, $\alpha_i > 0$, investors seeking to maximize Sharpe ratios of their portfolios reallocate capital from index and factor funds towards alpha-seeking managers, with higher alpha beliefs resulting in higher capital allocations to active managers.

In Proposition 1, what is important is not the total return on the alpha-seeking fund: the active manager must be providing returns, α_i, *in excess* of market index and factor exposures (μ_m and μ_f, respectively in equation (5)). In this simple example, the weights are direct functions of appraisal ratios (excess returns divided by volatility) of index, factors, and alpha; the larger the risk-adjusted performance of each component, the larger that component's weight is in the portfolio.

The maximum squared Sharpe ratio that is obtainable is given by $S = \sqrt{\mu'\Sigma^{-1}\mu}$, and the squared Sharpe ratio simplifies to

$$S^2 = {\mu_m^2}/{\sigma_m^2} + {\mu_f^2}/{\sigma_f^2} + {\alpha_i^2}/{\sigma_i^2}. \tag{6}$$

Stated in words, the maximum squared Sharpe ratio depends on the individual assets' squared reward-to-risk ratios. This was originally derived by Treynor and Black (1973) for the market return and alpha measured relative to the single CAPM factor. Thus:

Proposition 2: The marginal contribution to the portfolios Sharpe ratio of the alpha-seeking manager to index and factor funds depends on the appraisal ratio of the active manager's return *in excess of index and factors* to the active manager's idiosyncratic volatility.

The important variable is not the return of the alpha-seeking fund, or the return of that fund in excess of the market. Put another way, the active fund can have a positive return due only to index or factors, or both, but unless the alpha-seeking fund beats *both* the index and factor strategies, it is irrelevant for the investor.

1.2 Incorporating Uncertainty

The purpose of this section is to gain some intuition about incorporating parameter uncertainty on the information ratios. For simplicity, we assume there is one risky asset, y, with Sharpe ratio (or information ratio) $S = \frac{\mu}{\sigma}$. We have data available in the time series $\{y_t\}$ that is net of fees. The cost of transacting the strategy is c.[9] We assume that the standard deviation, σ, is known.

The Appendix describes the full model with uncertainty on all parameters.

1.2.1 Prior and Posteriors of the Sharpe Ratio

We assume that the returns, y, are normally distributed so the likelihood function is:

[9] We net fees from the expected returns provided to the optimization because fees are paid with certainty, but risks and returns are subject to uncertainty. The cost, c, can also be interpreted as a holding cost, and more generally, as a utility certainty equivalent cost for information or access to the underlying investment strategy.

$$p(y|S,\sigma^2) \sim N(\sigma S - c, \sigma^2) \tag{7}$$

We impose the following prior on the Sharpe ratio:

$$p(S|\sigma^2) \sim N(S_0, \tau^2), \tag{8}$$

where S_0 is the prior mean and τ^2 is the prior variance. In typical applications, factors should have lower prior means than alpha-seeking strategies, and prior standard deviations for factors should be lower, on average, than alpha funds.

We can derive the posterior $p(S|\mathcal{Y},\sigma^2) \propto p(\mathcal{Y}|S,\sigma^2)p(S|\sigma^2)$, where the posterior distribution is given by:

$$p(S|\mathcal{Y},\sigma^2) \propto \exp\left(-\frac{1}{2\sigma^2}\sum\left(y_t - (\sigma S - c)\right)^2\right)\exp\left(-\frac{1}{2\tau^2}\sum(S - S_0)^2\right),$$

This can be solved in closed form, and the posterior distribution of *S* is given by:

$$p(S|\mathcal{Y},\sigma^2) \sim N(\mu_S, \sigma_S^2), \tag{9}$$

where the posterior mean, μ_S, and variance, σ_S^2, of the posterior distribution of S are:

$$\mu_S = \left(\frac{T}{T + \frac{1}{\tau^2}}\right)\frac{\bar{y} + c}{\sigma}(\bar{y} + c) + \left(\frac{\frac{1}{\tau^2}}{T + \frac{1}{\tau^2}}\right)S_0$$

$$\sigma_S^2 = \left(T + \frac{1}{\tau^2}\right)^{-1},$$

where $\bar{y}$ is the sample mean of *y*.

The intuition for equation (9) is as follows. Under normality, the sample distribution of the standard mean estimator is a normal distribution, $\hat{\mu} = \bar{y} \sim N(\mu_0, \sigma^2/T)$, where μ_0 is the population mean and σ^2/T is the sample variance of $\hat{\mu}$. Thus, by the delta-method and Slutsky's theorem, the distribution of the Sharpe ratio, $S = \frac{\mu}{\sigma}$, has a variance of $1/T$.[10] The inverse of this expression is the first term in the posterior variance

[10] The convergence rate of $1/\sqrt{T}$ leads to substantial impacts on uncertainty, as noted by Fama and French (2018) and others. This is another reason why the use of a prior is so important for practical application.

of σ_S^2. This is combined with the inverse of the variance of the prior of *S* in equation (9). This also occurs in a general Bayesian regression set up, where the variance is the inverse of the sum of the inverses of the sample estimator variance and the prior variance—except now, because we work in terms of the Sharpe ratio, the sample variance is 1/*T*.

In equation (9), the posterior mean, μ_S, takes the form of a weighted average of the sample estimate, $\sigma\bar{y} + c$, and the prior mean, S_0. That is, we can rewrite equation (9) for μ_S as:

$$\mu_S = \text{weight} \times (\bar{y} + c) + (1 - \text{weight}) \times S_0, \tag{10}$$

or

$$\mu_S = \text{weight} \times \text{data estimate} + (1 - \text{weight}) \times \text{prior}.$$

The weighted average form of data estimates and prior in equation (10) is standard in Bayesian estimators. In this case, because we have specified the distribution of the returns, *y*, to be net of fees, the posterior mean of *S* is expressed as a gross-of-fee return, $\sigma\bar{y} + c$. Note that as $\tau \to \infty$, we have decreasing confidence in the prior and all the weight is placed on the data estimate. In the limit $\tau \to \infty$, the posterior mean and variance of *S* converge to $\mu_S \to \sigma\bar{y} + c$ and $\sigma_S^2 \to \frac{1}{T}$, which are the sample mean estimator and variance, respectively, of the Sharpe ratio, *S*.

1.2.2 Utility Problem with Parameter Uncertainty

To apply the framework with Sharpe ratio or information ratio uncertainty to a portfolio allocation problem, we require the predictive distribution:

$$p(y|\,\mathcal{Y}) = \int_S \; p(y|\mathcal{Y}, S)p(S|\mathcal{Y})dS, \tag{11}$$

where $p(y|\mathcal{Y}, S)$ is the likelihood and $p(S|\mathcal{Y})$ is the posterior of *S*.

The Bayesian predictive distribution accounts for uncertainty about the unknown parameters; the posterior mean and variance of the unknown parameter, *S*, enter the predictive distribution. In particular, the greater the parameter uncertainty, the higher the variance of the posterior, σ_S^2, and hence the higher the variance of the predictive distribution, σ^{*2} because the parameter uncertainty is an additional source of risk; the predictive distribution accounts for both parameter uncertainty and sampling variation. Note that the popular Black and Litterman (1991) procedure stops after deriving the posterior distribution and does not compute the predictive distribution.

For mean-variance utility—the most popular objective function in practice—we need just the mean and variance of the predictive distribution, $p(y|\mathcal{Y})$.[11] In our simple case, this can be derived in closed form by noting that

$$y = \big(y - (\sigma S - c)\big) + (\sigma S - c),$$

where $y - (\sigma S - c) \sim N(0, \sigma^2)$ and $S \sim N(\mu_S, \sigma_S^2)$ which is derived in equation (9). Since we have conjugate normality, the predictive y follows a normal distribution,

$$y \sim N(\mu^*, \sigma^{*2}), \tag{12}$$

where

$$\mu^* = \sigma E(S|\,\mathcal{Y}) - c = \sigma\mu_S - c,$$

$$\sigma^{*2} = \sigma^2(1 + \sigma_S^2).$$

In equation (12), the predictive mean depends directly on the posterior mean of the Sharpe ratio, μ_S. Thus any assumption on the prior of the Sharpe ratio or an effect of updating through properties of the data likelihood will affect the attractiveness of the active strategy. The predictive variance depends on the data variance, σ^2, but it also depends on parameter uncertainty through σ_S^2. For factor strategies where the prior

[11] With a general utility function, U, we would find the portfolio weight w to maximize the utility function under the predictive distribution:

$$\max_w \int_y U(w, y) p(y|\mathcal{Y}) dy,$$

where $U(w, y)$ is our objective function which is a function of the portfolio weights, w, and the asset returns, y.

dispersion on the Sharpe ratio is relatively low, this decreases the predictive variance and makes those strategies relatively more attractive.

2. Data

Section 2.1 describes the market cap index benchmark and the factor portfolios. Section 2.2 describes the alpha-seeking funds. To illustrate the methodology, we work only with factors and alpha-seeking funds on US equities.

2.1 Index and Factors

We take the market cap index to be US large capitalization equity as measured by the S&P 500 index.

For factors, we use long-only factor indexes tracking minimum volatility (Ang et al., 2006), return momentum (Jegadeesh and Titman, 1993), value (Basu, 1977), small size (Banz, 1981) and quality (Sloan, 1996) listed in the following table:

Factor	**Index**
Minimum Volatility	MSCI USA Minimum Volatility Index
Momentum	MSCI USA Momentum Index
Value	MSCI USA Enhanced Value Index
Small Size	S&P Small Cap 600 Index
Quality	MSCI USA Sector Neutral Quality Index

Our data sample for the factor index returns is from December 2010 to December 2020 at the monthly frequency. Like the model setup in Section 1 (see, for example, equations (1) and (2)), we model these factors as orthogonal to the market return. This is done running a regression of the factor index returns onto the US large cap equity index over the full sample. The residuals are then used to represent the factors.

2.2 Alpha-Seeking Funds

We construct a list of potential alpha-seeking funds by first taking funds from December 2010 to December 2020 at the monthly frequency from the Morningstar database for US listed equity mutual funds. We take funds in the US equity style box categories with capitalizations of Large, Mid, and Small, and styles of Value, Growth, Blend, with Technology sector funds. We report summary statistics of the funds in Exhibit 1. Most of the funds fall into US Large Growth and Blend categories, consistent where the majority of AUM falls for active funds (see, for example, recent statistics in Madhavan, Sobczyk, and Ang, 2020). We compute monthly frequency active returns as the gross fund return minus the stated primary benchmark index as specified by Morningstar. Fees, used in net return calculations, are as of December 31, 2020. Exhibit 1 reports the average monthly active return of these funds is positive, at 0.59% per month, with an average reward-to-risk ratio of 0.15. The reward-to-risk ratios range from -0.05 for Mid-Cap Blend funds to 0.42 for Technology managers.

Over the 120 months of active returns ending December 2020, there is a positive relation between average active returns and active risk: in a regression of active returns on active risk, we estimate a coefficient of 0.34. While the positive sign is consistent with Grinold (1994), the R^2 is only 0.14, indicating that there are many active managers which may generate high returns with low active risk.

We construct excess returns of the alpha-seeking funds relative to commonly used systematic factors in the academic literature to facilitate forming our priors for the alpha-seeking funds. We take the Fama and French (1993) factors of the market factor (MKT) and the one-month T-bill (RF), size (SMB), and value (HML). We also use Kenneth French's construction of the Jegadeesh and Titman (1993) momentum factor (UMD). We augment these factors with quality minus junk factor constructed by Asness, Frazzini, and Pedersen (2019) (QMJ) and the betting-against-beta factor (BAB) of Frazzini and Pedersen (2014). While these factors have long histories and are backed by published academic studies, they are not investible. Hence, we use these to inform the priors on the factors, but when we allocate, we hold the investible factor indexes listed in Section 2.1. We take advantage of the fact that we can use a longer time series, specifically of 25 years ending December 2020 at the monthly frequency to set the priors. The methodology, however, accommodates any well-defined priors, and any method of setting the priors. (See the Appendix for further details.)

3. Empirical Results

Section 3.1 estimates factor loadings of the active funds on the factors. In Section 3.2, we describe the resulting posterior and predictive distributions. These are used as inputs into a portfolio allocation problem and we describe results in Section 3.3.

3.1 Factor Loadings

Exhibit 2 reports the results of regression active fund returns onto the factors over the last ten years ending in December 2020. These regressions use returns of the fund in excess of the Large Cap equity benchmark on the left-hand side (LHS) and the orthogonalized factors on the right-hand side (RHS). In our benchmark results, we assume that these factor loadings are held fixed in order to isolate the effects of the IR assumptions on the asset allocation.[12]

Because both the LHS and RHS are in excess returns, the coefficients in Exhibit 2 are generally small in absolute value. The numbers in parentheses are averages of the t-statistics across the fund-by-fund regressions. There are no significant factor loadings for Large Cap Blend funds at the 5% significance level, but US Large Growth and Value funds do exhibit multiple significant factor exposures: to all factors except value for Large Growth funds, and especially to the value factor for Large Value funds. Perhaps surprisingly, Small Growth funds exhibit anti-momentum exposures for this set of funds; generally, growth funds load

[12] We find few differences when we allow parameter uncertainty in the factor loadings compared to the baseline results, which is shown in the Appendix. One of the reasons is that second moments tend to be estimated more reliably than means (see Merton, 1980).

significantly on the momentum factor in a wider population (see, for example, Ang, Madhavan, and Sobczyk, 2017). The Small Growth space is unusual in this respect; looking at the bottom row of Exhibit 2 across all funds, our mutual fund universe tends to have positive exposures to momentum and value, and also tends to hold larger and higher quality stocks.

Exhibit 3 reports data on the factors (see Section 2.1) over the last 25 years of data ending at December 2020 which we use to set the priors on factors. The prior mean IRs are set as the full data averages. When we bootstrap, we draw samples with replacement from the full sample, and we use the standard deviation of the bootstrapped samples as the standard deviations of the IRs. The highest IR is for minimum volatility, with an IR of 0.24 and the lowest is for size at 0.05. The standard deviations are relatively tight from 0.03 to 0.05.

3.2 Posterior and Predictive Results

After setting the priors, we run the procedure with No-U-Turn (NUTS) sampling, to generate posterior and predictive information ratios, alphas, and the covariance matrix. We use Automatic Differentiation Variational Inference (ADVI) to initialize the sampler, setting the target acceptance ratio to 80%. We use 25 chains with 5,000 draws each to obtain 125,000 draws in total. For posterior predictive analysis, we generate 50,000 simulated returns.[13]

To visualize the broad impact of the Bayesian procedure, we examine a scatter plot comparing the OLS-based data IR and the Bayesian IR (posterior mean) in Exhibit 4. Across all Morningstar categories, the Bayesian procedure attenuates the IR in data—as expected, high IRs in data may be due to noise and the

---

[13] If the Bayesian procedure is too computationally expensive for a large selection of funds, as an alternative, one can use ADVI with 10,000 steps to perform variational inference on the model as the ADVI's computation time is faster than Monte Carlo methods like NUTS. Note that variational inference methods like ADVI only provide an approximation to the posterior distribution. One can then compute the portfolio optimization procedure to select a short list of funds that are on the efficient frontier. Using this short list, researchers can perform the full NUTS sampler to confirm the approximated results from ADVI.

procedure shrinks these towards zero (see equations (9) and (10)). In Exhibit 4, we plot the linear fit which has a slope of 0.23. This is significantly less than one, which we can visually see on Exhibit 4 where we plot a 45 degree line the dashed diagonal line. This implies that the procedure shrinks IRs extreme IRs towards zero.

Shrinkage and Out-of-Sample Forecasts

The shrinkage does improve out-of-sample forecasts. Exhibit 5 reports the result where we split the data into two halves, from 2010 to 2015 and from 2016 to 2020. In the first half, we perform our procedure producing a posterior mean of IRs for each fund. Next, we sort the funds into three bins (high-middle-low) according to their posterior IRs. We examine the realized IRs of each fund in each bin over the second sample. If there is no predictive power, we should have 1/3 transition probabilities in each row. In the top panel, the Bayesian procedure shows large entries around or exceeding 0.5 in the diagonals. For the top performers in the first row—which are arguably the most important for a fund allocator, over 91% of the funds in the top tercile end in the top or middle terciles in the out-of-sample. For comparison, we also compute the Markov transition matrix across the two samples using IRs predicted using OLS regressions. In this case, there is little predictive power with the entries being close to a random sample of 1/3. Clearly, the Bayesian procedure helps in prediction.

Characterizing Posterior Distributions

Exhibit 6 reports the distribution of the posterior IRs and compares them with point estimates of IRs from OLS regressions (empirical IRs). We report the mean, standard deviation, and 3 and 97 percentiles of the posterior IR distribution for the factors and several selected alpha-seeking funds. Because we impose priors for index and factor funds, their posterior IRs reported in Exhibit 6 are expected to differ from their empirical counterparts due to shrinkage. For example, over the last ten years, value and size factors have underperformed with IRs of -0.33 and -0.09, respectively, but the posterior means are positive reflecting the prior assumption and that at the beginning of the 10-year period, value and size do exhibit outperformance. Likewise, the posterior IRs of the active funds are attenuated. For example, the American Century Equity Income I fund has an empirical IR of 0.30 over the past ten years. The posterior IR (3, 97)-percentile bounds are positive, but the mean posterior IR is shrunk to 0.22. In general, we observe that the standard deviations of the posterior IRs are smaller than the standard deviations set in the prior due to the contribution of the observable data and that prior standard deviations deliberately err on the non-informative side.

Posteriors as a Function of Priors

In Exhibit 7, we illustrate how the posterior moments of three funds change as we change the prior mean on the IR. In each row, we change the prior mean of the IR from zero (top row), 0.35 (middle row), and 0.69 (bottom row). Each column corresponds to a different fund. The blue vertical line is the mean of the fund alphas in data, which is graphed at the same position in all subplots for each fund in the columns. The green and the red densities correspond to the prior IR and the posterior IR, respectively. We hold the prior variance constant at 0.22 in this exercise to isolate the effect on the prior mean.

In the top row of Exhibit 7, the prior IR is zero and the posterior IR has a mean close to the empirical IR. For example, for fund Edgewood Growth Inst (Ticker: EGFIX), the data IR is 0.26 and the posterior IR has a mean of 0.25 in the first row. In the second row, we set the prior IR to have a mean of 0.35. In each case, as expected, the posterior mean increases with the prior mean, but are still fairly close to the data average. For example, for fund Edgewood Growth Inst in the second row with the prior IR of 0.35, the posterior IR is now 0.48 which is still close to the data IR of 0.26, but higher than the posterior IR with mean 0.25 in the first row. The posterior distributions also narrow compared to the first row. In the last row when the prior IR is highest, there is little effect of the data because the priors are much larger than the empirical distribution and the posterior IRs are close to the prior IRs. Thus, only for priors widely outside empirical experience does the data have little effect.

Predictive Distributions

The predictive distributions also reflect the effect of informative priors on the IR mean, which we report in Exhibit 8. The means of the style factors (momentum, quality, size, minimum volatility, and value) are stated in excess of the S&P 500. The last decade ended December 2020 saw negative performance of size and value in excess of the market, while the predictive means are positive at 1.4% and 0.3%, respectively, reflecting the effect of the priors. Going the other way, minimum volatility had a strong positive return in the data, and the Bayesian procedure shrinks that mean down to 5.1%. The alphas of the funds are stated in excess of the S&P 500 and the style factors. All fund alphas are generally attenuated by the priors of the IRs, but the effect can be small. There is little difference for the data and predictive standard deviations, which in many instances are the same to the third decimal place.

3.3 Portfolio Construction

Taking the predictive moments in Exhibit 8, we perform a mean-variance optimization exercise. In this section, we construct portfolios to maximize portfolio active net of fee return subject to a long only constraint and limit the number of holdings ten or fewer with a standard mean-variance objective function. In our benchmark case, we take active risk relative to the Large Cap benchmark of the S&P 500 Index, and we compute net-of-fee returns assuming fees as of December 2020. We use a risk aversion coefficient of 32, which can be calibrated to set a particular level of active risk.

Exhibit 9 reports the equity portfolio holdings for the baseline case. The three largest holdings are: the minimum volatility factor, American Century Disciplined Core Value, and Fidelity Advisor Growth Opportunities with weights of 22%, 12%, and 11%, respectively. Using the predictive moments from our procedure, we estimate this portfolio yields an excess return of 3.8% and with an active risk of 2.0%, both which represent an excess return to active risk ratio of 1.9 above the Large Cap benchmark. In this exercise, there is only one factor, momentum, with no direct allocation to the market factor. Minimum volatility is being used as a portfolio ballast: as active risk increases (not reported), the optimizer uses the higher risk budget to hold more aggressive positions in the alpha-seeking funds. These high-risk positions are effectively funded by larger positions in lower risk minimum volatility.

4. Conclusion

We show the separate roles played by index, factors, and active funds—and importantly show how investors can incorporate prior information on the three return sources of return that can be practically implemented in portfolio construction. Investors allocate to active funds only if they have excess returns higher than both index and factors. Factors are likely to have longer data samples and their premiums are based on economic rationales, whereas historical data for active funds may be shorter and their sources of returns may be more transitory. In addition, priors on information ratios (IRs) may have higher means for true alpha-seeking funds and while the mean prior IR for factors may be lower, the prior standard deviation on the factor IRs may be tighter than the prior standard deviation for active funds. Using the newly developed No-U-Turn (NUTS) Sampler, we construct posterior IRs and use the predictive moments in a portfolio construction exercise allocating to index, factors, and active funds.

There are many possible extensions from our work. We only incorporate beliefs and information on performance, and specify priors on IRs. Investors may also have preferences towards vehicles—like ETFs vs. traditional mutual funds, should take into account taxes, lean towards more transparent strategies, or may prefer delegated or direct portfolio management. Index, factors, and alpha funds each have different advantages and disadvantages along these dimensions beyond beliefs and historical data on IRs that can also be incorporated into the allocation problem.

Appendix A: Full Model

We describe the full model where all parameters carry uncertainty, the priors are specified in terms of Sharpe ratios or information ratios for all index, factors, and alpha assets, and we generate the full posterior distribution and moments of the predictive distribution. While the intuition is similar to the analytical example of the previous section, we must use numerical techniques in the full model.

We make one change in notation from Section 1 and use the vector f to incorporate both market index returns and factors. Thus, f is understood to be $[r_m, f]$ in the notation of Section 1. The returns, r, section are gross-of-fees and in excess of the fund's stated performance benchmark, and we explicitly handle management fees in the posterior predictive distribution.

A.1 Likelihood

We specify the j-th factor return, $j = 1..N_f$, to follow:

$$f_j = \mu_{f,j} + \varepsilon_{f,j}, \qquad \varepsilon_{f,j} \sim N(0, \sigma_{f,j}^2)$$

We stack the factor returns into a $N_f \times 1$ vector $f \equiv [f_1, \ldots, f_{N_f}]'$ and write:

$$f = \mu_f + \varepsilon_f,\ \ \varepsilon_f \sim N(0_{N_f \times 1}, \Sigma_f)\,. \tag{A.1}$$

The factor covariance matrix Σ_f can be dense, but assume it is diagonal for simplicity. The roots of diagonal elements of Σ_f are volatilities of each factor strategy which we denote as σ_f:

$$\sigma_f \equiv \sqrt{diag(\Sigma_f)} = \left[\sigma_{f,1}, \ldots, \sigma_{f,N_f}\right]' \quad ,$$

and thus the Sharpe ratios S_f are defined as the elementwise division (denoted by ./) of mean returns μ_f and standard deviations of factor returns σ_f:

$$S_f = \mu_f \,./\sigma_f \tag{A.2}$$

The i-th alpha strategy returns, $i = 1..N_r$, are modeled as:

$$r_i = \alpha_i + \beta_i f + \varepsilon_i, \qquad \varepsilon_i \sim N(0, \sigma_i^2)$$

where the factor loadings β_i is a $1 \times N_f$ row vector. The alpha for strategy i is the scalar α_i. Since f includes both index and factors, a fund with $\alpha_i > 0$ indicates the fund has a return in excess of the market and factors. The standard deviation of idiosyncratic return for this strategy is σ_i.We stack the scalar returns into a $N_r \times 1$ vector of alpha strategy returns $r = \left[r_1\ , \dots , r_{N_r}\right]'$ so can write:

$$r\ = \alpha + Bf + \varepsilon,\ \varepsilon \sim N(0_{N_r \times 1}, \Sigma_r), \tag{A.3}$$

where $B = \left[\beta_1\ , \dots , \beta_{N_r}\right]'$ is the $N_r \times N_f$ matrix of factor loadings, α is the $N_r \times 1$ vector of alphas, and Σ_r is the $N_r \times N_r$ covariance matrix. The factor covariance matrix Σ_r is assumed to be dense, which can allow the to be correlated, though in our setting the data generating process for the alphas is assumed produce alphas that on average are uncorrelated. Like the factor strategies, we define the Sharpe ratio or information ratio vector to be

$$S_r = \alpha\ ./\ \sigma,$$

where $\sigma = \left[\sigma_1\ , \dots , \sigma_{N_r}\right]'$ and ./ denotes element-by-element division.

We place priors on Sharpe ratios or information ratios, so we rewrite the original linear system of observations in equations (A.1) and (A.3) as:

$$\begin{bmatrix} f \\ r \end{bmatrix} = \begin{bmatrix} \mu_f \\ \alpha \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ B & 0 \end{bmatrix} \begin{bmatrix} f \\ r \end{bmatrix} + \begin{bmatrix} \varepsilon_f \\ \varepsilon_r \end{bmatrix}$$

in terms of S_f and S_r:

$$\begin{bmatrix} f \\ r \end{bmatrix} = \begin{bmatrix} I & 0 \\ B & I \end{bmatrix} \begin{bmatrix} \sigma_f . S_f \\ \sigma_r . S_r \end{bmatrix} + \begin{bmatrix} I & 0 \\ B & I \end{bmatrix} \begin{bmatrix} \varepsilon_f \\ \varepsilon_r \end{bmatrix}, \tag{A.4}$$

where . denotes element-by-element multiplication and the error terms between f and r are uncorrelated:

$$\begin{bmatrix} \varepsilon_f \\ \varepsilon_r \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_f & 0 \\ 0 & \Sigma_r \end{bmatrix} \right).$$

Note that because alpha-seeking funds load on the factors, there may be significant unconditional correlations between shocks to factors and alpha-seeking returns.

In our formulation in equation (A.4), the covariance matrices enter the mean. This echoes the formulation of the original CAPM, where the covariance (beta) affects a stock's expected return. Whereas working with Sharpe ratios or information ratios is economically intuitive, the presence of variances in the mean prevent us from using standard Gibbs sampling techniques, as done by authors like Pastor and Stambaugh (1999, 2000), Avramov (2004), and others who do not link the mean and Sharpe ratios (or information ratios).[14] Nevertheless, it is clear that given Sharpe ratios $S = [S_f, S_r]$, covariance matrices $[\Sigma_f, \Sigma_r]$, and factor loadings B, the likelihood function is completely pinned down.

A.2 Priors on Reward-to-Risk Ratios

We assume that the prior for each Sharpe ratio or information ratio, $S = [S_f, S_r]$, is normal:

$$S_{r,i} \sim N\left(\mu_{S,r}, \sigma_{S,r}^2\right),\ i = 1, \dots, N_r, \qquad \text{(A.5)}$$

$$S_{f,j} \sim N\left(\mu_{f,r}, \sigma_{S,f}^2\right),\ j = 1, \dots, N_f$$

We specify priors for the covariance matrices with the LKJ(η) – half-Cauchy(γ) distribution. We write the covariance matrix as:

$$\Sigma = \begin{bmatrix} \Sigma_f & 0 \\ 0 & \Sigma_r \end{bmatrix} = \tau \Omega \tau\ , \qquad \text{(A.6)}$$

where τ is a diagonal matrix of volatility scales of the diagonal elements of Σ, $\tau_j = \sqrt{\Sigma_{jj}}$, and Ω is a correlation matrix, $\Omega_{i,j} = \Sigma_{i,j} / (\tau_i \tau_j)$.

The volatility priors follow a half-Cauchy distribution:

$$\tau_{\mathrm{i}} \sim \mathrm{Cauchy}(\gamma), \text{ with } \tau_i > 0. \qquad \text{(A.7)}$$

[14] The popular Normal-Inverse Wishart distribution on Sharpe ratios and covariance matrices does yield a posterior distribution of Sharpe ratios that belongs to the exponential family, but there is no known way to sample from it. Metropolis-Gibbs sampling schemes may also take a long time to converge. The convergence of the NUTS sampler is very fast, described below, and in addition also handles incomplete data, and data of funds with different sample lengths.

As γ increases, the mean of the prior volatility increases. The main benefit of using the half-Cauchy is that it performs well with small numbers close to zero and yet is quite fat-tailed (see Gelman, 2006; Polson and Scott, 2012).

The LKJ distribution is specified on the correlation matrix:

$$\Omega \sim \mathrm{LKJ}(\eta), \tag{A.8}$$

with the LKJ defined as

$$\mathrm{LKJ}(\Sigma|\eta) \propto \det(\Sigma)^{\eta-1}.$$

Note that with $\eta = 1$, we have a uniform distribution. For $\eta > 1$, the prior favors correlations closer to zero (or the correlation matrix is close to unity) and for $\eta < 1$, the mass of the prior shifts towards correlations closer to ± 1. Put another way, the LKJ distribution is uniform over the space of all positive definite correlation matrices and η controls how close the samples are to the identity matrix (see Joe, 2006; Lewandowski, Kurowicka, and Joe, 2009. As is well known in portfolio optimization, extreme correlations produce large swings in portfolio weights (see, for example, Best and Grauer, 1991), and the LKJ prior is attractive because it enables us to draw positive definite correlation matrices with high probability and directly controls the amount of shrinkage in the covariance estimation.

The prior for factor loadings B is also chosen to be i.i.d normal elementwise:

$$\beta_{i,j} \sim N\left(\mu_{\beta(i,j)}, \sigma^2_{\beta(i,j)}\right). \tag{A.9}$$

A.3 Generating the Posterior

Let $\theta \equiv [S, B, \Sigma]$ denote all parameters and $\mathcal{Y}$ denote all observed data, then the posterior is given by:

$$p(\theta|\mathcal{Y}) = \frac{p(\mathcal{Y}|\theta)\pi(\theta)}{\int p(\mathcal{Y}|\theta)\pi(\theta)d\theta}$$

where $\pi(\theta)$ is the prior distribution of parameters specified above, or

$$p(\theta|\mathcal{Y}) \propto p(\mathcal{Y}|\theta)\pi(\theta). \tag{A.10}$$

We use Hoffman and Gelman's (2014) No-U-Turn Sampler (NUTS) to generate parameter draws from this posterior. NUTS is an implementation of Hamiltonian Monte Carlo (HMC) method (see Neal, 2011). Compared to the classical Metropolis-Hastings Random Walk MCMC, NUTS-HMC draws parameters with larger successive distances. Thus, NUTS-HMC requires significantly fewer iterations to reach convergence. There is additional time saved to compute the posterior as NUTS usually accepts draws with more than 80% probability while MCMC typically has acceptance probabilities less than 20%. Lastly, NUTS and HMC are less likely to get stuck by local minimum.

The high efficiency of NUTS means we can deal with high dimensional problems in terms of handling large numbers of funds and factors. This efficiency also allows us to incorporate missing data imputation into the sampling procedure using standard data augmentation techniques (see, for example, Tanner and Wong, 1987; Kong, Liu and Wong, 1994). Under the hood, the missing data follows the same likelihood function as observed data, and thus each parameter draw is accompanied by a draw of missing data from the likelihood function conditional on those drawn parameters.

A.4 Predictive Moments

Let y^{pred} be the predicted value for observed data, then the posterior predictive distribution is given the predictive distribution after integrating out parameter uncertainty:

$$p(y^{pred}|\mathcal{Y}) = \int p(y^{pred}, \theta|\mathcal{Y})d\theta = \int p(y^{pred}|\theta)p(\theta|\mathcal{Y})d\theta. \tag{A.11}$$

The same principle to derive equation (11) with uncertainty only on one Sharpe ratio can be used to update the moments of the predictive distribution with uncertainty on all parameters using iterative expectations. Following equation (11), the mean and variance of the predictive distribution are:

$$\mu^* = E(y|\,\mathcal{Y}) = E[E(y|S,\mathcal{Y})|\mathcal{Y}],$$

$$\sigma^{*2} = var(y|\,\mathcal{Y}) = E[var(y|S,\mathcal{Y})|\mathcal{Y}] + var[E(y|S,\mathcal{Y})|\mathcal{Y}].$$

In more general cases, the predictive distribution is not normally distributed.

The posterior predictive for $y \equiv [f, r]$ conditional on observed data proceeds per usual Bayesian procedures. Computationally, it is minimally costly to simulate the observed variables (index, factor returns, and fund returns) given each parameter draw by the NUTS sampler.[15] For mean-variance utility, we compute the predictive mean and variance from the sample counterparts from the generated predicted values. Since we can generate as many predicted values as we want, we can limit sampling error (versus the true posterior predictive mean and variance) to arbitrary degrees. We sample the net-of-fees posterior predictive return distribution by simply subtracting the fees (c_f, c_r) from the posterior predictive distribution:

$$p\left(\begin{bmatrix} f - c_f \\ r - c_r \end{bmatrix} | \mathcal{Y}\right) = p\left(\begin{bmatrix} f \\ r \end{bmatrix} | \mathcal{Y}\right) - \begin{bmatrix} c_f \\ c_r \end{bmatrix}, \tag{A.12}$$

where c_f are fees on the factor funds and c_r are fees on the alpha-seeking funds, respectively.

A.5 Specifying Priors

The ability to set priors is both an advantage and a disadvantage of Bayesian methods; the methodology incorporates any well-defined prior. Our priors for our empirical results are set as follows.

For the factors (see Section 2.1), we bootstrap the last 25 years of data ending at December 2020. The prior mean IRs are set as the full data averages. In the bootstrap, we draw samples with replacement from the full sample, and we use the standard deviation of the bootstrapped samples as the standard deviations of the IRs.

For empirical Bayes priors for the active fund returns, we use the last ten years of data ending at December 2020. We regress fund excess returns onto the long-short academic factors listed in Section 2.2. From the OLS regressions, we use the intercept and residual risk as our "empirical Bayes" prior means. We define residual risk as the standard deviation of residuals from the OLS regressions. This yields a Normal

[15] We can sample any deterministic function of observed variables easily for each parameter draw, including Value at Risk, optimal asset allocation statistics, and other stress testing statistics.

prior with 0.36 mean and 0.17 standard deviation (both annualized) as our prior baseline. For the volatilities, we set the Half-Cauchy γ prior to 1.0 to match the median of the empirical information ratios from OLS.

When we allow the idiosyncratic fund returns to be correlated, the η parameter for LKJ prior is set to 3 to match the interquartile range of the OLS-based return residual correlations. We assume normal information ratio, or in the case of the market factor, Sharpe ratio, priors based on a long run empirical examination of the factor's performances described above.

Appendix B: Convergence Diagnostics

For our main empirical results, we perform the following diagnostics to test for convergence in mean, variance, and autocorrelation of the sampling:

Geweke (1992) Test: This test examines whether early sections in the chain has the same mean as the later sections by computing the following z-score-like statistics:

$$\frac{E\left(x_{early} - x_{last}\right)}{\sqrt{V\left(x_{early}\right) + V\left(x_{last}\right)}},$$

Where x_{early} is a section in the earliest 10% of chain, and x_{last} is a section in the last 50% of the chain. We partition the first 10% and the last 50% into 20 segments and compute 20 Geweke test scores to see if they oscillate between -1 and 1, which indicates good convergence.

Rhat Test: The classic Gelman and Rubin (1992) Rhat test compares multiple independent chains. The idea is that when all chains have converged, the within-chain variances and between-chain variance should be identical. The diagnostic is computed as:

$$\hat{R} = \frac{\hat{V}}{W},$$

where W is the within-chain variance and $\hat{V}$ is the posterior variance estimate for the pooled (from multiple chains) chain. This score should be close to unity if all chains have converged. A recent improved version of the test is proposed by Vehtari et al. (2019). The rule of thumb is that [0.95, 1.05] would be the range indictive of convergence.

Effective Samples: This diagnostic from Gelman et al. (2013) estimates the number of effective parameters after accounting for autocorrelation induced by the sampling method. It is computed as:

$$\hat{n}_{eff} = \frac{nm}{1 + 2\sum_{p=1}^{P} \hat{\rho}_p},$$

where m is the number of chains, n the number of samples per chain, and $\hat{\rho_t}$ the estimated autocorrelation at lag p, and P is the first odd positive integer such that the sum of $\hat{\rho}_P$ and $\hat{\rho}_{P+1}$ is negative.

Longer and More Chains: In this diagnostic, we simply extend both the length and number of the simulated chains to examine whether our results of interest significantly change.

Results

For total sample sizes of half a million, we find that most variables have more than 90,000 effective samples. All Rhat statistics are extremely close to one. For IR and Sharpe ratios, the Geweke statistics sit comfortably within the [-1, 1] range which is indicative of convergence. To further enhance our confidence, we increase the total number of sequence to two million with 25 chains and 80,000 samples each. These results barely change from the samples of half a million.

Appendix C: Simulation Exercise

To demonstrate the appropriateness of the estimation technique, we run the following simulation exercise:

1. Simulate data from the general version of the model
2. Estimate the model with true factor loadings and covariances known
3. Vary the sample size to examine the finite-sample properties: how many observations are needed to accurately recover the true parameters?
4. Vary prior means for Sharpe ratios: how much does specifying "wrong" priors matter for recovering the true Sharpe ratios?
5. Vary prior means for Sharpe ratios: how sensitive is the posterior predictive distribution of alpha to a "wrong" prior?

We run all simulations using 10 Markov chains with 5,000 samples each and burn the first 500 samples.

C.1 Data Generating Process

We generate factors f and fund return r_i according to the following model:

$$f_t = \mu_f + \epsilon_{f,t}, \qquad \epsilon_{f,t} \sim N\left(0, \Sigma_f\right)$$

$$r_{i,t} = \alpha_i + \beta_i f_t + \epsilon_{i,t}, \qquad \epsilon_{i,t} \sim N(0, \sigma_i^2)$$

The priors of information ratios are given by:

$$S_f \equiv \frac{\mu_f}{\sigma_f} \sim N(\mu_{S_f}, \sigma_{S_f}^2)$$

$$S_i \equiv \frac{\alpha_i}{\sigma_i} \sim N(\mu_{S_i}, \sigma_{S_i}^2)$$

We assume the factors as a group are orthogonal to the fund returns. We assume the joint distribution of returns conditioning on loading matrix B is:

$$\begin{pmatrix} I & 0 \\ -B & I \end{pmatrix} \begin{pmatrix} f_t \\ r_t \end{pmatrix} = \begin{pmatrix} \mu_f \\ \alpha \end{pmatrix} + \begin{pmatrix} \epsilon_{f,t} \\ \epsilon_{i,t} \end{pmatrix},$$

where

$$\begin{pmatrix}\epsilon_{f,t}\\ \epsilon_{i,t}\end{pmatrix} \sim N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\Sigma_f & 0\\ 0 & \Sigma_r\end{pmatrix}\right)$$

and

$$\mu_f = \sqrt{diag}(\Sigma_f)S_f$$

$$\alpha = \sqrt{diag}(\Sigma_r)S_r$$

C.2 Simulation Settings

Let the dimension of funds be four and factors three. The true information/Sharpe ratios are given by:

Factor1	Factor 2	Factor 3	Fund 1	Fund 2	Fund 3	Fund 4
0.7357	-0.0954	1.2163	0.3436	0.1397	0.9436	0.9298

The true Σ (only upper triangle is shown) is given by:

	Factor 1	Factor 2	Factor 3	Fund 1	Fund 2	Fund 3	Fund 4
Factor 1	0.1322	0.7815	-0.3712				
Factor 2		8.5905	0.8693				
Factor 3			2.4907				
Fund 1				0.4938	0.2801	0.8145	0.2870
Fund 2					0.1885	0.5078	0.3399
Fund 3						2.1158	0.9988
Fund 4							1.7645

The factor loadings are given by:

	Factor 1	Factor 2	Factor 3
Fund 1	0.1048	0.1046	0.0864
Fund 2	-0.0120	0.0125	-0.0322
Fund 3	0.0842	0.2391	0.0076
Fund 4	-0.0566	0.0036	-0.2075

We set the prior for all Sharpe/information ratios to be $N(0.4,10^2)$.

C.3 Robustness Checks

Simulating from the parameters in Section B.2, we perform the following exercises:

Effect of Sample Size: The first exercise we do is to examine the effect of sample sizes. We use the following grid of sample sizes: [50, 100, 250, 300, 500, 1000, 5000, 10000], which we can interpret as months. We find that the error—the distance from the true Sharpe ratios and estimated—stabilizes at close to zero after 200 samples. This is also true for other similar statistics, including true vs estimated means of the factors and funds.

Sensitivity of the Prior: For these exercises, we set the sample size to be 200 and examine the sensitivity of the estimation results to the priors. We use the following grid for prior means on the IRs: [-20, -2, -1, 0, 0.1, 0.5, 1, 2, 5, 10, 50]. These values range from extremely low (-20) to extremely high (50). We find that there is little sensitivity to the prior—even very large priors do not affect, for these parameter values and for the sample size of 200, the posterior means. The estimated posterior mean of information ratios are generally within 0.05 in absolute value, but only for the most extreme priors extend to 0.10 in absolute value. The same is true for posterior mean asset return (alphas). Exhibit A-1 compares the estimated vs. true alphas for the simulated funds for various prior means on the IRs showing that in all cases, the estimated means lie close to the true means even for very dissimilar prior means.

Convergence Diagnostics: We compute the convergence diagnostics in Appendix A for all simulation exercises. All exercises converge without issues, including for simulations with sample sizes less than 250 and when the priors for the IRs are Normal(0.4, 10^2). The effective sample size ranges from 32% to 62% of the total samples. All Rhat statistics are very close to 1.0 (up to eight decimal places). Graphs of the Geweke test statistics for 20 intervals (each with size 1% of the simulated sequences) against the last 50% of the sequences show fluctuatations in a much tighter bound than [-1,1] which indicate good convergence.

Appendix D: Uncertainty in Factor Loadings

In Exhibit A-2, we show the effect of allowing uncertainty in the factor loadings with different priors. This estimation takes the same priors of the IRs used in the main text and Exhibits 4 to 8, except we also extend parameter uncertainty to the factor loadings. We graph three active funds in each row, which are Edgewood Growth Institutional, JP Morgan US Value I, and Putnam Multi-Cap Core Y. Each column represents a factor, with the first column representing the market S&P 500 benchmark. Across the board we observe no significant deviation in terms of point estimates from the factor loadings estimated with OLS. This is expected as we use relatively uninformative priors for factor loadings and the fact that second moments are estimated more precisely than first moments (see Merton, 1980). Thus, there is little effect of allowing for uncertainty in factor loadings vs. holding the factor loadings constant. We also find that changing the priors of IRs does not affect the estimates of factor loadings in Exhibit A-2.

References

Aliaga-Diaz, R., G. Renzi-Ricci, A. Daga, and H. Ahluwalia, 2020. Portfolio Optimization with Active, Passive, and Factors: Removing the Ad Hoc Step, Journal of Portfolio Management 47, 40-51.

Ang, A. 2014. *Asset Management: A Systematic Approach to Factor Based Investing*. New York, NY: Oxford University Press.

Ang, A., G. Bekaert, and J. Liu, 2005. Why Stocks May Disappoint. Journal of Financial Economics 76, 471-508.

Ang, A., R. H. Hodrick, Y. Xing, and X. Zhang, 2006. The Cross Section of Volatility and Expected Returns, Journal of Finance 61, 259-299.

Ang, A., W. N. Goetzmann, and S. Schaefer, 2011. Efficient Market Theory and Evidence: Implications for Active Management. Foundations and Trends 5, 158-242.

Ang, A., A. Madhavan, and A. Sobcyzk, 2017. Estimating Time-Varying Factor Exposures, Financial Analysts Journal 73, 41-54.

Asness, C. S., 2004. An Alternative Future: Part I. Journal of Portfolio Management 30, 94-103.

Asness, C. S., A. Frazzini, and L. H. Pedersen, 2019. Quality Minus Junk, Review of Accounting Studies 24, 34-112.

Avramov, D., 2004. Stock Predictability and Asset Pricing Models. Review of Financial Studies 17, 699-738.

Banz, R., 1981. The Relationship Between Return and Market Value of Common Stocks, Journal of Financial Economics 9, 3-18.

Basu, S., 1977. Investment Performance of Common Stocks in Relation to their Price-Earnings Ratios: A Test of the Efficient Market Hypothesis, Journal of Finance 32, 663-82.

Bawa, V. S., S. J. Brown, and R. W. Klein, 1979. *Estimation Risk and Optimal Portfolio Choice*. Amsterdam: North-Holland.

Bellord, E., J. Livnat, D. Porter, and M. B. Tarlie, 2019. Optimal Holdings of Active, Passive and Smart Beta Strategies. Journal of Investment Management 17, 40-62.

Best, M. J., and R. R. Grauer, 1991. On the Sensitivity of Mean Variance-Efficient Portfolios to Changes in Asset Means: Some Analytical and Computational Results. Review of Financial Studies 4, 315-342.

Black, F., and R. Litterman, 1991. Asset Allocation: Combining Investor Views with Market Equilibrium. Journal of Fixed Income 1, 7-18.

Barry, C. B., 1974. Analysis Under Uncertain Means, Variances, and Covariances, Journal of Finance 29, 515-522.

Brown, S. J., 1979. The Effect of Estimation Risk on Capital Market Equilibrium, Journal of Financial and Quantitative Analysis 14, 215-220.

Carson, B., S. Shores, and N. Nefouse, 2017. Life-Cycle Investing and Smart Beta Strategies. Journal of Retirement 5, 66-82.

Corum, A. A., A. Malenko, and N. Malenko, 2020. Corporate Governance in the Presence of Active and Passive Delegated Investment, SSRN Working paper 3681905.

Duane, A., A. D. Kennedy, B. J. Pendleton, and D. Roweth, 1987. Hybrid Monte Carlo. Physics Letters B 195, 216-222.

Fama, E. F., and K. R. French, 1993. Common Risk Factors in the Returns on Stocks and Bonds. Journal of Financial Economics 60, 3-43.

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

Fama, E. F. and K. R. French, 2018. Long-Horizon Returns. Review of Asset Pricing Studies 8, 232-252.

Frazzini, A., and L. H. Pedersen, 2014. Betting Against Beta. Journal of Financial Economics 111, 1-25.

Garlappi, L., R. Uppal, and T. Wang, 2007. Portfolio Selection with Parameter and Model Uncertainty: A Multi-Prior Approach. Review of Financial Studies 20, 41-81.

Gelman, A., and D. B. Rubin. 1992. Inference from Iterative Simulation Using Multiple Sequences. Statistical Science 7, 457–511.

Gelman, A., 2006. Prior Distributions for Variance Parameters in Hierarchical Models. Bayesian Analysis 1, 515–533.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, 2013. *Bayesian Data Analysis*. Boca Raton, FL: CRC Press.

Geman, S., and D. Geman, 1984. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. IEEE Transactions on Pattern Analysis and Machine Intelligence 6, 721-741.

Geweke, J., 1992. Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments, in Bernardo, J. M., J. O. Berger, A. P. Dawid, and A. F. M. Smith, eds., *Bayesian Statistics 4*, 169-193. Oxford University Press.

Grinold, R. C., 1994. Alpha is Volatility Times IC Times Core. Journal of Portfolio Management 20, 9-16.

Grinold, R. C., and R. N. Kahn, 2000. The Efficiency Gains of Long-Short Investing. Financial Analysts Journal 56, 40-53.

Hansen, L. P., and R. Jagannathan, 1991. Implications of Security Market Data for Models of Dynamic Economies. Journal of Political Economy 99, 225-262.

Hoffman, M. D. and A. Gelman, 2014. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. Journal of Machine Learning Research 15, 1593-1623.

Homescu, C., 2015. Better Investing Through Factors, Regimes, and Sensitivity Analysis. ssrn.com/abstract=2557236

Jegadeesh, N., and S. Titman, 1993. Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency, Journal of Finance 56, 699-720.

Joe, H., 2006. Generating Random Correlation Matrices Based on Partial Correlations. Journal of Multivariate Analysis 97, 2177-2189.

Kong, A., J. S. Liu, and W. H. Wong, 1994. Sequential Imputations and Bayesian Missing Data Problems. Journal of the American Statistical Association 89, 278-288.

Lewandowski, D., D. Kurowicka, H. Joe, 2009. Generating Random Correlation Matrices Based on Vines and Extended Onion Method. Journal of Multivariate Analysis 100, 1989-2001.

Lintner, J., 1975. Inflation and Security Returns. Journal of Finance 30, 259-280.

Maccheroni, F., M. Marianacci, and A. Rustichini, 2006. Ambiguity Aversion, Robustness, and the Variational Representation of Preferences. Econometrica 74, 1147-1498.

Madhavan, A., A. Sobczyk, and A. Ang, 2020. Alpha vs. Alpha: Selection, Timing, and Factor Exposures from Different Factor Models, Journal of Portfolio Management 46, 90-103.

Markowitz, H., 1952. Portfolio Selection. Journal of Finance 7, 77-91.

Merton, R. C., 1980. On Estimating the Expected Return of the Market: An Exploratory Investigation, Journal of Financial Economics 29, 449-470.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, and A. H. Teller, 1953. Equation of State Calculations by Fast Computing Machines 21, 1087-1092.

Neal, R. M., 1994. An Improved Acceptance Procedure for the Hybrid Monte Carlo Algorithm. Journal of Computational Physics 111, 194-203.

Neal, R. M., 2011. MCMC Using Hamiltonian Dynamics, in Brooks, S., A. Gelman, G. L. Jones, and X. L. Meng, eds., *Handbook of Markov Chain Monte Carlo*, Chapman & Hall/CRC, 113-161.

Pastor, L. and R. F. Stambaugh, 1999. Costs of Equity Capital and Model Mispricing. Journal of Finance 54, 67-121.

Pastor, L. and R. F. Stambaugh, 2000. Comparing Asset Pricing Models: An Investment Perspective. Journal of Financial Economics 56, 335-381.

Pedersen, L. H., A. Babu, and A. Levine, 2020. Enhanced Portfolio Optimization, forthcoming Financial Analysts Journal.

Polson, N. G., and J. G. Scott, 2012. On the Half-Cauchy Prior for a Global Scale Parameter. Bayesian Analysis 7, 887-902.

Roll, R., 1977. A Critique of the Asset Pricing Theory's Tests Part I: On Past and Potential Testability of the Theory. Journal of Financial Economics 4, 129-176.

Ross, S. A., 1976. The Arbitrage Theory of Capital Asset Pricing. Journal of Economic Theory 13, 341-360.

Routledge, B. R., and S. E. Zin, 2010. Generalized Disappointment Aversion and Asset Prices. Journal of Finance 65, 1303-1332.

Rouwenhorst, K. G., 2005. The Origins of Mutual Funds, in Goetzmann, W. N. and K. G. Rouwenhorst, eds., *The Origins of Value: The Financial Innovations that Created Modern Capital Markets*, Oxford University Press, 249-270.

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

Sloan, R., 1996. Do Stock Prices Reflect Information in Accruals and Cash Flows About Future Earnings? Accounting Review 71, 289-316.

Tanner, M. A., and W. H. Wong, 1987. The Calculation of Posterior Distributions by Data Augmentation. Journal of the American Statistical Association 82, 528-540.

Tu, J., and G. Zhou, 2004. Data-Generating Process Uncertainty: What Difference Does It Make in Portfolio Decisions? Journal of Financial Economics 72, 385-421.

Treynor, J. L. and F. Black, 1973. How to Use Security Analysis to Improve Portfolio Selection. Journal of Business 46, 66-88.

Aki Vehtari, Andrew Gelman, Daniel Simpson, Bob Carpenter, and Paul-Christian Bürkner, 2019. Rank-normalization, folding, and localization: An improved Rˆ for assessing convergence of MCMC.

Exhibit 1: Fund Universe

Morningstar Category	Average active return	Average active risk	Average return to risk ratio	Average fee	Number of funds
Large Blend	-0.14%	2.99%	-0.01	0.68%	116
Large Growth	0.80%	4.02%	0.14	0.78%	155
Large Value	-0.07%	3.52%	0.04	0.72%	138
Mid-Cap Blend	-0.20%	4.97%	-0.06	0.93%	30
Mid-Cap Growth	1.75%	4.97%	0.31	0.91%	76
Mid-Cap Value	-0.02%	3.76%	0.02	0.80%	42
Small Blend	0.35%	4.26%	0.10	0.94%	62
Small Growth	1.88%	5.00%	0.35	0.98%	86
Small Value	0.54%	4.20%	0.15	0.96%	48
Technology	3.34%	7.85%	0.42	0.95%	16
ALL	0.63%	4.11%	0.13	0.82%	769

Source: Morningstar as of December 31, 202

Exhibit 2: Factor Loadings of Active Funds

Morningstar Category	Intercept	US Equities	Momentum	Small Size	Quality	Value	Minimum Volatility
US Fund Large Blend	0.00 (0.02)	0.00 (0.10)	0.002 (0.06)	0.022 (0.71)	-0.043 (-0.16)	0.058 (0.74)	-0.032 (-0.57)
US Fund Large Growth	0.001 (0.08)	0.015 (0.37)	0.206** (2.03)	0.032 (0.57)	-0.074 (-0.34)	0.043 (0.42)	-0.165* (-1.85)
US Fund Large Value	0.00 (0.16)	-0.002 (-0.21)	0.019 (0.35)	0.004 (-0.01)	0.047 (0.50)	0.073 (0.82)	-0.048 (-0.75)
US Fund Mid-Cap Blend	-0.002 (0.01)	0.016 (-0.01)	0.028 (0.16)	0.023 (-0.37)	0.013 (0.19)	0.035 (0.26)	-0.15 (-1.45)
US Fund Mid-Cap Growth	0.014 (0.91)	0.007 (-0.25)	0.204* (1.71)	0.017 (0.22)	-0.022 (-0.04)	-0.043 (-0.44)	-0.079 (-0.90)
US Fund Mid-Cap Value	0.005 (0.54)	0.008 (0.06)	-0.051 (-0.68)	0.006 (0.18)	0.151 (1.05)	0.017 (0.08)	-0.184** (-2.21)
US Fund Small Blend	0.003 (0.29)	0.012 (0.10)	-0.045 (-0.69)	-0.138** (-2.34)	-0.033 (-0.12)	0.007 (0.02)	0.059 (0.73)
US Fund Small Growth	0.015 (0.85)	0 (-0.05)	0.127 (0.93)	-0.104* (-1.74)	-0.059 (-0.31)	-0.102 (-0.78)	-0.006 (0.02)
US Fund Small Value	0.002 (0.18)	0.048 (1.32)	0.013 (0.17)	-0.118** (-2.08)	0.079 (0.46)	0.054 (0.47)	-0.059 (-0.61)
US Fund Technology	0.026 (1.07)	0.044 (0.73)	0.308** (2.17)	0.235** (2.51)	0.135 (0.24)	-0.011 (-0.19)	-0.383** (-2.41)
All	0.004 (0.32)	0.009 (0.13)	0.081 (0.72)	-0.012 (-0.22)	-0.008 (0.05)	0.021 (0.25)	-0.078 (-0.89)

Average coefficients that are significant at the 10%, 5%, and 1% levels are marked by *, **, and ***, respectively.

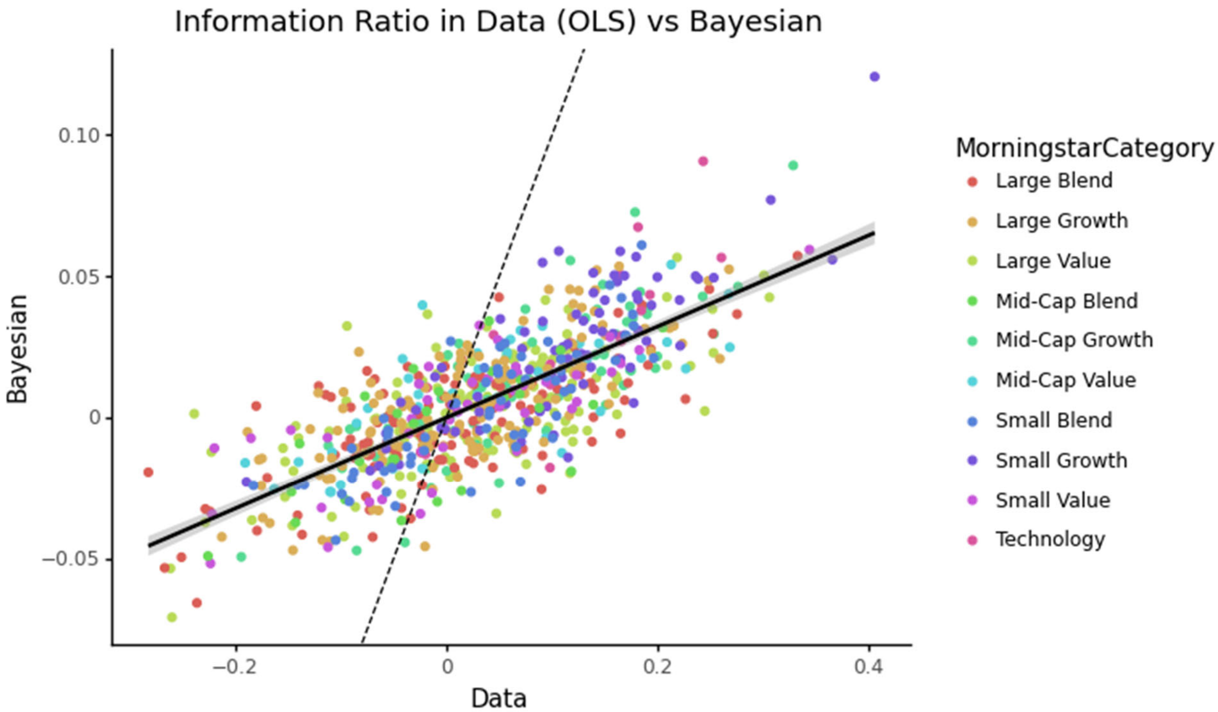
Source: Authors' calculations using factor, index and fund return data from Morningstar and Bloomberg. Past performance does not guarantee future results.

Exhibit 3: Empirical Information Ratios for Factor Indexes

Factor	Mean IR	Standard deviation of IR
US Equity Market	0.45	0.14
Minimum Volatility	0.94	0.21
Momentum	0.59	0.14
Value	0.28	0.17
Size	0.17	0.14
Quality	0.55	0.14

Source: Authors' calculations using factor, index and fund return data from Morningstar and Bloomberg. Past performance does not guarantee future results.

Exhibit 4: Information Ratios in Data vs Bayesian Procedure



Source: Authors' calculations using factor, index and fund return data from Morningstar and Bloomberg. Past performance does not guarantee future results. The dashed line is the 45 degree line.

Exhibit 5: Out-of-Sample Fund Ranks

Transition Matrix, Bayesian			
	Top 1/3	Middle 1/3	Bottom 1/3
Top 1/3	0.53	0.38	0.09
Middle 1/3	0.24	0.46	0.30
Bottom 1/3	0.24	0.15	0.61
Transition Matrix, OLS			
	Top 1/3	Middle 1/3	Bottom 1/3
Top 1/3	0.30	0.39	0.31
Middle 1/3	0.34	0.28	0.38
Bottom 1/3	0.36	0.32	0.31

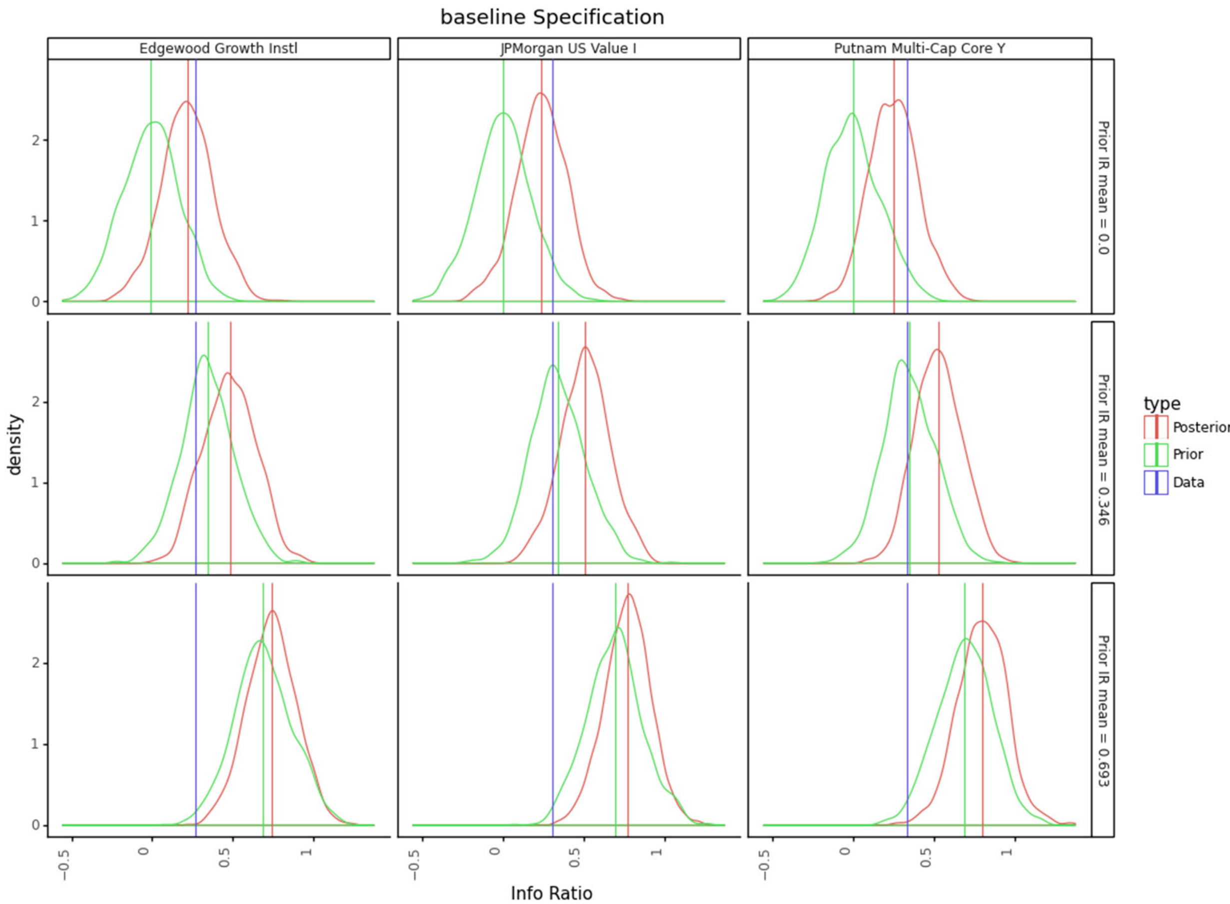
Source: Authors' calculations using factor, index and fund return data from Morningstar and Bloomberg. Past performance does not guarantee future results.

Exhibit 6: Posterior Distribution of Information Ratios (Annualized)

	Empirical IR	Mean of posterior IR	Std of posterior IR	Posterior IR 3%	Posterior IR 97%
S&P 500	1.069	0.551	0.097	0.367	0.727
Momentum	0.479	0.470	0.097	0.288	0.653
Quality	0.004	0.478	0.097	0.292	0.662
Size	-0.091	0.156	0.097	-0.025	0.339
Minimum Volatility	1.19	0.851	0.145	0.578	1.125
Value	-0.328	0.055	0.121	-0.173	0.282
Pioneer Equity Income Y	0.316	0.339	0.152	0.055	0.627
JPMorgan Equity Income I	0.847	0.461	0.152	0.173	0.752
T. Rowe Price Instl Mid-Cap Equity Gr	0.557	0.395	0.152	0.111	0.682
BlackRock Technology Opportunities Instl	0.573	0.398	0.152	0.114	0.689
PIMCO StocksPLUS Small Institutional	0.167	0.305	0.152	0.014	0.585

Source: Authors' calculations using factor, index and fund return data from Morningstar and Bloomberg. Past performance does not guarantee future results.

Exhibit 7: Fund Posterior Information Ratios Under Different Priors



Source: Authors' calculations using factor, index and fund return data from Morningstar and Bloomberg. Past performance does not guarantee future results.

Exhibit 8: Predictive Moments (Annualized Monthly Returns)

	Data (10yr) mean	Predictive mean	Data (10yr) std	Predictive std
S&P 500	0.145	0.076	0.136	0.139
Momentum	0.026	0.025	0.054	0.054
Quality	0.000	0.011	0.024	0.024
Size	-0.008	0.014	0.090	0.090
Minimum Volatility	0.070	0.051	0.059	0.060
Value	-0.019	0.003	0.057	0.057
Pioneer Equity Income Y	0.006	0.013	0.031	0.031
JPMorgan Equity Income I	0.015	0.012	0.029	0.029
T. Rowe Price Instl Mid-Cap Equity Gr	0.007	0.007	0.030	0.030
BlackRock Technology Opportunities Instl	0.050	0.036	0.071	0.073
PIMCO StocksPLUS Small Institutional	0.023	0.018	0.030	0.031

Source: Authors' calculations using factor, index and fund return data from Morningstar and Bloomberg. Past performance does not guarantee future results.

Exhibit 9: Resulting Portfolio Allocation

Panel A: Baseline Allocation

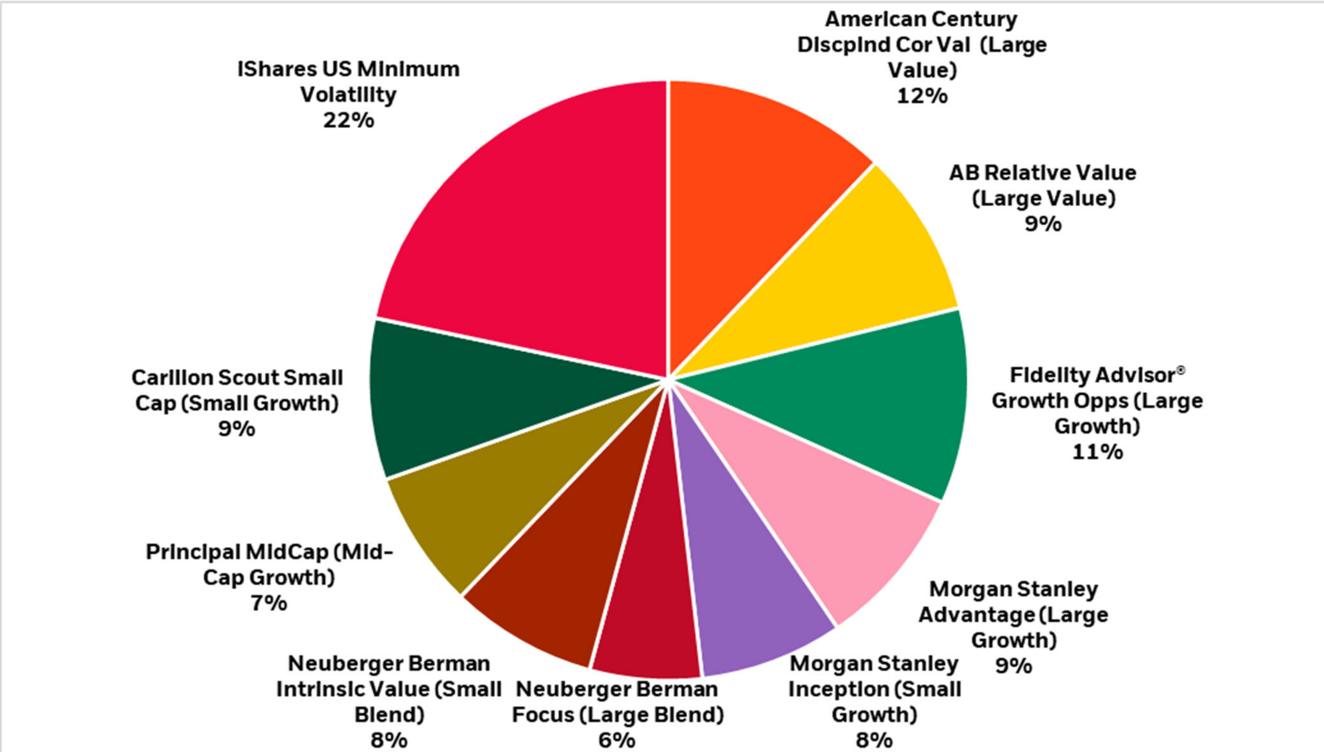
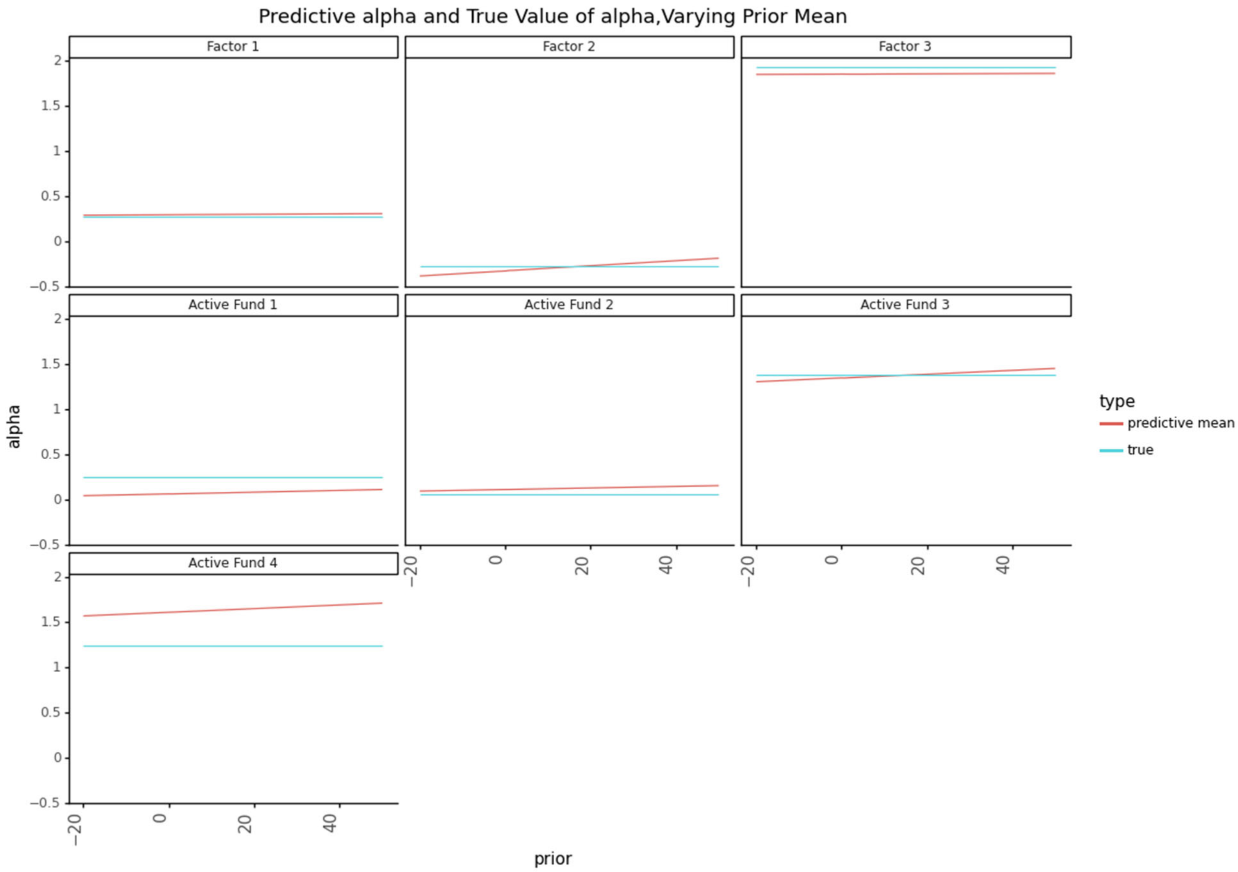
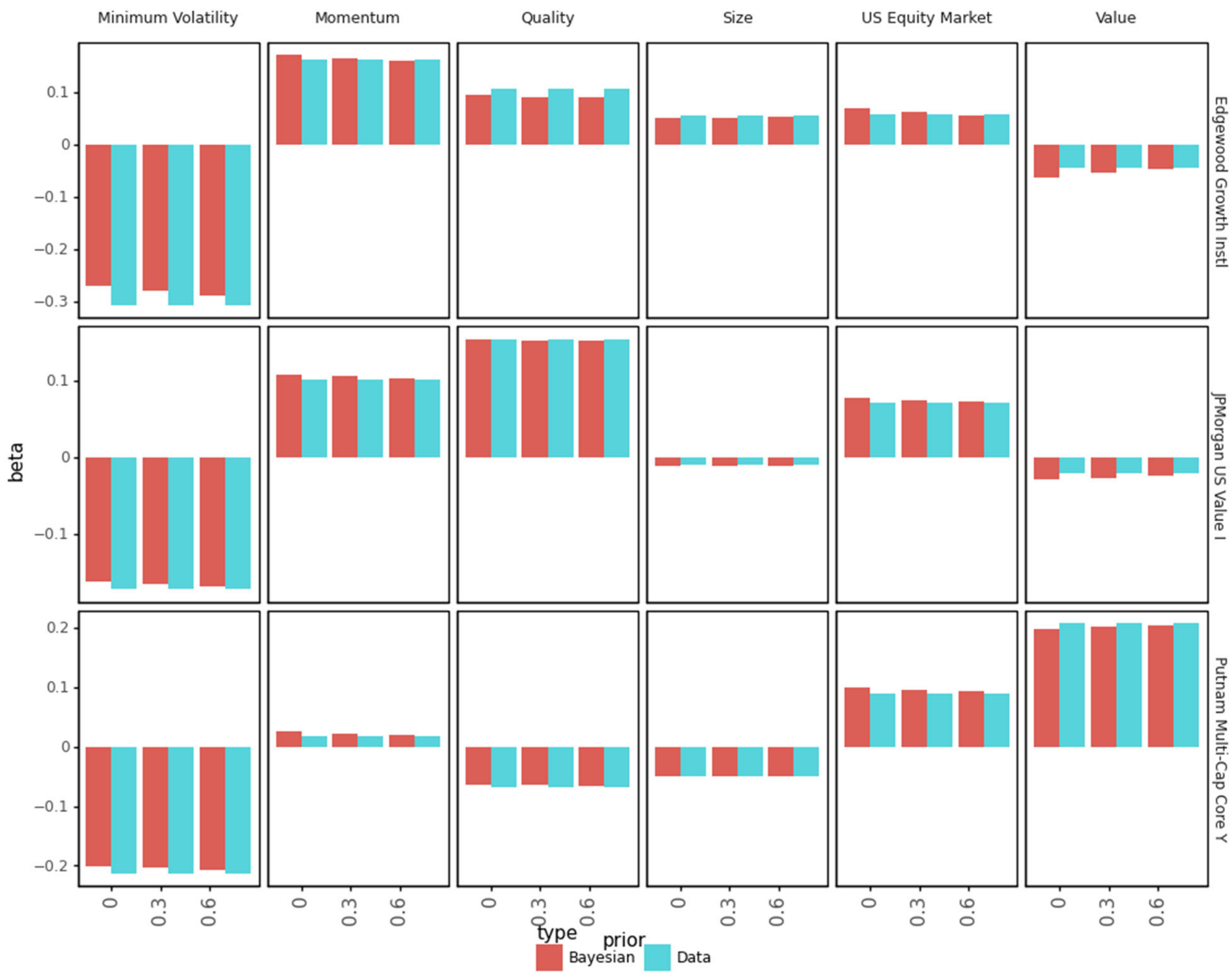


Exhibit A-1: Robustness Checks on Estimated Minus True Alphas in Simulation Exercise



The x-axis is various prior mean for information ratios we used. The y-axis is the information ratio. The red lines represent the posterior predictive mean results from our Bayesian procedure. The blue lines plot the true value. Each panel represents either an active fund or a factor. Source: Authors' calculations and simulations.

Exhibit A-2: Posterior Factor Loadings of Funds vs OLS Estimates Under Different Priors and Specifications



The x-axis shows different priors for active fund IRs. The y-axis shows factor loadings (betas). Each row represents an active fund and each column represents a factor/index. Source: Authors' calculations using factor, index and fund return data from Morningstar and Bloomberg. Past performance does not guarantee future results.