Financial Analysts Journal

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ufaj20

Index + Factors + Alpha

Andrew Ang, Linxi Chen, Michael Gates & Paul D. Henderson

To cite this article: Andrew Ang, Linxi Chen, Michael Gates & Paul D. Henderson (2021) Index + Factors + Alpha, Financial Analysts Journal, 77:4, 45-64, DOI: 10.1080/0015198X.2021.1960782

To link to this article: https://doi.org/10.1080/0015198X.2021.1960782

Financial Analysts Journal | A Publication of **CFA Institute**
https://doi.org/10.1080/0015198X.2021.1960782

Research

Index + Factors + Alpha

Andrew Ang, Linxi Chen, Michael Gates, and Paul D. Henderson

Andrew Ang is managing director at BlackRock, New York City. Linxi Chen is vice president at BlackRock, New York City. Michael Gates is managing director at BlackRock, San Francisco. Paul D. Henderson is director at BlackRock, New York City.

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, such as exposure to 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 will be smaller than those for alpha strategies, whereas prior means for alpha strategies may be larger than those for factor funds.

Disclosure: The authors report no conflicts of interest.

PL Credits: 2.0

Market-cap index, factors, alpha—these well-known sources of return each have extensive histories in academic literature and in practice.[1] The seminal role of the market in the first asset pricing model, the CAPM, has been a major topic in finance since Sharpe (1964) and Lintner (1975). Today's smart beta industry directly targets style factors, such as 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 commingled 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"), factor, 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. If 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 the tighter the prior standard deviation on the active funds' information ratios (IRs) and the more the active risk budget will be allocated to alpha versus 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 no tracking error budget would allocate 100% of the portfolio to index funds. An investor with a tracking error target but no conviction in alpha would allocate only to index and factor strategies.

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 from his 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 return requires specialized skills to take advantage of market inefficiencies or dislocations. Ang, Goetzmann, and Schaefer (2011) concluded 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 that 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 IRs of alpha-seeking strategies would typically have more dispersion but may have a higher prior mean than that 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 because the costs of factor strategies are, on average, significantly lower than those of alpha funds.

We employ a recent advance in Bayesian computation methods, the no-U-turn sampler (NUTS) 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 versus 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.

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 the traditional Markowitz (1952) mean–variance approach 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 usual, but we introduce a step before the optimization that 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 (2021) showed how traditional mean–variance approaches can be adapted to incorporate 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 applied 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 selected alpha-seeking funds from the Morningstar universe of active funds consisting of large-, mid-, and small-cap funds in value, growth, and blend styles and technology funds. We used standard long-only factor investment strategies tracking minimum-volatility, momentum, value, small size, and quality factors, all of which are easily accessible in low-cost and transparent exchange-traded funds (ETFs). These factor strategies have longer sample lengths than the alpha funds have. We set different prior beliefs on factor and alpha IRs, which we compared with the resulting posterior and predictive means and variances. In a second stage, we fed these updated means and variances into a regular mean–variance utility maximization to determine optimal index, factor, and alpha allocations.

Our approach is most related to that of Pástor and Stambaugh (1999, 2000); Baks, Metrick, and Wachter (2001); Tu and Zhou (2004); and Avramov (2004), who formulated asset allocation models in a Bayesian setting.[5] This literature does not directly specify priors in terms of Sharpe ratios or information ratios, although Pástor and Stambaugh's prior is proportional to residual variance. Consistent with most asset pricing models that 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 Markov chain Monte Carlo (MCMC) methods used in 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]

Theoretical Setting

In this section, we first summarize the different roles of index, factor, and alpha-seeking strategies. Investors should allocate to alpha only if those strategies are generating returns in excess of index *and* factors. To get a flavor of formal modeling, we then 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. We leave full expositions and proofs to Appendix A.

Alpha Is Excess of Index *and* Factors.

Consider three assets: index, factor, and pure alpha-seeking strategies. (We also refer to the market index as a benchmark portfolio.) The alpha-seeking return, r_i, loads on both the market and factor portfolios, and it has a mean return, α_i, *in excess* of market index and factor exposures. Then, we have the following:

> **Proposition 1:** If there is no alpha present ($\alpha_i = 0$), then the holdings of the active manager are 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 toward 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 outperform relative to the exposures of index and factors. We can also characterize the maximum Sharpe ratio achievable by allocating to index, factors, and alpha, which depends on the individual assets' squared reward-to-risk ratios.

> **Proposition 2:** The marginal contribution to the portfolio's Sharpe ratio of the alpha-seeking manager in addition to that of index and factor funds depends on the appraisal ratio (excess returns divided by volatility) of the active manager's return *in excess* of index and factors to the active manager's idiosyncratic volatility.

Put another way, the active fund can have a positive return due to index or factors or due to both, but unless the alpha-seeking fund beats *both* the index and factor strategies, it is irrelevant for the investor. Given these results, we assume that there are active managers with positive alpha in excess of index and factors. The next section explores the intuition behind a methodology to allocate to all three components.

Incorporating Uncertainty.

For simplicity, we assume there are past returns of one risky asset series, $\{y_t\}$, that are net of fees with Sharpe ratio (or information ratio) $S = \frac{\mu}{\sigma}$, where μ and σ are the asset's mean and standard deviation of returns, respectively. The cost of transacting the strategy is c.[7] We assume that the standard deviation, σ, is known and estimate the mean parameter, μ. Appendix B describes the full model with uncertainty on all parameters.

Priors and posteriors of the Sharpe ratio. We assume that the returns, y, are normally distributed, so the likelihood function is

$$p(y|S,\sigma^2) \sim N(\sigma S - c, \sigma^2). \tag{1}$$

We impose the following prior on the Sharpe ratio:

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

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 those for alpha funds.

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

$$p(S|\mathcal{Y},\sigma^2) \sim N(\mu_S, \sigma_S^2), \tag{3}$$

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

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

and

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

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

The intuition for Equation 3 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, the distribution of the Sharpe ratio, $S = \mu/\sigma$, has a variance of $1/T$.[8] The inverse of this expression is the first term in the posterior variance of σ_S^2. This term is combined with the inverse of the variance of the prior of S in Equation 3. A similar expression also occurs in a general Bayesian regression setup, 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 3, 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 3 for μ_S as

$$\mu_S = \text{Weight} \times (\bar{y}+c) + (1-\text{Weight}) \times S_0, \qquad (4)$$

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 4 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 $\sigma_S^2\to 1/T$, which are the sample mean estimator and variance, respectively, of the Sharpe ratio, S.

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:

$$p(y\mid\mathcal{Y}) = \int_S p(y|\mathcal{Y},S)p(S\mid\mathcal{Y})dS, \qquad (5)$$

where $p(y|\mathcal{Y},S)$ is the likelihood and $p(S\mid\mathcal{Y})$ is the posterior of S. The Bayesian predictive distribution accounts for uncertainty about the unknown parameters. 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.[9]

For mean–variance utility—the most popular objective function in practice—we need just the mean and variance of the predictive distribution, $p(y|\mathcal{Y})$.[10]

In our simple case, the predictive y follows a normal distribution,

$$y \sim N(\mu^*,\sigma^{*2}), \qquad (6)$$

where $\mu^* = \sigma E(S\,|\mathcal{Y}) - c = \sigma\mu_S - c$ and $\sigma^{*2} = \sigma^2(1+\sigma_S^2)$.

In Equation 6, 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 dispersion on the Sharpe ratio is relatively low, this fact decreases the predictive variance and makes those strategies relatively more attractive.

Data

In this section, we first describe the market-cap index benchmark and the factor portfolios. Then, we describe the alpha-seeking funds. To illustrate the methodology, we work only with US equity factor and alpha-seeking funds.

Index and Factors.

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

For factors, we used long-only factor indexes tracking minimum volatility (Ang, Hodrick, Xing, and Zhang 2006), return momentum (Jegadeesh and Titman 1993), value (Basu 1977), small size (Banz 1981), and quality (Sloan 1996), listed in **Table 1**.

Table 1. Factors and Corresponding Indexes

Factor	Index
Minimum volatility	MSCI USA Minimum Volatility (USD) Index
Momentum	MSCI USA Momentum Index
Value	MSCI USA Enhanced Value Index
Small size	S&P SmallCap 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. As in the model setup in the section "Theoretical Setting," we modeled these factors as orthogonal to the market return. This is done by running a regression of the factor index returns on the US large-cap equity index over the full sample. The residuals are then used to represent the factors.

Alpha-Seeking Funds. We constructed 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 took funds in the US equity style box categories with capitalizations of large, mid, and small and styles of value, growth, and blend, as well as technology sector funds. We report summary statistics of the funds in **Table 2**. Most of the funds fall into US large growth and blend categories, which account for the majority of assets under management for active funds (see, for example, recent statistics in Madhavan, Sobczyk, and Ang 2020). We computed 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 31 December 2020. Table 2 reports that the average monthly active return of these funds is positive, at 0.63% per month, with an average reward-to-risk ratio of 0.13. The reward-to-risk ratios range from −0.06 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 estimated 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 that may generate high returns with low active risk.

We constructed 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 took the Fama and French (1993) factors: the market factor (MKT), the one-month T-bill (RF), size (SMB), and value (HML). We also used Kenneth French's construction of the Jegadeesh and Titman (1993) momentum factor (UMD). We augmented these factors with the quality minus junk factor (QMJ) constructed by Asness, Frazzini, and Pedersen (2019) 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 investable. Hence, we used these to inform the priors on the factors, but when we allocated, we held the investable factor indexes listed in the section "Index and Factors." We took advantage of the fact that we could use a longer time

Table 2. Fund Universe, 31 December 2011 to 31 December 2021

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.

series—specifically, 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 Appendix B for further details.)

Empirical Results

In this section, we first discuss the estimation of factor loadings of the active funds on the factors. Next, we describe the resulting posterior and predictive distributions. Then, we describe the results from using these distributions as inputs into a portfolio allocation problem.

Factor Loadings. **Table 3** reports the results of regressing active fund returns on the factors over the 10 years ending December 2020. These regressions used 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 assumed that these factor loadings were held fixed in order to isolate the effects of the IR assumptions on the asset allocation.[11]

Because both the LHS and RHS are in excess returns, the coefficients in Table 3 are generally small in absolute value. The numbers in parentheses are

Table 3. Factor Loadings of Active Funds, 31 December 2011 to 31 December 2021

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

*Significant at the 5% level.

**Significant at the 1% level.

Notes: Numbers in parentheses are averages of the *t*-statistics across the fund-by-fund regressions.

Source: Authors' calculations using factor, index, and fund return data from Morningstar and Bloomberg.

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 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 Table 3 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.

Table 4 reports data on the factors (see the section "Index and Factors"), over the last 25 years of data ending December 2020, that we used to set the priors on factors. The prior mean IRs were set as the full data averages. When we bootstrapped, we drew samples with replacement from the full sample, and we used the standard deviations of the bootstrapped samples as the standard deviations of the IRs. The highest IR is for minimum volatility, at 0.94, and the lowest is for size, at 0.17. The standard deviations are relatively tight, from 0.14 to 0.21.

Posterior and Predictive Results.

After setting the priors, we ran the procedure with no-U-turn sampling to generate posterior and predictive information ratios, alphas, and the covariance matrix. We used automatic differentiation variational inference (ADVI) to initialize the sampler, setting the target acceptance ratio to 80%. We used 25 chains with 5,000 draws each to obtain 125,000 draws in total. For posterior predictive analysis, we generated 50,000 simulated returns.[12]

Table 4. Empirical Information Ratios for Factor Indexes, 31 December 2011 to 31 December 2021

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

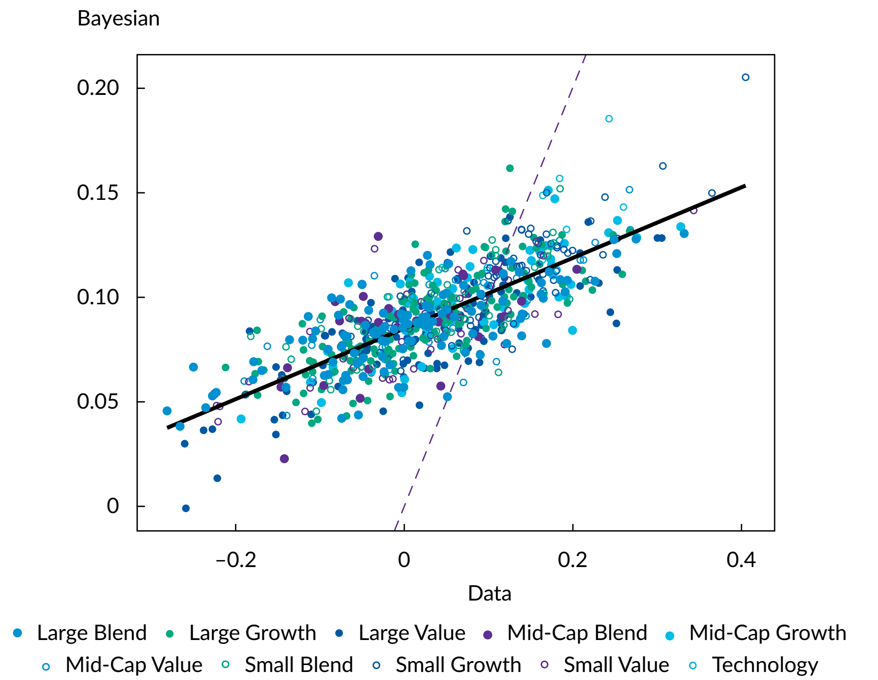
Source: Authors' calculations using factor, index, and fund return data from Morningstar and Bloomberg.

To visualize the broad impact of the Bayesian procedure, we examine a scatterplot comparing the IR based on ordinary least-squares (OLS) data and the Bayesian IR (posterior mean) in **Figure 1**. 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 procedure shrinks these toward zero (see Equations 3 and 4). In Figure 1, we plot the linear fit, which has a slope of 0.23. This is significantly less than 1, which we can see in Figure 1, where we also plot a 45-degree line as a dashed diagonal line. This slope implies that the procedure shrinks extreme IRs toward zero.

Shrinkage and out-of-sample forecasts. The shrinkage does improve out-of-sample forecasts.[13] **Table 5** reports the results when we split the data into two halves, from 2010 to 2015 and from 2016 to 2020. For the first half, we performed our procedure producing a posterior mean of IRs for each fund. Next, we sorted the funds into three bins (high/middle/low) according to their posterior IRs. We examined 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 Panel A of Table 5, 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 results. For comparison, we also computed 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. Table 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 3rd and 97th percentiles of the posterior IR distribution for the factors and several selected alpha-seeking funds. Because we imposed priors for index and factor funds, the resulting posterior IRs reported in Table 6 should be expected to differ from their empirical counterparts

Figure 1. Information Ratios in Data vs. Bayesian Procedure, 31 December 2011 to 31 December 2021



Note: The dashed line is the 45-degree line.
Source: Authors' calculations using factor, index, and fund return data from Morningstar and Bloomberg.

because of shrinkage. For example, over the last 10 years ending December 2020, 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 did exhibit outperformance. Likewise, the posterior IRs of the active funds are attenuated on average. For example, the JPMorgan Equity Income I fund has an empirical IR of 0.85 over the 10-year sample period. The posterior IR 3rd and 97th percentile bounds are positive, but the mean posterior IR is shrunk to 0.46. In general, we observe that the standard deviations of the posterior IRs are smaller than the standard deviations set in the prior owing to the contribution of the observable data and that prior standard deviations deliberately err on the noninformative side.

Table 5. Out-of-Sample Fund Ranks, 31 December 2011 to 31 December 2021

	Top 1/3	Middle 1/3	Bottom 1/3
A. Transition matrix, Bayesian			
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
B. Transition matrix, OLS			
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.

Posteriors as a function of priors. In **Figure 2**, 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) to 0.35 (middle row) to 0.69 (bottom row). Each column corresponds to a different fund. The heavy black 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 dotted blue and the solid green 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 Figure 2, the prior IR is zero and the posterior IR has a mean close to the empirical IR. For example, for institutional shares of the Edgewood

Table 6. Posterior Distribution of Information Ratios (Annualized), 31 December 2011 to 31 December 2021

	Empirical IR	Mean of Posterior IR	Std. Dev. of Posterior IR	Posterior IR 3rd Percentile	Posterior IR 97th Percentile
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.

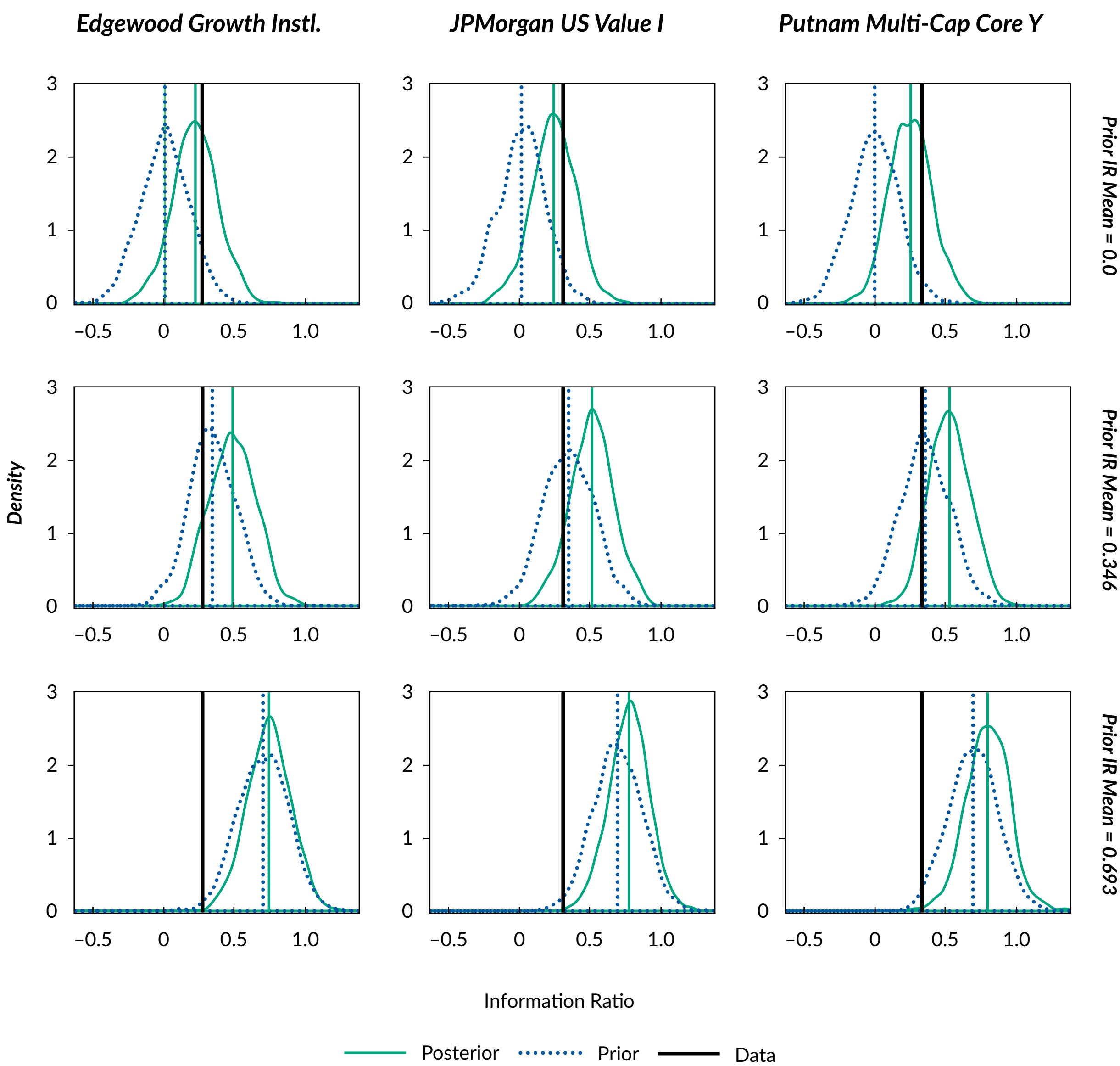
Growth Fund (ticker: EGFIX), the data IR is 0.26 and the posterior IR has a mean of 0.25 in the first row. In the second row, we set the prior IR to have a mean of 0.35. In each case, as expected, the posterior mean increases with the prior mean, but both means are still fairly close to the data average. For example, for EGFIX in the second row with a prior IR of 0.35, the posterior IR is now 0.48, which is still close to the data IR of 0.26 but is higher than the posterior IR with mean 0.25 in the first row. The posterior distributions also narrow compared with the first row. In the last row, where 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 do 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 **Table 7**. The means of the style factors (momentum, quality, size, minimum volatility, and value) are stated in excess of the S&P 500. The 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. In contrast, 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. For the effect of allowing uncertainty in the factor loadings, see the Online Supplemental Material.

Portfolio Construction. Taking the predictive moments in Table 7, we performed a mean–variance optimization exercise. We constructed portfolios to maximize portfolio active net-of-fee return subject to a long-only constraint and limited the number of holdings to 10 or fewer with a standard mean–variance objective function. In our benchmark case, we took active risk relative to the large-cap benchmark of the S&P 500 Index, and we computed net-of-fee returns assuming fees as of December 2020. We used a risk aversion coefficient of 32, which can be calibrated to target a particular level of active risk.

Figure 3 reports the equity portfolio holdings for the baseline case. The three largest holdings are the minimum-volatility factor (USMV), American Century Disciplined Core Value, and Fidelity Advisor Growth

Figure 2. Fund Posterior Information Ratios under Different Priors, 31 December 2011 to 31 December 2021



Source: Authors' calculations using factor, index, and fund return data from Morningstar and Bloomberg.

Opportunities, with weights of 22%, 12%, and 11%, respectively. Using the predictive moments from our procedure, we estimated this portfolio yields an excess return of 3.8% with an active risk of 2.0%, both of 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.

Conclusion

We showed the separate roles played by index, factor, and active funds—and importantly showed how investors can incorporate prior information on the three sources of return that can be practically implemented in portfolio construction. Investors

Table 7. Predictive Moments (Annualized Monthly Returns), 31 December 2011 to 31 December 2021

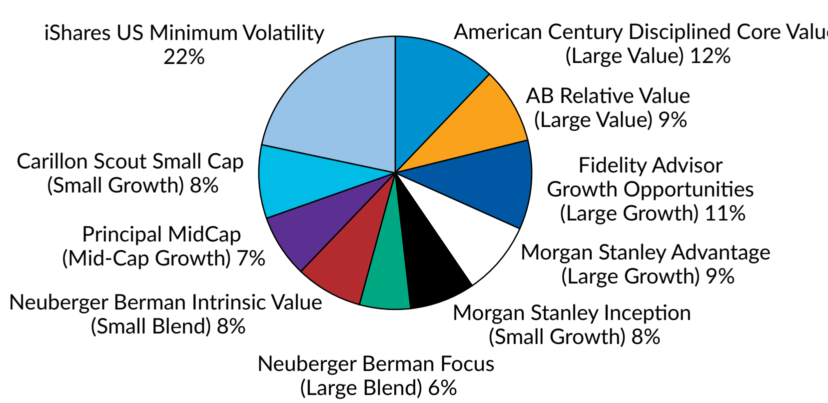
	Data (10-yr.) Mean	Predictive Mean	Data (10-yr.) Std. Dev.	Predictive Std. Dev.
S&P 500	0.145	0.076	0.136	0.139
Momentum	0.026	0.025	0.054	0.054
Quality	0.000	0.011	0.024	0.024
Size	–0.008	0.014	0.090	0.090
Minimum volatility	0.070	0.051	0.059	0.060
Value	–0.019	0.003	0.057	0.057
Pioneer Equity Income 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.

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 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 sampler, we constructed posterior IRs and used the predictive moments in a portfolio construction exercise allocating to index, factor, and active funds.

There are many possible extensions from our work. We can extend the application from long-only to long–short investing. We only incorporated beliefs and information on performance and specified priors on IRs. Investors may also have preferences for certain vehicles—such as ETFs versus traditional mutual funds—may take into account taxes, may lean toward more transparent strategies, or may prefer delegated or direct portfolio management. Index, factor, and alpha funds each have different advantages and

Figure 3. Resulting Portfolio Allocation: Baseline Allocation, 31 December 2011 to 31 December 2021



Source: Authors' calculations using factor, index, and fund return data from Morningstar and Bloomberg.

disadvantages along these dimensions beyond beliefs and historical data on IRs that can also be incorporated into the allocation problem.

Appendix A. Allocating to Index, Factor, and Alpha-Seeking Strategies

In this appendix, we prove Propositions 1 and 2. We work with three assets: index (r_m), factor (f), and pure alpha-seeking (r_i) strategies, all of which are specified in excess of the risk-free rate.

Assume that the factor, f, is uncorrelated with the market, r_m:

$$r_m = \mu_m + \varepsilon_m, \tag{A1}$$

and

$$f = \mu_f + \varepsilon_f, \tag{A2}$$

with orthogonal zero mean shocks ε_m and ε_f with variances σ_m^2 and σ_f^2, respectively.[14]

The alpha-seeking fund, r_i, loads on both the market and factor portfolios:[15]

$$r_i = \alpha_i + r_m + f + \varepsilon_i, \tag{A3}$$

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 A3 to represent a fund with a market beta of 1, and the fund has unit factor exposure f, as well as unique alpha insights in excess of the factor returns, α_i.

The covariance of $(r_m \, f \, r_i)'$ from Equations A1–A3 is given by

$$\Sigma = \begin{bmatrix} \sigma_m^2 & 0 & \sigma_m^2 \\ 0 & \sigma_f^2 & \sigma_f^2 \\ \sigma_m^2 & \sigma_f^2 & \sigma_m^2 + \sigma_f^2 + \sigma_i^2 \end{bmatrix}. \tag{A4}$$

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

$$w \propto \begin{bmatrix} \mu_m / \sigma_m^2 - \alpha_i / \sigma_i^2 \\ \mu_f / \sigma_f^2 - \alpha_i / \sigma_i^2 \\ \alpha_i / \sigma_i^2 \end{bmatrix}. \tag{A5}$$

From Equation A5, we have Proposition 1.

The key result is that the active manager alpha, α_i, must be *in excess* of market index and factor exposures (μ_m and μ_f, respectively, in Equation A5). In this 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

$$S^2 = \mu_m^2 / \sigma_m^2 + \mu_f^2 / \sigma_f^2 + \alpha_i^2 / \sigma_i^2. \tag{A6}$$

Stated in words, the maximum squared Sharpe ratio depends on the individual assets' squared reward-to-risk ratios. This result was originally derived by Treynor and Black (1973) for the market return and alpha measured relative to the single CAPM factor. This proof gives Proposition 2.

Appendix B. Full Model Likelihood

We specify the jth factor return, $j = 1,\ldots,N_f$, to follow

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

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

$$f = \mu_f + \varepsilon_f,\ \varepsilon_f \sim N(0_{N_f \times 1}, \Sigma_f)\,. \tag{B1}$$

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:

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

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:

$$S_f = \mu_f \,./\, \sigma_f. \tag{B2}$$

The ith alpha strategy returns, $i = 1, \ldots, N_r$, are modeled as

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

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

$$r = \alpha + Bf + \varepsilon,\ \varepsilon \sim N(0_{N_r \times 1}, \Sigma_r), \tag{B3}$$

where $B = [\beta_1, \ldots, \beta_{N_r}]'$ 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 factors to be correlated, though in our setting, the data-generating process for the alphas is assumed to produce alphas that on average are uncorrelated. As with the factor strategies, we define the Sharpe ratio or information ratio vector to be

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

where $\sigma = [\sigma_1, \ldots, \sigma_{N_r}]'$ 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 B1 and B3 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:

$$\begin{bmatrix} f \\ r \end{bmatrix} = \begin{bmatrix} I & 0 \\ B & I \end{bmatrix} \begin{bmatrix} \sigma_f . S_f \\ \sigma_r . S_r \end{bmatrix} + \begin{bmatrix} I & 0 \\ B & I \end{bmatrix} \begin{bmatrix} \varepsilon_f \\ \varepsilon_r \end{bmatrix}, \tag{B4}$$

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 B4, 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 prevents us from using standard Gibbs sampling techniques, as done by such authors as Pástor and Stambaugh (1999, 2000), Avramov (2004), and others who did not link the mean and Sharpe ratios (or information ratios).[16] 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.

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:

$$\begin{aligned} S_{r,i} &\sim N(\mu_{S,r}, \sigma_{S,r}^2),\ i = 1, \ldots, N_r; \\ S_{f,j} &\sim N(\mu_{f,r}, \sigma_{S,f}^2),\ j = 1, \ldots, N_f. \end{aligned} \tag{B5}$$

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

$$\Sigma = \begin{bmatrix} \Sigma_f & 0 \\ 0 & \Sigma_r \end{bmatrix} = \tau\Omega\tau, \tag{B6}$$

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:

$$\tau_i \sim \text{Cauchy}\ (\gamma), \text{ with } \tau_i > 0. \tag{B7}$$

As γ increases, the mean of the prior volatility increases. The main benefit of using the half-Cauchy distribution 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:

$$\Omega \sim \text{LKJ}(\eta), \tag{B8}$$

with the LKJ defined as

$$\text{LKJ}(\Sigma \mid \eta) \propto \det(\Sigma)^{\eta-1}.$$

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 toward 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 independently and identically distributed (i.i.d.) normal elementwise:

$$\beta_{i,j} \sim N\left(\mu_{\beta(i,j)}, \sigma^2_{\beta(i,j)}\right). \tag{B9}$$

Generating the Posterior

Let $\theta = [S,B,\Sigma]$ 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

$$p(\theta|\mathcal{Y}) \propto p(\mathcal{Y}|\theta)\pi(\theta). \tag{B10}$$

We use Hoffman and Gelman's (2014) no-U-turn sampler (NUTS) to generate parameter draws from this posterior. NUTS is an implementation of the Hamiltonian Monte Carlo (HMC) method (see Neal 2011). Compared with 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 in computing the posterior because 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 in local minimums of posterior densities.

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 follow 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.

Predictive Moments

Let y^{pred} be the predicted value for observed data; then the posterior predictive distribution is given by integrating out parameter uncertainty from the joint conditional distribution of predicted value and parameters:

$$\begin{aligned} p(y^{pred}|\mathcal{Y}) &= \int p(y^{pred}, \theta \mid \mathcal{Y})d\theta \\ &= \int p(y^{pred}|\theta)p(\theta \mid \mathcal{Y})d\theta. \end{aligned} \tag{B11}$$

The same principle for deriving Equation B11 with uncertainty on only one Sharpe ratio can be used to update the moments of the predictive distribution with uncertainty on all parameters using iterative expectations. Following Equation B11, the mean and variance of the predictive distribution are, respectively,

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

and

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

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, and fund returns) given each parameter draw by the no-U-turn sampler.[17] 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:

$$p\left(\begin{bmatrix} f - c_f \\ r - c_r \end{bmatrix} | \mathcal{Y}\right) = p\left(\begin{bmatrix} f \\ r \end{bmatrix} | \mathcal{Y}\right) - \begin{bmatrix} c_f \\ c_r \end{bmatrix}, \qquad \text{(B12)}$$

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

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 the section "Index and Factors"), we bootstrapped the last 25 years of data ending December 2020. The prior mean IRs were set as the full data averages. In the bootstrap, we drew samples with replacement from the full sample, and we used the standard deviation of the bootstrapped samples as the standard deviations of the IRs.

For empirical Bayes priors for the active fund returns, we used the last 10 years of data ending December 2020. We regressed fund excess returns on the long–short academic factors listed in the section "Alpha-Seeking Funds." From the OLS regressions, we used the intercept and residual risk as our "empirical Bayes" prior means. We defined residual risk as the standard deviation of residuals from the OLS regressions. This procedure yielded a normal 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 residual standard deviations from OLS.

When we allowed the idiosyncratic fund returns to be correlated, the η parameter for the LKJ prior was set to 3 to match the interquartile range of the OLS-based return residual correlations. We assume normal information ratio priors 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 C. Convergence Diagnostics

For our main empirical results, we performed 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 have the same mean as the later sections by computing the following *z*-score-like statistics:

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

where x_{early} is a section in the earliest 10% of the chain and x_{last} is a section in the last 50% of the chain. We partitioned the first 10% and the last 50% into 20 segments and computed 20 Geweke test scores to see whether these segments oscillate between –1 and 1, which indicates good convergence.

R-hat test: The classic Gelman and Rubin (1992) R-hat test compares multiple independent chains.

The idea is that when all chains have converged, the within-chain variance 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 was proposed by Vehtari, Gelman, Simpson, Carpenter, and Burkner (2019). The rule of thumb is that [0.95, 1.05] would be the range indicative of convergence.

Effective samples: This diagnostic from Gelman, Carlin, Stern, Dunson, Vehtari, and Rubin (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 is the number of samples per chain, $\widehat{\rho_p}$ is 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 extended both the length and number of the simulated chains to examine whether our results of interest significantly changed.

Results

For total sample sizes of half a million, we found that most variables have more than 90,000 effective samples. All R-hat statistics are extremely close to 1. For IRs and Sharpe ratios, the Geweke statistics sit comfortably within the [–1, 1] range, which is indicative of convergence. To further enhance our confidence, we increased the total number of sequences to 2 million with 25 chains and 80,000 samples each. These results barely differed from the samples of half a million.

Appendix D. Simulation Exercise

To demonstrate the appropriateness of the estimation technique, we ran 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 ran all simulations using 10 Markov chains with 5,000 samples each, and we treated the first 500 samples as burn-in (not considered in inferences but merely used for mitigating the effects of initial start values).

Data-Generating Process

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

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

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

The priors of information ratios are given by

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

and

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

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

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

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

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

Simulation Settings

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

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

The true Σ (only the upper triangle is shown) is given in the following table:

	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 as follows:

	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 $N(0.4, 10^2)$.

For robustness checks, see the Supplemental Online Material.

Editor's Note

Submitted 18 September 2020

Accepted 23 July 2021 by William N. Goetzmann

Notes

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) applied a regime-switching framework; Carson, Shores, and Nefouse (2017) used all three sources of return in target-date funds; and Bellord, Livnat, Porter, and Tarlie (2019) developed an expected shortfall methodology with a tracking error limit. Similar to our study, Aliaga-Diaz, Renzi-Ricci, Daga, and Ahluwalia (2020) treated index, factors, and active strategies as distinct sources of returns and proposed that the investor has three different risk aversion coefficients with respect to each of the three; they acknowledged the difficulty of calibrating risk aversion coefficients. We used only one risk aversion coefficient, as is commonly used in practice. They also did not consider parameter uncertainty. Corum, Malenko, and Malenko (2020) also considered allocations between index and active funds in a governance setting. None of these papers considered 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.

3. This is an implementation of Hamiltonian Monte Carlo, as developed by Duane, Kennedy, Pendleton, and Roweth (1987) and Neal (1994, 2011). These algorithms are much faster methods of constructing the posterior distribution of parameters than traditional Markov chain Monte Carlo (MCMC) methods, such as Metropolis, Rosenbluth, Rosenbluth, and Teller (1953) or Gibbs sampling (Geman and Geman 1984).

4. Our procedure also generates the whole predictive distribution, which could be used directly in general convex optimization problems (see, for example, Boyd and Vandenberghe 2004) or potentially in nonconvex utility functions, such as disappointment aversion or loss aversion (see, for example, Ang, Bekaert, and Liu 2005; Routledge and Zin 2010).

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 (1991) also derived posterior means and variances while imposing prior beliefs around equilibrium views from the CAPM. However, they did not consider the predictive distribution. Note that the predictive distribution depends on the posterior mean and variance, as well as sampling error. Black and Litterman used only posterior means and variances, without accounting for sampling error.

7. 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.

8. 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.

9. Note that the popular Black and Litterman (1991) procedure stops after deriving the posterior distribution and does not compute the predictive distribution.

10. 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. Equation 6 can be derived by noting that $y = \left[y-(\sigma S-c)\right]+(\sigma S-c)$, with $y-(\sigma S-c) \sim N(0,\sigma^2)$ and $S \sim N(\mu_S,\sigma_S^2)$, which is derived in Equation 3.

11. We found few differences when we allowed parameter uncertainty in the factor loadings compared with the baseline results, which we show in the Online Supplemental Material. One of the reasons is that second moments tend to be estimated more reliably than means (see Merton 1980; Chopra and Ziemba 1993).

12. 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 because ADVI's computation time is faster than Monte Carlo methods, such as NUTS. Note that variational inference methods, such as ADVI, provide only 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 no-U-turn sampler to confirm the approximated results from ADVI.

13. There are several procedures that could be used to obtain more accurate forecasts, such as employing rolling and expanding samples. We deliberately keep the analysis simple in this article.

14. The factor dynamics in Equation A2 relative to the index portfolio are less restrictive than they seem. 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) referred to the long–short portfolio as a "characteristic portfolio."

15. These results will be the same if Equation A3 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) showed 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, factor, and alpha funds will be unchanged.

16. 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 no-U-turn sampler is very fast, as described in the "Generating the Posterior" section, and NUTS also handles incomplete data and data of funds with different sample lengths.

17. 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.

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* 46 (4): 39–51.

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

Ang, A., G. Bekaert, and J. Liu. 2005. "Why Stocks May Disappoint." *Journal of Financial Economics* 76 (3): 471–508.

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., R. H. Hodrick, Y. Xing, and X. Zhang. 2006. "The Cross-Section of Volatility and Expected Returns." *Journal of Finance* 61 (1): 259–99.

Ang, A., A. Madhavan, and A. Sobczyk. 2017. "Estimating Time-Varying Factor Exposures." *Financial Analysts Journal* 73 (4): 41–54.

Asness, C. S. 2004. "An Alternative Future: Part I." *Journal of Portfolio Management* 30 (5): 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 (3): 699–738.

Baks, K. P., A. Metrick, and J. Wachter. 2001. "Should Investors Avoid All Actively Managed Mutual Funds? A Study in Bayesian Performance Evaluation." *Journal of Finance* 56 (1): 45–85.

Banz, R. 1981. "The Relationship between Return and Market Value of Common Stocks." *Journal of Financial Economics* 9 (1): 3–18.

Barry, C. B. 1974. "Analysis under Uncertain Means, Variances, and Covariances." *Journal of Finance* 29 (2): 515–22.

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 (3): 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 (2): 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 (2): 315–42.

Black, F., and R. Litterman. 1991. "Asset Allocation: Combining Investor Views with Market Equilibrium." *Journal of Fixed Income* 1 (2): 7–18.

Boyd, S., and L. Vandenberghe. 2004. *Convex Optimization*. Cambridge, UK: Cambridge University Press.

Brown, S. J. 1979. "The Effect of Estimation Risk on Capital Market Equilibrium." *Journal of Financial and Quantitative Analysis* 14 (2): 215–20.

Carson, B., S. Shores, and N. Nefouse. 2017. "Life-Cycle Investing and Smart Beta Strategies." *Journal of Retirement* 5 (2): 66–82.

Chopra, V. K., and W. T. Ziemba. 1993. "The Effect of Errors in Means, Variances, and Covariances on Optimal Portfolio Choice." *Journal of Portfolio Management* 19 (2): 6–11.

Corum, A. A., A. Malenko, and N. Malenko. 2020. "Corporate Governance in the Presence of Active and Passive Delegated Investment." European Corporate Governance Institute Finance Working Paper 695/2020 (31 August).

Duane, A., A. D. Kennedy, B. J. Pendleton, and D. Roweth. 1987. "Hybrid Monte Carlo." *Physics Letters B* 195 (2): 216–22.

Fama, E. F., and K. R. French. 1993. "Common Risk Factors in the Returns on Stocks and Bonds." *Journal of Financial Economics* 33 (1): 3–56.

——. 2018. "Long-Horizon Returns." *Review of Asset Pricing Studies* 8 (2): 232–52.

Frazzini, A., and L. H. Pedersen. 2014. "Betting against Beta." *Journal of Financial Economics* 111 (1): 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 (1): 41–81.

Gelman, A. 2006. "Prior Distributions for Variance Parameters in Hierarchical Models." *Bayesian Analysis* 1 (3): 515–33.

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.

Gelman, A., and D. B. Rubin. 1992. "Inference from Iterative Simulation Using Multiple Sequences." *Statistical Science* 7 (4): 457–511.

Geman, S., and D. Geman. 1984. "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images." *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-6 (6): 721–41.

Geweke, J. 1992. "Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments." In *Bayesian Statistics 4*, edited by J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, 169–93. Oxford, UK: Oxford University Press.

Grinold, R. C. 1994. "Alpha Is Volatility Times IC Times Score." *Journal of Portfolio Management* 20 (4): 9–16.

Grinold, R. C., and R. N. Kahn. 2000. "The Efficiency Gains of Long–Short Investing." *Financial Analysts Journal* 56 (6): 40–53.

Hansen, L. P., and R. Jagannathan. 1991. "Implications of Security Market Data for Models of Dynamic Economies." *Journal of Political Economy* 99 (2): 225–62.

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: 1593–623.

Homescu, C. 2015. "Better Investing through Factors, Regimes and Sensitivity Analysis." Working paper (25 January). Available at 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* 48 (1): 65–91.

Joe, H. 2006. “Generating Random Correlation Matrices Based on Partial Correlations.” *Journal of Multivariate Analysis* 97 (10): 2177–89.

Kong, A., J. S. Liu, and W. H. Wong. 1994. “Sequential Imputations and Bayesian Missing Data Problems.” *Journal of the American Statistical Association* 89 (425): 278–88.

Lewandowski, D., D. Kurowicka, and H. Joe. 2009. “Generating Random Correlation Matrices Based on Vines and Extended Onion Method.” *Journal of Multivariate Analysis* 100 (9): 1989–2001.

Lintner, J. 1975. “Inflation and Security Returns.” *Journal of Finance* 30 (2): 259–80.

Maccheroni, F., M. Marinacci, and A. Rustichini. 2006. “Ambiguity Aversion, Robustness, and the Variational Representation of Preferences.” *Econometrica* 74 (6): 1447–498.

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 (5): 90–103.

Markowitz, H. 1952. “Portfolio Selection.” *Journal of Finance* 7 (1): 77–91.

Merton, R. C. 1980. “On Estimating the Expected Return of the Market: An Exploratory Investigation.” *Journal of Financial Economics* 29: 449–70.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, and A. H. Teller. 1953. “Equation of State Calculations by Fast Computing Machines.” *Journal of Chemical Physics* 21 (6): 1087–92.

Neal, R. M. 1994. “An Improved Acceptance Procedure for the Hybrid Monte Carlo Algorithm.” *Journal of Computational Physics* 111 (1): 194–203.

——. 2011. “MCMC Using Hamiltonian Dynamics.” In *Handbook of Markov Chain Monte Carlo*, edited by S. Brooks, A. Gelman, G. L. Jones, and X. L. Meng, 113–61. Boca Raton, FL: Chapman & Hall/CRC.

Pástor, Ľ., and R. F. Stambaugh. 1999. “Costs of Equity Capital and Model Mispricing.” *Journal of Finance* 54 (1): 67–121.

——. 2000. “Comparing Asset Pricing Models: An Investment Perspective.” *Journal of Financial Economics* 56 (3): 335–81.

Pedersen, L. H., A. Babu, and A. Levine. 2021. “Enhanced Portfolio Optimization.” *Financial Analysts Journal* 77 (2): 124–51.

Polson, N. G., and J. G. Scott. 2012. “On the Half-Cauchy Prior for a Global Scale Parameter.” *Bayesian Analysis* 7 (4): 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 (2): 129–76.

Ross, S. A. 1976. “The Arbitrage Theory of Capital Asset Pricing.” *Journal of Economic Theory* 13 (3): 341–60.

Routledge, B. R., and S. E. Zin. 2010. “Generalized Disappointment Aversion and Asset Prices.” *Journal of Finance* 65 (4): 1303–32.

Rouwenhorst, K. G. 2005. “The Origins of Mutual Funds.” In *The Origins of Value: The Financial Innovations That Created Modern Capital Markets*, edited by W. N. Goetzmann and K. G. Rouwenhorst, 249–70. Oxford, UK: Oxford University Press.

Sharpe, W. F. 1964. “Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk.” *Journal of Finance* 19 (3): 425–42.

Sloan, R. G. 1996. “Do Stock Prices Fully Reflect Information in Accruals and Cash Flows about Future Earnings?” *Accounting Review* 71 (3): 289–315.

Tanner, M. A., and W. H. Wong. 1987. “The Calculation of Posterior Distributions by Data Augmentation.” *Journal of the American Statistical Association* 82 (398): 528–40.

Treynor, J. L., and F. Black. 1973. “How to Use Security Analysis to Improve Portfolio Selection.” *Journal of Business* 46 (1): 66–88.

Tu, J., and G. Zhou. 2004. “Data-Generating Process Uncertainty: What Difference Does It Make in Portfolio Decisions?” *Journal of Financial Economics* 72 (2): 385–421.

Vehtari, A., A. Gelman, D. Simpson, B. Carpenter, and P.-C. Burkner. 2019. “Rank-Normalization, Folding, and Localization: An Improved $\hat{R}$ for Assessing Convergence of MCMC (with Discussion).” *Bayesian Analysis* 16 (2): 667–718.