View Fusion Vis-à-Vis a Bayesian Interpretation of Black–Litterman for Portfolio Allocation

Trent Spears, Stefan Zohren, and Stephen Roberts

Trent Spears
is a doctoral student with the Machine Learning Research Group and the Oxford-Man Institute of Quantitative Finance at the University of Oxford in Oxford, UK.
trent@robots.ox.ac.uk

Stefan Zohren
is an associate professor (research) with the Machine Learning Research Group and the Oxford-Man Institute of Quantitative Finance at the University of Oxford in Oxford, UK.
zohren@robots.ox.ac.uk

Stephen Roberts
is the RAEng Professor of Machine Learning and the former director of the Oxford-Man Institute of Quantitative Finance at the University of Oxford in Oxford, UK.
sjrob@robots.ox.ac.uk

KEY FINDINGS

- The authors consider a Bayesian interpretation of the Black-Litterman model, and estimate views and view uncertainties with machine learning models.
- In some cases, traders or investors may hold multiple correlated views, including uncertainties, for the same underlying assets over the same time period. The authors extend the Black–Litterman model using techniques from the information fusion literature, aiming to combine a collection of views into an improved, single predictor.
- The authors present empirical work subsuming the arbitrage pricing theory that also accounts for transaction costs. Results calculated over a 30-year period present compelling supporting evidence for view fusion for portfolio strategy.

ABSTRACT

The Black–Litterman model extends the framework of the Markowitz modern portfolio theory to incorporate investor views. The authors consider a case in which multiple view estimates, including uncertainties, are given for the same underlying subset of assets at a point in time. This motivates their consideration of data fusion techniques for combining information from multiple sources. In particular, they consider consistency-based methods that yield fused view and uncertainty pairs; such methods are not common to the quantitative finance literature. They show a relevant, modern case of incorporating machine learning model-derived view and uncertainty estimates, and the impact on portfolio allocation, with an example subsuming arbitrage pricing theory. Hence, they show the value of the Black–Litterman model in combination with information fusion and artificial intelligence–grounded prediction methods.

The Black–Litterman model extends the Markowitz portfolio optimization framework to incorporate investor beliefs about asset returns (Black and Litterman 1991). Such beliefs are termed *views*. In its original formulation, the Black–Litterman framework is such that a given view induces an update of a prior assumption about the mean parameter of a Gaussian return distribution. Making use of views in this way can result in nontrivial differences to the estimated weights of a portfolio allocation strategy, and hence the relative investment outcome.

Views can be obtained from diverse sources, including market pundit opinions, Reserve Bank statements, and as output of statistical models. Consequently, in practice, a portfolio manager or trader may possess multiple views about the same underlying asset, or subset of assets, over the next investment period. A contribution of this work is to both consider and extend the Black–Litterman framework to incorporate multiple such views.

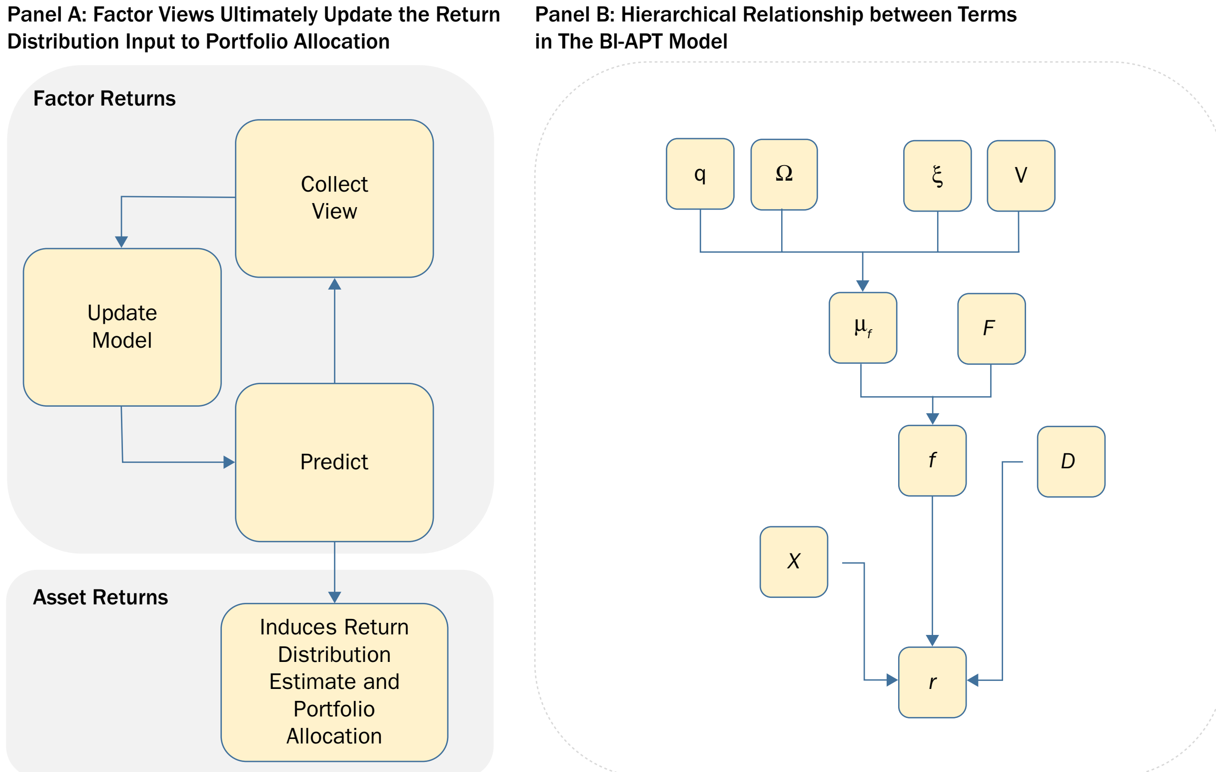
It is natural to formalize elements of the Black–Litterman model with respect to ideas from Bayesian statistics. See Walters (2014) for a review summary, with (Satchell and Scowcroft 2000; Kolm and Ritter 2017; Kolm, Ritter, and Simonian 2021) intriguing supplements. We note that the modeling of this article presupposes the canonical Bayes-based Black–Litterman formalism. This approach is useful for practitioners in allowing for a strong theoretical grounding of the model prior in the capital asset pricing model (CAPM) (Sharpe 1964) and for the relative ease with which one can incorporate market views. Specifically, views are given as a function of asset mean return estimates and accompanied by an uncertainty estimate that encapsulates a notion of confidence about that mean. The specification of the view uncertainty is a domain for debate in the literature. In contrast, the uncertainty specification is of critical importance to the mechanics of the Bayesian estimation framework. The literature often tacitly assumes that uncertainty is to be specified in excess of the given mean estimate. A prevailing method is to set the uncertainty term in proportion to a quantity such as an estimate of the underlying return covariance. The proportionality constant can then be tuned akin to a model hyperparameter. Such ad hoc specification can be unsatisfying; fortunately, scientific modeling offers an alternative.

Machine learning methods are enjoying a resurgence as oracles for financial markets, particularly given the advances underlying modern deep learning. Indeed, such methods have utility for forecasting time series, and predictions are often naturally partnered with uncertainty estimates. Such uncertainties are typically categorized in the literature as *aleatoric* or *epistemic*. Aleatoric uncertainty relates to the modeled random error of the assumed underlying data generating process, while epistemic uncertainty encompasses error relating to the model specification, which also subsumes uncertainties about parameter estimates. These two categories of uncertainty are recognized across multiple research domains and the models contained therein. In any case, the estimation of views by statistical forecasting models, coupled with the uncertainty estimate, allows for a more principled approach to setting view parameters when modelling with Black–Litterman. We describe this further throughout this work. This interpretation stands in stark contrast to existing methods—with one popular technique described earlier—that are more ad hoc. Our approach has the further benefits of preserving the intent of the canonical Black–Litterman model and has utility for automating an aspect of the investment decision-making process.

Further, our extended modelling framework, we consider combining multiple views for expected returns over the next time increment with respect to view correlation that may be nonzero. Although we consider the single-step case, we are inspired by the relationship between Black–Litterman and a state space modeling framework as recently described in (van der Schans and Steehouwer 2017). Indeed, at the core of our approach are techniques from the information fusion literature, that allow us to combine views in a principled way. In this work, we present methods for fusing views assuming both zero and nonzero correlation between them and consider a conservative fusion method for sets of views that may be inconsistent with one another (according to some distance metric between distributions). In their relative simplicity, the principles of these fusion methods are understandable, and the resulting fused estimate is explainable. Some practitioners may find this advantageous relative to alternative black box methods. In our empirical work, we find fusion-based methods

EXHIBIT 1

Schematic of the Black–Litterman Methodology as It Applies to the APT Framework



NOTES: In Panel A, asset returns are linearly projected onto factor returns, and the algorithm at the level of the factor model is such that at each time-step views are collected before the factor returns model is updated and predictions gleaned. Panel B shows the hierarchical model suggestive of the relationship between terms in the BL-APT model.

that improve on of the mean and median of our three view-generating models in a statistically meaningful way. This is interesting to the investment practitioner that cannot know ex ante which of a set of views might outperform; a fusion-based method may be selected instead.

To show the utility of data fusion and in connection with recent work by Kolm and Ritter (2017), we offer an application of the Black–Litterman model assuming a multifactor model for asset returns. Factor models present in finance with various specifications; see Giglio, Kelly, and Xiu (2022) for a recent review. We consider a conditional factor model, as inspired by the arbitrage pricing theory (APT) of Ross (1976). In the next section, before we further review view fusion, we expound on the Black–Litterman APT as a Bayesian hierarchical model; a schematic for the model is given in Exhibit 1. We carefully derive the solution for the optimal portfolio weights under an unconstrained Markowitz portfolio allocation as a further contribution to the literature. Extending our empirical work, we offer a preliminary model accounting for transaction costs, with details carefully specified. This model permits a reasonable proof of concept that could be built upon for industrial application or as future academic work.

Notes for the dataset construction are offered next. To the extent the raw data are available in the public domain (via the Wharton Research Data Service), these notes could lead to a straightforward replication. We then use the ideas and results of the preceding sections to give empirical results that showcase view fusion whereby estimates—including measures of uncertainty—are sourced from a collection of machine learning models. We discuss the ways in which fusion-based methods lead to investment outperformance relative to methods based on a singular view. We also contrast our methods with the difficult benchmark of a buy-and-hold investment in the S&P 500. We conclude and summarize avenues for future work in the final section.

BLACK–LITTERMAN

A Recap

Assume there exist n investable assets within a market with excess return vector $\boldsymbol{r}$ whose distribution is modeled as multivariate normal over discrete time increments, denoted $\boldsymbol{r} \sim \text{MVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Here, $\boldsymbol{\mu}$ denotes the mean vector of returns, and $\boldsymbol{\Sigma}$ denotes the return covariance. Then in a Markowitz modern portfolio theory framework (Markowitz 1952), assuming a single-period investment setting, the investor portfolio weights $\boldsymbol{w} = (w_1, \ldots, w_n)$ are a solution to the mean–variance optimization problem:

$$\max_{w \in \mathcal{C}} \mathbb{E}(\boldsymbol{w}'\boldsymbol{r}) - \frac{\gamma}{2}\text{var}(\boldsymbol{w}'\boldsymbol{r}) \quad (1)$$

Here, $\mathcal{C}$ is a set of weight constraints, if any, and $\gamma > 0$ is the investor risk aversion parameter, denoting preferences balancing return and risk. In the unconstrained case, differentiating Equation 1, setting it to 0 and solving for the optimal portfolio weights $\boldsymbol{w}^*$ in terms of $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ yields

$$\boldsymbol{w}^* = (\gamma \Sigma)^{-1} \mu \quad (2)$$

On analyzing the second derivative one finds that this solution is indeed a maximum. In practice, these weights could be solved for—and a portfolio subsequently implemented—upon estimation of the return distribution parameters, for example, upon substitution of a computed sample mean vector and sample covariance matrix, given historical data.

The Black–Litterman model (Black and Litterman 1991) extends this framework to incorporate (subjective) investor beliefs about asset returns, that are called views. A single view is assumed to be a function of real-valued vectors of length n denoting portfolio weights $\boldsymbol{p}$ and expected asset returns $\boldsymbol{\mu}$, a real-valued scalar denoting the portfolio value q, and a strictly positive real-valued scalar ϵ denoting uncertainty in the portfolio value, that can be written

$$\boldsymbol{p}'\mu = q + \epsilon \quad (3)$$

To be explicit, let us call a collection of k views a range, for natural numbers k. Then similar to the original presentation of Black and Litterman, a range of views can be arranged thus:

$$\boldsymbol{P}\mu = \boldsymbol{q} + \epsilon, \qquad \epsilon \sim N(\mathbf{0}, \Omega) \quad (4)$$

Here, $\boldsymbol{P} \in \mathbb{R}^{k \times n}$ is a real-valued matrix of portfolio weights, where the kth row is given by $\boldsymbol{p}_k'$ as defined earlier; $\boldsymbol{q} \in \mathbb{R}^k$ is a real-valued vector collecting k portfolio values; and $\boldsymbol{\Omega} \in \mathbb{R}^{k \times k}$ is assumed to be a diagonal matrix of portfolio value estimate uncertainties, where we have further assumed independence between views.

Given the above formulation, a view can express both absolute and relative relationships between mean asset returns. The case of an absolute (normalized) view corresponds to $\boldsymbol{p}$ taking a unit value at the ith entry, and zero otherwise, so that the left-hand side of Equation 3 equals μ_i. Any other (nontrivial) view we call relative; a common relative view, for example, is a belief on the return spread $\mu_i - \mu_j = q$. This could be expressed by setting alternate positive and negative unit values at the i and

j index of $\boldsymbol{p}$, and zero elsewhere. In the work that follows, we assume absolute views. We further assume that $\boldsymbol{P}$ is of full rank.

Given these preliminaries, the Black–Litterman model is such that the existing return distribution parameter assumption is updated given views. We understand such updating in the context of Bayesian statistics, as prior authors have noted (Satchell and Scowcroft 2000; Walters 2014; Kolm and Ritter 2017; van der Schans and Steehouwer 2017).

The Bayesian specification assumes a prior distribution $p(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ denotes the target parameter of the return distribution that we wish to express via a parametric probabilistic model. Most commonly, $\boldsymbol{\theta} = \boldsymbol{\mu}$. The original authors–and many since–set the prior given an estimate of the CAPM model (Sharpe 1964) and hence underwriting the prior with a foundational theoretical result.

Next, the view is related to the target parameter via a likelihood function $p(\boldsymbol{q}|\boldsymbol{\theta})$ such that, by the laws of conditional probability and Bayes' theorem, an updated posterior distribution for the target parameter is given by

$$p(\boldsymbol{\theta}|\boldsymbol{q}) \propto p(\boldsymbol{q}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \tag{5}$$

Finally, we can write a posterior predictive distribution for the asset returns given the views as

$$p(\boldsymbol{r}|\boldsymbol{q}) = \int p(\boldsymbol{r}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{q})d\boldsymbol{\theta} \tag{6}$$

Hence, the expectation and covariance of the random quantity $\boldsymbol{r}|\boldsymbol{q}$ can be computed analytically—or otherwise estimated if the integral is intractable—before substitution into, say, Equation 2 to yield a solution to the portfolio allocation problem.

Black–Litterman in the Context of Factor Models

In line with recent work by Kolm and Ritter (2017), we offer an application of the Black–Litterman model assuming a factor model for asset returns, and we conveniently conform with some of their notation.

For context, we note that in many investment studies of equity markets the empirical estimation of the return covariance is challenging, especially when the universe of equities under consideration is large. The linear factor model that we outline below, whereby $k << n$, is relatively practical and efficient. This has contributed to its popularity within the literature, and bolsters the case for industrial use (Fabozzi, Focardi, and Kolm 2006). We consider a conditional factor model, as inspired by the APT of Ross (1976). In the single-period investment setting, we model

$$\boldsymbol{r} = \boldsymbol{X}\boldsymbol{f} + \epsilon_r, \quad \epsilon_r \sim N(\boldsymbol{0}, \boldsymbol{D}) \tag{7}$$

Here, $\boldsymbol{r}$ is the n-dimensional random vector containing the cross section of excess returns for n equities over the next time increment $[t, t+1]$; $\boldsymbol{X}$ is a nonrandom $n \times k$ matrix of observable factor exposures estimated at t, for k the number of explanatory asset factors; $\boldsymbol{D}$ is a covariance matrix for the returns, also known at t and assumed to be diagonal; and $\boldsymbol{f}$ is a k-dimensional latent factor vector that we estimate at each time step by cross-sectional regression. Within a Black–Litterman framework, consider that our parameter of interest is now $\boldsymbol{\theta} = \boldsymbol{\mu}_f$ such that

$$\boldsymbol{f} = \boldsymbol{\mu}_f + \epsilon_f, \quad \epsilon_f \sim N(\boldsymbol{0}, \boldsymbol{F}) \tag{8}$$

We refer to the elements of $\boldsymbol{\mu}_f$ as the factor risk premiums.

Black–Litterman in the context of APT, henceforth labelled BL-APT, amounts to a Bayesian hierarchical model. A schematic for the model is given in Panel B of Exhibit 1. We note that in comparison to the Black–Litterman model applied to the return mean, an analogue to the CAPM prior does not exist. In the work that follows, we set the prior parameters as a simple function of recent data, per the discussion in the Results section. On the other hand, view estimates for the factor risk premiums at time *t* induce return forecasts (with uncertainties) over the next time period. These parameter estimates can then be utilized to estimate portfolio weights.

In the next section, we carefully show the derivation for the optimal portfolio weights of an unconstrained BL-APT as a reference contribution to the literature.

Solving for Optimal Weights: Hierarchical Bayesian Black–Litterman for APT

We can express our model for asset returns, factor returns, and views carefully as a hierarchical Bayesian model, with a suggestive schematic shown in Exhibit 1. We derive the optimal portfolio allocation after finding $p(\boldsymbol{r}|\boldsymbol{q})$, as defined in the following.

We begin by writing the joint distribution of random variables of interest as a conditional probability:

$$p(\boldsymbol{r},\boldsymbol{f},\theta,\boldsymbol{q}) = p(\boldsymbol{r},\boldsymbol{f}\,|\,\theta,\ \boldsymbol{q})p(\theta,\boldsymbol{q})$$

Substituting $p(\boldsymbol{\theta},\ \boldsymbol{q}) = p(\boldsymbol{\theta}|\boldsymbol{q})p(\boldsymbol{q})$ and dividing both sides by $p(\boldsymbol{q})$ yields

$$p(\boldsymbol{r},\ \boldsymbol{f},\theta\,|\,\boldsymbol{q}) = p(\boldsymbol{r},\boldsymbol{f}\,|\,\theta,\boldsymbol{q})\ p(\theta\,|\,\boldsymbol{q})$$

Integrating both sides over $\boldsymbol{\theta}$, we find that

$$p(\boldsymbol{r},\boldsymbol{f}\,|\,\boldsymbol{q}) = \int \underbrace{p(\boldsymbol{r},\boldsymbol{f}\,|\,\theta,q)}_{=p(\boldsymbol{r},\boldsymbol{f}|\,\theta)} p(\theta\,|\,\boldsymbol{q})d\theta$$

Again, integrating both sides, this time with respect to $\boldsymbol{f}$, we have that

$$p(\boldsymbol{r}\,|\,\boldsymbol{q}) = \int\int p(\boldsymbol{r},\boldsymbol{f}\,|\,\theta)p(\theta\,|\,\boldsymbol{q})d\theta d\boldsymbol{f} \tag{9}$$

With one more simplification, writing $p(\boldsymbol{r},\ \boldsymbol{f}|\boldsymbol{\theta}) = p(\boldsymbol{r}|\boldsymbol{f},\ \boldsymbol{\theta})p(\boldsymbol{f}|\boldsymbol{\theta}) = p(\boldsymbol{r}|\boldsymbol{f})p(\boldsymbol{f}|\boldsymbol{\theta})$, we find

$$p(\boldsymbol{r}\,|\,\boldsymbol{q}) = \int p(\boldsymbol{r}\,|\,\boldsymbol{f}) \underbrace{\int p(\boldsymbol{f}\,|\,\theta)p(\theta\,|\,\boldsymbol{q})d\theta}_{=p(\boldsymbol{f}|\boldsymbol{q})} d\boldsymbol{f} \tag{10}$$

Because we assume $\boldsymbol{\theta} = \boldsymbol{\mu}_f$, there holds the well-known facts that the posterior $p(\boldsymbol{\theta}|\boldsymbol{q})$ is normal and that the posterior predictive $p(\boldsymbol{f}|\boldsymbol{q})$ is normal. Hence, it is clear that $p(\boldsymbol{r}|\boldsymbol{q})$ is also normal, with its expectation and covariance given in closed form by the following three-step process.

First, choosing the prior $p(\boldsymbol{\theta}) \sim N(\boldsymbol{\xi},\ \boldsymbol{V})$, by Bayes' rule for linear Gaussian systems, as in Murphy (2012), we have that the posterior $p(\boldsymbol{\theta}|\boldsymbol{q})$ is multivariate normal with covariance and mean

$$\text{var}(\theta\,|\,\boldsymbol{q}) = (\boldsymbol{V}^{-1} + \boldsymbol{\Omega}^{-1})^{-1},$$
$$\mathbb{E}(\theta\,|\,\boldsymbol{q}) = \text{var}(\theta\,|\,\boldsymbol{q})\cdot(\boldsymbol{V}^{-1}\xi + \boldsymbol{\Omega}^{-1}\boldsymbol{q})$$

Second, recall the convolution integral for sums of random variables: If, $\boldsymbol{X}_i \overset{i.i.d.}{\sim} N(\mu_i, \Sigma_i), i = 1,2,$ then

$$f(\boldsymbol{z}) := \int f_{X_1}(\boldsymbol{x}) f_{X_2}(\boldsymbol{z}-\boldsymbol{x}) d\boldsymbol{x} \sim N(\mu_1 + \mu_2, \Sigma_1 + \Sigma_2) \tag{11}$$

Writing the inner integral of Equation 10 with the change of variable $\boldsymbol{f} \mapsto \boldsymbol{f} - \boldsymbol{\theta}$, we find

$$p(\boldsymbol{f}|\boldsymbol{q}) = \int \underbrace{p(\boldsymbol{f}-\theta|\theta)}_{\sim N(\boldsymbol{0},\boldsymbol{F})} p(\theta|\boldsymbol{q}) d\theta$$

so that by Equation 11 we have that $p(\boldsymbol{f}|\boldsymbol{q})$ is normally distributed with expectation and covariance given by

$$\mathbb{E}(\boldsymbol{f}|\mathbf{q}) = (\mathbf{V}^{-1} + \Omega^{-1})^{-1} \cdot (\boldsymbol{V}^{-1}\xi + \Omega^{-1}\boldsymbol{q}) = \mathbb{E}(\theta|\mathbf{q}),$$
$$\text{var}(\boldsymbol{f}|\mathbf{q}) = (\mathbf{V}^{-1} + \Omega^{-1})^{-1} + \boldsymbol{F} = \text{var}(\theta|\boldsymbol{q}) + \boldsymbol{F}$$

Third, we solve the outer integral of Equation 10 (deferring some of the details of the algebraic manipulations to the appendix, for brevity). Writing

$$p(\boldsymbol{r}|\boldsymbol{q}) = \int p(\boldsymbol{r}|\boldsymbol{f}) p(\boldsymbol{f}|\boldsymbol{q}) d\boldsymbol{f} \tag{12}$$

we collect and expand the terms in the exponent up to a factor of $-\frac{1}{2}$, complete the square in $\boldsymbol{f}$, and integrate what is recognizable as a Gaussian integral. By the recursive nature of the integrals, it is clear that we also retrieve a Gaussian for $\boldsymbol{r}|\boldsymbol{q}$. Hence, as supported by further calculations within the appendix, we have that the posterior predictive distribution $p(\boldsymbol{r}|\boldsymbol{q})$ is multivariate normal with covariance and mean

$$\text{var}(\boldsymbol{r}|\boldsymbol{q}) = [\boldsymbol{D}^{-1} - \boldsymbol{D}^{-1}\boldsymbol{X}[\boldsymbol{X}^T\boldsymbol{D}^{-1}\boldsymbol{X} + \text{var}(\boldsymbol{f}|\boldsymbol{q})^{-1}]^{-1}\boldsymbol{X}^T\boldsymbol{D}^{-1}]^{-1},$$
$$\mathbb{E}(\boldsymbol{r}|\boldsymbol{q}) = \text{var}(\boldsymbol{r}|\boldsymbol{q})\boldsymbol{D}^{-1}\boldsymbol{X}[\boldsymbol{X}^T\boldsymbol{D}^{-1}\boldsymbol{X} + \text{var}(\boldsymbol{f}|\boldsymbol{q})^{-1}]^{-1}\text{var}(\boldsymbol{f}|\boldsymbol{q})^{-1}\mathbb{E}(\boldsymbol{f}|\boldsymbol{q})$$

By Equation 2, we write the optimal weights for a single-step Black–Litterman portfolio allocation as

$$\begin{aligned} \boldsymbol{h}^*_{blb} &= \gamma^{-1}\text{var}(\boldsymbol{r}|\boldsymbol{q})^{-1}\mathbb{E}(\boldsymbol{r}|\boldsymbol{q}) \\ &= \gamma^{-1}\boldsymbol{D}^{-1}\boldsymbol{X}[\boldsymbol{X}^T\boldsymbol{D}^{-1}\boldsymbol{X} + \text{var}(\boldsymbol{f}|\boldsymbol{q}^{-1}]^{-1}\text{var}(\boldsymbol{f}|\boldsymbol{q})^{-1}\mathbb{E}(\boldsymbol{f}|\boldsymbol{q}) \end{aligned} \tag{13}$$

This expression differs meaningfully with other optimal weights given in the literature. This is the case in that we have carefully accounted for all of the specified terms within the hierarchy of the returns model.

An Optimal Weight Model under Transaction Costs

In this section, we consider a more realistic objective function for solving for portfolio weights, in that it incorporates transaction costs. We do this by making a minimal adjustment to the Markowitz objective of Equation 1:

$$\max_{\boldsymbol{w}\in\mathcal{C}} \quad \mathbb{E}(\boldsymbol{w}'\boldsymbol{r}) - \frac{\gamma}{2}\text{var}(\boldsymbol{w}'\boldsymbol{r}) - TC \tag{14}$$

with TC denoting transaction costs measured in percentage of invested wealth. Of the many ways to account for transaction costs, we broadly follow the single-period model recently presented in Jensen et al. (2022). Despite the differences in our application domains, our analysis and results contrast somewhat consistently, and certainly interestingly, with theirs.

First, we choose a transaction cost model inspired by Gârleanu and Pedersen (2013) so that transaction costs measured in dollars, TC^d, is given by

$$TC_t^d = \frac{1}{2}\, \Delta_t' \Lambda_t \Delta_t$$

Here, $\boldsymbol{\Delta}_t$ is the portfolio turnover vector for a time increment, defined as

$$\Delta_t \coloneqq \Pi_t(\boldsymbol{w}_t - \boldsymbol{R}_{t-1}\boldsymbol{w}_{t-1})$$

and the market impact vector $\boldsymbol{m}$ is given by

$$\boldsymbol{m}_t = \frac{1}{2}\Lambda_t \Delta_t$$

We use the subscript t to explicitly account for time dependence. The real scalar Π_t denotes total wealth investable at time t. Clearly, given this specification dollar transaction costs erode dollar returns quadratically with increasing wealth. The term $\boldsymbol{R}_{t-1}$ denotes a diagonal matrix whose diagonal elements account for the realized returns for the assets held in the portfolio at $t - 1$, over the time step $[t - 1, t)$, and adjusted for wealth delta. We assume that the term $\boldsymbol{\Lambda}_t$ is a diagonal matrix whose elements depend on the daily (dollar) volume traded in each underlying asset. Given the recent estimates of Frazzini, Israel, and Moskowitz (2018), we assume a 10-basis-point market impact when trading 1% of the daily dollar volume $\boldsymbol{L}_t$ of US equity underlyings whether long or short, so

$$\begin{aligned} \boldsymbol{m}_t &= 10 \text{ bps} \cdot \mathbf{1} = \frac{1}{2}\Lambda_t \times 1\%\boldsymbol{L}_t \\ \Rightarrow \Lambda_t &= \frac{1}{5}\boldsymbol{L}_t^{-1} \end{aligned} \tag{15}$$

In the work that follows, we set the values for daily dollar volumes at their six-month rolling average. Further, we assume that investor wealth grows proportionally with the market; specifically, we make the somewhat arbitrary assumption that at each time step the total investable capital is 1/10 of the sum of the daily dollar volumes of the assets in our trade universe, as known at the most recent time step.

Hence, Equation 14 can be written, noting that $TC = TC^d/\Pi$, as

$$\max_{\boldsymbol{w}_t \in \mathcal{C}} \mathbb{E}(\boldsymbol{w}_t'\boldsymbol{r}_t) - \frac{\gamma}{2}\text{var}(\boldsymbol{w}_t'\boldsymbol{r}_t) - \frac{\Pi_t}{2}(\boldsymbol{w}_t - \boldsymbol{R}_{t-1}\ \boldsymbol{w}_{t-1}^*)'\Lambda_t(\boldsymbol{w}_t - \boldsymbol{R}_{t-1}\boldsymbol{w}_{t-1}^*) \tag{16}$$

and conveniently solved in closed form for optimal weights:

$$\boldsymbol{w}_t^* = (\gamma\Sigma_t + \Pi_t\Lambda_t)^{-1}(\mu_t + \Pi_t\Lambda_t\boldsymbol{R}_{t-1}\boldsymbol{w}_{t-1}^*) \tag{17}$$

Hence, the optimal portfolio weights for the BL-APT model can be found by substituting estimates for $\mathbb{E}(\boldsymbol{r}|\boldsymbol{q})$ and $\mathbb{E}(\boldsymbol{r}|\boldsymbol{q})$ into Equation 17, as an alternative to Equation 13 when taking transaction costs into account.

VIEW FUSION

Preliminaries

The Black–Litterman model—sans solving for the portfolio weights—can be expressed via a state space representation; this was recently presented in van der Schans and Steehouwer (2017) in the context of modeling temporal dependency in asset returns and parameter view estimates.[1] In a simple case, consider ranges of views generated from a source at discrete time increments $k = 0, 1, \ldots$. We can infer the underlying target mean m_k in an online way, via the linear time-varying system of equations:

$$\boldsymbol{\mu}_{k+1} = \boldsymbol{\mu}_k + \epsilon_k^{\mu} \tag{18}$$

$$\boldsymbol{q}_k = \boldsymbol{P}_k \boldsymbol{\mu}_k + \epsilon_k^{q} \tag{19}$$

Here, ϵ_k^{μ} and ϵ_k^{q} model independent zero-mean white-noise processes with respective covariances $\boldsymbol{\Psi}_k$ and $\boldsymbol{\Sigma}_k$. We call Equations 18 and 19 the state and measurement equations, respectively. Solving the state space representation of Black–Litterman, assuming absolute views, we have the probability distribution for $\boldsymbol{\mu}_k | \boldsymbol{q}_k$, analogous to Equation 5 and solved equivalently. The predictive distribution for $\boldsymbol{\mu}_{k+1} | \boldsymbol{q}_k$, analogous to Equation 6 can also be solved equivalently. On the other hand, within the Black–Litterman literature, per the model we described earlier, the prior is typically reinitialized to $p(\boldsymbol{\mu}_0)$ at each time step, rather than dynamically updated given a sequence of view estimates. In the following work, we follow this approach and leave consideration of prior updates based on a larger subset of recent data as future work. In any case, such state space modeling naturally inspires an extension of the Black–Litterman framework to handle the combination—or fusion—of multiple views held about a particular portfolio.

Information Fusion

To this point, we have described the Black–Litterman model assuming a singular source for a range of views. Next, we consider the realistic—though in the context of Black–Litterman modeling, unexplored—case of multiple sources generating views simultaneously for each time increment and analyze the impact on the predictive distribution. We extend Equation 19 thus (for brevity, keeping to the case of absolute views):

$$\boldsymbol{q}_{k,s} = \boldsymbol{\mu}_{k,s} + \epsilon_{k,s}^{(q)} \tag{20}$$

where $\boldsymbol{q}_{k,s}$ denotes the view is from source s, $1 \leq s \leq S$, and each $\epsilon_{k,s}^{(q)}$ is a zero-mean white-noise process with covariance $\boldsymbol{\Sigma}_{k,s}$. Multiple views could realistically arise in the context of implementing a systematic trading or asset management strategy, whereby multiple predictions across multiple (potentially, broadly diverse) models need to be fused before capital allocation.

Information fusion has been common to the state space literature since the work of Willner, Chang, and Dunn (1976). However, fusion is not limited to the state space framework and instead applies more generally to the problem of fusing multiple estimates of an underlying random variable.[2] Correspondingly, data fusion can take

[1] There exists extensive literature on state space modeling, with applications across many domains. See, for example, Date and Ponomareva (2010) for an instructive review of filtering methods for mathematical finance.

[2] Various approaches exist in the data fusion literature for incorporating information from multiple sources; introductory reviews can be found in Khaleghi et al. (2013) and Castanedo (2013).

place at the level of the state or measurement equations, though we assume the former for our empirical work in the following.

Typically, fusion techniques amount to solving for an optimal fused estimate $(\hat{\mu}, \hat{\Sigma})$ for the true underlying values $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ given a collection of S underlying estimates $\{(\hat{\mu}_s, \hat{\Sigma}_s) : 1 \leq s \leq S\}$, with the mean parameter a weighted linear combination of the underlying estimates:

$$\hat{\mu} = \sum_s \boldsymbol{W}_s \ \hat{\mu}_s. \tag{21}$$

(We have omitted the subscript k, for ease of notation.) Optimal fusion is such that (i) $\hat{\Sigma}$ is minimal in some sense—typically with respect to matrix trace or determinant—and that (ii) $\hat{\Sigma}$ is consistent. Within the fusion literature, a consistent estimate is such that $\hat{\Sigma} \geq \Sigma$, in the sense that $\hat{\Sigma} - \Sigma$ is positive semidefinite. That is, the consistent estimate subsumes a covariance no less than the covariance $\boldsymbol{\Sigma}$ of the distribution of the true underlying process; see, for example, Uhlmann (1995). We note that this definition is independent of the well-known definition of a consistent estimator from the statistical literature, whereby a parameter is called consistent if it converges in probability to the true value of the parameter. The consistency property is crucial for many investors conscious of risk, in particular those that do not want to understate risk.[3] Another benefit of consistency is that probabilistic bounds for the state of the system can be deduced, as discussed in Reece and Roberts (2010).

Given the mean and covariance of a bivariate random variable, recall the concentration ellipse (Dempster 1969) defined as the locus of points $\{\boldsymbol{x}:(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = 1\}$. For visual intuition, we depict such ellipses, estimated using the fusion methods described in the following subsections, in Exhibit 2. These results are shown for the case of three underlying data sources, assumed to each yield a noisy estimate of the same target.

Precision-weighted (PW) fusion. We first consider PW fusion for the case that the cross-covariances between sources is known to be 0. Suppose that for each s, $\hat{\mu}_s$ is an unbiased estimate, and each $\hat{\Sigma}_s$ is a known error covariance estimate that is consistent. It follows that the fused estimate given in Equation 21 is an unbiased estimate if and only if $\boldsymbol{\Sigma}_s \boldsymbol{W}_s = \boldsymbol{I}$, for weight matrices $\boldsymbol{W}_s$, and $\boldsymbol{I}$ the identity matrix. To find the weight estimates, we solve

$$\min_{\boldsymbol{W}} \ \text{trace}(\hat{\Sigma}) \text{ subject to} \sum_s \boldsymbol{W}_s = \boldsymbol{I}$$

Let $\hat{\Sigma}^{ij}$ denote the cross-covariance estimate between the unique sources s_i and s_j, and write

$$\hat{\Sigma} = \sum_s \boldsymbol{W}_s \hat{\Sigma}_s \boldsymbol{W}_s^T + \sum_{s_i \neq s_j} \boldsymbol{W}_{s_i} \hat{\Sigma}^{ij} \boldsymbol{W}_{s_j}^T \tag{22}$$

Assuming the cross-covariances are everywhere 0, solving for the optimal weights one finds that, on weight substitution,

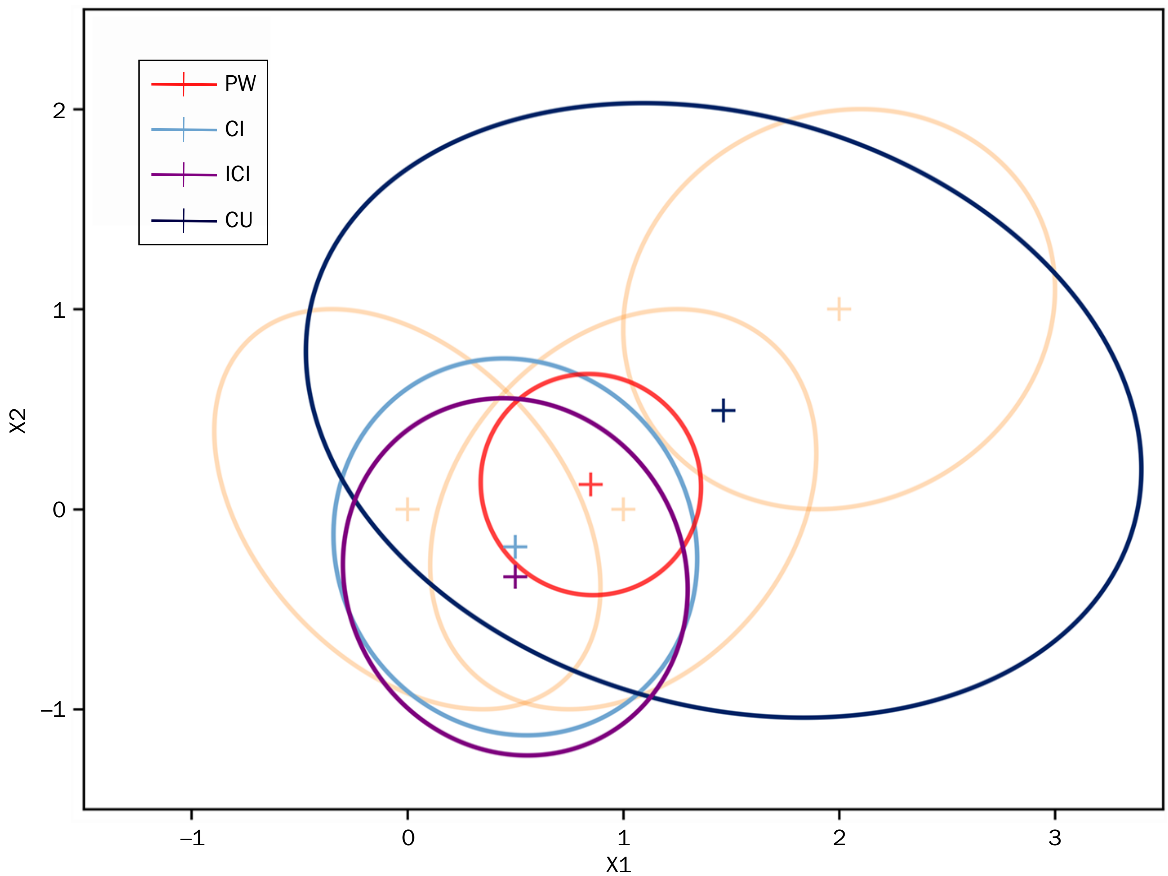
$$\hat{\Sigma} = (\hat{\Sigma}_1^{-1} + \cdots + \hat{\Sigma}_S^{-1})^{-1} \tag{23}$$

$$\hat{\mu} = \hat{\Sigma}(\hat{\Sigma}_1^{-1}\hat{\mu}_1 + \cdots + \hat{\Sigma}_S^{-1}\hat{\mu}_S) \tag{24}$$

[3] In the case of (multiperiod) Kelly investing, to which the fusion of this article could directly apply, investment size is inversely proportional to variance. But the investor faces almost sure ruin if they sufficiently overinvest (which can happen if uncertainty is estimated to be too small) relative to the optimum strategy (Thorp 2006).

EXHIBIT 2

Concentration Ellipses for Three Example Bivariate Gaussian Measurements of a Single State $x = (x_1, x_2)$ and the Optimal Fused State Estimates



NOTE: Single state $\boldsymbol{x} = (x_1, x_2)$ is light orange.

Further, details of this well-known result can be found in Maybeck (1979) and Willner, Chang, and Dunn (1976). The analogous result for fusing two sources that have known, nonzero correlation is given in Bar-Shalom (1981), though for our financial modeling framework this is less useful because the cross-correlations are not estimated. On the other hand, the case of unknown, potentially nonzero cross-correlation is the domain of the covariance intersection (CI) method, as presented in the following subsection.

CI. One could reasonably expect that views on the same underlying asset returns, as functions of similar or equal underlying data, exhibit nonzero cross-correlation. However, the cross-correlation is not necessarily known, or easily estimable. Such a setting is the domain of application for the CI fusion method of Julier and Uhlmann (1997). The insight of this approach is that, given Equation 22, the covariance ellipses for $\{\hat{\Sigma}_s\}$ contain $\hat{\Sigma}$ in their intersection for any values of the cross-covariance terms. Hence, the method yields

$$\hat{\Sigma}^{-1} = \sum_s \omega_s \hat{\Sigma}_s^{-1} \tag{25}$$

$$\hat{\mu} = \hat{\Sigma} \sum_s \omega_s \hat{\Sigma}_s^{-1} \hat{\mu}_s \tag{26}$$

for scalars $\omega_s \in [0, 1]$ with $\Sigma_s\omega_s = 1$ because a convex combination of covariances ensures $\hat{\Sigma}$ encloses the intersection region. The ω_s terms can be computed via a standard optimization routine subject to minimizing the trace of $\hat{\Sigma}$. Further, it has been shown that the CI method applied to a pair of consistent estimates yields a consistent estimate (Uhlmann 1995). Finally, Equations 25 and 26 as given earlier extend the canonical case of two sources, as offered in Chen, Arambel, and Mehra (2002).

CI is a straightforward method that holds for any cross-covariance, but a cost of its flexibility is that the estimate can be too conservative for many practical use cases. We next consider inverse CI (ICI) that attempts to retain the consistency benefit of CI, while being less conservative.

ICI. The method of ICI is similar in approach to that of CI but achieves a less conservative estimate while still maintaining consistency. The algorithm was presented recently in Noack et al. (2017). A central insight, inspired by Sijs and Lazar (2012), is to first consider estimates pairwise such that for s = 1, 2:

$$\hat{\mu}_s = \hat{\Sigma}_s((\tilde{\Sigma}_s)^{-1}\tilde{\mu}_s + \Gamma^{-1}\gamma) \tag{27}$$

$$\hat{\Sigma}_s = ((\tilde{\Sigma}_s)^{-1} + \Gamma^{-1})^{-1} \tag{28}$$

This is to say, that each estimate is itself the PW fusion of a sub-estimate from the collection $\{(\tilde{\mu}_s, \tilde{\Sigma}_s) : 1 \le s \le 2\}$, for sub-estimates uncorrelated between sources, and a shared set of information denoted $(\boldsymbol{\gamma}, \boldsymbol{\Gamma})$, also uncorrelated with each sub-estimate; we proceed assuming such a decomposition exist. If the shared information is known, then fusion can be solved optimally by subtracting the shared covariance from the PW fusion covariance for the original pair of estimates (Noack et al. 2017). But because it is not known, the second insight is to instead solve for, and subtract, an upper bound for the shared covariance, valid for any true underlying shared information. Further, regardless of the shared information, it is shown that a consistent, tight covariance estimate—if it exists—is of the form

$$\hat{\Sigma}^{-1} = \sum_{i=1}^{2}\hat{\Sigma}_s^{-1} - \left(\sum_{s=1}^{2}\omega_s\hat{\Sigma}_s\right)^{-1} \tag{29}$$

for scalars ω_s as in the case of CI, with the ICI covariance estimate achieving this form exactly (Noack et al. 2017). With a modicum of algebra, it follows that the fused mean is given by

$$\hat{\mu} = \sum_{s=1}^{2} C_s\hat{\mu}_s$$

$$\text{where } C_s = \hat{\Sigma}\cdot(\hat{\Sigma}_s^{-1} - \omega_s\left(\sum_{s=1}^{2}\omega_s\hat{\Sigma}_s)^{-1}\right) \tag{30}$$

It is important to recognize that this formula does not generalize to more than two information sources in an obvious way—the ICI algorithm fuses information sources pairwise. However, ICI can be applied recursively. This follows from the assumption that any pairwise-fused estimate has a decomposition analogous to Equations 27 and 28 (Noack et al. 2017). For further results and details, we also recommend Noack, Sijs, and Hanebeck (2017).

Covariance union (CU). The fusion methods outlined earlier assume that each source's estimate is itself consistent with respect to the target. We relax this assumption by requiring that at least one of the source estimates is consistent. A given estimate could be inconsistent, as evidenced by, say, mean components sufficiently different between estimates, given relatively smaller covariance estimates. This could be apparent, for example, when the Mahalanobis distance between some pair of estimates exceeds a critical threshold. Further, we may not know precisely which of the underlying estimates is consistent, or it may be (in some sense) costly to resolve. The CU algorithm has been proposed as a fusion method for this setting (Uhlmann 2003). CU constructs a fused estimate guaranteed to be consistent because the fused estimate is consistent with respect to each underlying estimate:

$$\hat{\Sigma} \geq \hat{\Sigma}_s + (\hat{\mu} - \hat{\mu}_s)(\hat{\mu} - \hat{\mu}_s)^T, \quad 1 \leq s \leq S$$

Although the idea behind CU is relatively simple, the numerical implementation for estimating the fused estimate is more challenging. In an attempt to simplify the computation, we offer the following details. First, CU can be implemented using nonconvex optimization software such as SolvOpt (Kuntsevich and Kappel 1997). We follow Bochardt et al. (2006) for implementing a SolvOpt-based solution for calculating the results below. Second, a key step is to define the coefficient vector that we wish to solve for $\boldsymbol{x}$. The first n elements of $\boldsymbol{x}$ are the elements of $\hat{\mu}$, and the remaining $\frac{n(n+1)}{2}$ elements of $\boldsymbol{x}$ are the elements of the upper triangle of $\hat{\Sigma}$. Then $\boldsymbol{x}$ becomes the optimization target for SolvOpt. We seek to minimize the determinant of $\hat{\mu}$ subject to the constraint that

$$\forall s, \quad \hat{\Sigma}_s^* := \hat{\Sigma} - \hat{\Sigma}_s - (\hat{\mu} - \hat{\mu}_s)(\hat{\mu} - \hat{\mu}_s)^T \geq 0 \tag{31}$$

A particular feature of the SolvOpt software is that it requires a constraint $c(\boldsymbol{x})$ such that, for any input, $c(\boldsymbol{x}) \leq 0$. Because $\hat{\Sigma}_s^*$ is required to be positive semidefinite, it must have nonnegative eigenvalues for all choices of s, and the constraint is satisfied for a choice of c such that

$$-c(\boldsymbol{x}) := \min\{\{\text{eigenvalues}(\hat{\Sigma}_s^*) : \forall s\}\} \geq 0$$

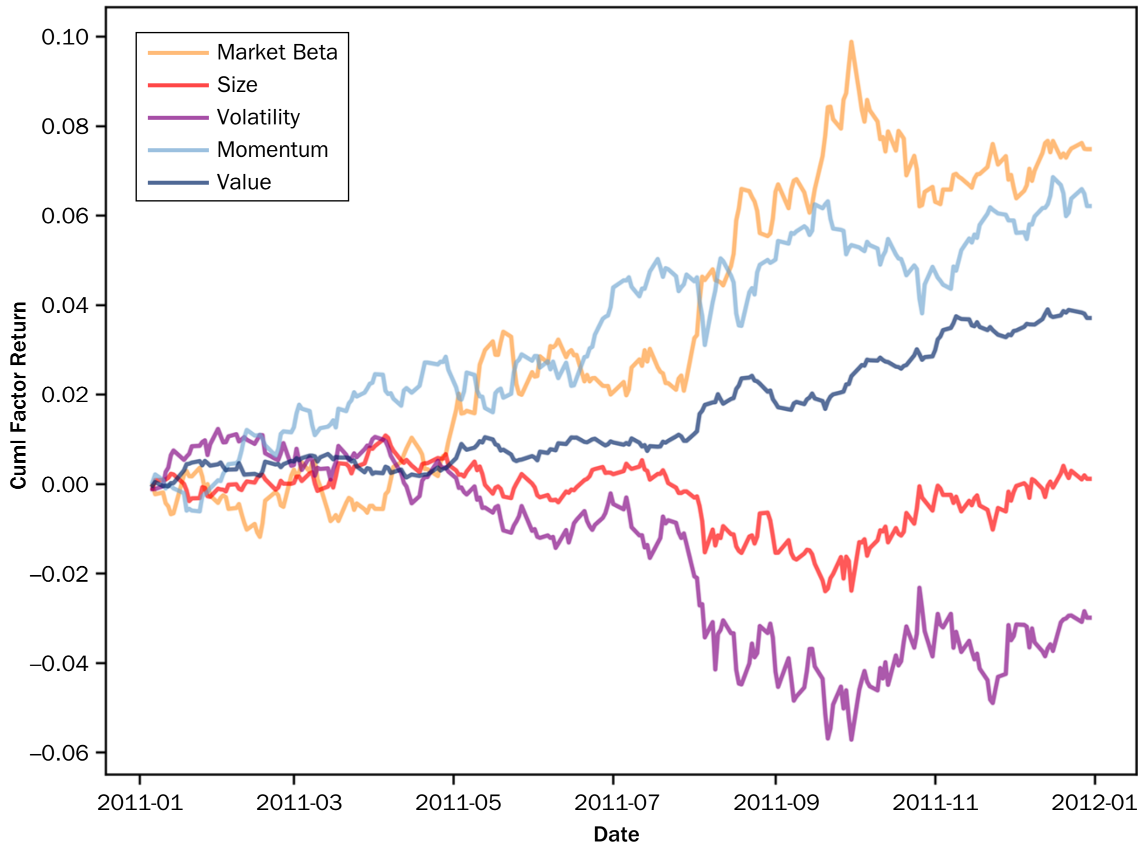
Finally, for the interested reader, the works of Bochardt et al. (2006) and Julier, Uhlmann, and Nicholson (2004) offer further efficient approximation methods for implementing CU, as required for particular practical use cases.

DATA

We describe the dataset construction for replicating the empirical example for BL-APT of Kolm and Ritter (2017). This construction matches the original descriptions as closely as possible, and we make note of any additional required assumptions made. Although we have attempted to carefully reproduce the database, our results need not match exactly. Not in the least, we have no guarantee that the underlying database that we call is the same five years later, even if our data replication process is perfectly faithful. A visual comparison can be made between our factor values against the source given in Exhibit 3, which demonstrates a reasonable replication.

The data is sourced from the Center for Research in Security Prices, LLC (CRSP) US Stock Database and the Institutional Brokers' Estimate System (IBES) database, with access granted via the Wharton Research Data Service (WRDS). The original work sourced daily US equity market data from 1992–2015; we extend our dataset until 2022 to include the market crash of March 2020, and subsequent market rebound, for interest's sake.

EXHIBIT 3
Cumulative Factor Returns to Risk Premiums, Year of 2011



NOTES: Depiction inspired by Figure 1 of Kolm and Ritter (2017). A single year is shown for ease of visualization. The year is arbitrarily chosen.

For each day, set n = 2,000 and select the top n stocks within the US market sorted by market capitalization. The dataset is filtered for US dollar–denominated common stock only—there are no closed end funds, Real Estate Investment Trust (REITs), exchange-traded fund (ETFs), unit trusts, depository receipts, warrants, and so on. We also collected the S&P 500 (excess) return time series, stock market capitalizations, and earnings per share, adjusted for splits.

The dataset requires the creation of five risk factors: market beta, size, volatility, momentum, and value. The reasonableness of our dataset construction is per the output of Exhibit 3, which accords with expectations. We also create 70 industry factors; these are one-hot encoded based on the industry given by Standard Industrial Classification (SIC). Calculating size is relatively straightforward in CRSP as the product of shares outstanding and the daily close price. Value was determined by analyzing IBES data and could be merged back to the CRSP database based on the Committee on Uniform Securities Identification Procedures (CUSIP) identifier. We note that risk factor construction is often nontrivial. Momentum was codable using CRSP data and the definitions of Asness, Moskowitz, and Pedersen (2013); we calculated the daily compounded 12-month returns sans the most recent one month of data. We created the market beta and volatility factors using the relevant CRSP data and the details of Kolm and Ritter (2017). Indeed, over a two-year rolling window, we estimate a collection of linear functions by regressing each asset's daily excess returns against the S&P 500 excess returns. The market beta is the regression gradient, whereas the volatility factor is set to be the regression mean square error.

The data are constructed for bimonthly rebalancing to facilitate the experiments of the following section.

RESULTS

View Models

To proxy multiple sources of view generation, we implemented various models with utility for forecasting financial time series. Indeed, machine learning models have proven useful for not only prediction but also for estimating prediction uncertainty. We implement an Autoregressive integrated moving average (ARIMA), boosted regression, and Gaussian process model for generating views, and calculate model-derived uncertainty estimates. Relevant details are included in the appendix because time-series prediction modeling is not the primary focus of this article. We denote the three view models ARIMA, Boost, and Gaussian Process (GP), respectively. We also implement the four fusion methods outlined above.

Parameter Choices

Per Exhibit 1 and the outline above, the BL-APT model requires the specification of $\boldsymbol{q}$, $\boldsymbol{F}$, $\boldsymbol{\Omega}$, $\boldsymbol{\xi}$, $\boldsymbol{V}$, and $\boldsymbol{D}$. We set $\boldsymbol{q}$ as the mean prediction estimate given by a view model, with the view models as previously described. The accompanying epistemic uncertainty estimate is set to $\boldsymbol{\Omega}$, and $\boldsymbol{F}$ is set equal to the aleatoric uncertainty estimate. The prior hyperparameter $\boldsymbol{\xi}$ is estimated as an average of $\boldsymbol{f}$ over a recent rolling lookback window of the most recent 20 observations. This window size was arbitrarily decided and is the same for all rolling windows introduced hereafter. The total variance of each underlying factor is calculated relative to out-of-sample data, and we yield an approximation to epistemic uncertainty by the estimation method as described for the ARIMA model in the appendix. We set $\boldsymbol{V}$ as this estimate. Our estimate for $\boldsymbol{D}$ is a diagonal covariance matrix equal to $\hat{\sigma}^2\boldsymbol{I}$, where $\hat{\sigma}^2$ is the mean square error yielded from estimating Equation 7, and $\boldsymbol{I}$ is the identity matrix. Finally, we set the relative risk aversion parameter to a constant value of 10.

Transaction Costs

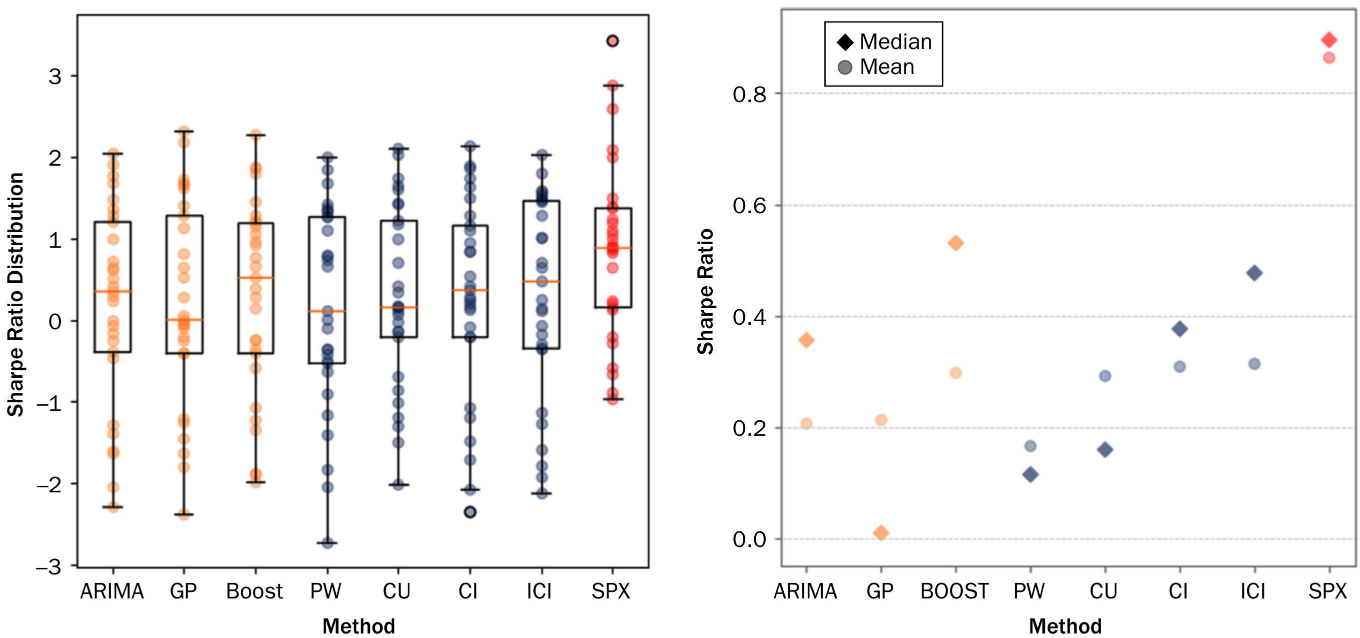
An effect of accounting for transaction costs and carefully setting the transaction cost model parameters is to greatly reduce portfolio turnover relative to the unconstrained Markowitz approach with costs assumed nil. Indeed, without transaction costs, it is well-known that this latter approach usually result in unrealistic and unreasonably large leverage suggested for the optimal portfolio and high portfolio turnover. Recent results of Jensen et al. (2022) show that incorporating an appropriate transaction cost model goes some way to mitigate these limitations in empirical studies. We implement the transcation cost model outlined above to the same ends. We set our portfolio rebalance period to be bimonthly, which is sufficiently long—given our assumed wealth process and the potential return being significantly larger than the cost of rebalancing—such that the reward of updating weights is not dominated by the cost of trading.

Results: Benchmark

We have tabulated performance results by year and model over a 29-year period and include them in the appendix. The method S&P 500 (TR) denotes the total

EXHIBIT 4

View-Based Portfolio Excess Return Performance Net of Trading Costs



NOTE: Benchmark SPX strategy is gross of (negligible) transaction costs, but total returns are calculated in excess of the risk-free rate.

return of the S&P 500 for the year. The annual returns may appear slightly less than commonly reported because they are net of the risk-free rate. The S&P 500 (TR) represents a useful benchmark in that it represents the market buy-and-hold strategy, accessible to the modern-day investor and with a potential for exceptionally competitive fees, especially in the more recent years. For comparison's sake, all results dependent on an underlying view model are reported based on normalized underlying data such that ex post annualized return volatility is equal to that of the benchmark portfolio. We make a note that, on perusing the exhibits, there is obvious decay in the performance of the view-based models over the 30-year period, as evidenced, for example, by the significant decline in Sharpe ratios over time.

Results: Statistics

We depict the Sharpe ratio statistics from the tabulated data in Exhibit 4. The leftmost chart shows box-and-whisker plots for the annual Sharpe ratio for the collection of portfolios over the 29-year period. Investment in the S&P 500 (TR) is clearly superior to our underlying modeling approach. On the other hand, there are improvements that could increase the view model performance that we deem beyond the scope of this work. For an easier comparison of centrality for the various method's Sharpe distributions, we show each portfolio's median and mean Sharpe ratio over the 29-year period in the rightmost chart. This visual depiction summarizes how fusion- and non-fusion-based methods performed relative to each other through time (and relative to the market total return portfolio). We observe that fusion-based medians are increasing through PW (0.11), CU (0.16), CI (0.38), and ICI (0.48), with the means showing a similar but less pronounced trend. The set of mean Sharpe ratios for the single view–based methods have a smaller range than the medians (0.21–0.30 versus 0.01–0.53, respectively); similarly for the fusion-based methods (0.17–0.31 versus 0.11–0.48). When considering the mean Sharpe ratio over the full data sample, ICI outperforms among the set of fusion-based methods and the

EXHIBIT 5

P-Values and Confidence Intervals for Tests Comparing 29 Years of Annual Sharpe Ratios Pairwise between Strategies

Pairwise Comparison		Ljung–Box	Shapiro–Wilk	Paired *t* P-Values (diff of means)	BCa CIs (diff of means)	BCa CIs (diff of med.)	Wilcoxon P-Values (med. of diffs)
SPX	ICI	0.14	0.21	**0.06**	**[0.15, 0.98]**	[–0.13, 0.90]	**0.07**
	CI	0.14	0.45	**0.05**	**[0.15, 0.99]**	[–0.07, 0.70]	**0.07**
	CU	0.12	0.75	**0.05**	**[0.19, 1.01]**	**[0.31, 1.03]**	**0.06**
	PW	0.17	0.36	**0.03**	**[0.29, 1.22]**	**[0.07, 1.30]**	**0.06**
	ARIMA	0.13	0.41	**0.03**	**[0.25, 1.09]**	**[0.17, 0.92]**	**0.03**
	GP	0.17	0.51	**0.03**	**[0.20, 1.09]**	**[0.34, 1.10]**	**0.03**
	Boost	0.16	0.32	**0.05**	**[0.18, 1.01]**	[–0.10, 0.83]	**0.08**
ICI	CI	0.80	**0.03**	0.47	[–0.08, 0.08]	[–0.26, 0.28]	0.31
	CU	0.88	0.12	0.41	[–0.11, 0.14]	**[0.01, 0.67]**	0.63
	PW	0.94	0.90	0.15	[–0.03, 0.32]	**[0.04, 1.00]**	0.19
	ARIMA	0.70	0.75	**0.08**	**[0.02, 0.21]**	[–0.26, 0.36]	**0.10**
	GP	0.33	0.65	**0.09**	**[0.01, 0.19]**	**[0.25, 0.73]**	0.12
	Boost	0.68	0.19	0.46	[–0.21, 0.25]	[–0.51, 0.39]	0.58
CI	CU	0.56	0.85	0.43	[–0.12, 0.14]	**[0.05, 0.51]**	0.55
	PW	0.67	0.56	0.15	[–0.02, 0.32]	[–0.29, 0.65]	0.18
	ARIMA	0.47	**0.00**	**0.06**	**[0.03, 0.19]**	[–0.17, 0.22]	0.93
	GP	0.11	0.74	0.13	[–0.01, 0.20]	**[0.30, 0.67]**	0.16
	Boost	0.35	0.91	0.47	[–0.17, 0.24]	[–0.52, 0.10]	0.50
CU	PW	0.29	0.28	0.17	[–0.04, 0.28]	[–0.57, 0.43]	0.17
	ARIMA	0.48	0.89	0.23	[–0.05, 0.22]	**[–0.46, –0.02]**	0.20
	GP	0.75	0.22	0.23	[–0.06, 0.20]	[–0.13, 0.37]	0.21
	Boost	0.73	0.98	0.49	[–0.21, 0.19]	[–0.78, 0.01]	0.52
PW	ARIMA	0.95	0.64	0.36	[–0.17, 0.11]	[–0.65, 0.24]	0.65
	GP	0.95	0.58	0.33	[–0.19, 0.08]	[–0.31, 0.66]	0.41
	Boost	0.17	0.45	**0.09**	**[–0.27, –0.03]**	**[–1.12, –0.14]**	**0.10**
ARIMA	GP	**0.01**	0.94	0.46	[–0.08, 0.07]	**[0.22, 0.54]**	0.45
	Boost	0.20	**0.06**	0.27	[–0.29, 0.09]	[–0.51, 0.19]	0.73
GP	Boost	0.74	0.34	0.27	[–0.26, 0.07]	**[–0.96, –0.32]**	0.64

NOTES: Bold font with yellow shading indicates significance at the 10% level. The rightmost four columns are one-sided tests.

single view–based models. Similarly to the median, ICI outperforms among the set of fusion-based methods and sits closely to the best of the single view–based models. We test for statistical significance of differences in these measures of distribution centrality per the results in Exhibit 5. Despite the small sample sizes, there is some interesting supporting evidence for the outperformance of fusion-based methods.

We test for the difference of means between methods with a collection of pairwise one-sided *t*-tests. Such tests make a critical assumption of independence in the pairwise Sharpe ratio differences as indexed by year. We note that despite individual time series of model performance statistics potentially displaying serial correlation, it seems reasonable to suggest that the performance between models over the years does not. This is clearly a critical claim. We examine the autocorrelation function for the 28 possible pairwise comparisons and find one case of rejecting the null hypothesis of no autocorrelation in a Ljung–Box test at the 10% level of significance. We further assume normality of the differences and test this with a (one-sided) Shapiro–Wilk test for normality, also at the 10% level of significance. We find that this assumption is rejected for three cases. Finally, we assume the nonexistence of outliers in the difference data and validate this claim on visual inspection of box and whisker plots.

On testing for the difference of means with a paired *t*-test, we find, of particular interest, that ICI outperforms the ARIMA and GP models, whereas CI outperforms the ARIMA only (though in this case the Shapiro–Wilks test leads us not to rely on some of the *t*-test P-values). To support testing for significance in the difference of means between groups, we bootstrap two-sided bias-corrected and accelerated (BCa) bootstrap 90% confidence intervals for paired data (Efron 1987). We find evidence that ICI outperforms the ARIMA and GP models, and again the CI outperforms the ARIMA model. We also test for the difference in medians between groups with BCa confidence intervals, but they are broadly less conclusive. This is potentially owing to the small sample size. On the other hand, we have evidence that both ICI and CI outperforms GP on this metric.

Alternatively to testing for the difference of medians across all samples, we can test for the median of the differences between pairwise annual Sharpe ratios. To this end, we utilize the Wilcoxon signed-rank test for paired samples (Wilcoxon 1945). In addition to assuming independence, a critical assumption is that of symmetry about the distributions of paired differences. Given the results of the Shapiro–Wilk tests and visual inspection of what appear to be reasonable quantile–quantile plots, we continue assuming symmetry is satisfied. We perform one-sided tests at the 10% level of significance and find evidence that ICI outperforms the ARIMA model for this comparison.

It is interesting that the CI and ICI methods achieve more compelling outperformance than the CU and PW fusion methods, relative to the underlying view models. The assumption of PW that the underlying views have 0 cross-covariance may be suboptimal such that the method is yielding inconsistent estimates and in turn impacting trading performance. On the other hand, perhaps CU could be utilized more sparingly and strategically, such as when the underlying views have relatively high dispersion and so appear sufficiently contradictory. An alternative fusion method such as ICI could be employed otherwise.

Corroborating our intuition about the relative strength of the S&P 500 (TR) strategy relative to all others, we see that SPX improves on all models—both single view- and fusion-based—for almost all tests. Sufficiently strong view models are required to compete with the market portfolio but beyond the scope of this work. It is perhaps surprising that view- and fusion-based models net of transaction costs have net positive Sharpe ratios at all. Indeed, we estimate global Sharpe ratios between 0.10 (PW) and 0.34 (ICI) for the entirety of the 29-year period.

Closing Remark

The results presented offer preliminary evidence supporting fusion-based methods within a Black–Litterman framework subsuming a relatively complex investment strategy net of transaction costs. Fusion-based methods result in a weighted-average performance respectful of uncertainty estimates. These methods could certainly be ex ante preferred when all views are to be incorporated into an investment decision in the case that the best view model is unknown (and/or as in our presented work, when the method constituting the best view model is both largely unpredictable and changing through time). We have supporting evidence for a rare free lunch in financial investing by way of model fusion, essentially amounting to a benefit we could call model diversification. This could be of further application in alternative trading domains, perhaps with more sophisticated underlying view models, and we hope to inspire readers to consider their practical use.

CONCLUSION AND NEXT STEPS

In this work, we further characterize view uncertainty within a Bayesian Black–Litterman framework. In doing so, we offer optimal portfolio weights for a case of Black–Litterman in an Arbitrage Pricing Theory setting. Considering the Bayesian Black-Litterman model via an equivalent state space representation, we are inspired by the information fusion literature for combining correlated prediction mean and uncertainty estimates streamed from diverse sources. Hence, we demonstrate prediction and uncertainty fusion for multiple, simultaneous views within the Black–Litterman framework and describe consistent uncertainty estimation within financial investing contexts. In our empirical work, we demonstrate the utility of our approach for BL-APT. Future practical work could be based on more realistic investment strategies, not the least incorporating methods of drawdown control.

APPENDIX

FURTHER APT CALCULATIONS

This section adds further detail to the BL-APT calculations above. In determining the expectation and covariance for $\boldsymbol{r}|\boldsymbol{q}$ in the case of unknown mean but known covariance, given Equation 12 we collect and expand the terms in the exponent up to a factor of $-\frac{1}{2}$, complete the square in $\boldsymbol{f}$, and integrate what is recognizable as a Gaussian integral:

$$
\begin{aligned}
&(\boldsymbol{r}-\boldsymbol{X}\boldsymbol{f})^T\boldsymbol{D}^{-1}(\boldsymbol{r}-\boldsymbol{X}\boldsymbol{f})+(\boldsymbol{f}-\mu)^T\Sigma^{-1}(\boldsymbol{f}-\mu)\\
&\quad=\boldsymbol{r}^T\boldsymbol{D}^{-1}\boldsymbol{r}-2\boldsymbol{f}^T\boldsymbol{X}^T\boldsymbol{D}^{-1}\boldsymbol{r}+(\boldsymbol{X}\boldsymbol{f})^T\boldsymbol{D}^{-1}\boldsymbol{X}\boldsymbol{f}+\boldsymbol{f}^T\Sigma^{-1}\boldsymbol{f}-2\boldsymbol{f}^T\Sigma^{-1}\mu+\mu^T\Sigma^{-1}\mu\\
&\quad=\boldsymbol{r}^T\boldsymbol{D}^{-1}\boldsymbol{r}+\boldsymbol{f}^T\boldsymbol{H}\boldsymbol{f}-2\eta\boldsymbol{f}+\mu^T\Sigma^{-1}\mu
\end{aligned}
$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ denote the mean and covariance for $\boldsymbol{f}|\boldsymbol{q}$, $\boldsymbol{H} := \boldsymbol{X}^T\boldsymbol{D}^{-1}\boldsymbol{X} + \Sigma^{-1}$ and $\boldsymbol{\eta} := \boldsymbol{X}^T\boldsymbol{D}^{-1}\boldsymbol{r} + \Sigma^{-1}\mu$. Completing the square,

$$\boldsymbol{f}^T\boldsymbol{H}\boldsymbol{f}-2\boldsymbol{f}^T\eta=(\boldsymbol{f}-\nu)^T\boldsymbol{H}(\boldsymbol{f}-\nu)-\nu^T\boldsymbol{H}\nu,\qquad \nu=\boldsymbol{H}^{-1}\eta$$

whereby

$$\int\exp\left(-\frac{1}{2}(\boldsymbol{f}-\nu)^T\boldsymbol{H}(\boldsymbol{f}-\nu)\right)d\boldsymbol{f}=\sqrt{\frac{(2\pi)^k}{|\boldsymbol{H}|}}$$

Therefore,

$$p(\boldsymbol{r}\mid\boldsymbol{q})\propto\exp\left(-\frac{1}{2}\left[\boldsymbol{r}^T\boldsymbol{D}^{-1}\boldsymbol{r}+\mu^T\Sigma^{-1}\mu-\nu^T\boldsymbol{H}\nu\right]\right)=\exp\left(-\frac{1}{2}\left[\boldsymbol{r}^T\boldsymbol{D}^{-1}\boldsymbol{r}+\mu^T\Sigma^{-1}\mu-\eta^T\boldsymbol{H}^{-1}\eta\right]\right)\quad\text{(A1)}$$

and we retrieve the exponent of the Gaussian for $\boldsymbol{r}|\boldsymbol{q}$. Inspired by Kolm and Ritter (2017), we next make use of the following lemma.

Lemma 1. *If a multivariate normal random variable* θ *has density* $p(\theta)$ *and* $-2\log p(\theta) = \theta'\boldsymbol{H}\theta - 2\nu'\theta +$ (*terms without* θ) *then* $var(\theta) = \boldsymbol{H}^{-1}$ *and* $\mathbb{E}\theta = \boldsymbol{H}^{-1}\nu$.

Collecting quadratic terms in $\boldsymbol{r}$ from the exponent of Equation A1,

$$
\begin{aligned}
-\eta^T\boldsymbol{H}^{-1}\eta+\boldsymbol{r}^T\boldsymbol{D}^{-1}\boldsymbol{r}&=-[\boldsymbol{X}^T\boldsymbol{D}^{-1}\boldsymbol{r}+\Sigma^{-1}\mu]^T\boldsymbol{H}^{-1}[\boldsymbol{X}^T\boldsymbol{D}^{-1}\boldsymbol{r}+\Sigma^{-1}\mu]+\boldsymbol{r}^T\boldsymbol{D}^{-1}\boldsymbol{r}\\
&=-\boldsymbol{r}^T\boldsymbol{D}^{-1}\boldsymbol{X}\boldsymbol{H}^{-1}\boldsymbol{X}^T\boldsymbol{D}^{-1}\boldsymbol{r}+\boldsymbol{r}^T\boldsymbol{D}^{-1}\boldsymbol{r}+\text{lower-order terms in } \boldsymbol{r}\\
&=\boldsymbol{r}^T[-\boldsymbol{D}^{-1}\boldsymbol{X}\boldsymbol{H}^{-1}\boldsymbol{X}^T\boldsymbol{D}^{-1}+\boldsymbol{D}^{-1}]\boldsymbol{r}+\text{lower-order terms in } \boldsymbol{r}\qquad\text{(A2)}
\end{aligned}
$$

Making use of the lemma, we have that the posterior predictive distribution $p(\boldsymbol{r}|\boldsymbol{q})$ is multivariate normal with covariance

$$\begin{aligned}\mathrm{var}(\boldsymbol{r}|\mathbf{q}) &= [-\boldsymbol{D}^{-1}\boldsymbol{X}\boldsymbol{H}^{-1}\boldsymbol{X}^T\boldsymbol{D}^{-1} + \boldsymbol{D}^{-1}]^{-1},\\ &= [\boldsymbol{D}^{-1} - \boldsymbol{D}^{-1}\boldsymbol{X}[\boldsymbol{X}^T\boldsymbol{D}^{-1}\boldsymbol{X} + [(\mathbf{V}^{-1} + \Omega^{-1})^{-1} + \boldsymbol{F}]^{-1}]^{-1}\boldsymbol{X}^T\boldsymbol{D}^{-1}]^{-1}.\end{aligned}$$

Further, we collect the linear terms in $\boldsymbol{r}$ from Equation A2

$$-\boldsymbol{r}^T\boldsymbol{D}^{-1}\boldsymbol{X}^T\boldsymbol{H}^{-1}\Sigma^{-1}\mu - \mu^T\Sigma^{-1}\boldsymbol{H}^{-1}\boldsymbol{X}^T\boldsymbol{D}^{-1}\boldsymbol{r} = -2\mu^T\Sigma^{-1}\boldsymbol{H}^{-1}\boldsymbol{X}^T\boldsymbol{D}^{-1}\boldsymbol{r}$$

so that again by the lemma

$$\begin{aligned}\mathbb{E}(\boldsymbol{r} \mid \boldsymbol{q}) &= \mathrm{var}(\boldsymbol{r} \mid \boldsymbol{q})\boldsymbol{D}^{-1}\boldsymbol{X}\boldsymbol{H}^{-1}\Sigma^{-1}\mu\\ &= \mathrm{var}(\boldsymbol{r} \mid \boldsymbol{q})\boldsymbol{D}^{-1}\boldsymbol{X}[\boldsymbol{X}^T\boldsymbol{D}^{-1}\boldsymbol{X} + [(\boldsymbol{V}^{-1} + \Omega^{-1})^{-1} + \boldsymbol{F}]^{-1}]^{-1}\\ &\quad\cdot [(\boldsymbol{V}^{-1} + \Omega^{-1})^{-1} + \boldsymbol{F}]^{-1}(\boldsymbol{V}^{-1} + \Omega^{-1})^{-1}(\boldsymbol{V}^{-1}\xi + \Omega^{-1}\boldsymbol{q})\end{aligned}$$

View Fusion under Model-Based Forecasts

Suppose we model $y = f(\boldsymbol{x}) + \epsilon$ for some function f parameterized by θ, with $\epsilon \sim N(0, \sigma^2)$. Then we can write $y|\boldsymbol{x}, \theta \sim N(f(\boldsymbol{x}), \sigma^2)$. Assuming θ has prior distribution $p(\theta)$, then for a collection of data $\mathcal{D}$ we can update the prior via the likelihood analogous to Equation 5: $p(\theta|\mathcal{D}) \propto p(\mathcal{D}|\theta)p(\theta)$. Then for a new data point $\boldsymbol{x}^*$, the predictive distribution for the corresponding y^* can be written

$$p(y^*|\boldsymbol{x}^*, \mathcal{D}) = \int p(y^*|\boldsymbol{x}^*, \theta)p(\theta|\mathcal{D})d\theta$$

The prediction variance, $\mathrm{var}(y^*|\boldsymbol{x}^*, \mathcal{D})$, can be decomposed making use of the law of total variance:

$$\mathrm{var}(y^*|\boldsymbol{x}^*, \mathcal{D}) = \mathrm{var}_{\theta|\mathcal{D}}(\mathbb{E}(y^*|\boldsymbol{x}^*, \theta)) + \mathbb{E}_{\theta|\mathcal{D}}(\mathrm{var}(y^*|\boldsymbol{x}^*, \theta)) \quad \text{(A3)}$$

Here, the subscript on the outer variance and expectation terms on the right-hand side denotes that the integral is calculated with respect to the conditional density for $\theta|\mathcal{D}$. The first term on the right-hand side, we call the epistemic uncertainty, sometimes described as a model uncertainty associated with estimating the model parameters, and that is reducible with increasing data. The second term we call the aleatoric uncertainty; this models the underlying stochasticity of the data distribution and is irreducible.

Estimation of predictive distributions or the components of model uncertainty is not always tractable; consequently, various approximation methods exists for estimating these quantities. A review is beyond the scope of this article; we instead refer interested readers to the recent work of Hüllermeier and Waegeman (2021) for further details.

Brief overview of view models. To yield the results above, we first implement three diverse view models. For completeness, we provide the implementation details in the following.

We implement a Gaussian Process model using scikit-learn in Python (Pedregosa et al. 2011). The kernel function is given by the sum of a radial basis function and a white-noise kernel; see, for example, Roberts et al. (2013) for context on Gaussian processes for time-series modeling. We refit the model on a bimonthly basis on a rolling data window of one year. The predict method is used to generate prediction estimates and an accompanying (total) uncertainty estimate. The uncertainty can be decomposed into epistemic and aleatoric components by setting the latter as the noise-level parameter

EXHIBIT A1

Performance Metrics, 1993–2000

Portfolio Method	Cuml Ret	Ret Vol	Sharpe	IR	Sortino	Max DD	Cuml Ret	Ret Vol	Sharpe	IR	Sortino	Max DD
				1993						1994		
S&P 500 (TR)	7.11	8.59	0.82	–1.94	1.21	–8.84	–2.03	9.80	–0.21	–1.42	–0.28	–8.91
ARIMA	–17.50	8.59	–2.04	–2.68	–2.01	–16.43	20.00	9.80	2.04	1.73	5.65	–18.76
GP	–15.42	8.59	–1.79	–2.60	–1.78	–15.90	21.39	9.80	2.18	1.78	5.91	–19.92
Boost	–9.14	8.59	–1.06	–1.84	–1.18	–10.40	12.53	9.80	1.28	1.04	2.73	–13.63
PW	–12.04	8.59	–1.40	–2.20	–1.44	–12.37	13.90	9.80	1.42	1.16	3.07	–14.73
CU	–11.11	8.59	–1.29	–1.98	–1.47	–13.29	20.65	9.80	2.11	1.74	5.76	–19.13
CI	–20.11	8.59	–2.34	–2.81	–2.22	–18.60	18.24	9.80	1.86	1.59	6.13	–17.71
ICI	–16.51	8.59	–1.92	–2.46	–1.82	–16.14	19.94	9.80	2.03	1.70	8.08	–18.73
				1995						1996		
S&P 500 (TR)	26.80	7.81	3.43	–2.57	5.72	–24.15	16.31	11.75	1.38	–1.60	1.96	–18.96
ARIMA	–12.74	7.81	–1.63	–4.48	–1.67	–14.11	14.18	11.75	1.21	–0.33	2.78	–19.34
GP	–11.30	7.81	–1.45	–4.33	–1.52	–13.81	9.49	11.75	0.81	–0.59	1.65	–16.73
Boost	–15.45	7.81	–1.98	–4.56	–2.06	–14.93	10.90	11.75	0.93	–0.49	1.62	–18.25
PW	–21.30	7.81	–2.73	–5.36	–2.46	–20.11	9.34	11.75	0.79	–0.60	1.34	–19.69
CU	–11.64	7.81	–1.49	–4.53	–1.57	–14.14	20.40	11.75	1.74	0.02	4.69	–22.35
CI	–13.29	7.81	–1.70	–4.87	–1.77	–14.72	25.06	11.75	2.13	0.29	6.07	–25.21
ICI	–12.45	7.81	–1.59	–4.78	–1.67	–14.07	21.15	11.75	1.80	0.06	4.78	–23.72
				1997						1998		
S&P 500 (TR)	25.34	18.09	1.40	–1.04	2.04	–23.74	22.47	20.25	1.11	–0.51	1.57	–22.93
ARIMA	26.74	18.09	1.48	–0.05	3.39	–21.90	12.62	20.25	0.62	–0.34	0.91	–16.82
GP	31.28	18.09	1.73	0.17	3.71	–25.04	10.57	20.25	0.52	–0.38	0.73	–16.72
Boost	32.70	18.09	1.81	0.22	4.57	–26.85	5.64	20.25	0.28	–0.58	0.41	–19.44
PW	30.50	18.09	1.69	0.13	3.69	–26.26	15.04	20.25	0.74	–0.24	1.06	–15.90
CU	28.95	18.09	1.60	0.05	3.75	–23.06	29.10	20.25	1.44	0.22	2.41	–27.24
CI	23.15	18.09	1.28	–0.22	2.61	–18.85	17.15	20.25	0.85	–0.18	1.24	–16.74
ICI	28.57	18.09	1.58	0.04	3.58	–22.90	26.14	20.25	1.29	0.12	2.20	–24.05
				1999						2000		
S&P 500 (TR)	16.15	18.04	0.90	–3.27	1.34	–14.82	–12.80	22.18	–0.58	–0.50	–0.81	–20.07
ARIMA	–28.79	18.05	–1.60	–1.72	–1.71	–27.22	–5.65	22.17	–0.25	–0.10	–0.33	–26.38
GP	–21.79	18.05	–1.21	–1.58	–1.40	–22.85	–8.89	22.17	–0.40	–0.23	–0.51	–26.27
Boost	19.04	18.05	1.06	–0.12	1.64	–25.24	–29.64	22.17	–1.34	–0.83	–1.55	–38.33
PW	–9.48	18.05	–0.53	–1.13	–0.68	–20.84	–10.81	22.17	–0.49	–0.32	–0.64	–27.58
CU	–3.66	18.05	–0.20	–0.94	–0.27	–21.42	1.46	22.17	0.07	0.15	0.09	–24.08
CI	–19.29	18.05	–1.07	–1.39	–1.27	–20.96	6.15	22.17	0.28	0.34	0.39	–23.62
ICI	–32.13	18.05	–1.78	–1.89	–1.89	–30.89	–6.63	22.17	–0.30	–0.12	–0.38	–27.23

NOTE: Max DD: Maximum drawdown.

EXHIBIT A2

Performance Metrics, 2001–2008

Portfolio Method	Cuml Ret	Ret Vol	Sharpe	IR	Sortino	Max DD	Cuml Ret	Ret Vol	Sharpe	IR	Sortino	Max DD
	2001						**2002**					
S&P 500 (TR)	–14.12	21.51	–0.67	–2.25	–0.92	–30.91	–23.24	25.99	–0.89	–1.43	–1.28	–33.81
ARIMA	36.09	21.51	1.68	1.72	2.57	–35.22	26.07	25.99	1.00	1.49	1.53	–34.82
GP	36.18	21.51	1.68	1.67	2.60	–34.88	33.51	25.99	1.29	1.73	2.10	–37.44
Boost	25.65	21.51	1.19	1.34	1.64	–30.70	31.67	25.99	1.22	1.65	1.99	–36.37
PW	23.71	21.51	1.10	1.28	1.48	–29.85	33.17	25.99	1.28	1.68	2.09	–37.85
CU	21.26	21.51	0.99	1.24	1.39	–26.00	4.75	25.99	0.18	0.83	0.24	–30.08
CI	35.16	21.51	1.63	1.66	2.62	–33.08	22.01	25.99	0.85	1.37	1.28	–31.67
ICI	34.26	21.51	1.59	1.64	2.52	–33.28	26.35	25.99	1.01	1.48	1.57	–35.08
	2003						**2004**					
S&P 500 (TR)	25.64	17.04	1.50	–2.42	2.32	–28.43	9.78	11.07	0.88	–1.37	1.28	–12.49
ARIMA	–23.61	17.03	–1.39	–2.48	–1.53	–23.77	14.26	11.07	1.29	0.26	2.62	–15.21
GP	–27.81	17.03	–1.63	–2.70	–1.73	–26.21	15.63	11.07	1.41	0.35	3.08	–16.01
Boost	–31.93	17.03	–1.87	–2.76	–1.95	–30.85	12.73	11.07	1.15	0.16	2.41	–14.50
PW	–31.18	17.03	–1.83	–2.80	–1.86	–29.23	14.71	11.07	1.33	0.30	3.02	–14.74
CU	–20.25	17.03	–1.19	–2.60	–1.28	–22.22	12.98	11.07	1.17	0.17	2.22	–15.10
CI	–20.35	17.03	–1.19	–2.32	–1.36	–21.51	12.22	11.07	1.10	0.13	2.18	–14.11
ICI	–21.49	17.03	–1.26	–2.39	–1.40	–22.16	16.73	11.07	1.51	0.43	3.21	–15.57
	2005						**2006**					
S&P 500 (TR)	2.37	10.26	0.23	–3.66	0.33	–9.87	10.49	10.02	1.05	–3.14	1.59	–12.93
ARIMA	3.65	10.27	0.36	–0.24	0.52	–9.98	–3.92	10.02	–0.39	–1.17	–0.58	–15.33
GP	–0.49	10.27	–0.05	–0.57	–0.07	–11.40	–2.43	10.02	–0.24	–1.03	–0.36	–14.76
Boost	1.49	10.27	0.15	–0.42	0.22	–11.54	–4.06	10.02	–0.41	–1.14	–0.58	–15.78
PW	0.10	10.27	0.01	–0.53	0.01	–12.06	–6.36	10.02	–0.63	–1.33	–0.87	–16.78
CU	1.58	10.27	0.15	–0.44	0.23	–10.15	3.40	10.02	0.34	–0.59	0.61	–13.37
CI	1.62	10.27	0.16	–0.41	0.22	–10.65	–2.07	10.02	–0.21	–1.02	–0.32	–14.31
ICI	1.23	10.27	0.12	–0.45	0.17	–9.78	1.52	10.02	0.15	–0.74	0.27	–12.77
	2007						**2008**					
S&P 500 (TR)	2.07	15.96	0.13	–4.06	0.17	–10.66	–39.34	40.88	–0.96	–0.77	–1.31	–47.78
ARIMA	–1.06	15.96	–0.07	–0.33	–0.10	–12.74	17.00	40.88	0.42	1.12	0.60	–42.30
GP	–1.52	15.96	–0.10	–0.35	–0.14	–13.27	46.05	40.88	1.13	1.69	1.83	–40.38
Boost	–3.80	15.96	–0.24	–0.48	–0.35	–15.23	39.48	40.88	0.97	1.34	1.56	–43.34
PW	1.83	15.96	0.11	–0.21	0.18	–14.36	56.22	40.88	1.38	1.81	2.61	–46.29
CU	–13.65	15.96	–0.86	–0.87	–1.09	–17.35	67.57	40.88	1.65	2.27	3.13	–51.49
CI	–3.35	15.96	–0.21	–0.44	–0.31	–11.66	47.33	40.88	1.16	1.79	2.16	–42.41
ICI	–0.99	15.96	–0.06	–0.31	–0.08	–9.66	60.76	40.88	1.49	2.33	2.91	–49.16

EXHIBIT A3

Performance Metrics, 2009–2021

Portfolio Method	Cuml Ret	Ret Vol	Sharpe	IR	Sortino	Max DD	Cuml Ret	Ret Vol	Sharpe	IR	Sortino	Max DD
	2009						2010					
S&P 500 (TR)	27.09	27.22	1.00	–0.83	1.45	–41.01	15.47	18.02	0.86	–1.81	1.23	–19.54
ARIMA	17.54	27.22	0.64	–0.13	1.13	–21.65	–41.11	18.02	–2.28	–2.30	–2.19	–37.28
GP	7.78	27.22	0.29	–0.43	0.48	–18.56	–42.99	18.02	–2.38	–2.42	–2.33	–38.98
Boost	–33.40	27.22	–1.23	–1.76	–1.51	–35.25	–33.99	18.02	–1.89	–2.07	–1.97	–33.55
PW	–9.72	27.22	–0.36	–1.00	–0.55	–19.77	–36.86	18.02	–2.04	–2.17	–2.06	–34.57
CU	19.15	27.22	0.70	–0.08	1.30	–21.64	–36.30	18.02	–2.01	–2.14	–2.03	–35.08
CI	25.99	27.22	0.95	0.12	1.52	–31.85	–37.32	18.02	–2.07	–2.14	–1.98	–34.19
ICI	27.58	27.22	1.01	0.17	1.74	–27.00	–38.24	18.02	–2.12	–2.16	–1.99	–33.98
	2011						2012					
S&P 500 (TR)	4.78	23.24	0.21	0.83	0.28	–18.64	15.63	12.73	1.24	–0.23	1.89	–14.23
ARIMA	–29.76	23.24	–1.28	–1.08	–1.35	–34.76	22.54	12.73	1.77	0.46	3.23	–19.67
GP	–29.17	23.24	–1.26	–1.03	–1.33	–33.83	29.53	12.73	2.32	0.86	5.28	–25.51
Boost	–8.30	23.24	–0.36	–0.41	–0.45	–18.13	28.89	12.73	2.27	0.83	5.22	–25.12
PW	–8.41	23.24	–0.36	–0.42	–0.48	–16.08	25.43	12.73	2.00	0.65	4.23	–24.06
CU	–15.96	23.24	–0.69	–0.56	–0.80	–29.10	15.64	12.73	1.23	0.09	1.79	–15.94
CI	–34.31	23.24	–1.48	–1.19	–1.53	–32.76	19.15	12.73	1.50	0.28	2.45	–18.43
ICI	–26.26	23.24	–1.13	–0.92	–1.18	–33.43	18.62	12.73	1.46	0.25	2.45	–16.55
	2013						2014					
S&P 500 (TR)	28.68	11.05	2.60	–1.62	3.96	–22.78	13.48	11.33	1.19	1.09	1.67	–18.27
ARIMA	2.61	11.05	0.24	–1.74	0.34	–11.82	8.11	11.33	0.72	–0.35	1.74	–10.11
GP	–4.37	11.05	–0.40	–2.45	–0.51	–16.53	7.28	11.33	0.64	–0.42	1.33	–8.94
Boost	7.30	11.05	0.66	–1.47	0.96	–8.04	16.40	11.33	1.45	0.18	3.85	–15.84
PW	–1.12	11.05	–0.10	–2.36	–0.14	–14.72	14.34	11.33	1.27	0.05	3.14	–14.43
CU	–11.24	11.05	–1.02	–2.96	–1.17	–19.21	1.81	11.33	0.16	–0.79	0.29	–7.87
CI	4.64	11.05	0.42	–1.57	0.63	–9.25	6.04	11.33	0.53	–0.47	1.26	–9.70
ICI	–3.97	11.05	–0.36	–2.40	–0.47	–15.92	7.34	11.33	0.65	–0.39	1.43	–9.50
	2015						2016					
S&P 500 (TR)	2.57	15.46	0.17	0.91	0.23	–12.04	11.94	13.07	0.91	–1.32	1.30	–20.80
ARIMA	29.50	15.46	1.91	1.56	2.82	–24.32	–0.11	13.07	–0.01	–0.90	–0.01	–11.81
GP	25.13	15.46	1.62	1.34	2.52	–24.29	0.75	13.07	0.06	–0.84	0.11	–12.31
Boost	28.98	15.46	1.87	1.47	3.17	–26.31	–3.25	13.07	–0.25	–1.08	–0.39	–10.62
PW	28.45	15.46	1.84	1.56	2.77	–26.20	–11.84	13.07	–0.91	–1.60	–1.22	–12.92
CU	31.35	15.46	2.03	1.83	2.96	–26.59	–1.96	13.07	–0.15	–1.01	–0.25	–12.55
CI	27.01	15.46	1.75	1.45	2.56	–22.93	2.62	13.07	0.20	–0.73	0.38	–13.80
ICI	22.50	15.46	1.46	1.12	2.42	–23.22	6.23	13.07	0.48	–0.52	1.03	–13.95
	2017						2018					
S&P 500 (TR)	19.16	6.67	2.89	–1.04	4.55	–17.04	–4.81	17.03	–0.28	–0.77	–0.37	–19.81
ARIMA	–1.08	6.67	–0.16	–2.14	–0.24	–7.76	4.96	17.03	0.29	0.50	0.40	–14.97
GP	0.06	6.67	0.01	–2.03	0.02	–5.84	–0.59	17.03	–0.03	0.26	–0.04	–14.22
Boost	–3.86	6.67	–0.58	–2.50	–0.79	–10.74	6.59	17.03	0.39	0.61	0.56	–16.52
PW	–7.71	6.67	–1.16	–2.99	–1.41	–12.44	13.39	17.03	0.79	0.87	1.17	–17.03
CU	–0.88	6.67	–0.13	–2.16	–0.20	–6.45	7.20	17.03	0.42	0.57	0.57	–12.90
CI	2.51	6.67	0.38	–1.75	0.61	–6.02	2.18	17.03	0.13	0.37	0.17	–12.58
ICI	4.71	6.67	0.71	–1.48	1.32	–4.78	–5.87	17.03	–0.34	0.05	–0.44	–14.05

(continued)

EXHIBIT A3 *(continued)*

Performance Metrics, 2009–2021

Portfolio Method	Cuml Ret	Ret Vol	Sharpe	IR	Sortino	Max DD	Cuml Ret	Ret Vol	Sharpe	IR	Sortino	Max DD
	2019						2020					
S&P 500 (TR)	26.03	12.45	2.09	−0.79	3.02	−24.34	22.51	34.37	0.65	−2.26	0.90	−41.22
ARIMA	22.69	12.45	1.37	−0.16	3.17	−22.74	17.36	34.37	0.51	−0.17	0.61	−38.09
GP	20.72	12.45	1.66	−0.30	3.69	−21.68	−1.95	34.37	−0.06	−0.59	−0.07	−36.61
Boost	23.24	12.45	1.87	−0.15	4.15	−20.27	26.22	34.37	0.76	0.03	1.01	−39.32
PW	16.83	12.45	1.35	−0.50	3.33	−19.62	22.99	34.37	0.67	−0.04	0.86	−37.70
CU	17.75	12.45	1.43	−0.50	2.73	−19.82	3.84	34.37	0.11	−0.47	0.13	−36.68
CI	23.50	12.45	1.89	−0.15	3.68	−21.02	8.55	34.37	0.25	−0.36	0.29	−36.89
ICI	19.15	12.45	1.54	−0.42	2.47	−19.28	8.74	34.37	0.25	−0.36	0.30	−36.94
	2021											
S&P 500 (TR)	26.10	13.07	2.00	1.55	2.97	−23.86						
ARIMA	−5.78	13.07	−0.46	−1.65	−0.60	−12.93						
GP	−2.48	13.07	−0.20	−1.47	−0.27	−10.21						
Boost	6.64	13.07	0.53	−0.81	0.84	−12.56						
PW	−5.24	13.07	−0.42	−1.55	−0.53	−12.61						
CU	−0.34	13.07	−0.03	−1.37	−0.04	−9.50						
CI	−1.12	13.07	−0.09	−1.34	−0.13	−9.80						
ICI	−2.15	13.07	−0.17	−1.40	−0.24	−10.50						

estimated for the white-noise kernel. By Equation A3, the epistemic uncertainty is then the difference between the total uncertainty and the aleatoric uncertainty estimates.

We implement a boosted regression model making use of CatBoost (Prokhorenkova et al. 2018). We refit the model on a bimonthly basis on a rolling data window of two years. The model yields an aleatoric uncertainty estimate owing to modeling this uncertainty directly via the loss function. An estimate for epistemic uncertainty is approximated by the variance of the underlying ensemble of predictions. Further details of this method can be found in Ustimenko, Prokhorenkova, and Malinin (2020).

We estimate an ARIMA model for view generation as a well-known baseline model. Further, this model naturally yields an aleatoric uncertainty estimate. To approximate epistemic uncertainty for this model, we are inspired by Algorithm 3 of Lahlou et al. (2022). We note that our method is somewhat ad hoc and has subtle differences: We estimate prediction error variance on a recent out-of-sample lookback window of size six months, estimating epistemic uncertainty as the difference between this term and today's aleatoric uncertainty estimate, to a minimum of 1e-8. We refit the model on a bimonthly basis on a rolling data window of one year.

Performance Results by Year

Performance metrics (Exhibits A1–A3) for market total return (in excess of the risk-free rate), against a collection of Black–Litterman models with varying underlying prediction methods to generate and/or fuse views. Methods are normalized by return volatility to share equal ex post annual return volatility of the market index.

ACKNOWLEDGMENTS

The authors thank the Oxford-Man Institute of Quantitative Finance for its generous support. Stephen Roberts thanks the Royal Academy of Engineering. Trent Spears thanks Professor Nir Vulkan and Dr. Jan-Peter Calliess for their helpful comments and suggestions.

REFERENCES

Asness, C. S., T. J. Moskowitz, and L. H. Pedersen. 2013. "Value and Momentum Everywhere." *The Journal of Finance* 68 (3): 929–985.

Bar-Shalom, Y. 1981. "On the Track-to-Track Correlation Problem." *IEEE Transactions on Automatic Control* 26 (2): 571–572.

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

Bochardt, O., R. Calhoun, J. K. Uhlmann, and S. J. Julier. 2006. "Generalized Information Representation and Compression Using Covariance Union." In *Proceedings of the 2006 9th International Conference on Information Fusion*, Florence, Italy, 10–13 July 2006, pp. 1–7, doi: 10.1109/ICIF.2006.301773.

Castanedo, F. 2013. "A Review of Data Fusion Techniques." *The Scientific World Journal* 2013: 1–19.

Chen, L., P. Arambel, and R. Mehra. 2002. "Fusion Under Unknown Correlation—Covariance Intersection as a Special Case." In *Proceedings of the Fifth International Conference on Information Fusion. FUSION 2002*, Annapolis, MD, USA, 08–11 July 2002, 2: 905–912, doi: 10.1109/ICIF.2002.1020908.

Date, P., and K. Ponomareva. 2010. "Linear and Non-Linear Filtering in Mathematical Finance: A Review. *IMA Journal of Management Mathematics* 21: 195–211.

Dempster, A. P. 1969. *Elements of Continuous Multivariate Analysis*. Reading, Mass: Addison-Wesley Pub. Co.

Efron, B. 1987. "Better Bootstrap Confidence Intervals." *Journal of the American Statistical Association* 82 (397): 171–185.

Fabozzi, F. J., S. M. Focardi, and P. N. Kolm. 2006. "Incorporating Trading Strategies in the Black–Litterman Framework." *The Journal of Trading* 1 (2): 28–37.

Frazzini, A., R. Israel, and T. J. Moskowitz. 2018. "Trading Costs." SSRN 3229719.

Gârleanu, N., and L. H. Pedersen. 2013. "Dynamic Trading with Predictable Returns and Transaction Costs." *The Journal of Finance* 68 (6): 2309–2340.

Giglio, S., B. Kelly, and D. Xiu. 2022. "Factor Models, Machine Learning, and Asset Pricing." *Annual Review of Financial Economics* 14 (1): 337–368.

Hüllermeier, E., and W. Waegeman. 2021. "Aleatoric and Epistemic Uncertainty in Machine Learning: An Introduction to Concepts and Methods." *Machine Learning* 110: 457–506.

Jensen, T. I., B. T. Kelly, S. Malamud, and L. H. Pedersen. 2022. "Machine Learning and the Implementable Efficient Frontier." Research paper no. 22-63, Swiss Finance Institute.

Julier, S. J., and J. K. Uhlmann. 1997. "A Non-Divergent Estimation Algorithm in the Presence of Unknown Correlations." In *Proceedings of the 1997 American Control Conference*, Albuquerque, NM, USA, 06–06 June 1997, 4: 2369–2373. doi: 10.1109/ACC.1997.609105.

Julier, S. J., J. K. Uhlmann, and D. Nicholson. 2004. "A Method for Dealing with Assignment Ambiguity." In *Proceedings of the 2004 American Control Conference* 5: 4102–4107.

Khaleghi, B., A. Khamis, F. O. Karray, and S. N. Razavi. 2013. "Multisensor Data Fusion: A Review of the State-of-the-Art." *Information Fusion* 14 (1): 28–44.

Kolm, P. N., and G. Ritter. 2017. "On the Bayesian Interpretation of Black–Litterman." *European Journal of Operational Research* 258 (2): 564–572.

Kolm, P. N., G. Ritter, and J. Simonian. 2021. "Black–Litterman and Beyond: The Bayesian Paradigm in Investment Management." *The Journal of Portfolio Management* 47 (5): 91–113.

Kuntsevich, A., and F. Kappel. 1997. "SolvOpt: The Solver for Local Nonlinear Optimization Problems." Institute for Mathematics, University of Graz.

Lahlou, S., M. Jain, H. Nekoei, V. Butoi, P. Bertin, J. Rector-Brooks, M. Korablyov, and Y. Bengio. 2022. "DEUP: Direct Epistemic Uncertainty Prediction." *arXiv* 2102.08501.

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

Maybeck, P. S. 1979. *Stochastic Models, Estimation, and Control*, Vol. 1. New York: Academic Press.

Murphy, K. P. 2012. *Machine Learning: A Probabilistic Perspective*. Cambridge, Mass: MIT Press.

Noack, B., J. Sijs, and U. Hanebeck. 2017. "Inverse Covariance Intersection: New Insights and Properties." In *Proceedings of the 2017 20th International Conference on Information Fusion (Fusion)*, Xi'an, China, 10–13 July 2017, pp. 1–8, doi: 10.23919/ICIF.2017.8009694.

Noack, B., J. Sijs, M. Reinhardt, and U. D. Hanebeck. 2017. "Decentralized Data Fusion with Inverse Covariance Intersection." *Automatica* 79: 35–41.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. "Scikit-Learn: Machine Learning in Python." *Journal of Machine Learning Research* 12: 2825–2830.

Prokhorenkova, L., G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin. 2018. "Catboost: Unbiased Boosting with Categorical Features." In *Proceedings of the 32nd International Conference on Neural Information Processing Systems: (NeurIPS 2018)*, Montreal, Canada, 03–08 December 2018, pp. 6639–6649. Red Hook, NY: Curran Associates, Inc.

Reece, S., and S. Roberts. 2010. "Generalised Covariance Union: A Unified Approach to Hypothesis Merging in Tracking." *IEEE Transactions on Aerospace and Electronic Systems* 46: 207–221.

Roberts, S., M. Osborne, M. Ebden, S. Reece, N. Gibson, and S. Aigrain. 2013. "Gaussian Processes for Time-Series Modelling." *Philosophical Transactions: Series A, Mathematical, Physical, and Engineering Sciences* 371: 20110550.

Ross, S. A. 1976. "The Arbitrage Theory of Capital Asset Pricing." *Journal of Economic Theory* 13 (3): 341–360.

Satchell, S., and A. Scowcroft. 2000. "A Demystification of the Black–Litterman Model: Managing Quantitative and Traditional Portfolio Construction." *Journal of Asset Management* 1: 138–150.

Sharpe, W. F. 1964. "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk." *The Journal of Finance* 19 (3): 425–442.

Sijs, J., and M. Lazar. 2012. "State Fusion with Unknown Correlation: Ellipsoidal Intersection." *Automatica* 48 (8): 1874–1878.

Thorp, E. O. 2006. "The Kelly Criterion in Blackjack, Sports Betting, and the Stock Market." *Handbook of Asset and Liability Management* 1: 385–428.

Uhlmann, J. K. "Dynamic Map Building and Localization: New Theoretical Foundations." PhD thesis, University of Oxford, 1995.

——. 2003. "Covariance Consistency Methods for Fault-Tolerant Distributed Data Fusion." *Information Fusion* 4: 201–215.

Ustimenko, A., L. Prokhorenkova, and A. Malinin. 2020. "Uncertainty in Gradient Boosting via Ensembles." *arXiv* 2006.10562.

van der Schans, M., and H. Steehouwer. 2017. "Time-Dependent Black–Litterman." *Journal of Asset Management* 18: 371–387.

Walters, J. 2014. "The Black–Litterman Model in Detail." SSRN 1314585.

Wilcoxon, F. 1945. "Individual Comparisons by Ranking Methods." *Biometrics Bulletin* 1 (6): 80–83.

Willner, D., C.-B. Chang, and K.-P. Dunn. 1976. "Kalman Filter Algorithms for a Multi-Sensor System." In *Proceedings of the 1976 IEEE Conference on Decision and Control including the 15th Symposium on Adaptive Processes*, Clearwater, FL, USA, 01–03 December 1976, pp. 570–574, doi: 10.1109/CDC.1976.267794.