Generalized Black-Litterman with Decision Fusion

Xinyu Huang* Massimo Guidolin † Emmanouil Platanakis ‡ David P. Newton §

Xiaoxia Ye ¶

January 9, 2023

Highlights:

- The Black-Litterman asset allocation framework is extended with a decision fusion system.
- An entropy-based multiple classifiers combination system in the framework of Dempster-Shafer Theory of Evidence is proposed for modelling investors' views.
- Reference portfolio is an optimal combination of the minimum-variance and 1/N portfolio.
- Dynamic covariance structure is estimated via diversification-based Graphical Lasso with Wishart Stochastic Volatility (WSV) modelling.
- The generalize Black-Litterman asset allocation strategy outperforms all benchmark strategies and outperform 1/N significantly.

Keywords: Black-Litterman model, Decision Fusion, Dempster-Shafer theory, Multivariate stochastic volatility, Portfolio optimization

JEL Classification: G11, G17

*xh357@bath.ac.uk, School of Management, University of Bath
†massimo.guidolin@unibocconi.it, Department of Finance, Bocconi University Milan
‡E.Platanakis@bath.ac.uk, School of Management, University of Bath
§D.P.Newton@bath.ac.uk, School of Management, University of Bath, Corresponding Author
¶Xiaoxia.Ye@liverpool.ac.uk, Management School, University of Liverpool

Generalized Black-Litterman with Decision Fusion

Abstract

The Black–Litterman asset allocation model provides an intuitive insight for combining market (equilibrium) views with investors' views within the Markowitz mean-variance framework, in which three key inputs are used: the reference portfolio, the investors' views, and the covariance matrices. The classical version uses the Capital Asset Pricing Model (CAPM) to calibrate the reference portfolio and assumes investor's views are correctly specified. However, the CAPM anomalies reported in the asset pricing literature suggest that holding a reference portfolio proportional to the market capitalization may be grossly inadequate, and investors often do not have perfect knowledge of the process associated with macro variables or stock returns, to make intelligent estimates of the future asset returns. Furthermore, it has been observed that the deterministic volatility models are too restrictive for many financial time series. In this paper we address three potential weaknesses that lie within the allocation framework. We formulate a decision-fusion model that provides a straightforward way for investors to quantify their uncertain knowledge and ignorance in the estimation of views. The reference portfolio is constructed using a portfolio combination strategy that accounts for the estimation error, and the specification error associated with the CAPM. Finally, we propose a diversification-based GLASSO-Wishart model for the analysis of the realized covariance matrix. Experimenting with four datasets, we construct portfolios whose out-of-sample Sharpe ratio is consistently and significantly higher than the naïve diversification, in the absence and presence of transaction costs. The newly obtained strategy outperforms eleven well-studied benchmark strategies in terms of risk-reward maximization.

1 Introduction

Selecting portfolios of financial assets to maximise returns for a given level of risk begins with the theoretically rather beautiful mean-variance framework laid down by Markowitz (1952) In his setup, the two key inputs for optimization are the means and the covariance matrix of asset returns. However, the implementation of the mean-variance framework is practically hindered by the need to estimate these inputs based on the time-series of historical returns. It is now well-known that, because of the resulting sampling errors, mean-variance portfolios frequently perform poorly out-of-sample. Optimized mean-variance portfolios often imply extreme long and short positions unless constraints are added. Furthermore, the optimized weights under the mean-variance framework based on the sample estimators of the first two moments are often highly concentrated, especially when no-short-selling constraints are absent (Best and Grauer (1991), Green and Hollifield (1992). In fact, it has been widely recognized that the benefits of optimal diversification may be completely eliminated by parameter estimation uncertainty (Board and Sutcliffe (1994), DeMiguel et al. (2009)). Nevertheless, progress has been made following Black and Litterman (1992), who allow the investors to blend market views (statistical information on asset returns) with investors' views (private information) in the framework of Markowitz's optimization. In this paper, we extend this line of research in addressing the weaknesses of the classical Markowitz's model and propose three methodological contributions that improve the performance of the Black and Litterman approach.

The original Black-Litterman model (henceforth, BLM) is an intellectual milestone in two respects. First, it provides a useful prior, the market portfolio, as a starting point for the estimation of asset expected returns. Views of the market portfolio, which are typically taken as the equilibrium expected returns calibrated via the Capital Asset Pricing Model (CAPM), can summarize non-sample information and are less sensitive to parameter estimation uncertainty than estimations based on time-series of returns. Second, it allows judgemental investors to provide views or forecasts on asset returns, which are combined with the equilibrium expected returns to reflect a mingling of statistical information and private information on asset returns. Investors should hold the capitalization-weighted portfolio when their views are consistent with the market. When investors have proprietary views on the absolute or relative performance of the assets, they should deviate from the market portfolio, according to the strength of their views. The blending of viewpoints enhances the process for return estimation. Note that in the original BLM, the equilibrium expected returns are obtained by an inverse optimization, which is essentially a function of degree of risk-aversion, covariance matrix, and market capitalization weights, rendering the estimation of covariance matrix as equally important for practical applications. Because of its rich theoretical foundation, BLM has led to the development of various modern portfolio selection framework (He and Litterman (2002), Idzorek (2007), Bertsimas et al. (2012), Walters (2014), Bessler et al. (2017), Platanakis et al. (2021)).

Within BLM, the process for discovering investors' views is very flexible. BLM does not require investors to provide return forecasts on all assets managed in the investment universe. The views can be expressed as absolute or relative performances and the ways of formulating them are miscellaneous. A commonly used quantitative approach is the factor view blending framework proposed by Fabozzi et al. (2006), where views are taken as the expected returns of long-short portfolios generated by using a cross-sectional momentum strategy. The framework was extended by Jones et al. (2007), Giacometti and Mignacca (2010), Harris et al. (2017), Kolm and Ritter (2020) and was also generalized by Cheung (2013), who proposes a ranking-free factor-mimicking approach with general distributions. A different approach presented in Beach and Orlov (2007) uses GARCH-derived views to capture stylized facts about asset returns (e.g., thick tails, volatility clustering) and dynamics of the central location measure and variances of returns. Fernandes et al. (2018) propose an autoregressive model with past returns and price-earnings data to update the conditional probability distribution of asset returns. Based on properties of an arbitrary return distribution, Xiao and Valdez (2015) extend Meucci (2005)'s market portfolio approach assuming views follow a generalized elliptical distribution. Meucci (2006) proposes a Copula Opinion Pooling approach to derive non-normal views, however, as point out in Palczewski and Palczewski (2019), the approach lacks a formal statistical background and cannot nest the original BLM for normal distribution of returns.

The literature of BLM has so far relied primarily on trading strategies and on the distributional properties of financial returns to estimate investors' views. Nevertheless, investors generally do not have perfect knowledge of the process associated with macroeconomic variables, or the stock returns, but instead need to make intelligent estimates on the future asset returns using all the information available to them (Ozoguz (2009)). Several studies show that investors' uncertainty over factors related to the state of economy (uncertain knowledge) can affect asset returns (Segal et al. (2015)). On the other hand, the exogenous characteristics of investors imply a possibility in which: (i) investors are unable to make forecasts based on the macroeconomic variables and historical data of asset returns, due to insufficient knowledge (referred as 'ignorance' in this study) and (2) investors may receive conflicting information from the data on which forecasts are based. In short, one should allow for uncertainty and ignorance over the information available to investors in the process of return estimation. A potential shortcoming of BLM is that the relationship between uncertain knowledge and the views has been a missing link; indeed, the model assumes that investors' views are correctly specified and then downweighs the forecasting information available in the views using the confidence levels. We notice that subsequent studies have largely ignored the role of investors, their uncertain knowledge and ignorance, though an exception is presented by Silva et al. (2017), who suggest a qualitative approach to incorporate investors' profile by using Verbal Decision Analysis.

The original BLM assumes that the vectors of equilibrium expected returns are correctly specified, which is generally not the case. For example, the equilibrium relationship hinges on the validity of CAPM, yet CAPM is empirically rejected (Zhou (2009)). The impact of specification error in the CAPM has long been recognized. This can be traced back to Roll (1977), who shows that the CAPM is not testable. Gibbons et al. (1989), Zhou (1991) and Engel and Rodrigues (1993) find the CAPM rejected by data and risk-aversion estimates derived from it are usually insignificantly different from zero. Thus, holding an equilibrium portfolio proportional to market capitalization is liable to be misleading. As suggested by Chen and Lim (2020) and Kolm et al. (2021), there are a number of factors in addition to the market factor that are pivotal for the derivation of the equilibrium expected returns. From an analytical point of view, Zhou (2009) argues that a capitalization-weighted portfolio is not sufficient to inform the prior because the historical asset returns contain beneficial information about future asset returns and dynamics of asset returns (regime change) which the equilibrium model completely overlooks. This leads to an incorporation of historical asset returns and a data-generating process into the BLM. The intuition is supported by Geyer and Lucivjanská (2016), who combine the equilibrium model with the expected returns from a predictive regression to form a data-driven prior. Subsequently, Kolm and Ritter (2020) propose a Bayesian extension to incorporate views and priors on factor risk premiums. Kolm et al. (2021) summarize literature contributions on the development of a data-driven prior.

BLM also assumes that asset returns follow multivariate normal distribution and have a constant conditional distribution. Both hypotheses are unrealistic when applied to financial asset returns. Ferson et al. (1987) and Jr et al. (1991) argue that investors perceive a time-varying conditional expectation of returns. Harvey (1989) shows that conditional covariances do change over time and there is overwhelming empirical evidence showing that return distributions are far from being Normal (Ang and Chen (2002)). Many studies generalize asset returns beyond the Gaussian (Normal) distribution (Meucci (2006); Giacometti et al. (2007); Xiao and Valdez (2015); Pang and Karan (2018); Palczewski and Palczewski (2019)). Nevertheless, little attention has been paid to relax the assumption of constant conditional distribution. Palomba (2008) constructs tracking error constrained BL portfolios in which conditional expected returns and covariance matrix are estimated by a Flexible Dynamic Conditional Correlation model. Harris et al. (2017) generalize the BLM to a fully-dynamic environment, adjusting covariance estimates using an Exponential Weighted Moving Average model and a Dynamic Conditional Correlations model. An alternative approach is presented by Sahamkhadam et al. (2021), who model dynamics in the realized covariance matrix using the joint distribution obtained from a GARCH-copula model to capture the asymmetries in the tail distribution. However, none of these studies allow for a stochastic variation in the conditional covariances of returns. It is well-documented that financial markets exhibit very complex dynamics, and the stylized facts are non-trivial; for example, fat-tailed distribution of the conditional mean innovations and leverage effect via the correlation between volatility and mean innovations (Jacquier et al. (2004)). It has been observed that the deterministic models are too restrictive for many financial time series (Harvey et al. (1994); Fridman and Harris (1998)). Consequently, stochastic volatility models receive much attention for portfolio selection problems owing to their greater flexibility over the GARCH family, analytical tractability, and parsimonious use of free parameters (see Moura et al. (2020)). We note, however, that many existing refinements of BLM stay closely to sample covariance or multivariate GARCH models and discussion around the stochastic volatilities is rather limited.

The objective of this paper is to address the three shortcomings, above, contained within BLM. First, we propose a portfolio combination strategy for the estimation of equilibrium expected returns. The performance of the new reference portfolio is strictly related to the prespecified benchmarks: naïve diversification, and the minimum-variance portfolio. Given that the strategy mix is a weighted-average of the two benchmarks and is derived from perceptions of a future risk-reward ratio, this leads investors to consistently update the prior via data and modify their existing holdings towards assets that they favour. Second, we characterize views as solutions to a decision-fusion system. We assume that investors hold a set of hypotheses on future market conditions and then narrow down their hypothesis set with the accumulation of evidence from various sources. The fusion system, which is overlaid with the Dempster-Shater Theory of Evidence (Dempster (1968); Shafer (1976)), can exploit complementary knowledge from different classifiers, taking advantage of the correct decisions produced by each classifier and dealing with conflicting decisions, eventually leading to more reliable views on the market. Finally, to reduce the sensitivity of our approach to parameter uncertainty, we propose a variation of the Wishart Stochastic Volatility model of Uhlig (1994), Uhlig (1997), which we call the GLASSO-Wishart model. Specifically, we restrict the covariance matrix of the Wishart distribution to a sparse parameterization using the Graphical LASSO (GLASSO) approach, in which the shrinkage is proportional to the diversification distribution among assets.

Our paper makes three methodological contributions to the literature on BLM. First, we propose using an entropy-based decision fusion system to identify investors' views. The novelty of the decision fusion system relative to the extant literature on BLM is to provide a straightforward way for investors to quantify uncertain knowledge and ignorance in a decision-making scenario. The framework accounts specifically for the exogenous characteristics of investors, e.g., the possibility that one cannot assign beliefs (a measure of support) to any atomic hypothesis due to insufficient information being available. In this case, the functions and the combination rules of the Dempster-Shafer Theory of Evidence enable investors to represent their lack of knowledge, without forcing them to overcommit when they are ignorant. In this paper, the evidence we obtain for one hypothesis is the individual belief produced by each classifier, which is a function of: the Euclidean distance between the reference value and the class probability estimated by each classifier; and the entropy-based weight, derived from the wealth of information each classifier has accessed. For each hypothesis, the individual beliefs produced by each of the classifiers are aggregated into a uniform one, representing the combination of confidences in this decision. We provide explicit expressions for combining the individual beliefs (evidence), and select parameters using Adaptive Moment Estimation. Finally, we show how the fused decision can be converted into the future asset returns. To our knowledge, this is the first study that generalizes BLM to incorporate a decision-fusion process.

Second, in our Bayesian setup, views of the reference portfolio are obtained through a novel portfolio combination technique. This is inspired by a series of papers by Kan and Zhou (see Kan and Zhou (2007), Kan et al. (2021)) where they derive optimal portfolio combination rules that maximize the expected utility under estimation risk. The intuition underlying their model is to diversify the estimation risk by holding a combination of risky portfolios (with a risk-free asset). In this study, we characterize the equilibrium model as an adaptive data-driven framework for which we provide explicit formulas obtaining the reference portfolio weights. We design two intuitive priors: the equally-weighted and the minimum-variance portfolio (portfolios relying on the expected returns are disregarded since the true expected returns are difficult to estimate and unduly affect the portfolio weights). For each prior, we determine the future value (reward-to-risk) of this prior by using sparse regression forecasts (LASSO, Tibshirani (1996)) and jointly combine the two priors into a new Bayesian prior (the reference portfolio). This leads investors to consistently modify existing holdings with data by selecting allocations which are derived from the historical sample mean, sample volatilities, and the expected reward-to-risk. Our approach, therefore, connects to quantitative asset allocation (e.g., Brodie et al. (2009)) and Bayesian decision making models.

Third, we introduce a diversification-based GLASSO-Wishart model, for analysis of the realized covariance matrices of asset returns, and show how estimation risk can be reduced by restricting the matrix of a Wishart distribution to a sparse parameterization with L-1 penalty. The new generalization follows from the Wishart Stochastic Volatility (WSV) model, which possess three key benefits. First, it naturally guarantees positive definiteness of the covariance matrix. Second, the model is well-suited for handling large systems of asset returns because the dynamics of WSV are controlled by just a single factor. Third, by imposing such a single factor structure, estimation errors can be reduced since fewer parameters are to be estimated. These advantages were first documented in the academic literature by Uhlig (1994), Uhlig (1997) and have thereafter been shown to be present in a high-dimensional data setting by Moura et al. (2020). To account further for estimation risk, we impose sparse parameterization on the Wishart specification based on the Graphical LASSO (GLASSO) model. Our choice of GLASSO is motivated by its computational efficiency and interpretability for interdependencies between assets. The choice of regularization parameter relies on a popular heuristic optimizer, the Genetic Algorithm. Our approach relates to a vast literature on the sparse covariance estimation (see Banerjee et al. (2008); Friedman et al. (2008)). However, our paper differs in that we introduce heterogenous regularizations based on the diversification distribution among assets.

We perform out-of-sample back-testing based on four Fama-French datasets that are typically used in the empirical asset management literature: the 17 industry portfolios, the 30 industry portfolios, the 10 size and short-term reversal portfolios and the 25 size and short-term reversal portfolios for the US equity market. For all datasets examined, we find a consistent and significant increase in Sharpe ratio compared with the 1/N strategy, both in the absence and presence of transaction costs. In terms of risk-reward maximization, we observe GBL outperforming the classic Black-Litterman, Markowitz mean-variance, and various allocation strategies that have been proposed in the literature to manage estimation risk, including the covariance shrinkage technique of Ledoit and Wolf (2003), the Bayes-Stein shrinkage rules of Jorion (1986), the non-optimization based rules of risk-parity and reward-to-risk timing rule of Kirby and Ostdiek (2012). The performance of GBL is also more stable compared with all benchmark strategies when short positions are allowed. Finally, relative to the GBL portfolios adopting an optimal portfolio combination or GLASSO-Wishart covariances alone, we find that GBL adopting both strategies perform better, particularly for datasets with a relatively large number of risky assets. Since incorporating more risky assets involves a higher level of estimation risk, our findings suggest that, by using an optimal combining strategy together with the GLASSO-Wishart model for generating BLM's inputs, we are able to reduce the sensitivity of the optimization to the estimation errors.

To check the sensitivity of our results to the different assumptions, we also perform robustness experiments for all allocation strategies and datasets considered after relaxing the assumptions we have made. The exercises include the alternative risk aversion level; alternative estimation window size; alternative reliability measure of equilibrium expected returns; alternative holding period; and alternative choice of base models. The key findings are qualitatively and quantitatively similar to the findings reported in this paper.

The remainder of the paper is structured as follows. Section 2 presents novel methods for modelling the investors' views, the reference portfolio, the time-varying covariance matrix, and the Black-Litterman optimization procedures adapted to incorporate these novel approaches. Section 3 presents the data. The realized, out-of-sample empirical performance of our methods is discussed in Section 4. Section 5 describes a series of robustness checks and section 6 concludes.

2 Methodology

In this section, we present the numerical methods for computing investors' views, reference portfolio, and covariance matrix, i.e., for all the inputs of our GBL methodology. We describe the process of incorporating these three components into our BLM framework. We begin with a short presentation of the BLM to pave the way for our contribution in the estimation of the three inputs.

2.1 Portfolio Optimization

Consider a portfolio choice problem in which an investor chooses the optimal portfolio among N risky assets. Let $\boldsymbol{r}_t$ be the returns of the N risky assets at time t, characterized by a conditional covariance matrix $\boldsymbol{\Sigma}_t$. As explained by Black and Litterman (1992), when CAPM holds, the equilibrium expected returns $\boldsymbol{\pi}_t$ satisfy the following equation:

$$\boldsymbol{\pi}_t = \delta \boldsymbol{\Sigma}_t \boldsymbol{w}_t^{ref}, \tag{1}$$

where δ is the risk-aversion coefficient and $\boldsymbol{w}_t^{ref}$ is the reference portfolio weights. Let $\boldsymbol{V}_t$ be the $\bar{K} \times 1$ vector summarizing $\bar{K}$ (absolute or relative) views at time t, the posterior expected returns $\boldsymbol{\mu}_{BL,t}$ can be obtained by combining the equilibrium expected returns, $\boldsymbol{\pi}_t$, with investors' views using a Bayesian technique, i.e.,

$$\boldsymbol{\mu}_{BL,t} = [(\tau \boldsymbol{\Sigma}_t)^{-1} + \boldsymbol{P}_t' \boldsymbol{\Omega}_t^{-1} \boldsymbol{P}_t]^{-1} [(\tau \boldsymbol{\Sigma}_t)^{-1} \boldsymbol{\pi}_t + \boldsymbol{P}_t' \boldsymbol{\Omega}_t^{-1} \boldsymbol{V}_t], \tag{2}$$

where $\boldsymbol{P}_t$ is a binary matrix (a matrix in which all the elements are either 0 or 1) identifying the number of assets associated with the views at time t. $\boldsymbol{\Omega}_t$ is a non-negative diagonal matrix measuring investors' degree of confidence in their views which, $\boldsymbol{\Omega}_t = \frac{1}{\delta} \boldsymbol{P}_t \boldsymbol{\Sigma}_t \boldsymbol{P}_t^{-1}$, and tau is a scalar parameter representing the reliability of the equilibrium expected returns, i.e., how close $\boldsymbol{\pi}_t$ is anticipated to fall relative to the true expected returns. We follow Bessler et al. (2017) and start with τ equal to 0.025 [1]. The posterior

[1] Bessler et al. (2017) show that Black–Litterman results are robust for the commonly used values for tau between 0.025 and 1.00. We report results for a reliability measure of 0.1625 for robustness checks, which corresponds to the mean of the range of values used by previous studies (see, Black and Litterman (1992), He and Litterman (2002), Idzorek (2007)).

conditional covariance matrix is then given by:

$$\boldsymbol{\Sigma}_t^{BL} = \boldsymbol{\Sigma}_t + [(\tau \boldsymbol{\Sigma}_t)^{-1} + \boldsymbol{P}_t' \boldsymbol{\Omega}_t^{-1} \boldsymbol{P}_t]^{-1} \tag{3}$$

Let $\boldsymbol{w}$ be the weights of a portfolio of the N risky assets. An investor who seeks to maximize the utility function will choose the portfolio allocations $\boldsymbol{w}$ by:

$$Max\ U \equiv \boldsymbol{\mu}_{BL,t}^{\top} \boldsymbol{w} - \frac{\delta}{2} \boldsymbol{w}^{\top} \boldsymbol{\Sigma}_t^{BL} \boldsymbol{w}, \quad s.t. \sum_{i=1}^{N} w_i = 1, \tag{4}$$

where $\boldsymbol{\mu}_{BL,t}$ and $\boldsymbol{\Sigma}_t^{BL}$ are respectively the posterior expected returns and covariance matrix. The budget constraint $\sum_{i=1}^{N} w_i = 1$ indicates that investors need to have 100% of the wealth invested in the risky assets.

2.2 Subjective returns modelling

This section details our approach for modelling investors' views. We begin with the theoretical foundations and intuition of Dempster-Shafer (henceforth, DS) Theory of Evidence, Then, we proceed to present the numerical methods of the decision-fusion approach.

The expression of investor's views can be interpreted as a decision-making process. As Evangelou et al. (2005) claimed, knowledge of the current situation is a prerequisite for any prospective decision whereby good decisions should be accomplished by a shared understanding of various knowledge domains. Silva et al. (2017) also indicate that the knowledge acquired from analyzing the objects from various perspectives could well enhance the confidence with which decisions are taken. Motivated by these earlier studies, we link BLM to a decision-making scenario and collect knowledge from multiple classifiers. In particular, we consider a process of fusion of classifier decisions which estimates a belief that next period's market condition belongs to a particular class {bullish, bearish, very bullish, very bearish}.

The classifier-combination framework presented in Al-Ani and Deriche (2002) uses a a cost function to minimize the mean square error between the predictions and the target classification outputs. We follow this line of work and develop a refined model within BLM. First, we perform feature engineering to select a set of features that are potentially useful for predicting the future performance of the stock market. This includes a set of macroeconomic indicators, financial indicators, and time-series of past returns, which have been tested and confirmed in recent literature (see, Flannery and Protopapadakis (2002), Welch and Goyal (2008), Rapach et al. (2010), and Gu et al. (2020)). Second, we apply feature-subsampling and apply each individual classifier to each set of sub-features. We combine decisions of such sub-feature classifiers while each of them makes a decision based on its sub-feature. The idea behind this approach is to enhance the diversity between classifiers and to improve the generalization of the outcome with a better bias-variance trade-off [2]. According to Large et al. (2019), for a multi-classifier combination to be strong and be able to provide a good way of achieving the near-optimal performance, it needs to contain classifiers that are good for making decisions in areas of their expertise but do not unduly overlap. By training each classifier on the sub-features, each classifier can access different sources of information in the data and allow investors to exploit the complementary knowledge provided by various classifiers. In Al-Ani and Deriche (2002), evidence from each classifier was collected from the same source of information and the contribution of each classifier is not disentangled. Our paper suggests using a weighting-scheme, which we perform an assessment for all features based on Shannon's entropy and summarize the wealth of information provided by each feature subset into one single index, the information volume (IV). The index identifies the amount of information available to each classifier, and is a key input in our model for constructing the evidence. Finally, we replace the gradient descent method used by Al-Ani and Deriche (2002) with the Adaptive Moment Estimation to

[2] Brown et al. (2005) argue that it is possible to reformulate any classification problem as a regression one by choosing to approximate the class posterior probabilities and show that mean squared error of an ensemble estimator depends critically on the amount of error correlated between individual classifiers.

optimize parameters in the system.

2.2.1 Dempster-Shafer Theory of Evidence

Originally proposed by Dempster (1968) and completed by Shafer (1976), the Dempster-Shafer Theory of Evidence has remained active for modelling and processing information uncertainty after years of investigation. Practical applications involves pattern recognition, transportation solution, and network system. In this section, we present the fundamentals of this framework.

Let Θ be the set of mutually exclusive set of atomic hypotheses, $\Theta = \{\theta_1, \theta_2, ..., \theta_K\}$, Θ is named the frame of discernment with a corresponding power set 2^Θ, and $\varnothing$ denotes the empty set:

$$2^\Theta = \{\varnothing, \{\theta_1\}, \{\theta_2\}, ..., \{\theta_K\}, \{\theta_1, \theta_2\}, ...\{\theta_1, \theta_2, \theta_3\}, ..., \Theta\}. \tag{5}$$

The basic belief assignment (BBA) is a function denoted by m, which distributes a belief value in $[0, 1]$ to every subset of Θ: i.e., $m : 2^\Theta \to [0, 1]$, and satisfies the following:

$$\sum_{\mathbb{G} \subseteq 2^\Theta} m(\mathbb{G}) = 1 \quad \text{and} \quad m(\varnothing) = 0, \tag{6}$$

where the value of $m(\mathbb{G})$ represents the degree of belief (evidence) assigned to the composite hypothesis $\mathbb{G}$. The degree of belief will reflect some ignorance due to inability to further subdivide the measure of support into finer subsets. Every subset $\mathbb{G}$ for which $m(\mathbb{G}) > 0$ (i.e., a non-empty subset) is called a focal element. For a composite hypothesis $\mathbb{Z}$ (where $\mathbb{G} \in \mathbb{Z}$), the belief function Bel which is derived from the BBA, can be used to characterize the measure of support to $\mathbb{Z}$: $Bel(Z) = \sum_{\mathbb{G} \in \mathbb{Z}} m(\mathbb{G})$. Consider a more general case where there are two independent BBA, $m_1(\mathbb{G})$, $m_2(\mathbb{Z})$, representing independent arguments of the evidence supporting $\mathbb{G}$ and $\mathbb{Z}$, Dempster's combination rule allows one to aggregate the independent BBA into one unified function:

$$m(\mathbb{M}) = \frac{\sum_{\mathbb{G} \cap \mathbb{Z}} m1(\mathbb{G}) \cdot m2(\mathbb{Z})}{1 - \chi}, \quad \chi = \sum_{\mathbb{G} \cap \mathbb{Z} = \varnothing} m1(\mathbb{G}) \cdot m2(\mathbb{Z}), \tag{7}$$

where χ measures the degree of conflict between $m_1(\mathbb{G})$ and $m_2(\mathbb{Z})$, and the denominator acts as a normalizing factor that attributes the BBA associated with the conflict to the empty set. To select the most reasonable hypothesis for a given problem, we need to transform the beliefs into pignistic probabilities (the probability that a rational person will assign to an option when make decisions) using the pignistic probability function ($BetP$):$BetP(m(\mathbb{M})) = \sum_{\mathbb{M}_e \in \mathbb{M}} \frac{m(\mathbb{M})}{|\mathbb{M}_e|}$, where $|m(\mathbb{M})_e|$ refers to the cardinality (number of cardinal members) of $|\mathbb{M}_e|$. Then, the subset with the highest pignistic probability is considered to be the predicted outcome.

2.2.2 System Model

Next, we explain the numerical work for implementing the decision fusion approach. This is an extension of Al-Ani and Deriche (2002) and can be summarized into the following steps, that we shall explain in detail in what follows:

I. For asset i in the asset menu, transform the return r_t^i into label l_t^i, and collect a number Q number of feature sets $\boldsymbol{X}^1, \boldsymbol{X}^2, ..., \boldsymbol{X}^Q$.

II. Fit a single classifier c^q to each training set $\boldsymbol{D}^{q \in Q}$, produce the class probabilities $\boldsymbol{y}^q$ for K possible class labels, i.e., $\boldsymbol{y}^q = [y_1^q, y_2^q ..., y_K^q]$, y_k^q is the class probability that c^q classifies the output to label k.

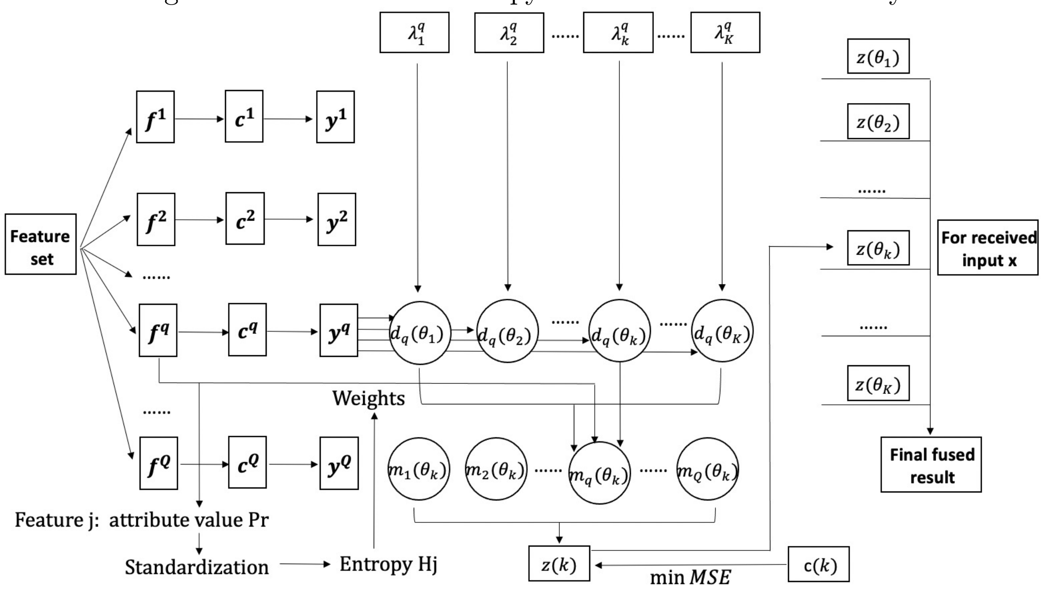
III. Calculate the Shanon's entropy of feature set $\boldsymbol{X}^q$ and derive the entropy-based weight ew^q.

IV. For a hypothesis set $\Theta = \{\theta_1, \theta_2, ..., \theta_K\}$, Θ, denote θ_k a focal element, calculate the belief in class label k assigned by classifier c^q (i.e., $m_q(\theta_k)$, based on: (1) ew^q; (2) the unnormalized ignorance g^q; (3) the Euclidean distance between the reference value λ_k^q and the class probability y_k^q.

V. Combine the belief in label k from all classifiers using Dempster's combination rule.

VI. Optimize λ_k^q and g^q using Adapt Moment Estimation by minimizing the classification error between the target output (one-hot-encoding of the true label) and the fused classification results.
VII. Choose the label with the maximum pignistic probability.
VIII. Based on results from step VII, the risk aversion level, and the covariance matrix, set entries of the view

Figure 1: Flow chart of entropy-based D-S Evidence Theory



Step I.

We assume that investors bear the hypothesis set {bullish, bearish, very bullish, very bearish}. The candidate labels set are initialized based on: (i) $\bar{r}^i_{t,t-\rho}$, which is the ρ-month moving average of returns (we consider $\rho = 120$ in this study);, (ii)z^i_t, which is given by $z = \frac{r^i_t - \bar{r}^i_{t,t-\rho}}{\sigma^i_{t,t-\rho}}$, here $\sigma^i_{t,t-\rho}$ represents the ρ-month moving standard deviation. As pointed out by Tarekegn et al. (2021), an inherent characteristic of many multi-label data is the class imbalance problem, the samples and their corresponding labels could be non-uniformly distributed over the data space (e.g., too many samples are labelled as "bearish", and very little samples being labelled as "very bullish"). To avoid the class probabilities being extremely low, we establish the threshold similar to Charte et al. (2019), where $l_{i,t} \in \{2, 1, -2, -2\}$ represents "very bullish", "bullish", "bearish", "very bearish" market conditions, respectively:

$$l_{i,t} = \begin{cases} 2 & \text{if } r^i_t \geq \bar{r}^i_{t,t-\hat{f}} \text{ and } z^i_t \geq 1 \\ 1 & \text{if } r^i_t \geq \bar{r}^i_{t,t-\hat{f}} \text{ and } 0 < z^i_t < 1 \\ -1 & \text{if } r^i_t < \bar{r}^i_{t,t-\hat{f}} \text{ and } -1 < z^i_t < 0 \\ -2 & \text{if } r^i_t < \bar{r}^i_{t,t-\hat{f}} \text{ and } z^i_t \leq -1. \end{cases} \tag{8}$$

At this point, consider an original feature set $\boldsymbol{X}$ with Ψ number of features and T observations. We partition the feature set $\boldsymbol{X}$ into Q overlapping feature sets: $\boldsymbol{X}^1, \boldsymbol{X}^2, ..., \boldsymbol{X}^Q$. Suppose the feature set $\boldsymbol{X}^Q$ contains $\bar{\Psi}$ features ($\bar{\Psi} < \Psi$): we concatenate $\boldsymbol{X}^q$ (a $T \times \bar{\Psi}$ matrix) with the labels l (a $T \times 1$ vector) to form a dataset $\boldsymbol{D}^q$, for each asset i.

Step II.

Let $\Theta = \{\theta_1, \theta_2, ..., \theta_K\}$ be the frame of discernment where θ_k represents the class label k. The dataset $\boldsymbol{D}^q$ obtained in Step I is used to train a classifier c^q and produce a vector of class probabilities $\boldsymbol{y}^q$ for for the samples in $\boldsymbol{X}^q$, i.e., $\boldsymbol{y}^q = [y^q_1, y^q_2, ..., y^q_K]$, where y^q_k is the probability that c^q believes $\boldsymbol{X}^q$ should belong to class label k

Then, we will be looking at the basic belief assignments (BBA) that summarizes the relative likelihood of each hypothesis. One of the most widely used measures for obtaining the belief mass is the distance-based measure, which was first introduced by Mandler and Schümann (1988). In their paper, the authors calculate the Euclidean distance between the learning datasets and the reference points to estimate the statistical distributions of intraclass and interclass distances. The major advantage of using distance measures is that they can be easily normalized by using some proper aggregation techniques. An extension would be to use the Hamming or the Minkowski distance.

In our application, we denote as λ_k^q the reference value assigned by classifier c^q to the label k. We use the Euclidean distance between λ_k^q and the classifier's probabilistic output y_k^q, to compute the belief mass $m_q(\theta_k)$. This represents the degree of belief in class label k given by classifier c^q. For the case in which c^q cannot provide any discriminant information about the class to which the feature vector $\boldsymbol{x}$ should belong to, we follow Al-Ani and Deriche (2002) and represent this lack of knowledge by using the unnormalized ignorance measure g^q.

Step III.

As explained earlier, we aim to enhance the diversity between the base models by applying feature sub-sampling, and one question arises automatically: when attributes/features of various alternatives are available to the participants, can we assume the decision made by each individual should be viewed as equal? A series of studies in Multiple-Attribute Group Decision-making (MAGDM) have shown that we cannot explicitly assume all attributes have equal weights when a group decision-makers receive attributes of different alternatives (Lotfi and Fallahnejad (2010)). This is because there might coexist attributes of a different (cost/benefit) and even a conflicting nature (He et al. (2016)). As Munda (1996) explains, one would need to consider the possibility of offsetting a cost on $\boldsymbol{x}_1$ by a sufficiently large benefit on another attribute $\boldsymbol{x}_2$. Aa a simple illustration, consider a classifier c^1 trained with the attribute set $[\boldsymbol{x}_1, \boldsymbol{x}_2]$ and another classifier c^2, trained with a different attribute set $[\boldsymbol{x}_1, \boldsymbol{x}_3]$. If $\boldsymbol{x}_3$ contains a sufficiently higher volume of information than $\boldsymbol{x}_2$, then, the evidence provided by classifier c^2 tends to be more valuable than the evidence provided by classifier c^1. Motivated by this argument, we implement an assessment framework for all attributes and develop an entropy-based weight drawn from the information entropy given by each attribute set.

Information entropy refers to a general measure of uncertainty, introduced by Shannon (1948). The concept of entropy represents the steppingstone of information theory, and has been employed in miscellaneous scientific fields such as transportation systems, expert systems, and physics. Nowadays, it has also been applied empirical work studying the dynamics of stock and foreign exchange markets (Oh et al. (2007), Pele et al. (2017)). According to Shannon, the number or quality of the information available in a decision-making setting is one of the key determinants of accuracy and reliability of decisions. In this sense, Shannon's Entropy can provide a natural measure of the quantity of the useful information provided by data when assessing a framework under alternative scenarios.

Given a feature set $\boldsymbol{X}$ with two sets of vectors: feature descriptions $\boldsymbol{X}^q(.t) = \{x_tj, j = 1, 2, \ldots, \tilde{\Psi}\}$ and observation description $\boldsymbol{X}_(j.) = \{x_tj, t = 1, 2, \ldots, \tilde{T}\}$, where $\tilde{Psi}$ represents the number of features and $\tilde{T}$ represents the number of observations for the training sample, the dissimilarity between each feature can be obtained by Shannon's entropy H_j, defined as:

$$H_p = -\frac{1}{\tilde{T}} \sum_{1}^{\tilde{T}} Pr_{tj} log_e Pr_{tj}. \tag{9}$$

A higher value of H_j implies stronger relevance of feature j to the classification task. For feature set $\boldsymbol{X}$ which contains x_{tj} as the t^{th} observation under the j^{th} feature, we define $Pr_{ij} = \frac{\tilde{x}_{tj}}{\sum_{t=1}^{\tilde{T}} \tilde{x}_{tj}}$ as the attribute value of the j^{th} feature, where $\tilde{x}_{tj}$ represents the standardized version of x_{tj} and $0 \leq Pr_{tj} \leq 1$. We define information volume (Iv) as the rescaled exponential of H_j so as to amplify the difference of contrast between features. Then, the Iv of the j[th] feature can be computed as:

$$Iv(j) = e^{-\frac{\sum_{i=1}^{\tilde{T}} Pr_{tj} log_e(Pr_{tj})}{log_e \tilde{T}}}. \tag{10}$$

At this point, we are free to develop an entropy-based weighting system, $ew^q = \frac{\sum_{j=1}^{\bar{\Psi}} Iv(j)}{\sum_{j=1}^{\Psi} Iv(j)}$ to be used in the combination system. The larger the entropy-based weight, the greater discriminant power of that attribute vector in the decision-making process, therefore the highest belief should be given to a given decision-maker (classifier).

Step IV.

The entropy-based weight ew_q developed in step III can be readily combined with the Euclidean distance $d_q(\theta_k)$ and the unnormalized ignorance g^q to form the basic belief assignment (BBA). This is the belief in class label k assigned by classifier c^q, i.e., $m_q(\theta_k)$, For c^q, we denote the belief committed neither to k nor to the complement of k as $m_q(\Theta)$. The BBA assigned by classifier c^q is defined as:

$$\begin{cases} d_q(\theta_k) = ||\lambda_k^q - y_k^q|| \\ m_q(\theta_k) = \frac{d_q(\theta_k) ew_q}{\sum_{k=1}^{K} d_q(\theta_k) + g^q} \\ m_q(\Theta) = \frac{g^q ew_q}{\sum_{k=1}^{K} d_q(\theta_k) + g^q} \end{cases} \tag{11}$$

Step V.

The belief in class label k constructed in step IV can be jointly combined from all classifiers according to Dempster's combination rule, assuming $I = 1, 2, ..., Q \backslash \{q\}, m_I = \oplus_{i \in I} m_i$:

$$\hat{z}(k) = m(\theta_k) = (\oplus_{q \in \{1,2,...,Q\}} m_q)(\theta_k) = \frac{[m_I(\theta_k)m_q(\theta_k) + m_I(\Theta)m_q(\theta_k) + m_I(\theta_k)m_q(\Theta)]}{1 - \sum_{\bar{p}} \sum_{\bar{s}, \bar{s} \neq \bar{p}} m_I(\theta_p) m_q(\theta_s)} \tag{12}$$

$$\hat{z}(M+1) = m(\Theta) = \oplus_{q \in \{1,2,...,Q\}} m_q(\Theta) = \frac{m_I(\Theta)m_q(\Theta)}{1 - \sum_{\bar{p}} \sum_{\bar{s}, \bar{s} \neq \bar{p}} m_I(\theta_p) m_q(\theta_s)} \tag{13}$$

We obtain binary value $\boldsymbol{c} = [c_1, c_2, ..., c_K]$ by applying one-hot-encoding on the true, observed label l. These represent the target classification output. For example, $c_k = 1$ if the attribute vector $\boldsymbol{x}$ truly falls into class label k, and $c_k = 0$ otherwise. Our objective is to minimize the mean square error (MSEr) between the combination output $\hat{z}(k)$) and target output $c(k)$: $MSEr \equiv ||\hat{\boldsymbol{z}}(k) - \boldsymbol{c}(k)||^2$. Given eq. (16), we are interested in minimizing the $MSEr$ w.r.t its parameters λ_k^q and g^q, i.e., $\frac{\partial MSEr}{\partial \lambda_k^q}$ and $\frac{\partial MSEr}{\partial g^q}$, which can be derived in closed form as:

$$\frac{\partial MSEr}{\partial \lambda_k^q} = \sum_{i=1}^{K+1} [\frac{\partial Er}{\partial \hat{z}(i)} (\sum_{j=1}^{K+1} \frac{\partial \hat{z}(i)}{\partial m_q(\theta_k)} \frac{\partial m_q(\theta_k)}{\partial \lambda_k^q} + \frac{\partial \hat{z}(i)}{\partial m_q(\Theta)} \frac{\partial m_q(\Theta)}{\partial \lambda_k^q})]. \tag{14}$$

$$\frac{\partial MSEr}{\partial g^q} = \sum_{i=1}^{K+1} [\frac{\partial Er}{\partial \hat{z}(i)} (\sum_{j=1}^{K+1} \frac{\partial \hat{z}(i)}{\partial m_q(\theta_k)} \frac{\partial m_q(\theta_k)}{\partial g^q} + \frac{\partial \hat{z}(i)}{\partial m_q(\Theta)} \frac{\partial m_q(\Theta)}{\partial g^q})]. \tag{15}$$

Step VI.

To minimize the value of eq.(14) and eq.(15), we initialize the reference value λ_k^q and the unnormalized ignorance g^q, then fine-tune them using the Adaptive Moment Estimation (ADAM) algorithm of Kingma and Ba (2014). ADAM is a first-order gradient-based optimizer of stochastic objective functions, it deals specifically with large datasets and/or high dimensional parameter space and has proved itself an efficient algorithm in many numerical examples. Kingma and Ba (2014), for example, shows by using three popular machine learning models: Logistic Regression, Multilayer Perceptron, and Convolutional Neural Network

(CNN), ADAM can achieve higher convergence rate than alternative optimizers. Mehta et al. (2019) also show that for CNN-based traffic sign classification tasks, ADAM can obtain considerably higher testing accuracy than stochastic gradient descent.

Let $\hat{C}_t$ be the partial derivative of $MSEr$ with respect to λ_k^q evaluated at the timestep t. The theoretical basis of ADAM is to update the parameter using the exponential moving averages of the gradient and the squared gradient. It realizes the benefits of two popular optimization algorithms: Adagrad and RMSprop. Similar to Adagrad, ADAM adapts the learning rate to different parameters of the estimator, such that parameters associated with more frequently occurring features can be updated at a small learning rate (and vice versa), hence, it is well-suited to deal with sparse features and gradients. On the other hand, ADAM adapts the learning rate based on the average of recent magnitudes of the gradients, which works very well for non-stationary setting (when the optimal function to be learned changes over time), similar to RMSprop.

Let u_t and v_t be the first and second moment estimates of the gradient $\hat{C}_t$. We start with $u_0 = 0$ and $v_0 = 0$ and update u_t and v_t using hyperparameters $b_1, b_2 \in [0, 1)$ as follows:

$$u_t = b_1 u_{t-1} + (1 - b_1)\hat{C}_t, v_t = b_2 v_{t-1} + (1 - b_2)\hat{C}_t^2, \tag{16}$$

where b_1, b_2 are the exponential decay rates for the moment estimates for gradient $\hat{C}_t$. However, if the moving averages are initialized as vectors of zeroes, then the moment estimates will bias towards zero, especially during initial time steps or when the decaying rates are small. Therefore, it is necessary to counteract the initialization bias with the bias-corrected first and second moment estimates: $\hat{u}_t = u_t - \frac{u_t}{1-b_1^t}$, $\hat{v}_t = v_t - \frac{v_t}{1-b_2^t}$. In In this way, we obtain ADAM's updating rule for this way, we obtain ADAM's updating rule for $\lambda_{k,t}^q$, where a is the learning rate that controls the step size of the descent of gradients: $\lambda_{k,t}^q = \lambda_{k,t-1}^q - a\frac{\hat{u}_t}{\sqrt{\hat{v}_t}+1e^{-4}}$. The above procedures can be repeated for g^q to generate an optimal estimate.

Algorithm 1 Meta Algorithm for Entropy-based Decision Fusion

1: **Initialization**
2: **Require**: $\lambda_{k,0}^q$, g_0^q: Initialized reference value and unnormalized ignorance
3: **Require**: $c(k)$: One-hot-encoding based on the label l
4: **Require**: $a \in [0, 1)$: Learning rates for λ_k^q
5: **Require**: $\_1, \_2 \in [0, 1)$: Exponential decay rates for λ_k^q
6: **for** $i \in N$ **do**
7: Fit classifier c^q to dataset $\boldsymbol{D}^q$ and produce class probabilities $\boldsymbol{y}^q$.
8: $Iv(j) = e^{-\frac{\sum_{i=1}^{\tilde{T}} Pr_{tj} log_e(Pr_{tj})}{log_e \tilde{T}}}, j \in [1 : \tilde{\Psi}]$ (Compute information volume for the j$^{\text{th}}$ feature)
9: Compute $m_q(\theta_k)$, $m_q(\Theta)$, $\hat{z}(k)$, $MSEr(\lambda_k^q, g^q)$.
10: $m_0 \leftarrow 0, v_0 \leftarrow 0$
11: $\overline{m}_0, \overline{v}_0 \leftarrow 0$
12: $t \leftarrow 0$
13: **while** $\lambda_{k,t}^q$ and $g_{k,t}^q$ not converged **do**
14: $t \leftarrow t + 1$
15: $\hat{C}_t \leftarrow \dfrac{\partial MSEr_t}{\partial \lambda_{k,t-1}^q}$
16: $u_t \leftarrow b_1 u_{t-1} + (1 - b_1)\hat{C}_t$
17: $v_t \leftarrow b_2 v_{t-1} + (1 - b_2)\hat{C}_t^2$
18: $\hat{u}_t \leftarrow u_t - \dfrac{u_t}{1 - b_1^t}, \hat{v}_t \leftarrow v_t - \dfrac{v_t}{1 - b_2^t}$
19: $\lambda_{k,t}^q \leftarrow \lambda_{k,t-1}^q - a\frac{\hat{u}_t}{\hat{v}_t^{1/2}+1e^{-4}}$
20: Repeat Step 14-19 for g^q
21: **end while**
22: **end for**
23: **Output** ew^q, λ_k^q, g^q

Step VII.

Next, we use the $\lambda^q_{k,t}$ and $g^q_{k,t}$ obtained in step VI to compute the belief function $m_q(\theta_k)$ for our test set. The forecast of the fused label is simply the label with maximum pignistic probability (BetP).

Step VIII.

Finally, we will transform the fusion results obtained in step VII into vectors of investors' views. Let $\boldsymbol{V}_t$ denote the vector of investor's views. We set the entries in $\boldsymbol{V}_t$ based on the selected label *eta*and of the covariance matrix (computed using the methodology described in section 2.3), following Donthireddy (2018): $\boldsymbol{V}_t = \boldsymbol{P}_t\boldsymbol{\pi}_t + (\eta \boldsymbol{P}_t\boldsymbol{H}_t\boldsymbol{P}_t^{\top})$, where $\boldsymbol{P}_t$ is a binary matrix identifying the number of securities associated with the view at time t, $\boldsymbol{S}_t$ is the time-varying conditional covariance matrix (which will be explained in section 2.4), and $\boldsymbol{pi}_t$ is the equilibrium expected returns that follows from eq(1). $\eta \in -2, -1, 1, 2$ can be taken as "very Bearish", "bearish", "bullish", "very Bullish" views, respectively, and indicates the chosen label with the maximum pignistic probability. We present in Algorithm 1 the meta-algorithm that could be used to obtain the outputs of this procedure.

2.2.3 Base model for Subjective return prediction

The base models are the individual classifiers that produce class probabilities $\boldsymbol{y}^q$ for our system model. We classify the base classifiers considered in this study into three broad categories: tree-based techniques, statistical techniques, and support vector machines[3]. In this section, we will introduce the intuition of each technique and briefly explain their functioning.

Tree-based models are popular techniques for modern machine learning practice that utilizes a decision tree structure. At the roof node, a decision tree (DT) chooses the feature with the most information gain and computes the values for this feature that splits the best. For each branch of the tree created, available data from prior splits are further partitioned by computing the information gain with respect to the remaining features. The algorithm refines the classification results recursively until some stopping rules are applied. Because shallow trees are easily described and visualized, DT attains good interpretability, which makes it particularly useful in the field of healthcare and finance where intuition and explanation of their predictions are societally important (see Ludwig and Baracaldo (2022)). Recently, Blanquero Bravo et al. (2019) propose a sparse optimal tree approach that fewer predictor variables can be used while preserving the classification accuracy.

Advanced tree-based techniques have also been explored. The gradient boosted decision tree (GBDT) of Friedman (2001) is an algorithm that favours choosing inaccurately labelled instances for learning and iteratively corrects the resulting error. It is, in general, ideal for reducing the bias without unduly affecting the variance in a classification task. In particular, a Random Forest (RF, see Breiman (2001)) is a combination of de-correlated DTs in which each DT contributes with a single vote to assign the most frequent class of the input vector. RF improves the diversity of the trees by making them grow from different training data subsets created through bootstrap aggregating (bagging), a technique used for training data creation that performs random resampling of the original dataset with replacement. Hence, some data may be used more than once in the training of classifiers while others may never be used. Therefore, greater classifier stability is achieved since RF is robust when facing variations in the input data and is not sensitive to noise or overtraining. Moreover, tree-based methods makes no assumptions on the relationship between features, which makes it well-suited when data have multi-collinearity among variables.

[3]Perception-based techniques (Artificial Neural Networks) are also powerful tools to analyze the data. However, their performance is unstable since the design of network architecture is not uniform and the parameter estimation uncertainty is non-trivial. The major criticism of neural networks is overfitting - as the model becomes more complex, it will tend to fit noise in the data rather than finding the signals. With limited data in practice, convoluted interactions among the predictors will only add complexity to the model without adding any corresponding benefits (Hawkins (2004)). This is particularly problematic for asset allocations due to the comparative dearth of data and the low signal-to-noise ratio (Gu et al. (2020)). Therefore, we exclude neural networks from this study.

The statistical algorithms used in this study are the k-Nearest Neighbour (kNN) and Logistic regression (LR). The KNN algorithm of Fix and Hodges (1989) is based on predicting the class of the sample by using class of its neighbours. First, it calculates the distance between unlabelled sample and each training sample using principal component analysis. Next, by selecting k number of training sample which are closest to the unlabelled sample (refer as k neighbours), it assigns class to the unlabelled sample using the class which owns the majority among the k neighbours. In fact, KNN does not need to train model for generalization, but memorizes the training dataset, and is therefore known as a “lazy learner”. The key advantages of kNN are that the algorithm makes no explicit assumption about the distribution of the data, is insensitive to outliers and is free of redundant information and/or missing values, which makes it powerful tool to analyze data. LR is an efficient statistical algorithm based on estimating the probability of being in one binary outcome versus the other. The model iteratively finds the best linear combination of the explanatory variables with the greatest probability of detecting the observed outcome, using components of linear regression reflected in the logit scale (Stoltzfus (2011)). LR is a very powerful tool, however, it requires strong assumptions on the independence between residuals and the variables; and independence among features (no multi-collinearity).

The Support Vector Machine (SVM) of Hastie et al. (2009) is based on the goal of finding an optimal separating hyperplane on the feature set, which the hyperplane is as far as possible from the closest members of both classes. The closest members of both classes are referred as support vectors and the distance between the hyperplane and the support vectors is referred as the margin. The classification result for linear SVM is in general, a linear combination of all the data points that lie on the margin. When the feature set must be mapped to a high-dimensional space before they can be divided linearly to separate the classes, the classification result is a combination of all data points that lie on the margin using a kernel function. Thus, SVM is compatible for handling high dimensional data.

2.3 Reference Portfolio

The theoretical basis of BLM is Bayesian learning. The market portfolio acts as a point of reference, as more data gradually become available, existing beliefs about expected returns are revised into posterior beliefs to account for the new information. However, the CAPM anomalies reported in the asset pricing literature suggest that holding a reference portfolio proportional to the market capitalization may be grossly inadequate. Furthermore, Zhou (2009) indicates that the nature of the assumed data-generating process are essential in Bayesian learning, which BLM completely disregards. If the equilibrium expected returns calibrated using the degree of risk aversion, covariance matrix, and market capitalization weights, were to be substantially different from the historical returns, investors could easily under- or overestimate the true expected returns; thus, leads to large estimation errors for implementing the optimization framework. On the other hand, the information inherited from data provides understanding of the general economy (macro fluctuation, regime changes), rendering historical data important for constructing the investors' prior.

For these reasons, we obtain two priors directly from the data: the equally-weighted (1/N), and the no-short-sell minimum variance portfolio, applying ten-year expanding data windows of asset returns. Both strategies have been intensively tested in the portfolio selection literature and are usually more competitive in out-of-sample validations than strategies relying on the expected return estimates, since expected return estimates are more sensitive to estimation errors compared to volatility estimates (Bianchi and Guidolin (2014); Fugazza et al. (2015)).

Given the time series of monthly excess returns, we generate reward-to-risk measures for strategy γ, which is defined as the in-sample mean divide by the variances estimates:

$$R^{\gamma} = \frac{Mean^{\gamma}}{Std^{\gamma}} = \frac{\hat{\boldsymbol{\mu}}_{\gamma}^{IS,\top} \hat{\boldsymbol{w}}_{\gamma}}{\hat{\boldsymbol{w}}_{\gamma}^{\top} \hat{\boldsymbol{\Sigma}}_{\gamma}^{IS} \hat{\boldsymbol{w}}_{\gamma}}, \tag{17}$$

where $\hat{\boldsymbol{\Sigma}}_{\gamma}^{IS}$ represents the in-sample covariance estimates for strategy γ. For 1/N strategy which does not require a covariance matrix, we use estimates of the in-sample standard deviation. Let M be the length of the estimation period, based on the expanding window approach, the time series of $\boldsymbol{R}^{\gamma}$ is a $(T-M)\times 1$ column vector, where R_t^{γ} is the in-sample reward-to-risk of strategy γ at period t.

Next, we develop a data-generating process that obtains regression forecasts for future reward-to-risk under each prior, based on a quantitative forecasting model, selected to be the least absolute shrinkage and selection operator (LASSO, see Tibshirani (1996))[4]. Let Φ be the non-negative penalty parameter. The regression forecast $\hat{R}_t^{\gamma}$ at period t can be obtained by minimizing the loss function:

$$L(\beta;\Phi)=\frac{1}{M}\sum_{t=\vartheta+1}^{M+\vartheta} R_t^{\gamma}-(\hat{\beta}_0+\tilde{\boldsymbol{x}}_t\hat{\boldsymbol{\beta}})+\Phi\sum_{t=1}^{\vartheta}|\hat{\beta}_t|, \tag{18}$$

where $\hat{\beta}_t$ represent the regression coefficients, ϑ denotes the number of lags (ϑ is selected to be 120 in this study, which is equivalent to include information of ten year), $\hat{\beta}=[\hat{\beta}_{t-1},\hat{\beta}_{t-2},...,\hat{\beta}_{t-\vartheta}]$, and $\tilde{\boldsymbol{x}}_t=[R_{t-1}^{\gamma},R_{t-2}^{\gamma},...,R_{t-\vartheta}^{\gamma}]$. Thus, the reward-to-risk predicted in period t is: $\hat{R}_{\vartheta+1}^{\gamma}=\hat{\beta}_0+\sum_{t=1}^{\vartheta}\hat{\beta}_t R_t^{\gamma}$. We split the training and the testing sample such that we generate forecasts for five trading years (from December 2016 to December 2020). Then, predictions for $\hat{\boldsymbol{R}}^{\gamma}$ and the corresponding weight matrix $\boldsymbol{w}^{\gamma}$ can be stored in a 120×1 column vector and a $120\times N$ matrix, respectively.

Based on the regression forecasts of future reward-to-risk ratios, we can combine the two priors into a new Bayesian prior. The resulting reference portfolio, which is taken as a shrinkage of the 1/N and the minimum-variance portfolio, can moderate parameter estimation uncertainty and provide a reasonable starting point for return estimation. The approach is also very flexible, one can select any shrinkage target and construct the combination without deriving analytical solutions.

Following eq.(18), we can generate a list of two candidate portfolios which, their predicted risk-to-reward ratios can be stored in a 120×2 matrix. At period t, we have a 1×2 vector of: $\hat{\boldsymbol{R}}_t=[\hat{R}_t^{1/N},\hat{R}_t^{MV}]$. The probability of each prior is obtained by normalizing their risk-to-reward ratios. Let $\hat{R}_t^{sum}=\hat{R}_t^{1/N}+\hat{R}_t^{MV}$, the reference portfolio weights can hence be set as: $\boldsymbol{w}_t^{ref}=\frac{\hat{R}_t^{1/N}}{\hat{R}_t^{sum}}\boldsymbol{w}_t^{1/N}+\frac{\hat{R}_t^{MV}}{\hat{R}_t^{sum}}\boldsymbol{w}_t^{MV}$, where $\boldsymbol{w}_t^{1/N}$ and $\boldsymbol{w}_t^{MV}$ are the weight compositions of the 1/N and minimum-variance portfolios, $\frac{\hat{R}_t^{1/N}}{\hat{\boldsymbol{R}}_t^{sum}}$ and $\frac{\hat{R}_t^{MV}}{\hat{\boldsymbol{R}}_t^{sum}}$ are the probability of the two priors, respectively. The approach is very general that can be applied to various situations, one can generate the framework to incorporate more than two candidate portfolios.

2.4 Wishart Stochastic Volatility with Graphical Lasso (GWSV)

The prior covariance matrix $\boldsymbol{\Sigma}_t$ is required at two stages of the Black-Litterman model, first, to reverse engineer the equilibrium expected returns $\boldsymbol{\pi}_t$ from the reference portfolio; second, to obtain the posterior covariance matrix $\boldsymbol{\Sigma}_t^{BL}$. The aforementioned paper, Harris et al. (2017), has relaxed the constant conditional volatility assumption by using multivariate GARCH models. In this study, we generalize their research to incorporate stochastic variations in the covariance matrices. We model the unobserved dynamic precision matrices as a latent Wishart process (see, Uhlig (1994) and Uhlig (1997)) and further shrink the one-step ahead prediction of conditional volatility using Graphical LASSO to bound the estimation error, while the shrinkage is proportional to the diversification distribution among the risky assets. Parameters in the newly obtained GLASSO-Wishart model are calibrated using the Genetic Algorithm (GA). The prior distribution for precision matrix is conjugate to the Gaussian distribution, so the likelihood function can be derived in closed form. In the following, we provide the relevant details.

Consider a N-dimensional return vector $\boldsymbol{r}_t$ with mean μ_t and conditional covariance matrix $\boldsymbol{H}_t$. WSV

[4]LASSO (a penalized regression technique) shrinks the large coefficients towards zero by applying L^1 norm regularization on the slope coefficients. It introduces bias in the coefficient estimates but reduces their variance, therefore mitigating the risk of overfitting and potentially improving the out-of-sample performance owing to a better realized bias-variance trade-off. This approach is also attractive because it simultaneously performs variable selection and allows a researcher to identify the pertinent and remove irrelevant predictors. In recent times, LASSO has received intensive scrutiny, especially when applied to stock return predictability, see, e.g., Chinco et al. (2019), Gu et al. (2020), and Lee et al. (2022).

defines a multiplicative law of motion for the stochastic inverse covariance matrix $\boldsymbol{H}_t^{-1}$ (Moura et al. (2020)):

$$\boldsymbol{H}_t^{-1} = \frac{\kappa+1}{\kappa} U(\boldsymbol{H}_{t-1}^{-1})' \boldsymbol{\psi}_t U(\boldsymbol{H}_{t-1}^{-1}), \;\; \boldsymbol{\psi}_t \sim \mathfrak{B}_N(\frac{\kappa}{2}, \frac{1}{2}), \tag{19}$$

where κ is the degree of freedom, $\mathfrak{B}_N$ represents a N-dimensional singular Dirichlet distribution, $\boldsymbol{\psi}_t$ contains random shocks drawn from $\mathfrak{B}_N$ and $U(\boldsymbol{H}_t^{-1})$ is the upper triangular matrix from the Cholesky decomposition of $\boldsymbol{H}_t^{-1}$. Suppose $\boldsymbol{H}_t^{-1}$ is initialized on a prior that follows a Wishart distribution: $\boldsymbol{H}_{t-1}^{-1}|\boldsymbol{r}_t \sim \boldsymbol{W}_N(\kappa, [\kappa \boldsymbol{S}_{t-1}]^{-1})$ and $E(\boldsymbol{H}_t^{-1}) = \boldsymbol{S}_{t-1}^{-1}$, the one-step-ahead prediction for covariance is: $\boldsymbol{S}_t = \frac{1}{\kappa+1}\boldsymbol{r}_t\boldsymbol{r}_t' + \frac{\kappa}{\kappa+1}\boldsymbol{S}_{t-1}$. The smoothing formula can be reformed as follows, where $\xi = \frac{\kappa}{\kappa+1}$ is the degree of time variation:

$$\boldsymbol{S}_t = \xi^t \boldsymbol{S}_0 + (1-\xi)\sum_{i=1}^{t-1} \xi^{i-1} \boldsymbol{r}_{t-i}\boldsymbol{r}_{t-i}'. \tag{20}$$

The parameter ξ can be characterized as a discounting factor given that $\xi = \frac{\kappa}{\kappa+1}$ and $\kappa > N+1$, so the smoothing formula in (20) implies a shrinkage between the initial covariance $\boldsymbol{S}_0$ and the Exponentially Weighted Moving Average (EWMA) covariance, where ξ governs both the smoothing and the shrinkage intensity. To estimate ξ, we follow Kim (2014) and apply a maximum likelihood approach, conditional on $\boldsymbol{S}_0$, hence, the choice of $\boldsymbol{S}_0$ is of crucial importance.

Choosing an intuitive initial covariance matrix may be challenging. A standard choice would be to adopt the sample covariance matrix $\boldsymbol{\Sigma}_t$, but a general concern is that the maximum-likelihood estimator (MLE) is often poorly behaved with overwhelming random noise. Uhlig (1997) advises to use a diagonal matrix where the diagonal is given by average squared residuals obtained from an AR(1) regression. Moura et al. (2020) also suggest using a diagonal matrix in which the main diagonal collects the sample variance estimates of individual asset returns to remove any noisy correlations. In our generalized BLM, we extend Friedman et al. (2008) by using the portfolio's diversification distribution (explained later in this section) to regularize the conditional correlations between assets.

In the literature on sparse covariance estimation, learning the structure of undirected Gaussian graphs is equivalent to estimating the inverse covariance matrix when assuming a Gaussian distribution for asset returns. As shown by Friedman et al. (2008), one can impose an L^1 norm of regularization on the negative log-likelihood to encourage sparsity of the inverse covariance matrix:

$$Min_{\boldsymbol{S}_0^{-1} \succ 0}\{-\log\det|\boldsymbol{S}_0^{-1}| + tr[\boldsymbol{\Sigma}\boldsymbol{S}_0^{-1}] + \left\|\mathfrak{J} * \boldsymbol{S}_0^{-1}\right\|_1\}, \tag{21}$$

where $\boldsymbol{\Sigma}$ is the sample covariance, tr is the trace operator, $\mathfrak{J} = \{\mathfrak{J}_i\}$ is the inverse-volatility based regularizing parameter for each variable and * denotes pairwise multiplication. $\left\|\boldsymbol{S}_0^{-1}\right\|_1$ is the sum of the absolute values of $\boldsymbol{S}_0^{-1}$. For $\mathfrak{J} > 0$, the log-determinant term guarantees the convexity of eq.(21) that a unique global positive definite minimizer will always exist. The dual of the penalized likelihood problem can be solved by using a block coordinate descent algorithm. Let $\boldsymbol{F}$ be the estimates of $\boldsymbol{S}_0$, consider a partition of $\boldsymbol{F}$:

$$\begin{pmatrix} \boldsymbol{F}_{11} & \boldsymbol{f}_{12} \\ \boldsymbol{f}_{12}' & f_{22}, \end{pmatrix}$$

where $\boldsymbol{F}_{11}$ is a $(N-1)\times(N-1)$ matrix that removes the $\hat{k}^{\text{th}}$ column and the $\hat{j}^{\text{th}}$ row, $\boldsymbol{g}_{12}$ is a $(N-1)$ vector produced by removing the diagonal element g_{22} from the $\hat{j}^{\text{th}}$ row, and g_{22} is a scalar. $\boldsymbol{\Sigma}$ is partitioned in the same fashion. The idea of using block coordinate is optimize over one row/column each time, restrict the remaining elements to be constant and recursively cycling through all rows and columns until convergence. Banerjee et al. (2008) show that based on sub-gradient notations and dual convexity, the solution to g_{12} is equivalent to repeatedly solving the following Lasso problem:

$$Min_{\alpha}\{\frac{1}{2}\left\|\boldsymbol{F}_{11}^{\frac{1}{2}}\hat{\boldsymbol{\alpha}} - \boldsymbol{F}_{11}^{\frac{1}{2}}\boldsymbol{\Sigma}_{12}\right\|^2 + \|\mathfrak{J} * \hat{\boldsymbol{\alpha}}\|_1\}, \tag{22}$$

where $\hat{\boldsymbol{\alpha}}$ is a $(N-1)$ solution vector that can be derived using the soft threshold operator:

$$\hat{\boldsymbol{\alpha}}_j \leftarrow Soft(\hat{\boldsymbol{U}}_{\hat{j}} - \boldsymbol{\Sigma}_{\hat{k}\neq\hat{j}}\hat{V}_{\hat{k}\hat{j}}, \hat{\boldsymbol{\alpha}}_{\hat{k}}, \mathfrak{J})/\hat{\boldsymbol{V}}_{\hat{j}\hat{j}}, \tag{23}$$

where $\hat{\boldsymbol{V}} = \boldsymbol{F}_{11}$, $\hat{\boldsymbol{U}} = \boldsymbol{\Sigma}_{12}$, $\boldsymbol{f}_{12} = \boldsymbol{F}_{11}\hat{\boldsymbol{\alpha}}$, and f_{22} denotes the variance which remains unchanged during the process. The estimate of $\boldsymbol{S}_0$ can be solved by updating column/row-wisely with $\boldsymbol{f}_{12}$.

To further improve the performance of Graphical LASSO method, we suggest an alternative approach, the diversification-based GLASSO, which follows Meucci (2009)'s analysis on the entropy of portfolio diversification distribution. The idea underlying our approach is to account for the fact that estimation errors are not the same for all stocks: the estimation errors are larger for stocks with larger variance (Levy and Levy (2014)). The diversification-based GLASSO regularizes conditional covariance to be directly proportional to the diversification measure: the greater the diversification index, the higher the exposure to risk, the larger the estimation error, and the stronger regularization that will take place.

To implement the approach, we need to determine a diversification measure for each asset in the portfolio. In this respect, Meucci (2009) argues that the commonly used diversification measures such as risk-based, factor-based, and weight-based measures (the Herfindal's index) cannot represent the structure of portfolio concentration profile in the correlated markets, and can only be applied in specific circumstances and/or under restrictive hypotheses. Instead, Meucci introduces a new measure, diversification distribution, to identify the diversification measure offered by each individual asset. In our paper, the strength of the regularization applied the asset i, or equivalently, the penalty weight $(\tilde{w}_i)$, is given by the diversification distribution of uncorrelated principal portfolios, which is then multiplied with the homogenous regularization parameter $\mathfrak{J}$ to yield the solution.

In markets characterized by correlated asset returns, the generic portfolio can be represented as a set of uncorrelated principal portfolios that expose to uncorrelated sources of risk, and the principal component decomposition of sample covariance matrix $\boldsymbol{\Sigma}$ can be used to identify the uncorrelated sources of risk among assets: $\boldsymbol{E}'\boldsymbol{\Sigma}\boldsymbol{E} = \boldsymbol{\Lambda}$ where E represents an orthogonal matrix whose columns $\boldsymbol{e}_1, \boldsymbol{e}_2, ..., \boldsymbol{e}_N$ contain N eigenvectors and $\boldsymbol{\Lambda}$ is a diagonal matrix whose diagonal entry is given by the corresponding eigenvalues ($\varphi_1 > \varphi_2 > ... > \varphi_N$). Each eigenvector represents an uncorrelated principal portfolio, and the corresponding eigenvalue denotes the corresponding portfolio return variance.

According to Meucci (2009), the original portfolio can then be replicated by combining N uncorrelated principal portfolios with the weights $\tilde{\boldsymbol{w}}$: $\tilde{\boldsymbol{w}} = \boldsymbol{E}^{-1}\hat{\boldsymbol{w}}$. We take the equal weights (1/N) as the reference for $\hat{\boldsymbol{w}}$. Define the variance concentration curve as: $\tilde{v}_n = \tilde{w}_n^2\varphi_n, n = 1, ..., N$, where $\tilde{v}_n$ denotes the amount of original portfolio variance explained by the n^{th} principal portfolio. Meucci's diversification distribution can be computed as $\hat{p}_n = \frac{\tilde{w}_n^2\varphi_n}{Var(R_w)}$, that is, the contribution to original portfolio's variance from n^{th} principal portfolio, where $Var(R_w) = \sum_{n=1}^{N}\tilde{v}_n$ denotes the total portfolio variance. Given $\sum_{n=1}^{N}\hat{p}_n = 1$ and $0 \leq \hat{p}_n \leq 1$, the diversification distribution $\hat{p}_n$ can be interpreted as a set of probability masses related to each principal portfolio. The penalty weight $\tilde{w}_i$ can be set equal to $\hat{p}_n$ so the regularization parameter can be computed as: $\mathfrak{J}_i = \mathfrak{J}\tilde{w}_i$. To avoid extreme low $\tilde{w}_i$, we set a lower constraint: $\bar{w}_i \geq \frac{0+1/N}{2}$ where N represents the number of assets in the asset menu.

We use the in-sample period (i.e., training period, which contains 70% of our data) to search for the optimal $\mathfrak{J}$ and follow this choice throughout the out-of-sample period. The optimal $\mathfrak{J}$ is found using the genetic algorithm (GA)[5]. The general workflow of GA is to randomly initialize the population in the search space (encoded as chromosomes representations) and produce iterative new generations of possible solutions by performing crossover and mutations. GA will continuously remove the individual with lowest fitness value until some stopping rules are applied, then reaches the near-optimal solution. For implementation of the GBL optimization model, we replace the sample covariance matrix $\boldsymbol{\Sigma}_t$ in the Black-Litterman optimization

[5]The theoretical basis of GA is rule of survival of fitness. Standard parameter values are chosen to implement the Genetic Algorithm Toolbox in MATLAB (number of individuals=80, maximum number of generations=100, precision of variable=20, generation gap=0.9, crossover probability=0.7, mutation probability=0.0017).

framework with the GLASSO-Wishart covariance matrix $\boldsymbol{S}_t$ as given by eq.(20).

3 Data

We evaluate the realized, out-of-sample performance of our generalized Black-Litterman strategy based on four Fama-French data sets that are typically used in the empirical asset allocation literature (see e.g., DeMiguel et al. (2009), Kapadia (2011), Bianchi and Guidolin (2014), Harris and Mazibas (2022)), all sampled at monthly frequencies. The first dataset consists of monthly return observations of the Fama French 17 industry portfolios (*Ind17*): consumer durables, fabricated products, clothes, transportation, construction, utilities, retail stores, steel works, consumption, oil, chemicals, automobiles, financials, food, mines, machinery and business equipment, others. We also employ the Fama-French 30 industry portfolios for the US equity market (*Ind30*)[6], the Fama-French 25 short-term reversal sorted portfolios (*FF25*) and the Fama-French 10 size and short-term reversal sorted portfolios (*FF10*). All data span the period August 1960–December 2020. Because returns are expressed in US dollars, the appropriate monthly risk-free rate is again obtained from the Ken French's website and corresponds to the 1-month T-bill rate. We summarize details of the datasets in Table 1.

Table 1: List of datasets

#	Dataset	Source	N	Period	Abbre
1	Industry portfolios	Ken French's website	17	1960.8-2020.12	*Ind17*
2	Size and short-term reversal sorted portfolios	Ken French's website	25	1960.8-2020.12	*FF25*
3	Industry portfolios	Ken French's website	30	1960.8-2020.12	*Ind30*
4	Size and short-term reversal sorted portfolios	Ken French's website	10	1960.8-2020.12	*FF10*

This table summarizes the datasets used in this paper. We report from left to right the number of assets (N) in each asset menu, the period spanned (Period) and the abbreviation (Abbre) used to refer to the dataset in the paper.

To forecast the investors' views for each dataset, we consider predictors that have been studied intensively and confirmed in the recent equity research. They were culled from the predictive variables appearing in Rapach et al. (2010), Dangl and Halling (2012) and Neely et al. (2014), among others. First, we use the 14 predictive variables constructed following Welch and Goyal (2008): the dividend price ratio (d/p), dividend yield (d/y), earnings price ratio (e/p), dividend payout ratio (d/e), stock variance (svar), book-to-market ratio (b/m), net equity expansion (ntis), treasury bills (tbl), long-term return (ltr), long term yield (lty), term spread (tms), default return spread (dfr), default yield spread (dfy), inflation (infl, which are lagged 2-month to account for the delay in the CPI releases). We further include three uncertainty measures of Ludvigson et al. (2021): the financial uncertainty (finunc), real uncertainty (realunc), macroeconomic uncertainty (macunc); and five macroeconomic variables that influence aggregate stock returns as identified in Flannery and Protopapadakis (2002): monetary aggregate (generally M1), producer price index (PPI), employment report (emp), balance of trade (bot), and housing starts (houst)[7]. The corresponding series are all obtained from Federal Reserve Economic Data (FRED). Please note that we have included the variable "inflation", and in order to avoid multi-collinearity, we exclude the sixth explanatory variable, the consumer price index (CPI) as identified in Flannery and Protopapadakis (2002). A summary of these explanatory variables are listed in Table 2.

In our two-classifier combination strategy, each classifier is trained with a subset of the predictors using the method as described in section 2.1.2. However, as indicated by Boivin and Ng (2006) and Bianchi et al.

[6]The industries include food, beer, smoke, games, books, consumer goods, clothes, health, chemicals, textiles, construction, steel work, fabricated products, electrical equipment, automobiles, aircraft/ship/railroad equipment, mines, coal, oil, communication, services, business equipment, paper, transportation, wholesale, retail, meals, finance, utilities, and other.

[7]For variable "bot": from 1960 to 1992 we use the Balance on Merchandise Trade, and from 1992 to 2020 we use the Trade Balance: Goods and Services, Balance of Payments Basis for Balance of Trade. Both series are obtained from FRED.

(2021), factor identification is influenced by the amount of information used, therefore, we also expand each set of predictors with 10-year lagged returns taken from the Fama French data. The series were selected to represent pure stock-level information. The use of return time series nests many models proposed in the literature (Cujean and Hasler (2017), Gu et al. (2020)) and allows us to use both macroeconomic factors and stock-level information in a study of joint predictive ability.

Table 2: List of Macroeconomic and Financial Predictors

#	Predictor	Description	Abbreviation
1	dividend price ratio	Difference between the log of dividends and the log of stock prices	d/p
2	dividend yield	Difference between the log of dividends and the log of lagged stock price	d/y
3	earnings price ratio	Difference between the log of earnings and the log of stock price	e/p
4	dividend payout ratio	Difference between the log of dividends and the log of earning	d/e
5	stock variance	Sum of squared daily returns on the S&P 500	svar
6	book-to-market ratio	Ratio of book value at the end of the previous year divided by the price at the end of the current month, for the Dow Jones Industrial Average	b/m
7	net equity expansion	Ratio of 12-month moving sums of net issues by NYSE listed stocks divided by the total end-of-year market capitalization of NYSE stocks	ntis
8	treasury bills	3-month treasury bill: secondary market rate	tbl
9	long-term return	Long term yield from Ibbotson's Stocks, Bonds, Bills and Inflation Year	ltr
10	long term yield	U.S. yield on long-term United States bonds series	lty
11	term spread	Difference between the long-term yield on government bonds and treasury bill	tms
12	default return spread	Difference between long-term corporate bond and long-term government bond return	dfr
13	default yield spread	Difference between BAA and AAA-rated corporate bond yield	dfy
14	inflation	Consumer price index, lagged 2-month to account for the delay in the releases	infl
15	real uncertainty	Uncertainty in real activity variables such as output and unemployment	realunc
16	financial uncertainty	Uncertainty in financial variables (firm-level income growth series)	fiunc
17	macroeconomic uncertainty	A measure of macro (economy-wide) uncertainty	macun
18	monetary aggregate	A measure of the most-liquid assets in the U.S. money supply: cash, checking accounts, traveller's checks, demand deposits, and other checkable deposits	M1
19	producer price index	Average change over time in the selling prices received by domestic producers for their output	PPI
20	employment report	The unemployment rate announced by FRED	emp
21	balance of trade	International trade deficit in goods and services	bop
22	housing starts	New privately-owned housing units started: total units, from U.S. Census Bureau and U.S. Department of Housing and Urban Development	houst

This table summarizes the macroeconomic and financial predictors used in this paper. We report from left to right the variable name, the description of the variable, and the abbreviation used to refer to the predictor. In the column 'Description', dividends and earnings refer to the 12-month moving sums of the dividends and earnings paid on the S&P 500 index, respectively. The macroeconomic, real, financial uncertainty indices are constructed using a statistical model with a large set of macroeconomic, sectoral, and financial data. The uncertainty indices and the methodology for constructing the indices are available from Sydney C Ludvigson's web site.

4 Empirical Results

In this section, we compare the realized performance of the GBL strategy with eleven popular benchmark strategies from the portfolio literature. We do so across the four different data sets listed in Table 1. The comparisons occur using the following performance criteria: the average return (Ret), average standard deviation (Std), Sharpe ratio (SR), Sortino ratio (SOR), and return over Value-at-Risk at 99% confidence level (ROV). The Sortino ratio (Sortino and Price (1994)) is defined as the sample mean of out-of-sample excess returns divided by their downside deviation. We summarize their computation methods in Table 3.

Table 3: List of Performance Metrics

Performance Metric	Computation	Abbreviation
Sharpe ratio	$\widehat{SR} = \frac{\hat{\mu} - r_f}{\hat{\sigma}}$	SR
Sortino ratio	$\widehat{SOR} = \frac{\hat{\mu} - r_f}{\hat{\sigma}^-}$ with $\hat{\mu}^- = \min(\hat{\mu}, 0)$, $\hat{\sigma}^- = \sqrt{Var(\hat{\mu}^-)}$	SOR
Return over Value-at-Risk	$\widehat{ROV} = \frac{\hat{\mu} - r_f}{VaR_{1-\hat{\alpha}}}$ with $VaR_{\hat{\alpha}}(\boldsymbol{w}) = \hat{\zeta}_{\hat{\alpha}}\sqrt{\boldsymbol{w}^T\boldsymbol{\Sigma}\boldsymbol{w}} - \boldsymbol{\mu}^T\boldsymbol{w}$, $\hat{\zeta}_{\hat{\alpha}} = -\hat{\phi}^{-1}(1 - \hat{\alpha})$	ROV
with: $\hat{\mu}$=sample mean of asset returns		$\boldsymbol{w}$=vector of portfolio weights
$\hat{\sigma}$=sample standard deviation of asset returns		$\boldsymbol{\mu}$=vector of return estimates
r_f=risk-free rate		$\boldsymbol{\Sigma}$=estimates of covariance matrix
$\hat{\phi}(\cdot)$=standard normal distribution function		$\hat{\alpha}$=confidence level

This table summarizes the three performance metrics and the formulas used to obtain each performance metric. The last column gives the abbreviation used to refer to the metric in the paper.

Table 4: List of Asset Allocation Strategies

#	Model	Abbreviation
Naive		
1	1/N with rebalancing (naïve strategy)	1/N
Classical allocation approach		
2	Risk-parity	RiskParity
3	Reward-to-risk-timing	RRT
4	Mean-variance	MeanVar-u
5	Maximum Sharpe ratio	MaxSR-u
6	Maximum Sharpe ratio with Ledoit&Wolf shrinkage covariance	LW-MSR
Bayesian approach to estimation error		
7	Bayes-Stein shrinkage	BS-u
8	Black-Litterman	BS-u
Generalized Black-Litterman (GBL) approach		
9	GBL with optimal combinations of portfolios & sample covariance	GBL-OS-u
10	GBL with 1/N & GLASSO-Wishart	GBL-NG-u
11	GBL with portfolio combination & GLASSO-Wishart	GBL-OG-u
Portfolio with short-sell constraints		
12	Mean-variance with short-sale constraints	MeanVar-c
13	No-short-sell maximum Sharpe ratio with Ledoit&Wolf shrinkage covariance	LW-MSR-c
14	Black-Litterman with short-sale constraints	BL-c
15	No-short-sell GBL with portfolio combination & GLASSO-Wishart	GBL-OS-c
16	No-short-sell GBL with 1/N & GLASSO-Wishart	GBL-NG-c
17	No-short-sell GBL with portfolio combination & GLASSO-Wishart	GBL-OG-c

This table lists the various asset allocation strategies investigated in this paper. The last column of this table gives the abbreviations used to refer to the strategy in the tables where the performance of various strategies is compared. For the Bayesian and GBL strategies, we assume investors seek to maximize the portfolio Sharpe ratio. For strategies 1-11, we impose a weight constraint which $-0.5 \leq w_i \leq 1$ and $\sum_{i=1}^{N} w_i = 1$ in the corresponding optimization problems. For strategies 12-17, we impose an additional nonnegativity constraint on the portfolio weights in the corresponding optimization problems.

As for the benchmarks, the first portfolio we will be examining is the one returned by 1/N naïve diversification, which is a simple strategy that requires neither estimation nor optimization. Next, we compare our GBL portfolios with a portfolio that completely ignores the information in the expected returns and only relies on the in-sample variance, the risk-parity portfolio. We also consider portfolios that require estimation of both sample mean and volatility estimates, i.e., the mean-variance (MeanVar) portfolio, maximum Sharpe ratio (MaxSR) portfolio with sample covariance, maximum Sharpe ratio (MaxSR) portfolio with covariance shrinkage (Ledoit and Wolf (2003)), reward-to-risk timing portfolio of Kirby and Ostdiek (2012), the Bayes-Stein (BS) shrinkage portfolio of Jorion (1986), and the classic Black-Litterman (BL) portfolios, constructed using (i) historical average returns as the investors' views; (ii) a reference portfolio of 1/N; (iii) sample covariance matrix. For the BL portfolio, we choose 1/N to be our reference portfolio because 1/N has been shown to outperform a host of optimal portfolio strategies for asset allocation (DeMiguel et al. (2009); Platanakis et al. (2021)). All portfolios discussed in this section are monthly rebalanced and are computed assuming a rather prudent, but standard risk aversion of 1. Since transaction cost is a non-trivial aspect that investors must consider when acquiring the allocations they select, we also evaluate the performance for all strategies after transaction costs, assuming that of 20 bps (the choice of 20 bps is motivated by Kan et al. (2021)). The list of benchmark strategies and their optimization framework are provided in Table 4 and Table 5.

To evaluate the suggested generalized Black-Litterman portfolios, we perform an out-of-sample testing and examine the performance of optimal portfolio strategies, for a period spans from January 2016 to December 2020. To obtain the views, we conduct a large-scale empirical analysis investigating stocks over 55 years from 1960 to 2015, which leads to 5-year out-of-sample results (a similar training period was used in Gu et al. (2020) where they investigate stock returns from 1957 to 2016). Motivated by DeMiguel et al. (2009), we use ten-year data windows of asset returns to estimate the equilibrium expected returns and the covariance matrix, however, we use an expanding window approach, which is expected to provide more stable inputs than using a rolling window approach. At period $t = 121$, we use data in the previous 120 months to estimate the equilibrium expected returns and covariance matrix. These inputs, together with the views, are used to determine the portfolio weight for the GBL strategy. The process is continued by maintaining the earliest returns but including the returns for the next period in the dataset, until data points are exhausted, so that we compute and evaluate the last realized out-of-sample performance with reference to November 2010: November 2020 to provide a forecast for December 2020.

The decision-fusion process developed in this paper optimally combines the classification results of q classifiers by DS Theory of Evidence. This is different from general ensemble learning algorithms (e.g., meta learners) where a linear model (e.g., logistic regression) is used to aggregate the classification predictions of base models. Intuitively, one should let q be sufficiently large in order to create diversity among classifiers. However, this would inevitably lead to a concern with computational intractability. According to the computational complexity theory, decision problems can be categorized into various complexity classes subject to the amount of computational resource (time and space) that algorithms take to solve, and only a problem solvable by a polynomial-time algorithm is deemed as tractable. Because we have N risky assets in the investment universe, as our combination system becomes more complicated, the optimization would become computationally inefficient, and expected running time would be non-trivial. On the other hand, the massive number of parameters associated with a large number of base classifiers are easily prone to estimation errors, due to the comparative dearth of the financial data. This would easily result in an overfitting and a decreased realized forecasting accuracy.The results reported in Al-Ani and Deriche (2002) also show that, the forecasting accuracy for a two-classifier combination tend to be higher than combinations using more than two classifiers. Motivated by this earlier literature, we consider in this study two classifiers-combination in each dataset ($q = 2$). Given the advantages of tree-based techniques explained in section 2.1.3, we focus on portfolios using random forest and gradient boosted decision trees as our base classifiers. We also conduct robustness experiments for alternative choices of base models (please refer to section 5 for more details regarding robustness checks).

Table 5: Overview of Benchmark Allocation Strategies

Strategy	Optimization function/Weight	Required parameters
Risk Parity	$w_i = \frac{1/\hat{\sigma}_i^2}{\sum_{i=1}^{N}(1/\hat{\sigma}_i^2)}$	Variances of asset returns
Reward-to-risk Timing	$w_i = \frac{\hat{\mu}_i/\hat{\sigma}_i^2}{\sum_{i=1}^{N}(\hat{\mu}_i/\hat{\sigma}_i^2)}$	Returns and variances of asset returns
Mean-variance	$max\, U = \boldsymbol{\mu}^{\boldsymbol{T}}\boldsymbol{w} - \frac{\delta}{2}\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{\Sigma}\boldsymbol{w}$	Risk-aversion level, covariance matrix and return estimates for all assets
Maximum Sharpe ratio (MSR)	$\max SR = \frac{\boldsymbol{\mu}^{\boldsymbol{T}}\boldsymbol{w} - r_f}{\sqrt{\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{\Sigma}\boldsymbol{w}}}$	Covariance matrix and return estimates for all assets, risk-free rate
MSR with Ledoit Wolf shrinkage covariance	$\max SR = \frac{\boldsymbol{\mu}^{\boldsymbol{T}}\boldsymbol{w} - r_f}{\sqrt{\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{\Sigma}_{\boldsymbol{LW}}\boldsymbol{w}}}$ with: $\hat{\kappa} = \frac{\hat{\pi} - \hat{\rho}}{\hat{\gamma}}$ $\varsigma = \max\{0, \min\{\frac{\hat{\kappa}}{T}, 1\}\}$ $\boldsymbol{\Sigma}_{LW} = \varsigma * \widehat{\boldsymbol{F}} + (1-\varsigma) * \boldsymbol{\Sigma}$	Covariance matrix, average sample correlation, shrinkage intensity, risk-free rate
Black-Litterman	$\max SR = \frac{{\mu_{BL}}^{\boldsymbol{T}} w - r_f}{\sqrt{w^T \Sigma_{BL} w}}$ with: $\boldsymbol{\pi} = \delta\, \boldsymbol{\Sigma}\boldsymbol{w}^{\boldsymbol{ref}}$ $\boldsymbol{\mu}_{BL} = [(\tau\boldsymbol{\Sigma})^{-1} + \boldsymbol{P}'\boldsymbol{\Omega}^{-1}\mathrm{P}]^{-1}[(\tau\boldsymbol{\Sigma})^{-1}\boldsymbol{\pi} + \boldsymbol{P}'\boldsymbol{\Omega}^{-1}\boldsymbol{V}]$, $\boldsymbol{\Sigma}_{BL} = \boldsymbol{\Sigma} + [(\tau\boldsymbol{\Sigma})^{-1} + \boldsymbol{P}'\boldsymbol{\Omega}^{-1}\boldsymbol{P}]^{-1}$	Risk-aversion coefficient, covariance matrix, return estimates for all assets, reference portfolio, risk-free rate
Bayes-stein Shrinkage	$\max SR = \frac{{\boldsymbol{\mu}_{\boldsymbol{BS}}}^{\boldsymbol{T}}\boldsymbol{w} - r_f}{\sqrt{\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{\Sigma}_{\boldsymbol{BS}}\boldsymbol{w}}}$ with: $g_{mv} = \frac{\mathrm{N}+2}{(\mathrm{N}+2) + \mathrm{M}\left(\boldsymbol{\mu}_{\boldsymbol{ml}} - \mu_g\boldsymbol{1}\right)^T\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu}_{\boldsymbol{ml}} - \mu_g\boldsymbol{1}\right)}$ $\boldsymbol{\mu}_{\boldsymbol{BS}} = (1 - g_{mv})\,\boldsymbol{\mu}_{\boldsymbol{ml}} + g_{mv}\mu_g\boldsymbol{1}$ $\tilde{\upsilon} = \frac{M+2}{(\boldsymbol{\mu}_{\boldsymbol{ml}} - \mu_g\boldsymbol{1})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_{\boldsymbol{ml}} - \mu_g\boldsymbol{1})}$ $\boldsymbol{\Sigma}_{BS} = \boldsymbol{\Sigma}\left(\frac{N+\tilde{\upsilon}+1}{N+\tilde{\upsilon}}\right) + \frac{\tilde{\upsilon}}{N(N+\tilde{\upsilon}+1)}\frac{\boldsymbol{1}\boldsymbol{1}^T}{\boldsymbol{1}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{1}}$	Risk-free rate, covariance matrix, length of estimation window, maximum-likelihood estimation for returns, global mean return

with: w_i=portfolio weight of asset i ($\boldsymbol{w}$=vector of portfolio weights)
$\hat{\sigma}_i^2$=volatility estimate of asset i
U=utility function of investor
$\boldsymbol{\mu}$=vector of return estimates
$\boldsymbol{\pi}$=vector of implied return estimates
$\boldsymbol{\Sigma}$=estimates of covariance matrix
τ= reliability measure of equilibrium expected returns
$\boldsymbol{\mu}_{\boldsymbol{ml}}$=vector of maximum-likelihood estimation for returns
$\boldsymbol{P}$ =unity matrix (in case of an absolute return estimate for each asset
$\widehat{\boldsymbol{F}}$=sample constant correlation matrix (shrinkage target)
$\hat{\pi}$=constant estimator for the sum of asymptotic variances of the entries of $\boldsymbol{\Sigma}$ scaled by $\sqrt{T}$
$\hat{\rho}$=sum of asymptotic covariances of the entries of the shrinkage target with the entries of $\boldsymbol{\Sigma}$ scaled by $\sqrt{T}$
$\hat{\gamma}$=misspecification of the (population) shrinkage targe
$\boldsymbol{r}_{\boldsymbol{f}}$=risk-free rate
$\boldsymbol{V}$=vector of investors' views
δ=risk version level
$\boldsymbol{\Omega}$=reliability of views
g_{mv}=shrinkage factor
$\mathbf{M}$=length of estimation window
μ_g= global mean return
$\tilde{\upsilon}$=prior precision
$\boldsymbol{w}^{\boldsymbol{ref}}$=reference portfolio weights
ς=shrinkage intensity

This table summarizes details of the benchmark strategies employed in this paper. We report from left to right the model, its optimization functions/methodology for obtaining the optimal portfolio weights, the weight constraint imposed, and the parameters required for implementing the optimization framework. For readability, we drop the explicit time index.

4.1 Fama French 17 Industry Portfolio

Table 6 reports the out-of-sample realized performance for Fama French 17 Industry portfolios. Among the seventeen allocation strategies that we examined, the no-short-position GBL strategy with portfolio combination and GLASSO-Wishart model (GBL-OG-c) performs the best for risk-reward maximization, followed by its variation which allows short-position (GBL-OG-u). The difference in Sharpe ratios vs. the 1/N is statistically significant for both strategies, and results are consistent when transaction costs are in absence or in presence. The reward-to-risk timing (RRT), maximum Sharpe ratio without short-position (MSR-c), MSR-c with a shrinkage covariance estimator ($LW - MSR - c$), and Black-Litterman strategy BL-u, BL-c) also perform well against alternative benchmark strategies, indicating that these are effective approaches to manage parameter uncertainty in the mean-variance framework. However, these strategies are less effective than GBL-OG-u and GBL-OG-c. Comparing to 1/N strategy, no statistically significant spread in the Sharpe ratio have been documented for MSR-c and $LW - MSR - c$ pre-transaction costs, for BL-u and BL-c post-transaction costs. In addition, the performance ratios are generally lower for RRT than for GBL-OG-u and GBL-OG-c.

Table 6: Out-of-sample Performance for dataset 17Ind (2016:1-2020:12)

	Ret	Std	SR	SOR	ROV	Ret	Std	SR	SOR	ROV
	Pre-Transaction Cost					Post-Transaction Cost				
Panel A: Classic allocation strategies										
Naïve	0.155	0.244	0.59	0.993	0.348	0.154	0.244	0.585	0.984	0.344
Riskparity	0.146	0.282	0.481	0.828	0.265	0.145	0.283	0.476	0.819	0.262
RRT	0.146	0.187	0.724**	1.183	0.467	0.145	0.187	0.719**	1.175	0.463
MeanVar-u	0.092	0.362	0.225	0.332	0.108	0.05	0.364	0.108	0.157	0.049
MaxSR-u	0.066	0.151	0.366	0.53	0.194	0.053	0.151	0.283	0.404	0.143
Lediot_MSR-u	0.059	0.156	0.311	0.438	0.159	0.047	0.156	0.23	0.321	0.113
Panel B: Bayesian approach to estimation risk										
BS_u	0.102	0.175	0.523	0.786	0.3	0.087	0.174	0.441	0.653	0.241
BL_u	0.544	0.661	0.807*	1.353	0.536	0.521	0.66	0.773	1.286	0.502
Panel C: Generalized Black-Litterman (GBL) approach										
GBL-OS-u	1.167	1.714	0.674*	1.202	0.409	1.066	1.709	0.618	1.086	0.362
GBL-NG-u	0.24	0.443	0.518	0.827	0.29	0.177	0.442	0.375	0.587	0.194
GBL-OG-u	0.393	0.464	0.824**	1.431	0.556	0.374	0.464	0.783**	1.351	0.514
Panel D: Portfolio constraints										
MaxSR-c	0.155	0.161	0.894	1.539	0.653	0.149	0.161	0.858	1.471	0.611
Lediot_MSR-c	0.155	0.178	0.812	1.409	0.557	0.149	0.178	0.776	1.339	0.519
BL-c	0.551	0.722	0.748**	1.226	0.478	0.531	0.72	0.722*	1.173	0.453
GBL-OS-c	1.504	2.092	0.714**	1.284	0.443	1.244	2.093	0.589	1.011	0.34
GBL-NG-c	0.318	0.478	0.643	1.095	0.386	0.286	0.476	0.579	0.978	0.335
GBL-OG-c	0.372	0.429	**0.841****	1.433	0.575	0.355	0.429	**0.802****	1.358	0.534

This table reports the annualized out-of-sample portfolio average return (Ret), standard deviation (Std), Sharpe ratio (SR), Sortino ratio (SOR), and return over Value-at-Risk at 99% confidence level (ROV) for the benchmark strategies and the GBL strategies for Fama French 17 Industry portfolio (Ind17). We present the significance level from a test of a zero difference between Sharpe ratio of each strategy vs. the 1/N, where ***, **, * represent 1%, 5%, and 10% test size levels, respectively. For Sharpe ratio, boldfacing indicates the best-performing asset allocation strategy when applied to a dataset.

Note that the decision-fusion model can be used to generate two additional strategies: GBL-OS and GBL-NG, i.e., (i) rules with optimal portfolio combination and sample covariance; (ii) rules with a reference portfolio of 1/N and GLASSO-Wishart covariance (the sample covariance and a reference portfolio of 1/N are used as inputs in our baseline BL strategy). We compare the realized performance for GBL-OS and GBL-NG with GBL-OG. Such comparisons can offer direct insight in the effect of adopting the two strategies. In panel D, we noted that GBL-OG-c outperforms both GBL-NG-c and GBL-OS-c (same for panel C when short-positions are allowed), suggesting that adopting portfolio combination together with GLASSO-

Wishart model for the realized expected returns and covariance estimates can generate a better performing portfolio than using one of these strategies alone to deal with estimation risk. The finding is also confirmed in Kan et al. (2021) that portfolios using the optimal combining coefficient with shrinkage estimators usually generates better performance.

Next, we consider portfolios of the remaining benchmark strategies. Although the optimization-based strategies (MaxSR-u, $LW-MSR-u$, BS-u) fail to improve the out-of-sample realized portfolio return, they usually results in a less volatile portfolio than 1/N with lower standard deviation, which is also consistent to the finding in Sahamkhadam et al. (2022). MaxSR-u performs the best in terms of minimizing portfolio variance, followed by $LW-MSR-u$, where the standard deviations are 38% and 36% lower than that of the naïve strategy, respectively.

4.2 Fama French 25 Size and Short-term Reversal Sorted Portfolios

Table 7: Out-of-sample Performance for dataset FF25 (2016:1-2020:12)

	Ret	Std	SR	SOR	ROV	Ret	Std	SR	SOR	ROV
	Pre-Transaction Cost					Post-Transaction Cost				
Panel A: Classic allocation strategies										
Naïve	0.166	0.219	0.71	1.175	0.452	0.165	0.219	0.706	1.167	0.448
Riskparity	0.171	0.232	0.69	1.155	0.433	0.17	0.232	0.686	1.147	0.43
RRT	0.162	0.207	0.731	1.2	0.472	0.161	0.207	0.727	1.193	0.469
MeanVar-u	0.064	0.334	0.161	0.254	0.075	0.016	0.333	0.017	0.027	0.008
MaxSR-u	0.047	0.09	0.403	0.787	0.223	0.032	0.089	0.234	0.442	0.118
Lediot_MSR-u	0.034	0.086	0.274	0.492	0.142	0.021	0.086	0.119	0.208	0.057
Panel B: Bayesian approach to estimation risk										
BS_u	0.081	0.106	0.66	1.379	0.421	0.066	0.105	0.52	1.058	0.305
BL_u	0.279	0.353	0.759	1.428	0.493	0.263	0.354	0.713	1.326	0.449
Panel C: Generalized Black-Litterman (GBL) approach										
GBL-OS-u	2.041	2.393	0.849**	1.503	0.575	1.81	2.392	0.752	1.316	0.478
GBL-NG-u	0.118	0.158	0.679	1.128	0.429	0.013	0.16	0.016	0.025	0.007
GBL-OG-u	1.067	1.181	**0.895****	1.587	0.628	0.995	1.162	**0.847****	1.466	0.575
Panel D: Portfolio constraints										
MaxSR-c	0.194	0.23	0.796	1.49	0.536	0.19	0.23	0.782	1.456	0.521
Lediot_MSR-c	0.189	0.267	0.668	1.066	0.412	0.186	0.267	0.658	1.047	0.403
BL-c	0.387	0.493	0.764	1.293	0.495	0.376	0.492	0.743	1.251	0.475
GBL-OS-c	2.404	2.912	0.822**	1.396	0.546	-0.405	5.923	-0.07	-0.079	-0.029
GBL-NG-c	0.332	0.579	0.555	0.853	0.316	0.293	0.579	0.487	0.736	0.267
GBL-OG-c	0.837	1.128	0.733	1.23	0.461	0.765	1.108	0.681	1.113	0.415

This table reports the annualized out-of-sample portfolio average return (Ret), standard deviation (Std), Sharpe ratio (SR), Sortino ratio (SOR), and return over Value-at-Risk at 99% confidence level (ROV) for the benchmark strategies and the GBL strategies for Fama French 25 portfolios sorted by size and short term reversal (FF25). We present the significance level from a test of a zero difference between Sharpe ratio of each strategy vs. the 1/N, where ***, **, * represent 1%, 5%, and 10% test size levels, respectively. For Sharpe ratio, boldfacing indicates the best-performing asset allocation strategy when applied to a dataset.

Table 7 reports the realized performance for Fama French 25 size and short-term reversal sorted portfolios. For the realized performance reported in panel A and B, we find that the non-optimization based strategies, i.e., naïve diversification, risk-parity and the reward-to-risk timing (RRT), outperform those optimization-based strategies (MeanVar-u, MaxSR-u, LW-MSR, BS-u). Moreover, none of the benchmark strategies can yield a significantly higher Sharpe ratio compared to the 1/N strategy. For RRT and BL-u which outperforms the 1/N, the difference in their Sharpe versus that of the 1/N is not statistically significant, and the potential gains which can be realized by investors are relatively marginal (for example, 0.731 for RRT vs. 0.710 for 1/N). Consistent with DeMiguel et al. (2009) and Harris and Mazibas (2022), our finding confirms the well-known hazard of using sample-based estimates of the moments of asset returns to implement Markowitz's mean-variance optimization.

We also note that for benchmark strategies, portfolios with short-position always underperform those without short-position and the difference in the realized performance is quite sizable, especially for MaxSR-u and LW-MSR framework (e.g., the realized Sharpe ratio of MaxSR can be improved from 0.403 to 0.796), suggesting that the estimation risk hindered in these optimization methods is so large that regularization on portfolio weights must be in place. However, the effect of imposing weight constraint is in general less effective for the GBL strategies, noting that all GBL strategies with short-position outperform their no-short-position counterparts, GBL is closer to find the optimal allocation points to maximize the portfolio Sharpe ratio with short-position allowed, suggesting that the sparsity introduced in the estimation of expected return and covariance using LASSO and Graphical LASSO approach can help to generate robust and stable portfolios.

For results reported in panel C and panel D, we observe similar patterns to dataset Ind17. First, we observe a better risk-return profile for the GBL strategies compared with the benchmark strategies. When transaction costs are not taken into account, half of GBL strategies deliver a Sharpe ratio over 0.82 while most benchmark strategies have Sharpe ratio well below 0.7. When transaction costs are in presence, GBL-OG-u remains the only strategy that can significantly outperform the 1/N among the seventeen allocation strategies that we examine. The magnitude of the potential gains that can actually be realized by investors are also substantial— GBL-OG-u delivers the highest out-of-sample Sharpe, Sortino, and ROV. Second, when decision-fusion is used together with portfolio combination (GBL-OS) or GLASSO-Wishart model (GBL-NG) alone, the resulting portfolio usually underperforms the GBL-OG, which further supports the argument that adopting decision-fusion, portfolio combination, together with the GLASSO-Wishart model can reduce the estimation risk than using decision-fusion with any of the two strategies.

4.3 Fama French Size 30 Industry Portfolio

Table 8 reports the realized performance for Fama French 30 Industry portfolios. We noted that the outperformance of GBL strategies over the benchmark and the BL strategies is greater comparing to Ind17 and FF25. Relative to the 1/N strategy, five GBL strategies yield significant spread in Sharpe ratio (GBL-NG-u, GBL-OS-u, GBL-OG-u, GBL-OG-c), two yield significantly higher Sharpe ratios post-transaction costs (GBL-OG-u and GBL-OG-c), while none of the benchmark strategies are able to outperform the 1/N significantly. Although BL-c results in significantly higher Sharpe ratios before and after transaction costs are considered, the improvement over 1/N is relatively minor, comparing to GBL-OG-u and GBL-OG-c (e.g., Sharpe ratio of 0.564 for 1/N, 0.686 for BL-c, 0.732 for GBL-OG-c, and 0.759 for GBL-OG-u, pre-transaction costs). For results reported in panel C and D, we can see that high turnover is the most troublesome aspect of the performance of GBL-OS strategies. Although GBL-OS-u has the highest performance ratios pre-transaction costs, the turnover is sufficiently high that there is a substantial deterioration in the performance of the strategy, the outperformance over 1/N is no long significant with transaction costs of 20 bps. Comparing with GBL-OS and GBL-NG, we find that GBL-OG leads to a reasonable turnover. When transaction costs are in presence, both GBL-OG-u and GBL-OG-c have significantly higher Sharpe ratio over 1/N at 5% level. Similar to the pattern observed in *Ind17* and *FF25*, we find that imposing a no-short-position constraint does not seem to be effective in improving performance for GBL strategy, indeed, GBL-NG-u, GBL-OS-u, and GBL-OG-u all outperformed their short-sale constrained version when transaction costs are in absence or in presence. To summarize our findings, the newly obtained GBL strategy can outperform eleven well-studied benchmark strategies by a clear margin in terms of risk-return metrics. Using regularization methods for estimating the equilibrium expected returns and covariance matrix is able to reduce the sensitivity of optimization to the estimation error. For dataset with a large number of risky assets (e.g., N=17), using optimal portfolio combination together with the GLASSO-Wishart model in the GBL strategy (GBL-OG) can yield superior performance than using one of the models alone (GBL-OS, GBL-NG), the outperformance comes with higher risk-reward ratios and more reasonable turn-overs.

Table 8: Out-of-sample Performance for dataset 30Ind (2016:1-2020:12)

	Ret	Std	SR	SOR	ROV	Ret	Std	SR	SOR	ROV
	Pre-Transaction Cost					Post-Transaction Cost				
Panel A: Classic allocation strategies										
Naïve	0.146	0.24	0.564	0.945	0.328	0.145	0.24	0.559	0.936	0.324
Riskparity	0.112	0.273	0.373	0.647	0.194	0.111	0.273	0.368	0.638	0.191
RRT	0.15	0.193	0.722**	1.187	0.465	0.149	0.193	0.717**	1.178	0.461
MeanVar-u	0.137	0.544	0.232	0.432	0.112	0.045	0.546	0.063	0.114	0.028
MaxSR-u	0.079	0.198	0.346	0.554	0.179	0.058	0.197	0.239	0.377	0.118
Lediot_MSR-u	0.078	0.218	0.31	0.518	0.157	0.058	0.217	0.219	0.36	0.106
Panel B: Bayesian approach to estimation risk										
BS_u	0.134	0.215	0.576	0.936	0.338	0.113	0.214	0.478	0.763	0.265
BL_u	0.553	0.691	0.785	1.486	0.513	0.523	0.689	0.743	1.391	0.473
Panel C: Generalized Black-Litterman (GBL) approach										
GBL-OS-u	2.028	2.692	0.749**	1.416	0.475	1.471	2.891	0.505	0.832	0.277
GBL-NG-u	0.251	0.177	1.360**	6.949	1.496	0.137	0.177	0.716	2.646	0.461
GBL-OG-u	0.493	0.59	**0.818*** **	1.488	0.547	0.458	0.59	**0.759** **	1.364	0.489
Panel D: Portfolio constraints										
MaxSR-c	0.379	0.39	0.944	1.713	0.695	0.373	0.39	0.929	1.682	0.676
Lediot_MSR-c	0.315	0.361	0.845	1.531	0.581	0.31	0.361	0.83	1.499	0.564
BL-c	0.871	1.194	0.721**	1.197	0.45	0.824	1.184	0.686*	1.123	0.42
GBL-OS-c	2.367	3.539	0.666**	1.169	0.401	2.001	3.562	0.559	0.955	0.316
GBL-NG-c	0.326	0.547	0.577	1.011	0.333	0.275	0.548	0.481	0.826	0.263
GBL-OG-c	0.477	0.594	0.784**	1.39	0.513	0.446	0.594	0.732**	1.286	0.464

This table reports the annualized out-of-sample portfolio average return (Ret), standard deviation (Std), Sharpe ratio (SR), Sortino ratio (SOR), and return over Value-at-Risk at 99% confidence level (ROV) for the benchmark strategies and the GBL strategies for Fama French 30 Industry portfolio (Ind30). We present the significance level from a test of a zero difference between Sharpe ratio of each strategy vs. the 1/N, where ***, **, * represent 1%, 5%, and 10% test size levels, respectively. For Sharpe ratio, boldfacing indicates the best-performing asset allocation strategy when applied to a dataset.

4.4 Fama French 10 Size and Short-term reversal Sorted Portfolios

Table 9 reports results for Fama French 10 portfolios sorted by size and short-term reversal. For the realized performance reported in panel A and B, again, we find that strategies that do not require optimization (1/N, risk-parity and RRT) tend to outperform those rely on an optimization framework (e.g., MeanVar-u, MaxSR-u, Lediot-MSR, BS-u). Furthermore, we find that RRT is in general, more effective in dealing with estimation risk than the alternative benchmark strategies, it outperforms 1/N at 5% significance level in two datasets we examine and the performance persists after accounting for the impact of transaction costs. Our finding is consistent to the argument made by Kirby and Ostdiek (2012) that there is substantial value of using sample information to guide portfolio selection decisions.

Next, we consider performance for the GBL strategy. For results reported in panel C and panel D, we find GBL-OG-u, GBL-OS-c, and GBL-NG-c to significantly outperform the 1/N at 5%, 1%, and 10% level, respectively, when transaction costs are not accounted for. When transaction costs of 20 bps are considered, GBL-OG-u and GBL-OS-c can outperform 1/N at 10% and 1%, respectively, suggesting that applying the decision-fusion model with optimal portfolio combination is a powerful tool to manage the estimation risk.

Table 9: Out-of-sample Performance for dataset FF10 (2016:1-2020:12)

	Ret	Std	SR	SOR	ROV	Ret	Std	SR	SOR	ROV
	Pre-Transaction Cost					Post-Transaction Cost				
Panel A: Classic allocation strategies										
Naïve	0.183	0.187	0.918	1.591	0.677	0.182	0.188	0.913	1.581	0.671
Riskparity	0.195	0.213	0.866	1.513	0.612	0.194	0.213	0.861	1.504	0.607
RRT	0.173	0.173	0.941	1.617	0.709	0.173	0.173	0.937	1.608	0.703
MeanVar-u	0.1	0.172	0.523	0.901	0.3	0.085	0.172	0.433	0.737	0.236
MaxSR-u	0.035	0.072	0.341	0.539	0.185	0.028	0.073	0.244	0.381	0.126
Lediot_MSR-u	0.046	0.088	0.399	0.646	0.22	0.038	0.088	0.31	0.496	0.163
Panel B: Bayesian approach to estimation risk										
BS_u	0.071	0.096	0.634	1.074	0.4	0.065	0.096	0.57	0.954	0.345
BL_u	0.336	0.342	0.95	1.609	0.704	0.328	0.343	0.925	1.558	0.673
Panel C: Generalized Black-Litterman (GBL) approach										
GBL-OS-u	0.899	0.942	0.943	1.621	0.685	0.863	0.936	0.911	1.544	0.647
GBL-NG-u	0.379	0.468	0.787	1.124	0.518	0.346	0.467	0.719	1.014	0.453
GBL-OG-u	0.817	0.789	**1.022****	1.856	0.789	0.791	0.788	**0.991***	1.784	0.748
Panel D: Portfolio constraints										
MaxSR-c	0.199	0.178	1.055	1.696	0.869	0.196	0.178	1.039	1.666	0.845
Lediot_MSR-c	0.227	0.225	0.963	1.505	0.73	0.223	0.225	0.945	1.473	0.707
BL-c	0.405	0.395	0.996	1.741	0.762	0.397	0.396	0.977	1.701	0.737
GBL-OS-c	1.054	0.931	1.121***	2.165	0.936	1.021	0.928	1.089***	2.079	0.886
GBL-NG-c	0.524	0.476	1.078*	2.07	0.876	0.497	0.476	1.021	1.941	0.794
GBL-OG-c	0.73	0.687	1.047	1.829	0.826	0.706	0.686	1.013	1.752	0.779

This table reports the annualized out-of-sample portfolio average return (Ret), standard deviation (Std), Sharpe ratio (SR), Sortino ratio (SOR), and return over Value-at-Risk at 99% confidence level (ROV) for the benchmark strategies and the GBL strategies for Fama French 10 portfolios sorted by size and short-term reversal (FF10). We present the significance level from a test of a zero difference between Sharpe ratio of each strategy vs. the 1/N, where ***, **, * represent 1%, 5%, and 10% test size levels, respectively. For Sharpe ratio, boldfacing indicates the best-performing asset allocation strategy when applied to a dataset.

For Ind17, FF25 and Ind30, we find that GBL-OG always outperforms those based on either the optimal portfolio combination or the GLASSO-Wishart model alone. However, this is not always the case for datasets with relatively small number of risky assets, i.e., FF10. Comparing results in panel C and E, we observe GBL-OG-u (based on GLASSO-Wishart covariance) to underperform GBL-OS-c (based on the sample covariance) before or after transaction costs are accounted for. The difference between GBL-OG (based on optimal portfolio combination) and GBL-NG (based on the 1/N) is not very obvious, GBL-OG-u underperforms GBL-NG-c pre-transaction costs but outperforms it post-transaction costs, indicating that the underperformance in GBL-OG is mainly driven by the use of GLASSO-Wishart model. Our finding is closely related to Kan et al. (2021) where the authors find portfolios relying on MacKinlay-Pastor single factor structure and an optimal combination coefficient perform worse in dataset with smaller number of risky assets. In our GLASSO-Wishart model, the dynamics of Wishart specification is also governed by one single factor, the discount factor. As discussed in Kan et al. (2021), in most cases, the single factor structure is not true in population and imposing such a structure introduces a bias. When the optimal portfolio combination strategy is used, the effect of estimation risk is already reduced substantially, and it is only beneficial to further impose the single factor structure if the gain from further reduced estimation risk outweighs the cost coming from introducing bias. When more risky assets are involved, the estimation risk are larger, and it is more likely for GBL-OG to outperform GBL-NG and GBL-OS.

5 Robustness Checks

In section 4, we have reported empirical results only for the base-case estimates of asset moments. To check the sensitivity of our results to various assumptions, we generate results for the Sharpe ratio, Sortino ratio, and return over Value-at-Risk for all allocation strategies and datasets considered after relaxing the various assumptions made above. These exercises are based on: (1) alterative risk aversion coefficient; (2) alternative estimation window size (conditional information that used to estimate asset moments); (3) alternative reliability measure for the equilibrium expected returns, specifically, we report for a reliability measure of 0.1625, which corresponds to the mean of the range of values as adopted by the pre-existing empirical work (see Black and Litterman (1992); He and Litterman (2002); Idzorek (2007); among many others); (4) alternative holding period for assets in the portfolio; and (5) alternative base models for the decision-fusion process. We summarize the various robustness checks in Table 10. Because of the large set of tables for the exercises, we provide an online supplementary document summarizing all the empirical results. The main finding is that the performance reported in the baseline set of results for various strategies does not depend significantly on above assumptions.

6 Conclusion

In this paper, we introduce and evaluate a novel GBL approach for modelling and forecasting asset moments. The main objective of this paper is to address three potential shortcomings within the classic BL framework: (1) the specification error associated with the CAPM equilibrium; (2) limited discussion in the dynamic conditional volatilities; and (3) little emphasis on the role of investors, their knowledge and uncertainty. Our model generalizes the BL framework to incorporate a decision-fusion process, where an entropy-based multiple classifier combination system nested within the Dempster-Shafer Theory of Evidence is used for modelling investors' views. We provide a non-analytical and efficient portfolio shrinkage technique to obtain the equilibrium expected return, which is able to alleviate the specification error associated with CAPM, and exploit information inherited in the data to generate more reliable trading strategies. Finally, we propose a diversification-based GLASSO-Wishart model for the analysis of the realized covariance matrix of asset returns. Empirically, we find that restricting the matrix of Wishart distribution to a sparse parameterization by using an L-1 penalty generates a substantial improvement over the 1/N portfolio, for datasets with relatively large numbers of risky assets.

We use the methodology to compute efficient investment portfolios and compare with a variety of benchmark strategies from the literature, such as the Markowitz's mean-variance, maximum-Sharpe ratio, Bayes-Stein shrinkage of Jorion (1986), covariance shrinkage of Ledoit and Wolf (2003), reward-to-risk timing of Kirby and Ostdiek (2012), naïve-diversification, and the risk-parity rule. Experiments with four datasets support the use of our GBL approach. First, we show that the realized performance, as measured by its Sharpe ratio, Sortino ratio, return over VaR, is considerably higher than alternative rules in the absence or presence of transaction costs. Second, we find the use of the GBL strategy can stabilize the optimization problem. By using regularization methods (LASSO, and Graphical LASSO) in the estimation of BLM's inputs, we reduce the sensitivity of the optimization to estimation errors, and the results are stable even when no constraint is in place. We strive to ensure that the GBL approach can be as free as possible from both the restrictions of theoretical models. Our results are robust to the holding period, the estimation window size, and the choice of risk aversion parameter, reliability measure, base model.

The GBL strategy can be extended in several directions. First, alternative classifiers-combination technique, (e.g., Bayesian fusion methods, fuzzy integrals) can also be used in this approach. Second, it will be of interest to examine the gains of our proposed strategy in other asset classes, such as bonds, commodities, and crypto-currencies.

References

Al-Ani, A. and Deriche, M. (2002). A new technique for combining multiple classifiers using the dempster-shafer theory of evidence. *Journal of Artificial Intelligence Research*, 17:333–361.

Ang, A. and Chen, J. (2002). Asymmetric correlations of equity portfolios. *Journal of financial Economics*, 63(3):443–494.

Banerjee, O., El Ghaoui, L., and d'Aspremont, A. (2008). Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data. *The Journal of Machine Learning Research*, 9:485–516.

Beach, S. L. and Orlov, A. G. (2007). An application of the black–litterman model with egarch-m-derived views for international portfolio management. *Financial Markets and Portfolio Management*, 21(2):147–166.

Bertsimas, D., Gupta, V., and Paschalidis, I. C. (2012). Inverse optimization: A new perspective on the black-litterman model. *Operations Research*, 60(6):1389–1403.

Bessler, W., Opfer, H., and Wolff, D. (2017). Multi-asset portfolio optimization and out-of-sample performance: an evaluation of black–litterman, mean-variance, and naïve diversification approaches. *The European Journal of Finance*, 23(1):1–30.

Best, M. and Grauer, R. (1991). On the sensitivity of mean-variance-efficient portfolios to changes in asset means: some analytical and computational results. *The Review of Financial Studies*, 4(2):315–342.

Bianchi, D., Büchner, M., and Tamoni, A. (2021). Bond risk premia with machine learning. *The Review of Financial Studies*, 34(2):1046–1089.

Bianchi, D. and Guidolin, M. (2014). Can long-run dynamic optimal strategies outperform fixed-mix portfolios? evidence from multiple data sets. *European Journal of Operational Research*, 236(1):160–176.

Black, F. and Litterman, R. (1992). Global portfolio optimization. *Financial Analysts Journal*, 48(5):28–43.

Blanquero Bravo, R., Carrizosa Priego, E. J., Molero Río, C., and Romero Morales, M. D. (2019). Sparsity in optimal randomized classification trees. *European Journal of Operational Research, 284 (1), 255-272.*

Board, J. L. and Sutcliffe, C. M. (1994). Estimation methods in portfolio selection and the effectiveness of short sales restrictions: UK evidence. *Management Science*, 40(4):516–534.

Boivin, J. and Ng, S. (2006). Are more data always better for factor analysis? *Journal of Econometrics*, 132(1):169–194.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

Brodie, J., Daubechies, I., De Mol, C., Giannone, D., and Loris, I. (2009). Sparse and stable markowitz portfolios. *Proceedings of the National Academy of Sciences*, 106(30):12267–12272.

Brown, G., Wyatt, J., Harris, R., and Yao, X. (2005). Diversity creation methods: a survey and categorisation. *Information fusion*, 6(1):5–20.

Charte, F., Rivera, A. J., del Jesus, M. J., and Herrera, F. (2019). Remedial-hwr: Tackling multilabel imbalance through label decoupling and data resampling hybridization. *Neurocomputing*, 326:110–122.

Chen, S. D. and Lim, A. E. (2020). A generalized black–litterman model. *Operations Research*, 68(2):381–410.

Cheung, W. (2013). The augmented black–litterman model: A ranking-free approach to factor-based portfolio construction and beyond. *Quantitative Finance*, 13(2):301–316.

Chinco, A., Clark-Joseph, A. D., and Ye, M. (2019). Sparse signals in the cross-section of returns. *The Journal of Finance*, 74(1):449–492.

Cujean, J. and Hasler, M. (2017). Why does return predictability concentrate in bad times? *The Journal of Finance*, 72(6):2717–2758.

Dangl, T. and Halling, M. (2012). Predictive regressions with time-varying coefficients. *Journal of Financial Economics*, 106(1):157–181.

DeMiguel, V., Garlappi, L., and Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The Review of Financial Studies*, 22(5):1915–1953.

Dempster, A. P. (1968). A generalization of bayesian inference. *Journal of the Royal Statistical Society: Series B (Methodological)*, 30(2):205–232.

Donthireddy, P. (2018). Black-litterman portfolios with machine learning derived views.

Engel, C. M. and Rodrigues, A. P. (1993). Tests of mean-variance efficiency of international equity markets. *Oxford Economic Papers*, pages 403–421.

Evangelou, C., Karacapilidis, N., and Khaled, O. A. (2005). Interweaving knowledge management, argumentation and decision making in a collaborative setting: the kad ontology model. *International Journal of Knowledge and Learning*, 1(1-2):130–145.

Fabozzi, F. J., Focardi, S. M., and Kolm, P. N. (2006). Incorporating trading strategies in the black-litterman framework. *The Journal of Trading*, 1(2):28–37.

Fernandes, B., Street, A., Fernandes, C., and Valladão, D. (2018). On an adaptive black–litterman investment strategy using conditional fundamentalist information: A brazilian case study. *Finance Research Letters*, 27:201–207.

Ferson, W. E., Kandel, S., and Stambaugh, R. F. (1987). Tests of asset pricing with time-varying expected risk premiums and market betas. *The Journal of Finance*, 42(2):201–220.

Fix, E. and Hodges, J. L. (1989). Discriminatory analysis. nonparametric discrimination: Consistency properties. *International Statistical Review/Revue Internationale de Statistique*, 57(3):238–247.

Flannery, M. J. and Protopapadakis, A. A. (2002). Macroeconomic factors do influence aggregate stock returns. *The review of financial studies*, 15(3):751–782.

Fridman, M. and Harris, L. (1998). A maximum likelihood approach for non-gaussian stochastic volatility models. *Journal of Business & Economic Statistics*, 16(3):284–291.

Friedman, J., Hastie, T., and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232.

Fugazza, C., Guidolin, M., and Nicodano, G. (2015). Equally weighted vs. long-run optimal portfolios. *European Financial Management*, 21(4):742–789.

Geyer, A. and Lucivjanská, K. (2016). The black–litterman approach and views from predictive regressions: Theory and implementation. *The Journal of Portfolio Management*, 42(4):38–48.

Giacometti, R., Bertocchi, M., Rachev, S. T., and Fabozzi, F. J. (2007). Stable distributions in the black–litterman approach to asset allocation. *Quantitative Finance*, 7(4):423–433.

Giacometti, R. and Mignacca, D. (2010). Using the black and litterman framework for stress test analysis in asset management. *Journal of asset management*, 11(4):286–297.

Gibbons, M. R., Ross, S. A., and Shanken, J. (1989). A test of the efficiency of a given portfolio. *Econometrica: Journal of the Econometric Society*, pages 1121–1152.

Green, R. C. and Hollifield, B. (1992). When will mean-variance efficient portfolios be well diversified? *The Journal of Finance*, 47(5):1785–1809.

Gu, S., Kelly, B., and Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5):2223–2273.

Harris, R. D. and Mazibas, M. (2022). Portfolio optimization with behavioural preferences and investor memory. *European Journal of Operational Research*, 296(1):368–387.

Harris, R. D., Stoja, E., and Tan, L. (2017). The dynamic black–litterman approach to asset allocation. *European Journal of Operational Research*, 259(3):1085–1096.

Harvey, A., Ruiz, E., and Shephard, N. (1994). Multivariate stochastic variance models. *The Review of Economic Studies*, 61(2):247–264.

Harvey, C. R. (1989). Time-varying conditional covariances in tests of asset pricing models. *Journal of Financial economics*, 24(2):289–317.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.

Hawkins, D. M. (2004). The problem of overfitting. *Journal of chemical information and computer sciences*, 44(1):1–12.

He, D., Xu, J., and Chen, X. (2016). Information-theoretic-entropy based weight aggregation method in multiple-attribute group decision-making. *Entropy*, 18(6):171.

He, G. and Litterman, R. (2002). The intuition behind black-litterman model portfolios. *Available at SSRN 334304*.

Idzorek, T. (2007). A step-by-step guide to the black-litterman model: Incorporating user-specified confidence levels. In *Forecasting expected returns in the financial markets*, pages 17–38. Elsevier.

Jacquier, E., Polson, N. G., and Rossi, P. E. (2004). Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics*, 122(1):185–212.

Jones, R. C., Lim, T., and Zangari, P. J. (2007). The black-litterman model for structured equity portfolios. *The Journal of Portfolio Management*, 33(2):24–33.

Jorion, P. (1986). Bayes-stein estimation for portfolio analysis. *Journal of Financial and Quantitative analysis*, 21(3):279–292.

Jr, B., N, J., and Mark, N. C. (1991). Testing the capm with time-varying risks and returns. *The Journal of Finance*, 46(4):1485–1505.

Kan, R., Wang, X., and Zhou, G. (2021). Optimal portfolio choice with estimation risk: No risk-free asset case. *Management Science*.

Kan, R. and Zhou, G. (2007). Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis*, 42(30):621–656.

Kapadia, N. (2011). Tracking down distress risk. *Journal of Financial Economics*, 102(1):167–182.

Kim, D. (2014). Maximum likelihood estimation for vector autoregressions with multivariate stochastic volatility. *Economics Letters*, 123(3):282–286.

Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kirby, C. and Ostdiek, B. (2012). It's all in the timing: simple active portfolio strategies that outperform naive diversification. *Journal of Financial and Quantitative Analysis*, 47(2):437–467.

Kolm, P. N. and Ritter, G. (2020). Factor investing with black–litterman–bayes: incorporating factor views and priors in portfolio construction. *The Journal of Portfolio Management*, 47(2):113–126.

Kolm, P. N., Ritter, G., and Simonian, J. (2021). Black–litterman and beyond: The bayesian paradigm in investment management. *The Journal of Portfolio Management*, 47(5):91–113.

Large, J., Lines, J., and Bagnall, A. (2019). A probabilistic classifier ensemble weighting scheme based on cross-validated accuracy estimates. *Data mining and knowledge discovery*, 33(6):1674–1709.

Ledoit, O. and Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of empirical finance*, 10(5):603–621.

Lee, J. H., Shi, Z., and Gao, Z. (2022). On lasso for predictive regression. *Journal of Econometrics*, 229(2):322–349.

Levy, H. and Levy, M. (2014). The benefits of differential variance-based constraints in portfolio optimization. *European Journal of Operational Research*, 234(2):372–381.

Lotfi, F. H. and Fallahnejad, R. (2010). Imprecise shannon's entropy and multi attribute decision making. *Entropy*, 12(1):53–62.

Ludvigson, S. C., Ma, S., and Ng, S. (2021). Uncertainty and business cycles: exogenous impulse or endogenous response? *American Economic Journal: Macroeconomics*, 13(4):369–410.

Ludwig, H. and Baracaldo, N. (2022). Federated learning: A comprehensive overview of methods and applications.

Mandler, E. and Schümann, J. (1988). Combining the classification results of independent classifiers based on the dempster/shafer theory of evidence. In *Machine Intelligence and Pattern Recognition*, volume 7, pages 381–393. Elsevier.

Markowitz, H. (1952). The utility of wealth. *Journal of political Economy*, 60(2):151–158.

Mehta, S., Paunwala, C., and Vaidya, B. (2019). Cnn based traffic sign classification using adam optimizer. In *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*, pages 1293–1298. IEEE.

Meucci, A. (2005). *Risk and asset allocation*, volume 1. Springer.

Meucci, A. (2006). Beyond blak-litterman: Views on non-normal markets. *Quantitative Finance*, 19(2):87–92.

Meucci, A. (2009). Managing diversification. *Risk*, pages 74–79.

Moura, G. V., Santos, A. A., and Ruiz, E. (2020). Comparing high-dimensional conditional covariance matrices: Implications for portfolio selection. *Journal of Banking & Finance*, 118.

Munda, G. (1996). Cost-benefit analysis in integrated environmental assessment: some methodological issues. *Ecological economics*, 19(2):157–168.

Neely, C. J., Rapach, D. E., Tu, J., and Zhou, G. (2014). Forecasting the equity risk premium: the role of technical indicators. *Management science*, 60(7):1772–1791.

Oh, G., Kim, S., and Eom, C. (2007). Market efficiency in foreign exchange markets. *Physica A: Statistical Mechanics and its Applications*, 382(1):209–212.

Ozoguz, A. (2009). Good times or bad times? investors' uncertainty and stock returns. *The Review of Financial Studies*, 22(11):4377–4422.

Palczewski, A. and Palczewski, J. (2019). Black–litterman model for continuous distributions. *European journal of operational research*, 273(2):708–720.

Palomba, G. (2008). Multivariate garch models and the black-litterman approach for tracking error constrained portfolios: an empirical analysis. *Global Business and Economics Review*, 10(4):379–413.

Pang, T. and Karan, C. (2018). A closed-form solution of the black–litterman model with conditional value at risk. *Operations Research Letters*, 46(1):103–108.

Pele, D. T., Lazar, E., and Dufour, A. (2017). Information entropy and measures of market risk. *Entropy*, 19(5):226.

Platanakis, E., Sutcliffe, C., and Ye, X. (2021). Horses for courses: Mean-variance for asset allocation and 1/N for stock selection. *European Journal of Operational Research*, 288(1):302–317.

Rapach, D. E., Strauss, J. K., and Zhou, G. (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *The Review of Financial Studies*, 23(2):821–862.

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):129–176.

Sahamkhadam, M., Stephan, A., and Östermark, R. (2021). Copula-based black–litterman portfolio optimization. *European Journal of Operational Research*.

Segal, G., Shaliastovich, I., and Yaron, A. (2015). Good and bad uncertainty: Macroeconomic and financial market implications. *Journal of Financial Economics*, 117(2):369–397.

Shafer, G. (1976). *A mathematical theory of evidence*, volume 42. Princeton university press.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423.

Silva, T., Pinheiro, P. R., and Poggi, M. (2017). A more human-like portfolio optimization approach. *European Journal of Operational Research*, 256(1):252–260.

Sortino, F. A. and Price, L. N. (1994). Performance measurement in a downside risk framework. *the Journal of Investing*, 3(3):59–64.

Stoltzfus, J. C. (2011). Logistic regression: a brief primer. *Academic emergency medicine*, 18(10):1099–1104.

Tarekegn, A. N., Giacobini, M., and Michalak, K. (2021). A review of methods for imbalanced multi-label classification. *Pattern Recognition*, 118:107965.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.

Uhlig, H. (1994). On singular wishart and singular multivariate beta distributions. *The Annals of Statistics*, 22(1):395–405.

Uhlig, H. (1997). Bayesian vector autoregressions with stochastic volatility. *Econometrica: Journal of the Econometric Society*, 65(1):59–73.

Walters, J. (2014). The black-litterman model in detail. *Available at SSRN 1314585.*

Welch, I. and Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, 21(4):1455–1508.

Xiao, Y. and Valdez, E. A. (2015). A black–litterman asset allocation model under elliptical distributions. *Quantitative Finance*, 15(3):509–519.

Zhou, G. (1991). Small sample tests of portfolio efficiency. *Journal of Financial Economics*, 30(1):165–191.

Zhou, G. (2009). Beyond black–litterman: Letting the data speak. *The Journal of Portfolio Management*, 36(1):36–45.