Sector Allocation with Absolute, Relative, and Mixed Views Generated by Regularized Regressors

Hosang Yoon, Amrit Prasad, and Li Cao
Haas School of Business, University of California, Berkeley, California 94720
(Dated: October 10, 2018)

We explore a sector asset allocation strategy using the Black-Litterman model with forecasted views generated by various regularized regressors. We utilize regression techniques such as Elastic Net, Multi-task Elastic Net, and Bayesian Ridge to forecast absolute and relative sector returns, using sector-specific fundamental data and macroeconomic conditions as input features. These returns, along with our level of confidence in them, are combined with a forecasted covariance matrix to perform a Bayesian update to provide the inputs to the subsequent portfolio optimization. The results point towards an unusually effective generalization power of the L^1 regularization. In particular, the Multi-task Elastic Net resulted in a very parsimonious model with a relatively low turnover. It is also robust to the introduction of more realistic conditions such as transactions costs, long-only constraints, and lagged inputs.

Asset allocation is the decision faced by portfolio managers to allocate fund across different assets to achieve optimal risk-adjusted returns. Since the pioneering works of Markowitz [1], mean-variance optimization has been the main workhorse for maximizing returns while optimally trading off the associated risks of portfolios. While this approach has been highly successful, one often runs into practical issues in implementing this optimization, as the portfolio weights are very sensitive to the expected returns provided as inputs, yet they are difficult to estimate accurately. This problem is summarized by the quote of the optimization process being an 'error maximization' [2]. Furthermore, even if one recognizes such an issue, there was no straightforward way of incorporating the uncertainty or level of confidence in the estimated inputs.

Black and Litterman [3, 4] proposed a solution to this problem by allowing the expected return to be a distributional input instead of a point-estimate input. In this model, one assumes a prior distribution of expected returns around the market equilibrium and modifies it using a conditional likelihood formed by investor views of absolute performance (e.g., asset x will have an excess return of ψ%) or relative performance (e.g., excess return of asset x over y will be ν%) of specific assets. The model then forms a Bayesian posterior distribution of expected returns and updated covariance matrices that can be used for subsequent portfolio optimization. It allows investors to combine their views regarding the returns of specific assets with the market equilibrium to render an intuitive, diversified portfolio with stable weights.

Traditionally, the views that form the conditional likelihood in the Black-Litterman model have been based on analyst reports and recommendations with subjective confidence levels and other heuristics. However, this need not be this way. More recently, there has been some literature exploring the systematic generation of asset views using, e.g., ensemble learning techniques [5, 6], logistic regression, support vector machines, and naïve Bayes classifiers [6]. In this report, we follow a similar path in exploring the generation of asset views using various machine learning techniques for the sector allocation problem. We introduce novelty by focusing on the model and feature selection abilities of regularized regressors, and how they can be used to semi-automatically generate not only absolute views, but also relative and mixed views on the assets. We also explore the use of multi-task regressors that generate views on multiple assets simultaneously instead of treating the assets as independent.

I. THE BLACK-LITTERMAN MODEL

We start with a brief overview of the Black-Litterman model. Closely following [7–9], we outline how each part of the model is formed, and discuss the implications of using the model.

A. Prior Distribution

The framework starts by assuming that the excess returns of assets, $\mathbf{r} \in \mathbb{R}^n$, follow a normal distribution $\mathbf{r} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and that the expected return $\boldsymbol{\mu}$ itself also follows a normal distribution. In the absence of additional information, the mean of the prior distribution for $\boldsymbol{\mu}$ is usually assumed to be the mean excess returns implied by the market equilibrium, in the sense of the capital asset pricing model (CAPM) [10]. Assuming a risk-aversion parameter δ of the market, which can be estimated from data, the equilibrium mean excess returns implied by the market capitalization weights $\mathbf{h}_{\mathrm{mkt}}$ are given by $\boldsymbol{\pi} \triangleq E[\boldsymbol{\mu}] = \delta \boldsymbol{\Sigma} \mathbf{h}_{\mathrm{mkt}}$. It remains to specify the covariance matrix of the prior distribution of $\boldsymbol{\mu}$. The Black-Litterman model asserts that it is proportional to $\boldsymbol{\Sigma}$, given how we have found the implied expected return from $\mathbf{h}_{\mathrm{mkt}}$ using $\boldsymbol{\Sigma}$, which itself was an estimated parameter with an estimation error [11]. To summarize, the prior distribution is assumed to be of the form

$$\boldsymbol{\mu} \sim N(\boldsymbol{\pi}, \tau \boldsymbol{\Sigma}). \tag{1}$$

The scalar dimensionless parameter τ characterizes the amount of uncertainty in $\boldsymbol{\pi}$ induced by the estimation errors in $\boldsymbol{\Sigma}$. Therefore, τ is of order $\frac{1}{N_s}$ where N_s is the number of samples used in the estimation of $\boldsymbol{\Sigma}$ [8].

B. Conditional Likelihood

The conditional likelihood is formed from additional investor views that deviate from the market equilibrium. It can be based on heuristics or forecasting methods, including linear regression or other machine learning techniques. These views can be absolute or relative, and each view is associated with a confidence level which allows one to control the extent to which the view affects the portfolio.

The conditional views are expressed using a pick matrix, $\mathbf{P} \in \mathbb{R}^{k \times n}$, where k is the number of views and n is the number of assets, and a suggested excess returns vector $\mathbf{q} \in \mathbb{R}^k$. Each row $\mathbf{P}_{i,\cdot} \in \mathbb{R}^n$ defines a portfolio corresponding to a view. For an absolute view expressed on asset j, $\mathbf{P}_{i,\cdot} = \mathbf{e}_j$, where $\mathbf{e}_j \in \mathbb{R}^n$ is a unit vector with 1 in the j-th coordinate. q_i is then the excess return suggested by the investor for that asset. For a relative view expressed between asset j and asset j', we have $\mathbf{P}_{i,\cdot} = \mathbf{e}_j - \mathbf{e}_{j'}$, and q_i is the suggested excess return of asset j over asset j'. While it is possible to define a more general portfolio expressing more complicated views, we limit ourselves to these two cases to explore a semi-automated generation of parsimonious set of views.

In addition to specifying the suggested excess returns of the view portfolios, we must also specify their conditional covariance matrix, $\boldsymbol{\Omega} \in \mathbb{R}^{k \times k}$. Its inverse $\boldsymbol{\Omega}^{-1}$ then expresses the confidence that the investor has in each view. Mathematically, we may summarize the views expressed as a conditional likelihood as

$$\mathbf{q}|\boldsymbol{\mu} \sim N\left(\mathbf{P}\boldsymbol{\mu}, \boldsymbol{\Omega}\right), \tag{2}$$

where the suggested excess returns, $\mathbf{q}$, of the view portfolios are assumed to be drawn from this distribution. $\boldsymbol{\Omega}$ is often assumed to be diagonal, since estimating the correlations of views accurately is difficult and is likely to introduce more errors.

C. Posterior Distribution

Given the prior distribution in Eq. (1) and the conditional distribution in Eq. (2), we obtain the posterior distribution of the expected excess return through a Bayesian update conditional on the views,

$$\begin{aligned}\boldsymbol{\mu}|\mathbf{q} \sim\; & \\ & N\Big(\left(\mathbf{P}'\boldsymbol{\Omega}^{-1}\mathbf{P} + (\tau\boldsymbol{\Sigma})^{-1}\right)^{-1} \left(\mathbf{P}'\boldsymbol{\Omega}^{-1}\mathbf{q} + (\tau\boldsymbol{\Sigma})^{-1}\boldsymbol{\pi}\right), \\ & \quad \left(\mathbf{P}'\boldsymbol{\Omega}^{-1}\mathbf{P} + (\tau\boldsymbol{\Sigma})^{-1}\right)^{-1} \Big) \triangleq N\left(\hat{\boldsymbol{\pi}}, \hat{\boldsymbol{\Sigma}}\right).\end{aligned} \tag{3}$$

Proof for Eq. (3) is given in Appendix A. Given this updated distribution of the expected mean, we also have an updated distribution of the returns. Assuming that the variations in $\mathbf{r}$ and $\boldsymbol{\mu}|\mathbf{q}$ are uncorrelated, we have

$$\mathbf{r}|\mathbf{q} \sim N\left(\hat{\boldsymbol{\pi}}, \boldsymbol{\Sigma} + \hat{\boldsymbol{\Sigma}}\right). \tag{4}$$

These moments are then used for subsequent portfolio optimization.

The advantages of using Black-Litterman model are evident from this exposition. First, the model allows one to specify the uncertainties in the estimated inputs to the portfolio optimization in a theoretically sound manner. Second, the model shrinks the posterior expected returns towards the market equilibrium when there is not much confidence in a specific view. This makes the optimal weights relatively stable over re-balances and mitigates the 'error-maximization' problem. Finally, the model provides a method for incorporating forecast models on *pairs* of assets into the traditional mean-variance optimization.

II. DATA AND IMPLEMENTATION

A. Investment Universe

We explore sector asset allocation within the US S&P 500 universe from April 1, 1998 to December 29, 2017. To do this, we aggregated the daily market capitalization of companies according to their Global Industry Classification Standard (GICS) sectors, and split the S&P 500 index proportional to the total market capitalization of each sector on each date. Prices, market capitalization, and index constituents data were obtained from Compustat. This decomposition of the S&P 500 index is plotted in Fig. 1.

For asset allocation, we treat each of these sub-indices as investable assets and perform portfolio optimization across them. We have formed this artificial sector index so that $\mathbf{h}_{\mathrm{mkt}}$ can be directly inferred from the 'prices' for each sector, which is an input to the prior distribution of the Black-Litterman model. However, this is not a limiting assumption, as we may equally as well use exchange traded fund (ETF) prices that track these sectors, with $\mathbf{h}_{\mathrm{mkt}}$ given as a separate input.

B. Covariance Estimation

There are a total of 11 sectors in GICS. Because this number is small, we estimate the covariance matrix directly using exponentially weighted averaging,

$$\mathbf{S}_t = \lambda\mathbf{S}_{t-1} + (1-\lambda)(\mathbf{r}_t - \bar{\mathbf{r}})(\mathbf{r}_t - \bar{\mathbf{r}})'. \tag{5}$$

Following [12], the decay parameter was set to $\lambda = 0.94$ with Eq. (5) computed at a daily frequency. In order

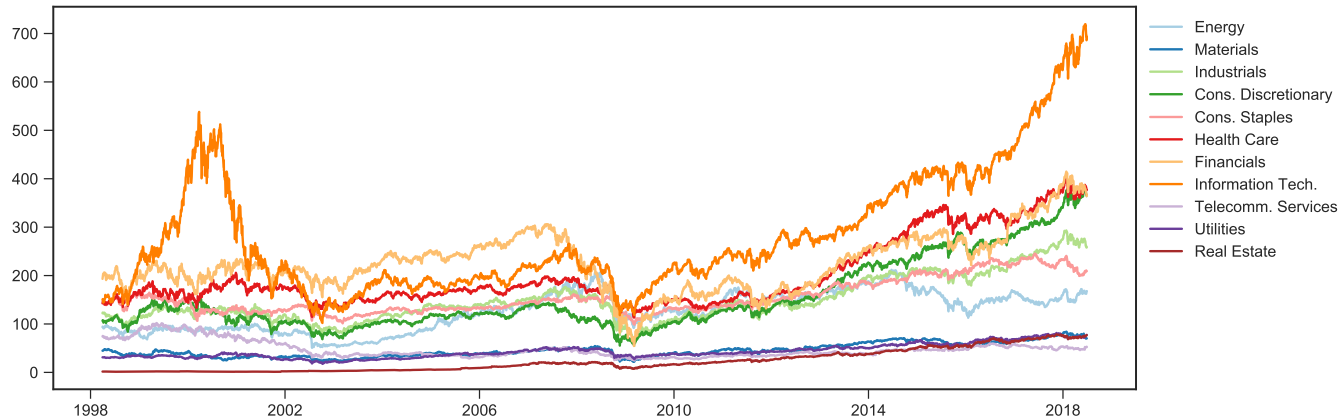


FIG. 1. S&P 500 index decomposed into GICS sectors based on the market capitalization of its constituents on each date.

to have a stable precision matrix $\mathbf{\Sigma}^{-1}$, we also employ Ledoit and Wolf shrinkage [13], given by

$$\mathbf{\Sigma}_t = \xi \frac{\mathrm{Tr}(\mathbf{S}_t)}{n}\mathbf{I}_n + (1-\xi)\mathbf{S}_t, \tag{6}$$

where ξ is chosen according to [13].

C. Predictive Features

To generate the views on excess returns of the sectors, we used the following two groups of predictive features.

- Sector-specific fundamental features: P/E ratio, P/B ratio, dividend yield, free cash flow yield, cash flow per share, cash flow from investment, cash flow from financing, sales, EBITDA, EV/EBITDA, current ratio, operating margin, profit margin, EBITDA margin, debt/asset.

- Macroeconomic features: Conference Board leading and coincidental/lagging index, CPI, core CPI, unemployment rate, initial jobless claims, industrial production, Fed funds rate, M2 money supply, Chicago Fed National Activity Index (CFNAI), housing unit starts, US dollar index, futures on US dollar index, futures on Crude oil, US trade-weighted dollar index, Treasury yields at various maturities.

Details of calculations performed on these features are provided in Table I. The total number of features derived from these data was $F = 32$. These data, obtained from Bloomberg, are available at monthly frequency and therefore we consider portfolio rebalancing at monthly frequency as well.

Based on the availability of the sector-specific features, we have excluded the Financials and Real Estate for forecasting models as large number of features were missing for the Financials sector and data was available only 2016 onward for the Real Estate sector. Nevertheless, we consider the returns of all $n = 11$ sector indices for the rest of the discussion.

D. Baseline Forecast Model

As the baseline forecast model that can be compared against the other models to be discussed in this report, we first explore an ordinary least squares (OLS) regression, performed independently for each sector. The OLS regression solves

$$\min_{\mathbf{w}} ||\mathbf{y} - \mathbf{X}\mathbf{w}||_2^2, \tag{7}$$

where $\mathbf{y} \in \mathbb{R}^T$ contains the cumulative excess returns until the next rebalancing date for each sector, $\mathbf{X} \in \mathbb{R}^{T\times(F+1)}$ contains the features at each rebalancing date and a column of ones, and $\mathbf{w} \in \mathbb{R}^{F+1}$ contains the weights for the features as well as the bias term.

For the OLS regression and for the rest of the models discussed in the paper, dates up to January 31, 2013 was used as the training/validation set for training and cross validation of hyperparameters, and dates from January 31, 2013 onward was used as the test set for the final performance evaluation.

Appendix B shows the results of this regression. We first note that while the R^2 values are high, this is likely to be a result of overfitting due to adding too many features that do not have predictive power. This is indicated by the very small number of parameters that were statistically significant. To implement this in a statistically sound manner, one needs to add the new features parsimoniously while testing for the statistical significance offered by a new feature. Nevertheless, we still keep this model as the baseline which we improve upon by employing the Black-Litterman model with various regressors.

E. Portfolio Implementation

At each rebalancing date, we have a set of forecasted excess returns $\mathbf{r}_t$ generated by a forecast model, and a covariance matrix $\mathbf{\Sigma}_t$. If the Black-Litterman model is not used, we directly use these moments for the subsequent optimization. If the Black-Litterman model is used, we use the moments of its posterior distribution. Regardless of whether the Black-Litterman model is used or not, we will denote the first and second moments to be used for the optimization as $\mathbf{R}$ and $\mathbf{V}$.

We seek a dollar-neutral long-short portfolio $\mathbf{h}$ that solves the optimization [14]

$$\begin{aligned} &\max_{\mathbf{h}} \mathbf{R}'\mathbf{h} - \frac{\gamma}{2}\mathbf{h}'\mathbf{V}\mathbf{h}, \\ &\text{s.t. } \mathbf{1}'\mathbf{h} = 0, \\ &\qquad ||\mathbf{h}||_1 \leq L_{\max}, \end{aligned} \tag{8}$$

where γ is our risk-aversion parameter, $\mathbf{1}$ is a vector of ones, and $L_{\max}$ is the maximum leverage of the portfolio, placed as a safety measure against extreme positions.

The optimization is performed using cvxpy [15]. After this step, to enable a comparison between our dollar-neutral portfolio and the market portfolio, we add the market capitalization weights of the market to the portfolio and use $\mathbf{h}_P = \mathbf{h}_L + \mathbf{h}_{\text{mkt}}$ as our fully invested portfolio. For the subsequent discussion, we assume $\gamma = 2$ and $L_{\max} = 10$.

III. VIEW GENERATION MODELS

We now explore the generation of conditional views using various approaches and machine learning algorithms. The approaches are categorized into three groups. In the absolute view group, we consider models that forecast the excess returns for each sector. In the relative view group, we consider models that forecast the differences of excess returns between two sectors. In the final group, we explore different ways of merging these two approaches.

A. Absolute View Models

1. *Elastic Net*

Elastic Net [16] performs linear regression with both L^1 and L^2 regularization. It seeks the solution of the optimization

$$\min_{\mathbf{w}} \frac{1}{2T}||\mathbf{y} - \mathbf{X}\mathbf{w}||_2^2 + \eta\left(\nu||\mathbf{w}||_1 + (1-\nu)\frac{1}{2}||\mathbf{w}||_2^2\right), \tag{9}$$

with hyperparameters η and ν chosen by cross-validation. Choosing $\nu = 0$ is equivalent to a Ridge regression and $\nu = 1$ is equivalent to a LASSO regression. By altering $\nu \in [0, 1]$ during cross validation, Elastic Net subsumes both Ridge and LASSO regressions. We perform 3-fold cross validation using scikit-learn [17].

We have cross-validated Eq. (9) for each sector excluding Financials and Real Estate, and report the results in Table I. We first note that, as a result of cross-validation, many sectors were considered not significant enough to be considered in a predictive model that performs better than a constant model. This is expressed by the results of the Elastic Net cross-validation resulting in all-zero weights. We also note that, even within the sectors that had non-zero weights, the features selected are fairly sparse, with cash flow related variables and CFNAI occurring as features in many of the sectors. The prevalence of cash flow related variables as predictive features is in line with the literature [18]. While it will be possible to perform a more careful statistical analysis to select the best set of features for a better predictive power, we will use this model as a building block for the relative and mixed view models as it provides very strong model and feature selection ability.

Once the i-th model for a sector is deemed significant after cross validation, we obtain an estimate of the inverse of the confidence in the model as $\Omega_{ii} = \frac{1}{cT}||\mathbf{y} - \mathbf{X}\mathbf{w}||_2^2$, i.e., the variance of the residuals scaled by c. The hyperparameter c, which is shared across all forecast models, represents the overall confidence we place on the forecast models as a dimensionless quantity, and is tuned at a later stage along with τ of the Black-Litterman model, as will be discussed in Sec. IV A.

2. *Multi-task Elastic Net*

As a possibly crude approach to predicting multi-sector returns instead of treating each sector as independent, we may stack the feature matrices [19] and target vectors for each sector together and perform a joint regression. This approach with a grouped L^1 regularization and L^2 regularization is known as Multi-task Elastic Net [20]. It seeks the solution of the optimization

$$\min_{\mathbf{W}} \frac{1}{2T}||\mathbf{Y} - \tilde{\mathbf{X}}\mathbf{W}||_F^2 + \eta\left(\nu||\mathbf{W}||_{2,1} + (1-\nu)\frac{1}{2}||\mathbf{W}||_F^2\right), \tag{10}$$

where $\mathbf{Y} \in \mathbb{R}^{T\times n}$, $\tilde{\mathbf{X}} \in \mathbb{R}^{T\times(\tilde{F}+1)}$, and $\mathbf{W} \in \mathbb{R}^{(\tilde{F}+1)\times n}$ are the stacked target vectors, feature matrices, and weight vectors, respectively, $||\mathbf{W}||_F^2 = \text{Tr}(\mathbf{W}\mathbf{W}')$, and $||\mathbf{W}||_{2,1} = \sum_i ||W_{i,\cdot}||_2$. In essence, the regularization groups weights for the different sectors together for a given feature and performs a sparse selection of features that has predictive power across all sectors. Once the model is trained, Ω_{ii} is obtained in a similar manner as in Sec. III A 1, individually for each sector.

The cross-validated set of weights are summarized in Fig. 2. Quite interestingly, the model has selected only one feature as having a significant predictive power across all sectors, and it was not a macroeconomic one but rather the change of free cash flow into the Utilities sec-

TABLE I. Coefficients of cross-validated Elastic Net regression. Empty coefficient indicates 0. Sectors with blank η, ν, and R^2 indicate that the optimal model chosen by cross validation for these sectors was such that the weights were all driven to 0 by the L^1 penalty. Δ indicates month-over-month percentage change (clipped at $\pm50\%$) for level variables or differences in rates for rate variables. DXY: US Dollar index, DX1: 1-month futures on DXY, WTI: WTI futures, USTW$: US trade-weighted dollar index.

	Energy	Mater.	Indust.	Cons. Discre.	Cons. Staples	Health	Info. Tech.	Tele. Serv.	Util.
η	0.004	0.001				0.001		0.001	0.005
ν	0.100	1.000				1.000		0.100	0.100
Intercept	0.004	0.000				0.001		-0.001	0.001
ΔP/E								-0.072	
ΔP/B								-0.173	
ΔDiv. yield								-0.216	
ΔFCF/P	0.032							0.062	-0.009
ΔCF									
ΔCFI						0.013		-0.044	
ΔCFF	0.031	0.020						-0.014	0.015
ΔSales									
ΔEBITDA								0.033	
ΔEV/EBITDA									
ΔCur. ratio									
ΔOper. margin								0.000	
ΔProf. margin								0.015	
ΔEBITDA margin									
ΔDebt/Asset								-0.023	
ΔLeading index									
ΔCoin./lag. index									
ΔCPI									
ΔCPI core									
ΔUnemployment									
ΔInit. jobless cl.	-0.010							-0.160	
ΔIndust. Prod.									
ΔFed funds r.									
ΔM2 money sp.									
CFNAI/10	0.050					0.014		0.023	0.016
ΔPriv. housing								0.040	
ΔDXY								-0.096	
DX1/DXY-1									
ΔWTI								-0.078	
ΔUSTW$								-0.031	
Δ6m yield								-0.001	
2-10y spread									
R^{2}[a]	0.056	0.014				0.015		0.161	0.029

[a] Note that R^2 is not a measure of forecasting power, and is used here only as an indicator of whether the weights of the model have been all driven to 0 or not.

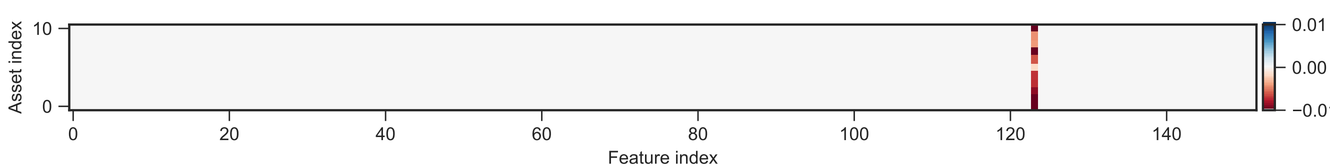


FIG. 2. Coefficients $\mathbf{W}'$ of cross-validated Multi-task Elastic Net regression represented as a color map. The only row of $\mathbf{W}$ with non-zero weights corresponds to ΔFCF/P for the Utilities sector. Optimal values of the hyperparameters were found to be $\eta = 0.012$ and $\nu = 0.5$.

tor. A potential economic interpretation is that this is because the companies in the Utilities sector determine the input cost to the other sectors. If the free cash flow into the Utilities sector can be a good proxy for this cost, this variable has predictive power in forecasting returns for the other sectors. The weights in $\mathbf{W}_{\Delta FCF/P,\cdot}$ are then the degree to which the sectors are affected by this effect.

While this approach did not start out in a parsimonious way, it still attains a strong sparsity due to the regularization of the Elastic Net. We will see that this particular model achieves meaningful results in Sec. IV B.

3. *Bayesian Ridge*

Bayesian Ridge regression [21] assumes a probabilistic model for the residuals and the weights. Specifically,

$$\begin{aligned} \mathbf{y}|\mathbf{X},\mathbf{w},\omega &\sim N\left(\mathbf{X}\mathbf{w},\omega^{-1}\mathbf{I}_T\right), \\ \mathbf{w}|\zeta &\sim N\left(\mathbf{0},\zeta^{-1}\mathbf{I}_{F+1}\right), \end{aligned} \tag{11}$$

where the prior distributions for the precision parameters ω and ζ are assumed to be from the gamma distribution, the conjugate prior for the precision of a normal distribution. The model starts from uninformative priors and updates the conditional distributions of ω and ζ using the data.

In Bayesian Ridge regression, the appropriate regularization is directly learned from the data through ζ rather than having to rely on cross validation. Another advantage of the method is that it generates an estimation of variance for each forecasted sample. We use this estimated variance divided by c as Ω_{ii} for this model.

4. *Other methods*

In principle, many other types of regressors from the machine learning literature may be used for forecasting the excess returns for the sectors. However, we found that the more complex algorithms such as Support Vector Regression or Gradient Boosted Regression Trees were unable to generalize on the out of sample data as effectively as the linear models. We suggest two reasons. The first is that the number of data points are limited as our features are only available at a monthly frequency. Secondly, the portfolio construction process is inherently low in signal-to-noise ratio. We thus focused our research on linear models with strong regularization. The remaining models will use the regressors discussed above, and consider combining these in creating relative and mixed view models.

B. Relative View Models

In these models, we forecast the differences in returns between pairs of sectors. Given the 9 sectors for which we have the features data, there are 36 unique pairs of sectors that may allow forecast opportunities. To select a sparse set of pairs of sectors, we use the Elastic Net cross validation which allowed strong feature and sector selection in Sec. III A 1. The sector-specific features for the pair of sectors and macroeconomic features are stacked together as the features for this regression.

The results of this procedure is shown in Table II. Similarly to the absolute view case, we observe that roughly half of the pairs were deemed insignificant by the procedure. We note that the pairs of sectors that had the most in-sample explanatory power have some economic interpretation. For instance, economic consumption theory suggests that consumption in discretionary goods and durable/staple goods have opposite patterns in their response to economic conditions. For another instance, Materials sector provides inputs to the Industrial sector. It therefore will make sense to try to predict the differentials in returns between these pairs given that sufficient information is present in the features for the two sectors.

TABLE II. Pairs of sectors selected by Elastic Net cross validation.

Long	Short	R^2 [a]
Industrials	Consumer Discretionary	0.194
Consumer Discretionary	Consumer Staples	0.183
Consumer Staples	Telecomm. Services	0.127
Consumer Discretionary	Telecomm. Services	0.120
Materials	Telecomm. Services	0.096
Materials	Industrials	0.084
Health Care	Telecomm. Services	0.048
Telecomm. Services	Utilities	0.047
Information Technology	Telecomm. Services	0.027
Materials	Consumer Staples	0.026
Industrials	Information Technology	0.023
Energy	Health Care	0.019
Energy	Consumer Staples	0.016
Energy	Information Technology	0.011
Materials	Information Technology	0.010
Materials	Health Care	0.003
Energy	Consumer Discretionary	0.003

[a] Note that R^2 is not a measure of forecasting power, and is used here only as an indicator of whether the weights of the model have been all driven to 0 or not.

C. Mixed View and Hybrid Models

In these models, we consider using the absolute views and relative views together. There are two approaches one may take in this regard, although both uses the same trained weights from Sec. III A and III B. In the first approach, the views from the absolute and relative excess return forecast models are simply stacked together as

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_{\text{abs}} \\ \mathbf{P}_{\text{rel}} \end{bmatrix}, \quad \mathbf{q} = \begin{bmatrix} \mathbf{q}_{\text{abs}} \\ \mathbf{q}_{\text{rel}} \end{bmatrix}, \quad \mathbf{\Omega} = \begin{bmatrix} \mathbf{\Omega}_{\text{abs}} & \\ & \mathbf{\Omega}_{\text{rel}} \end{bmatrix}. \tag{12}$$

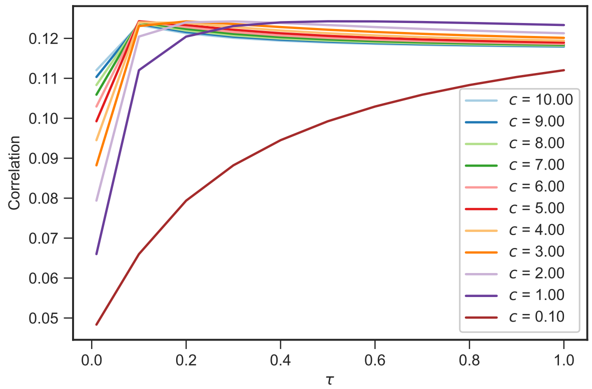


FIG. 3. Tuning of τ and c for the hybrid model. y-axis is the correlation of forecasted and realized excess returns.

We denote this approach the mixed view model.

In the second approach, we take the returns forecasted by the absolute view models as the mean of the prior distribution for the Black-Litterman model. More specifically, we replace μ_j with the return forecasted for sector j, and let τ model the overall error in the prior distribution. We denote this approach the hybrid model.

IV. RESULTS AND DISCUSSIONS

We now assess the performance of these models. We first perform out-of-sample backtests with portfolio weights obtained by the optimization in Eq. (8), using either forecasted excess returns and covariance matrices directly or after applying the Black-Litterman framework in absolute, relative, or mixed view settings. After that, we select a few promising models and consider the impact of applying more realistic implementation-related assumptions on the model performances.

A. Hyperparameters of the Black-Litterman Model

The hyperparameters of the Black-Litterman model are τ that represents the amount of error in the prior distribution as a dimensionless quantity, and c that represents the overall amount of confidence that we place on the forecast models as a dimensionless quantity. We tune these parameters for each set of models. An example is shown in Fig. 3 for the hybrid model. The fitness is measured by the correlation of forecasted and realized excess returns in the training/validation set [22]. We choose the values of τ and c such that τ is small enough to be of the order of magnitude expected for $\frac{1}{T_s}$, and c is large enough to get a high correlation, but not unnecessarily high. For example, in Fig. 3, we choose $\tau = 0.2$ and $c = 3$. Other models' τ and c were selected in a similar way.

B. Out-of-Sample Backtests

We performed out-of-sample backtests from January 31, 2013 to December 29, 2017 with monthly rebalancing when new sets of data become available. The results are summarized in Fig. 4.

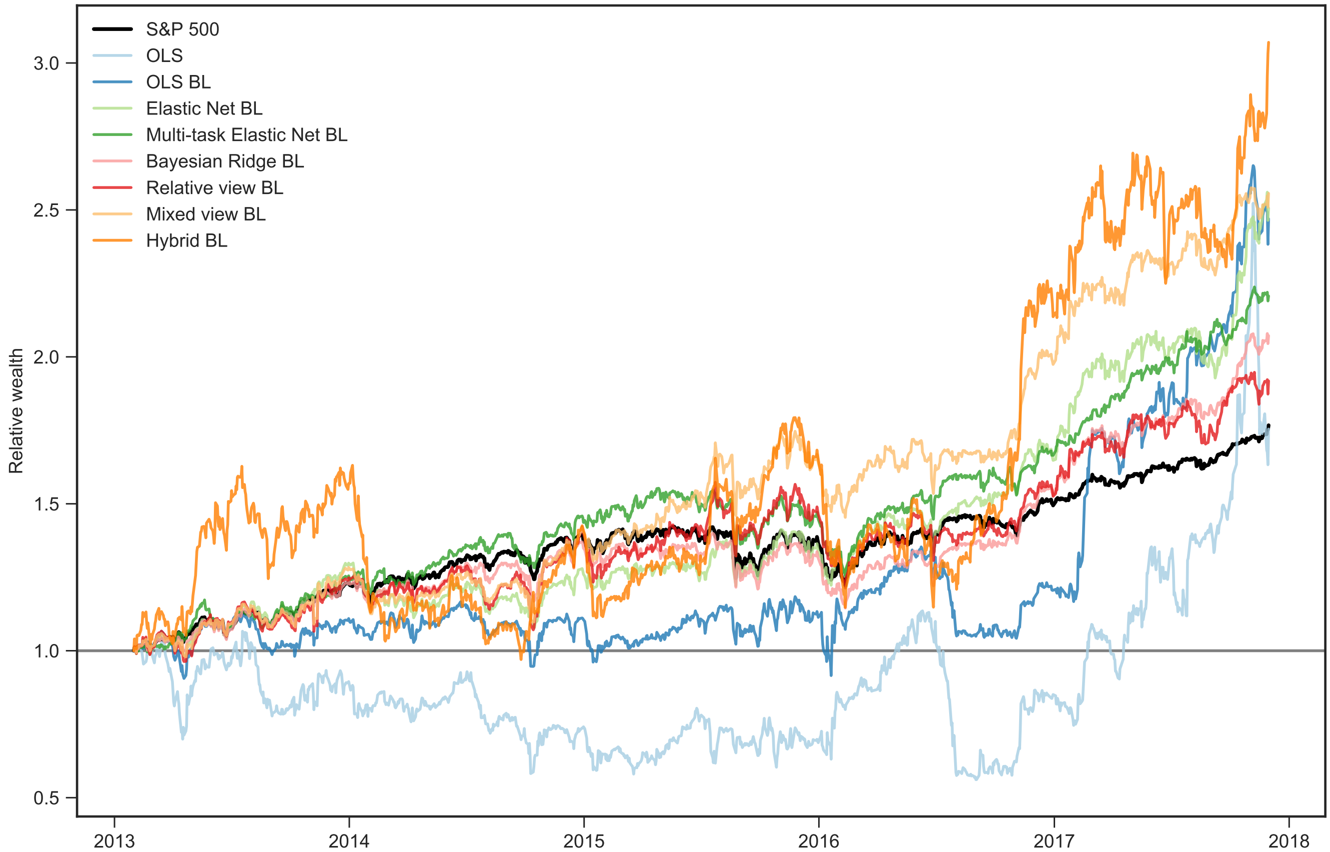
We first examine the baseline OLS model without any modification. We found in Sec. II D that this baseline OLS model included too many features without predictive power and overfitted the data significantly. As expected, this is reflected in the poor out-of-sample performance seen in Fig. 4. However, even this crude model saw significant improvement as the Black-Litterman framework was introduced. This is because we specify our confidence in the model and shrink the weights toward the market equilibrium weights if the level of confidence is low. Although the model still has a high turnover and invests too wildly, this demonstrates the stabilizing effect of the Black-Litterman model.

We next examine the absolute view models that forecast the excess returns of the sectors with respect to cash. We see that the three models in this category, Elastic Net, Multi-task Elastic net, and Bayesian Ridge, show promising performances. In particular, we observe that the Elastic Net-based models that utilize L^1 regularization in addition to L^2 regularization showed higher out-of-sample information ratio compared to the Bayesian Ridge model that only utilizes L^2 regularization.

L^1 regularization is a powerful technique for feature selection that can sometimes generalize better than L^2 regularization. This is because the number of training samples needed to 'learn well' in the presence of a large number of irrelevant features is logarithmic in the number of irrelevant features if L^1 regularization is used as opposed to a linear relation if L^2 regularization is used [23]. This means that when there are a large number of irrelevant features, the L^1 regularization tends to suffer less from them.

An added benefit of L^1 regularization is that it leads to models with better interpretability, as many of the weights are driven to 0. This can be understood in the context of performing gradient descent to find the optimal weights [24] . With L^1 regularization, the gradient of the penalty, $\nabla_{\mathbf{w}}||\mathbf{w}||_1$, drives the weights toward $\mathbf{0}$ with an equal magnitude regardless of where $\mathbf{w}$ is in the parameter space, except when it is exactly $\mathbf{0}$. On the other hand, with L^2 regularization, $\nabla_{\mathbf{w}}||\mathbf{w}||_2^2$ gradually becomes smaller as $\mathbf{w}$ is driven closer to $\mathbf{0}$. As a result, with L^1 regularization, only the features that are able to generate a gradient from the loss function that is at least as large as the gradient from the penalty will survive and have non-zero weights. In other words, L^1 regularization selects the dimensions that have enough curvature near the optimum and ignores ones that are flat, or equivalently, uninformative.

On the other hand, one benefit of the L^2 regularization is that it tends to provide more stable weights. This is likely the reason why the Bayesian Ridge model had a



	Annual return	Annual volatility	Annual tracking error	Annual turnover	Information ratio	Sharpe ratio	Maximum drawdown	Daily VaR (5%)
OLS	12.4%	43.7%	41.1%	7705%	0.00	0.49	-50.6%	4.1%
OLS BL	20.7%	23.8%	18.2%	2927%	0.38	0.91	-23.0%	2.2%
Elastic Net BL	21.1%	17.2%	10.1%	1178%	0.72	1.20	-16.9%	1.8%
Multi-task Elastic Net BL	17.9%	14.0%	6.6%	514%	0.70	1.25	-19.2%	1.5%
Bayesian Ridge BL	16.4%	14.7%	5.8%	849%	0.57	1.10	-18.2%	1.6%
Relative view BL	14.5%	19.4%	12.1%	2216%	0.14	0.80	-22.9%	2.0%
Mixed view BL	21.6%	18.1%	11.2%	2033%	0.69	1.17	-16.8%	1.6%
Hybrid BL	26.3%	33.3%	28.5%	3777%	0.40	0.87	-40.5%	3.4%

FIG. 4. Out-of-sample backtest results of the various models considered in this report. BL indicates the use of the Black-Litterman model. Information ratio was calculated as $E[r_p - r_b]/Std[r_p - r_b]$ and Sharpe ratio was calculated as $E[r_p]/Std[r_p]$ where r_p and r_b are the excess returns of the portfolio and the benchmark, respectively.

lower turnover as opposed to the Elastic Net model. This point will be revisited in Sec. IV E.

The Multi-task Elastic Net had a much lower turnover compared to the other models considered in this report. The likely reason for this is that because it regularized over the whole feature to multi-sector target mapping space, as opposed to considering each sector separately. As a result, it selected one feature that makes good forecasts for all the sectors. As only one feature was involved, there was less room for temporal variation in the weights, leading to a lower turnover. This demonstrates one direct benefit of parsimony in the investment optimization, as it appears to reduce the turnover of the models.

We now examine the relative, mixed, and hybrid model variants. The relative view model only uses the forecasted return differences for pairs identified in Table II in the views for the Black-Litterman model. The posterior moments would then suggest overweighting and underweighting the long and short sides of the pairs given a positive forecasted relative return. It appears that using the relative views only does add value but is not very effective, and also induces a high turnover. In fact, it performs worse than the crude unregularized OLS with the Black-Litterman model. Likely as a result, the mixed

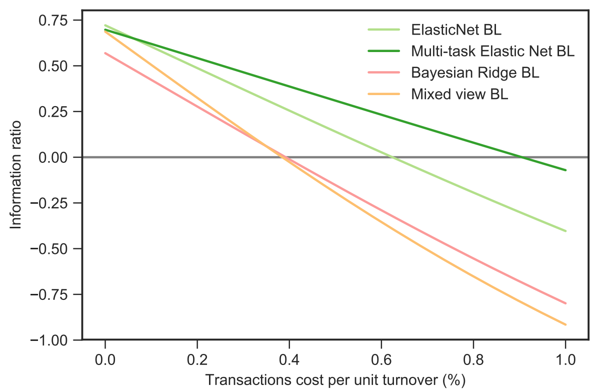


FIG. 5. Out-of-sample information ratios as a function of transactions cost during backtests.

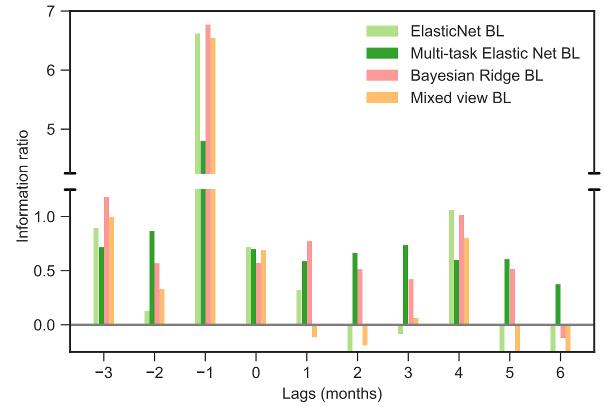


FIG. 6. Out-of-sample information ratios with different amounts of lag applied on the features data.

and hybrid models that added the relative views to the Elastic Net model's views did not add value in that the information ratio decreased while the turnover increased. We also experimented on using a smaller number of pairs from Table II, but it did not improve the results. Our conclusion is that generating relative views using regressors is not very helpful in improving the performance of the Black-Litterman model, and that it suffices to generate the absolute views well, parsimoniously.

C. Transactions Costs

We explore the impact of linear transactions costs on the performances of the models with information ratios exceeding 0.5 in Fig. 4. We assume that $c\mathbf{1}'|\Delta\mathbf{h}_P|$ amount of proportional transactions cost is dissipated at every rebalance, and adjust the backtest results accordingly. The results are summarized in Fig. 5.

We note that, as expected from the models' respective annual turnovers, the Multi-task Elastic Net is affected the least by the introduction of transactions costs. In fact, there is a close relationship between this plot and the annual turnover, given by

$$TO = \frac{\Delta IR}{\Delta c}(\text{tracking error}). \tag{13}$$

We also note that due to the lower turnover of the Bayesian Ridge model as compared to the Mixed view model, the performances of the two models cross over at $c \approx 0.35$. Overall, except for the Multi-task Elastic Net, these models exhibit high annual turnovers and demand low transactions costs to be profitable.

D. Long-only Constraint

We explore the effects of placing a long-only constraint $h_{P,i} \geq -h_{\mathrm{mkt},i}$ for each asset i, as is sometimes obligated for portfolio managers. The results are summarized in Table III.

We first note that there is a dramatic reduction in annual turnover for all models. This means that the original models were making significant negative bets that were varying swiftly across time, but now cannot be invested as such. However, most models saw a significant reduction in their information ratios as well, indicating that those negative bets were made for a reason.

We also note that while the information ratio of most models were affected considerably by the introduction of the long-only constraint, the Multi-task Elastic Net was almost unaffected although its annual return and tracking both slightly decreased. This may be due to how the weights in Fig. 2 were all in the same negative direction with similar weights on many sectors. This means that if the long-only constraint binds, it tends to distort the signal in a less severe manner than if the weights were a mix of positive and negative weights of different magnitudes.

E. Information Horizon

We performed information horizon analysis [14] by applying different amounts of lags or leads on the features data, re-training the models, and performing out-of-sample backtests. For fairness in comparison, τ and c were kept the same as in Sec. IV B. This analysis provides an indication of how fast the information content of a signal decays and how stable the training process is upon re-training with slightly different data. The results of this analysis are summarized in Fig. 6.

We first note that the information ratio has a very high peak if the input features are shifted 1 month from the future to the present. This is to be expected, as some of our input features almost directly measure trailing price changes, which are then very close to the target variable after the shift.

We then note that while the information ratios for the Multi-task Elastic Net and the Bayesian Ridge were relatively stable over different lag values, those for the Elastic Net and Mixed view model exhibited high fluctuations.

TABLE III. Out-of-sample backtest results with long-only constraints.

	Annual return	Annual volatility	Annual tracking error	Annual turnover	Information ratio	Sharpe ratio	Maximum drawdown	Daily VaR (5%)
OLS	15.0%	16.0%	9.5%	948%	0.22	0.95	-17.6%	1.5%
OLS BL	9.3%	15.1%	7.6%	802%	-0.39	0.66	-17.5%	1.6%
Elastic Net BL	14.7%	13.4%	4.3%	396%	0.43	1.09	-11.8%	1.4%
Multi-task Elastic Net BL	16.1%	12.3%	4.3%	204%	0.71	1.27	-14.0%	1.3%
Bayesian Ridge BL	13.1%	13.4%	3.4%	347%	0.12	0.98	-15.4%	1.4%
Relative view BL	10.6%	13.8%	4.5%	649%	-0.39	0.80	-19.4%	1.5%
Mixed view BL	13.7%	13.3%	4.1%	611%	0.23	1.03	-15.8%	1.4%
Hybrid BL	15.1%	15.4%	7.3%	389%	0.30	0.99	-20.8%	1.6%

This is likely due to different sectors and features being selected for the new data in the cross validation procedure outlined in Sec. III A and III B, which we perform in an automated fashion. This tells us that while automated feature and model selection is highly convenient, they do not guarantee temporal stability upon retraining. One way to mitigate this problem will be to use cross-validated Ridge regressors on the set of sectors and features selected from Sec. III A and III B, in place of using the cross-validated Elastic Net on the whole feature and sector set.

On the other hand, the information ratios for the Multi-task Elastic Net and the Bayesian Ridge were quite stable over different lags except for a strange peak at the 4 month lag, which also occurred for the other two models as well. We are unsure of the exact reason for this behavior, but it could to be due to delays in the time stamps of certain variables.

F. Alternative Risk Measures

We have so far considered mean-variance optimization as the primary tool for achieving optimally risk-adjusted returns, where risk is measured as the forecasted variance of the portfolio. In this section, we explore some modifications to this traditional measure of risk and how they affect the results of this report.

One alternative to the mean-variance optimization is mean-standard deviation optimization. An interesting property of this optimization is that it can be interpreted as a robust optimization of the form [25]

$$\begin{aligned} &\max_{\mathbf{h}} \ \mathbf{R}'\mathbf{h} - \epsilon ||\mathbf{V}^{1/2}\mathbf{h}||_2 \iff \\ &\max_{\mathbf{h}} \left[\min_{\mathbf{u}} \left(\mathbf{R} + \mathbf{V}^{1/2}\mathbf{u}\right)' \mathbf{h} \text{ s.t. } ||\mathbf{u}||_2 \le \epsilon \right], \end{aligned} \tag{14}$$

i.e., it is equivalent to maximizing the worst-case returns under some bounded uncertainty of the expected returns, where ϵ and $\mathbf{V}$ control the size and direction of this bound.

We find that the mean-standard deviation optimization is not very helpful to our problem, in particular together with the dollar-neutral constraint. If ϵ is large enough, the dollar-neutral constraint forces $\mathbf{h} \to \mathbf{0}$, as the objective function is driven toward negative values. This means that $\mathbf{h}_P \to \mathbf{h}_{\mathrm{mkt}}$. This case is shown in Appendix C, where $\epsilon = 0.3$ forced $\mathbf{h}_P \to \mathbf{h}_{\mathrm{mkt}}$ for many of the models for most rebalancing periods and caused the cumulative returns to track the S&P 500 index in parallel. On the other hand, $\mathbf{h}$ becomes non-zero if ϵ is slightly smaller, but they vary wildly over time with performance not particularly better than the mean-variance optimization. The results are summarized in Appendix C for the case $\epsilon = 0.1$, just a bit lower than $\epsilon = 0.3$ considered above. Lower ϵ leads to even worse results. Overall, the mean-standard deviation optimization does not seem compatible with our problem.

We next consider approximate risk parity portfolios. A risk parity portfolio is a portfolio where the marginal contributions to risk for each asset are all equal [26]. Because such a portfolio is unique for a given $\mathbf{V}$ irrespective of $\mathbf{R}$, we instead consider a relaxation of the problem that trades off the risk contribution variance with our original objective in Eq. (8). It is given by the optimization [27]

$$\max_{\mathbf{h},\zeta} \mathbf{R}'\mathbf{h} - \frac{\gamma}{2}\mathbf{h}'\mathbf{V}\mathbf{h} - \frac{\rho}{n}\sum_{i=1}^{n} \left(h_i \left(\mathbf{V}\mathbf{h}\right)_i - \zeta\right)^2 . \tag{15}$$

Note that the optimal value of ζ is $\zeta^* = \frac{1}{n}\sum_{i=1}^{n} h_i \left(\mathbf{V}\mathbf{h}\right)_i$ for any $\mathbf{h}$ in the feasible set, and therefore the third term is proportional to the variance of the risk contributions of the assets. In the limit $\rho \to \infty$, we obtain full risk parity, while with a finite ρ, we obtain an approximately mean-variance optimal portfolio with a tilt towards the risk-parity portfolio.

We show the results of this optimization in in Table VII with $\rho = 10,000$, which leads to a moderate tilt towards the risk-parity portfolio. Note that ρ needs to be quite large in order to have a meaningful effect on the optimization, as the scale of the variance of risk contribution is much smaller than the first two terms in Eq. (15). We observe that while the introduction of the risk-parity term somewhat reduced the turnover of the models, it traded off the information ratio of the portfolio in a less-than-optimal way. Whether such a result is desirable will depend on the objective of the user of the models.

V. CONCLUSION

We have explored the sector allocation problem within the S&P 500 universe using the Black-Litterman model with absolute, relative, and mixed views semi-automatically generated by various regression techniques. Absolute view models forecast returns for specific sectors, relative view models forecast returns for pairs of sectors, and mixed models employ both approaches. We found that L^1-regularized regressors combined with the Black-Litterman model exhibit strong out-of-sample performance. In particular, Multi-task Elastic Net, which was the more crude approach to modeling multi-sector returns, outperformed all other methods considered in this report. We have made efforts in trying to understand the model's unusual effectiveness by experimenting with more realistic backtesting conditions such as transactions costs, long-only constraints, and lagged inputs, as well as their interaction with different types of risk trade-offs.

ACKNOWLEDGMENTS

We thank Dr. Linda Kreitzman for providing guidance for the project and the Haas School of Business for providing the data sources needed for the project from Compustat, CRSP, and Bloomberg.

---

[1] H. Markowitz, The journal of finance **7**, 77 (1952).
[2] R. O. Michaud, Financial Analysts Journal **45**, 31 (1989).
[3] F. Black and R. B. Litterman, The Journal of Fixed Income **1**, 7 (1991).
[4] F. Black and R. Litterman, Financial Analysts Journal **48**, 28 (1992).
[5] D. Alexander and D. Svetlana, Review of Business and Economics Studies (2013).
[6] Pavan Reddy, doi:10.13140/rg.2.2.26727.96160 (2018).
[7] G. He and R. Litterman, Social Science Research Network (2002), 10.2139/ssrn.334304.
[8] A. Meucci, Social Science Research Network (2008), 10.2139/ssrn.1117574.
[9] J. Walters, Social Science Research Network (2011), 10.2139/ssrn.1314585.
[10] W. F. Sharpe, The journal of finance **19**, 425 (1964).
[11] The shape of the estimation error need not necessarily be proportional to $\mathbf{\Sigma}$. This is a simplifying assumption to parsimoniously express the error in terms of the only covariance matrix given as an input in the problem. In practice, τ needs to be chosen, e.g., in a cross-validation scheme, to optimize a desired metric.
[12] J. Morgan and Reuters, "RiskMetrics – Technical Document," (1996).
[13] O. Ledoit and M. Wolf, Journal of Multivariate Analysis **88**, 365 (2004).
[14] R. C. Grinold and R. N. Kahn, *Active Portfolio Management, 2/E* (McGraw-Hill Education, 2004).
[15] S. Diamond and S. Boyd, Journal of Machine Learning Research **17**, 1 (2016).
[16] H. Zou and T. Hastie, Journal of the Royal Statistical Society: Series B (Statistical Methodology) **67**, 301 (2005).
[17] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, Journal of Machine Learning Research **12**, 2825 (2011).
[18] E. E. Qian, R. H. Hua, and E. H. Sorensen, *Quantitative equity portfolio management: modern techniques and applications* (Chapman and Hall/CRC, 2007).
[19] We note that macroeconomic features are only included in $\tilde{\mathbf{X}}$ once to avoid multicollinearity. We forecast the returns for all 11 sectors in this case, as the existence of the macroeconomic factors, and the potential relation of features of other sectors to the Financials and Real Estates sectors, allows us to make meaningful forecasts for these sectors.
[20] G. Obozinski, B. Taskar, and M. Jordan, Statistics Department, UC Berkeley, Tech. Rep **2** (2006).
[21] D. J. C. MacKay, Neural Computation **4**, 415 (1992).
[22] Ideally, we need to have a separate validation set for this process. However, because the amount of data is limited, and also because the regressor was strongly regularized, we choose to perform this tuning on the training/validation set.
[23] A. Y. Ng, in *Proceedings of the twenty-first international conference on Machine learning* (ACM, 2004) p. 78.
[24] All models we have so far considered are convex, and therefore gradient descent yields the global optimum.
[25] K. Natarajan, D. Pachamanova, and M. Sim, Operations Research **57**, 1129 (2009).
[26] E. Qian, "Risk parity portfolios: Efficient portfolios through true diversification," (2005).
[27] X. Bai, K. Scheinberg, and R. Tutuncu, Quantitative Finance **16**, 357 (2016).

Appendix A: Proof of the Posterior Distribution

We assume that the covariance matrices are positive definite. Multiplying the probability density function of the prior distribution in Eq. (1) and the conditional likelihood in Eq. (2), the probability density function of the posterior distribution is of the form [8]

$$\begin{aligned} f_{\boldsymbol{\mu}|\mathbf{q}}(\boldsymbol{\mu}) &\propto \exp\left[-\frac{1}{2}(\mathbf{q}-\mathbf{P}\boldsymbol{\mu})'\,\boldsymbol{\Omega}^{-1}(\mathbf{q}-\mathbf{P}\boldsymbol{\mu})-\frac{1}{2}(\boldsymbol{\mu}-\boldsymbol{\pi})'(\tau\boldsymbol{\Sigma})^{-1}(\boldsymbol{\mu}-\boldsymbol{\pi})\right] \\ &= \exp\left[-\frac{1}{2}\Big(\boldsymbol{\mu}'\left(\mathbf{P}'\boldsymbol{\Omega}^{-1}\mathbf{P}+(\tau\boldsymbol{\Sigma})^{-1}\right)\boldsymbol{\mu}-2\left(\mathbf{P}'\boldsymbol{\Omega}^{-1}\mathbf{q}+(\tau\boldsymbol{\Sigma})^{-1}\boldsymbol{\pi}\right)'\boldsymbol{\mu}+\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{q}+\boldsymbol{\pi}'(\tau\boldsymbol{\Sigma})^{-1}\boldsymbol{\pi}\Big)\right]. \end{aligned}$$

Since the prior and conditional distributions are normal, the posterior distribution must also be normal. Therefore, we conclude that

$$\boldsymbol{\mu}|\mathbf{q} \sim N\Big(\left(\mathbf{P}'\boldsymbol{\Omega}^{-1}\mathbf{P}+(\tau\boldsymbol{\Sigma})^{-1}\right)^{-1}\left(\mathbf{P}'\boldsymbol{\Omega}^{-1}\mathbf{q}+(\tau\boldsymbol{\Sigma})^{-1}\boldsymbol{\pi}\right),\left(\mathbf{P}'\boldsymbol{\Omega}^{-1}\mathbf{P}+(\tau\boldsymbol{\Sigma})^{-1}\right)^{-1}\Big).$$

Appendix B: Results of Ordinary Least Squares Regression

TABLE IV. OLS regression coefficients on sector-specific fundamental variables. Standard errors are reported in parentheses. (*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$)

	Energy	Mater.	Indust.	Cons. Discre.	Cons. Staples	Health	Info. Tech.	Tele. Serv.	Util.
Intercept	0.02	0.01	0.01	0.02	-0.01	0.01	0.03	0.00	-0.01
	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)
ΔP/E	-0.07	0.20	0.63	-0.03	0.23	0.27	0.33	-0.21	0.13
	(0.18)	(0.20)	(0.44)	(0.12)	(0.27)	(0.40)	(0.21)	(0.12)	(0.17)
ΔP/B	-0.00	-0.15	-0.05	-0.12	-0.18	0.08	-0.55*	-0.29**	-0.01
	(0.24)	(0.20)	(0.27)	(0.14)	(0.16)	(0.18)	(0.25)	(0.10)	(0.19)
ΔDiv. yield	-0.07	0.02	0.30	0.12	0.18	0.16	-0.06	-0.50**	0.04
	(0.30)	(0.19)	(0.32)	(0.10)	(0.17)	(0.34)	(0.17)	(0.17)	(0.28)
ΔFCF/P	0.05	-0.25	0.08	-0.70***	-0.05	-0.16	-0.28	0.05	-0.02
	(0.04)	(0.15)	(0.13)	(0.16)	(0.07)	(0.35)	(0.22)	(0.04)	(0.02)
ΔCF	-0.17	0.64	0.12	1.29***	0.07	0.25	0.72	-0.02	-0.06
	(0.33)	(0.38)	(0.40)	(0.32)	(0.14)	(0.48)	(0.58)	(0.19)	(0.09)
ΔCFI	0.21	-0.07	0.03	0.09	-0.09*	0.05	-0.03	-0.05	-0.02
	(0.11)	(0.05)	(0.04)	(0.04)	(0.05)	(0.03)	(0.08)	(0.06)	(0.05)
ΔCFF	0.12*	0.02	0.01	0.01	-0.02	-0.00	0.00	-0.02	0.02
	(0.06)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.04)	(0.03)	(0.02)
ΔSales	-1.04*	0.39	-0.66	-0.96	1.19	-0.19	0.79	-0.27	-0.03
	(0.42)	(0.79)	(1.12)	(0.49)	(0.64)	(0.61)	(0.80)	(0.35)	(0.14)
ΔEBITDA	0.35	-0.66	1.12	-0.12	-0.48	-0.16	-0.77	0.25	0.50
	(0.44)	(0.54)	(1.02)	(0.41)	(0.60)	(0.73)	(0.49)	(0.28)	(0.34)
ΔEV/EBITDA	-0.06	-0.40	-0.09	-0.47*	0.27	-0.47*	-0.08	0.10	0.03
	(0.21)	(0.31)	(0.21)	(0.23)	(0.27)	(0.20)	(0.26)	(0.12)	(0.20)
ΔCur. ratio	0.04	0.17	0.12	0.13	-0.03	0.09	-0.33	0.09	-0.19
	(0.29)	(0.17)	(0.16)	(0.08)	(0.15)	(0.09)	(0.44)	(0.12)	(0.17)
ΔOper. margin	0.16	-0.52	-0.15	0.10	0.72	0.22	0.16	0.21	0.23
	(0.26)	(0.65)	(0.85)	(0.20)	(0.64)	(0.47)	(0.15)	(0.31)	(0.40)
ΔProf. margin	0.02	-0.03	0.10	0.04	-0.09	0.00	0.06	-0.00	-0.07
	(0.11)	(0.05)	(0.10)	(0.03)	(0.13)	(0.09)	(0.06)	(0.04)	(0.04)
ΔEBITDA margin	-0.73	0.99	-1.00	-0.56	0.27	-0.59	-0.14	-0.44	-0.27
	(0.53)	(1.12)	(1.57)	(0.50)	(0.88)	(0.76)	(0.53)	(0.61)	(0.44)
ΔDebt/Asset	0.01	0.12	-0.59	0.15	-0.07	0.04	-0.38	-0.08	-0.47
	(0.21)	(0.18)	(0.38)	(0.16)	(0.17)	(0.12)	(0.31)	(0.24)	(0.31)

TABLE V. OLS regression coefficients on macroeconomic variables.

	Energy	Mater.	Indust.	Cons. Discre.	Cons. Staples	Health	Info. Tech.	Tele. Serv.	Util.
ΔLeading index	1.05	1.04	0.13	0.26	-0.06	0.79	-0.57	-2.03	1.11
	(1.32)	(1.37)	(1.23)	(1.26)	(0.74)	(0.97)	(1.69)	(1.13)	(0.90)
ΔCoin./lag. index	-0.18	1.25	1.35	1.32	0.83	0.20	1.72	-0.47	-0.17
	(1.60)	(1.69)	(1.46)	(1.58)	(0.91)	(1.12)	(2.04)	(1.37)	(1.05)
ΔCPI	-3.72	0.44	1.48	-0.36	0.64	-1.26	-0.02	0.06	-0.86
	(2.12)	(2.16)	(1.84)	(2.00)	(1.22)	(1.42)	(2.65)	(1.73)	(1.40)
ΔCPI core	-2.23	-11.18	-9.05	-11.08	0.37	-3.09	-18.36*	-9.03	-0.51
	(6.98)	(7.11)	(6.40)	(6.35)	(4.02)	(4.98)	(9.09)	(5.96)	(4.74)
ΔUnemployment	-4.39	-3.77	-2.80	0.17	-1.95	0.75	5.18	1.94	-4.08
	(4.51)	(4.67)	(4.32)	(4.25)	(2.60)	(3.24)	(5.83)	(3.87)	(3.02)
ΔInit. jobless cl.	-0.01	-0.02	-0.07	0.07	0.03	0.01	-0.08	-0.24**	-0.02
	(0.11)	(0.11)	(0.09)	(0.09)	(0.06)	(0.07)	(0.13)	(0.09)	(0.07)
ΔIndust. Prod.	2.44	2.55	3.31*	2.46	2.00*	1.81	2.09	2.33	2.32*
	(1.53)	(1.60)	(1.41)	(1.41)	(0.84)	(1.06)	(1.92)	(1.28)	(0.99)
ΔFed funds r.	-0.94	-0.78	2.63	2.42	0.59	2.04	7.57	4.06	0.66
	(2.94)	(3.13)	(2.76)	(2.84)	(1.78)	(2.05)	(3.94)	(2.57)	(1.96)
ΔM2 money sp.	0.33	0.07	0.27	0.28	-0.31	0.35	-0.02	0.13	1.32**
	(0.73)	(0.77)	(0.66)	(0.67)	(0.42)	(0.50)	(0.95)	(0.63)	(0.48)
CFNAI/10	-0.09	-0.20	-0.21	-0.14	-0.16	-0.10	-0.04	-0.06	-0.23*
	(0.18)	(0.18)	(0.17)	(0.17)	(0.10)	(0.12)	(0.22)	(0.14)	(0.11)
ΔPriv. housing	0.01	0.03	0.08	0.03	0.08	0.07	0.01	0.09	0.07
	(0.08)	(0.08)	(0.07)	(0.08)	(0.05)	(0.06)	(0.10)	(0.07)	(0.06)
ΔDXY	0.03	-0.27	-0.28	0.02	0.06	-0.20	-0.92	-0.84	-0.28
	(0.82)	(0.85)	(0.77)	(0.80)	(0.47)	(0.59)	(1.10)	(0.69)	(0.55)
DX1/DXY-1	2.48	-0.23	-0.27	1.11	-0.58	1.53	0.28	-6.54	-4.53
	(3.88)	(4.04)	(3.42)	(3.61)	(2.22)	(2.68)	(4.82)	(3.42)	(2.62)
ΔWTI	0.09	0.08	-0.00	-0.04	0.02	-0.01	-0.09	-0.11*	0.01
	(0.08)	(0.07)	(0.06)	(0.06)	(0.04)	(0.05)	(0.09)	(0.06)	(0.05)
ΔUSTW$	0.07	0.42	0.32	-0.33	-0.16	-0.09	0.73	0.54	0.31
	(1.00)	(1.04)	(0.93)	(0.98)	(0.55)	(0.69)	(1.32)	(0.81)	(0.65)
Δ6m yield	0.01	0.02	-0.00	0.03	0.01	-0.00	-0.01	0.01	-0.01
	(0.04)	(0.04)	(0.04)	(0.04)	(0.02)	(0.03)	(0.05)	(0.03)	(0.03)
2-10y spread	-0.83	-0.07	-0.52	-0.01	-0.13	-0.95	-0.09	1.23	0.27
	(0.95)	(0.96)	(0.84)	(0.86)	(0.53)	(0.67)	(1.19)	(0.80)	(0.64)
R^2	0.20	0.18	0.19	0.26	0.18	0.17	0.20	0.29	0.22

Appendix C: Backtest results using mean-standard deviation optimization

TABLE VI. Out-of-sample backtest results using standard deviation as the risk metric with $\epsilon = 0.1$.

	Annual return	Annual volatility	Annual tracking error	Annual turnover	Information ratio	Sharpe ratio	Maximum drawdown	Daily VaR (5%)
OLS	12.1%	87.2%	82.7%	10541%	-0.01	0.55	-74.9%	7.9%
OLS BL	0.8%	87.0%	83.0%	9242%	-0.13	0.44	-76.1%	7.5%
Elastic Net BL	20.1%	61.5%	57.8%	6388%	0.11	0.61	-64.8%	6.0%
Multi-task Elastic Net BL	32.2%	39.3%	37.2%	3023%	0.43	0.91	-38.9%	3.4%
Bayesian Ridge BL	13.9%	34.5%	30.5%	3177%	0.04	0.55	-46.8%	3.5%
Relative view BL	12.9%	56.3%	51.3%	7724%	0.00	0.50	-77.6%	5.9%
Mixed view BL	28.2%	47.3%	43.0%	6933%	0.30	0.76	-55.7%	4.5%
Hybrid BL	12.9%	67.5%	62.8%	5505%	0.00	0.52	-79.3%	6.6%

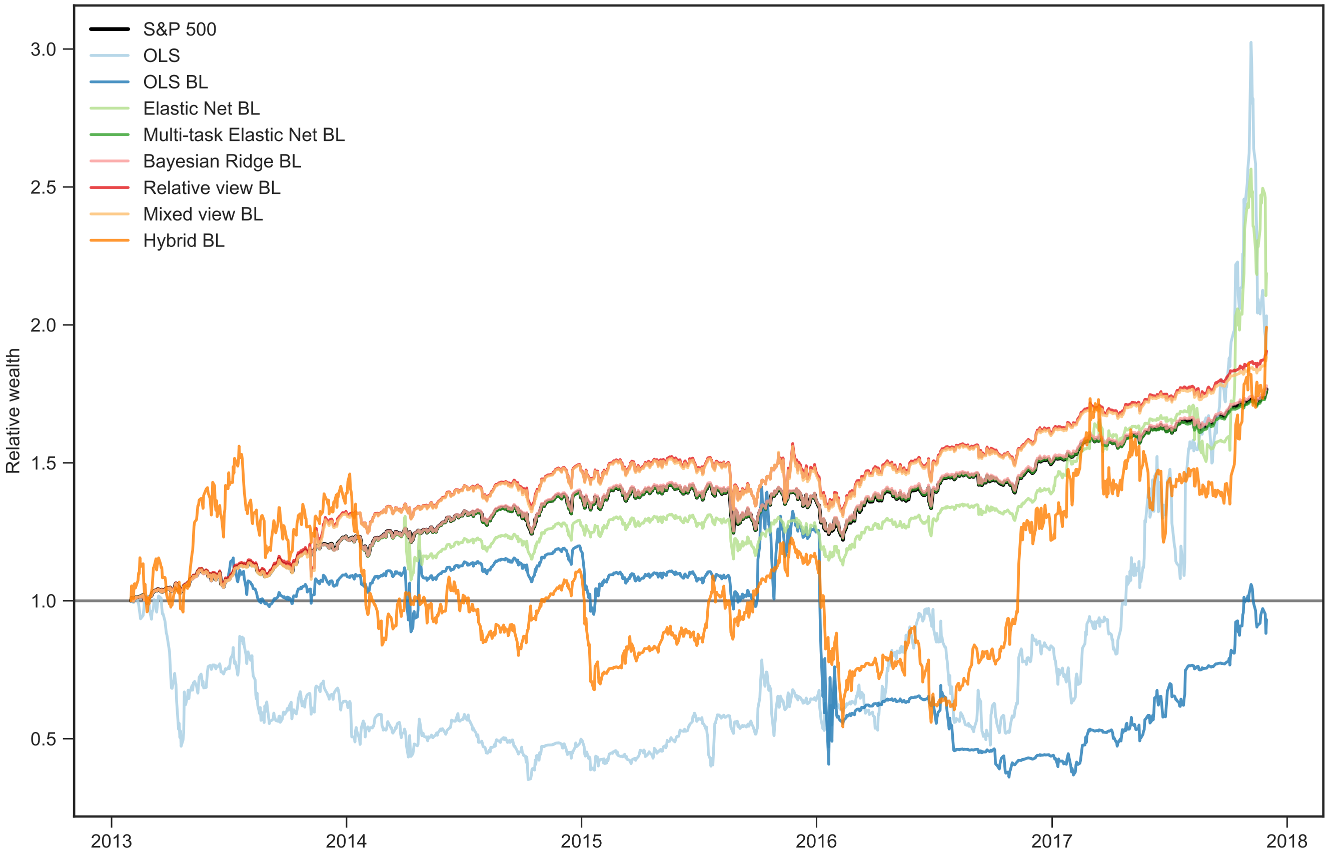


FIG. 7. Out-of-sample backtest results using standard deviation as the risk metric with $\epsilon = 0.3$.

Appendix D: Backtest results using approximate risk parity optimization

TABLE VII. Out-of-sample backtest results of approximate risk parity optimization with $\rho = 10000$.

	Annual return	Annual volatility	Annual tracking error	Annual turnover	Information ratio	Sharpe ratio	Maximum drawdown	Daily VaR (5%)
OLS	-3.3%	27.3%	24.1%	5432%	-0.63	0.01	-45.2%	2.8%
OLS BL	10.8%	17.9%	11.1%	1966%	-0.14	0.66	-34.8%	1.8%
Elastic Net BL	16.7%	15.0%	7.0%	970%	0.51	1.10	-14.5%	1.6%
Multi-task Elastic Net BL	16.0%	13.0%	4.5%	392%	0.67	1.21	-16.9%	1.4%
Bayesian Ridge BL	14.9%	13.7%	4.0%	627%	0.50	1.08	-18.3%	1.5%
Relative view BL	13.6%	15.6%	7.3%	1472%	0.12	0.90	-18.1%	1.7%
Mixed view BL	16.9%	14.9%	6.7%	1394%	0.55	1.12	-16.4%	1.5%
Hybrid BL	26.5%	23.8%	18.9%	2646%	0.62	1.11	-32.3%	2.4%