Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Performance Attribution for Portfolio Constraints

Andrew W. Lo; , Ruixun Zhang

informs
https://pubsonline.informs.org/journal/mnsc

MANAGEMENT SCIENCE
Articles in Advance, pp. 1–23
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Performance Attribution for Portfolio Constraints

Andrew W. Lo,[a,b] **Ruixun Zhang**[c,*]

[a] Sloan School of Management, Laboratory for Financial Engineering, and Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139; [b] Santa Fe Institute, Santa Fe, New Mexico 87501; [c] School of Mathematical Sciences, Center for Statistical Science, National Engineering Laboratory for Big Data Analysis and Applications, and Laboratory for Mathematical Economics and Quantitative Finance, Peking University, Beijing 100871, China

*Corresponding author

Contact: alo-admin@mit.edu, https://orcid.org/0000-0003-2944-7773 (AWL); zhangruixun@pku.edu.cn, https://orcid.org/0000-0002-7670-8393 (RZ)

Received: March 19, 2024
Revised: June 21, 2024
Accepted: August 22, 2024
Published Online in Articles in Advance: December 23, 2024

https://doi.org/10.1287/mnsc.2024.05365

Abstract. We propose a new performance attribution framework that decomposes a constrained portfolio's holdings, expected returns, variance, expected utility, and realized returns into components attributable to (1) the unconstrained mean-variance optimal portfolio; (2) individual static constraints; and (3) information, if any, arising from those constraints. A key contribution of our framework is the recognition that constraints may contain information that is correlated with returns, in which case imposing such constraints can affect performance. We extend our framework to accommodate estimation risk in portfolio construction using Bayesian portfolio analysis, which allows one to select constraints that improve—or are least detrimental to—future performance. We provide simulations and empirical examples involving constraints on environmental, social, and governance portfolios. Under certain scenarios, constraints may improve portfolio performance relative to a passive benchmark that does not account for the information contained in these constraints.

History: Accepted by Kay Giesecke, finance.

Funding: Research funding from the National Key Research and Development Program of China [Grant 2022YFA1007900], the National Natural Science Foundation of China [Grants 12271013, 72342004], Peking University's Fundamental Research Funds for the Central Universities, and the MIT Laboratory for Financial Engineering is gratefully acknowledged.

Supplemental Material: The online appendix and data files are available at https://doi.org/10.1287/mnsc.2024.05365.

Keywords: portfolio theory • performance attribution • constraints • information • ESG investing • socially responsible investing

1. Introduction

Constraints are ubiquitous in portfolio management. They are regularly imposed, both directly and indirectly, by portfolio managers, regulators, risk managers, trading desks, and investors. Because these constraints directly affect the portfolio-construction process, all stakeholders have become interested in quantifying how constrained portfolios deviate from the unconstrained optimal benchmark, using various metrics and concepts, such as unrealized alpha, opportunity cost, and implementation inefficiency. Measuring the impact of constraints on portfolio performance has become particularly important as socially responsible investing (SRI) and environmental, social, and governance (ESG) products have grown in popularity over the last decade because constraints are usually involved in the construction of these portfolios.

In this article, we develop a general framework in which the financial impact of constraints can be measured by attributing the performance of portfolios to contributions from individual constraints. Because a constrained portfolio contains a proper subset of securities of the unconstrained version, mathematical logic suggests that the constrained optimum is at best equal to the unconstrained optimum or, more likely, inferior. However, the nonsuperiority of constrained optima relies on a key assumption that is almost never explicitly stated: the constraint does not provide additional information regarding asset returns. In other words, constraints are assumed to be statistically independent of the returns. In some cases, such an assumption is warranted; for example, one can imagine constructing a subset of securities with Committee on Uniform Securities Identification Procedures (CUSIP) identifiers that contain prime numbers. Imposing such a constraint clearly reduces the risk-adjusted return of the optimized portfolio.

But what if the constraint is not independent of the returns? For example, consider the constraint to invest only in stocks whose prices will appreciate by more than 10% over the next 12 months. Apart from the infeasibility of imposing such a condition, it should be obvious that this constraint would, in fact, increase the risk-adjusted return of the optimized portfolio. Therefore, quantifying the impact of constraints rests entirely on whether and how the constraints are related to the performance characteristics of the securities under consideration.

To formalize this idea, we consider investors who construct portfolios by maximizing the standard mean-variance utility. We denote the optimal portfolio without constraints as the mean-variance optimal (MVO) portfolio, which is usually a simple and passive benchmark to which all investors have access. However, the portfolio obtained when imposing constraints will likely differ from the MVO portfolio. Therefore, we develop a methodology to decompose the constrained portfolio's holdings, expected returns, variance, expected utility, and realized returns into different components: those attributable to the MVO portfolio, the individual constraints treated as static, and the information contained in the constraints. This methodology yields a constraint attribution framework for evaluating the performance of a portfolio.

The key to our framework is to model the information content available in portfolio constraints. We assume that each constraint is based on a firm characteristic, $\mathbf{x} \equiv [x_1 \; x_2 \; \cdots \; x_N]'$, where x_i is the characteristic for the ith asset, such as its ESG score or a label representing its industry. By modeling $\mathbf{x}$ as a random variable and allowing it to be correlated with asset returns, we are able to provide an explicit decomposition of the performance of a portfolio attributable to information in its constraints, which depends critically on the expected value and covariance matrix of returns conditioned on $\mathbf{x}$. Furthermore, in the special case of normally and multivariate Student's t distributed returns, we demonstrate that the information contribution from a constraint is determined by the correlation between $\mathbf{x}$ and the individual asset returns. The excess return from information is positive when this correlation is positive and the constraint is binding. The excess variance of a portfolio is negative when the portfolio holdings of a shrinkage portfolio (defined in Equation (8)) and the holdings attributable to constraints are positively correlated, and the magnitude of the reduction in variance depends on the absolute value of the same correlation.

This simple but profound result highlights the mechanism through which a constraint contributes to the performance of a portfolio. Whereas a constraint treated as static must decrease a portfolio's expected utility, the information in the constraint can contribute either positively or negatively to a portfolio's expected utility and returns, depending on whether the characteristics of the constraint are positively or negatively correlated with asset returns. In this sense, constraints serve as an indirect mechanism for using information that is otherwise unavailable to investors in the passive benchmark.[1]

In addition, we demonstrate that our framework can accommodate estimation risk in the expected value and covariance matrix of asset returns used to construct portfolios. We do this using Bayesian portfolio analysis and modeling the dependence between firm characteristic $\mathbf{x}$ and the posterior predictive returns. This allows for performance attribution of portfolio constraints for out-of-sample returns. Furthermore, we establish an equivalence result between a Bayesian portfolio with constraints and an unconstrained portfolio with a certain prior on asset returns, which implies that our attribution results can also be interpreted as the influence of investors' views on returns (expressed as Bayesian priors) on the portfolio's performance.

We apply our framework to two common classes of portfolio constraints. The first occurs when investors restrict exposure to a certain factor, such as the average ESG score, market capitalization, beta, or book-to-market values of the portfolio. The second is exclusionary investing, in which certain assets are excluded from the portfolio based on criteria such as whether the firm belongs to an industry associated with "sin" stocks. We derive additional analytical results and provide simulated examples to illustrate the attribution in these scenarios, which naturally leads to a method for selecting constraints that may improve portfolio performance.

Finally, we provide an empirical application in the context of SRI and ESG investing. Their growing popularity and assets under management have triggered a backlash recently.[2] Our framework provides a potential solution to properly disclose the financial impact of the constraints imposed by these investments and reconcile SRI and ESG investing with fiduciary duty.

In particular, using real-world data sets, we quantify the financial impact of ESG constraints when the average portfolio ESG score is required to be above a certain threshold as well as exclusionary investing based on sin stocks and stranded assets. Whereas the expected utility contribution of these constraints, treated as static, is indeed negative, the contribution from the information contained in the constraints to portfolio performance is dynamic over time. This contribution is generally negative before 2007, implying that high ESG stocks delivered lower excess returns relative to the Fama–French five-factor model, on average, which is consistent with equilibrium theories of ESG returns (Pástor et al. 2021, Pedersen et al. 2021). However, after 2008, the information in the constraints starts to contribute positively to portfolio performance in certain years, reflecting the increasing attention toward SRI and ESG-related issues, an effect consistent with the Pástor et al. (2022) findings of shifted preferences.

We emphasize that our intention in this article is not to provide a measure for determining whether SRI and ESG investing deliver positive or negative excess returns. Instead, our primary objective is to illustrate how our framework can be used to attribute performance to any portfolio constraints and to the information contained in those constraints. We make our software publicly available to researchers and investors to facilitate the application of our performance attribution framework.

1.1. Contributions and Related Literature

Our article contributes to the literature in several respects. First and foremost, we provide an attribution framework that decomposes the performance of constrained portfolios relative to a static benchmark by quantifying the information content in constraints both in-sample and out-of-sample. The classic literature on style analysis and performance attribution includes Fama (1972), Brinson et al. (1986), and Sharpe (1992). In addition, the transfer coefficient (Clarke et al. 2002) and other measures based on the shadow cost have been proposed to measure the marginal cost of each constraint (Grinold 2005, Stubbs and Vandenbussche 2010, Menchero and Davis 2011, Goldberg 2021).

A key contribution of our decomposition framework to this literature is a quantitative measure of the information contained in each constraint. We show that recognizing the information contained in the constraints has important implications for performance analysis and understanding the impact of, for example, SRI and ESG investing.

Second, our framework is connected to the extensive literature on robust portfolios that accounts for estimation risk in the parameters of portfolio construction because imposing constraints is one way to deal with estimation risk. We discuss the relationship between this literature and our method in more detail in Section 3. In particular, portfolio constraints can mitigate estimation risk as shown by the no-short-sale constraints of Jagannathan and Ma (2003), the norm constraints of DeMiguel et al. (2009b), the gross-exposure constraints of Fan et al. (2012b), and the variance-based constraints of Levy and Levy (2014).

Although our framework can accommodate estimation risk, our goal is not to propose a new robust portfolio rule or a better way to deal with estimation risk. Instead, we propose to use our framework to decompose the performance of out-of-sample returns of existing robust portfolio rules. In fact, we show that, even without estimation error, there exists information in constraints that can affect portfolio performance. Whereas constraints undoubtedly play a role in reducing estimation risk and improving out-of-sample performance as documented by the literature cited above, we demonstrate a different mechanism through which they can affect a portfolio: the information contained in the constraints, which can contribute either positively or negatively to the performance of a portfolio depending on the statistical correlations of this information with returns. This is a key distinction from the prior literature on estimation risk and robust portfolio optimization. In practice, investors often impose constraints for business or regulation reasons other than reducing estimation risk, such as in SRI and ESG products, and these constraints can serve as an indirect mechanism to incorporate information. Our methodology provides a way to quantify this effect precisely, which, to the best of our knowledge, has not appeared in the existing literature.

Third, our framework and results differ in several respects from those of Brandt et al. (2009) and Hjalmarsson and Manchev (2012), who incorporate firm characteristics into the estimation of portfolio weights to derive robust portfolios. Their frameworks parameterize the portfolio weights of each stock as a (linear) function of the firm characteristic and then estimate the coefficients of this function. They do not model the impact of portfolio constraints, which is our focus. Also, the goal of their frameworks is to derive robust portfolio rules by avoiding the estimation of the return distribution. As discussed above, our framework has a very different goal. Whereas Brandt et al. (2009) and Hjalmarsson and Manchev (2012) convincingly demonstrate that their frameworks can produce robust portfolios, our results make a different contribution: we provide a quantitative measure of the information contained in portfolio constraints. This is relevant to both the asset pricing literature and the SRI and ESG literature because constraints are routinely used to construct such portfolios.

Fourth, from a technical perspective, our methodology involves solving a quadratic optimization problem subject to stochastic linear constraints. Whereas the literature on stochastic optimization is vast (Powell 2019), most of it is concerned with solving a sequential problem in which the stochasticity in the environment, objective, or constraints affects the optimal policy. The constraint is imposed on either the expected value of the random variables as in Jin et al. (2008) or the probability of a certain condition as in Bonami and Lejeune (2009). Our framework is concerned with the static portfolio selection problem, which is easy to solve, and the constraint is imposed sample-wise. Our main contribution, instead, is to quantify the information that these stochastic constraints contain with respect to the random variables involved in the objective function.

Finally, our results also contribute to the literature on the impact of SRI, ESG, and other nonfinancial objectives on investment returns. Because ESG measures from different data providers can lead to very different correlations (Berg et al. 2022), our findings highlight that the effect of a specific measure of SRI or ESG on investment performance depends on the information contained in the constraints created by these measures. These constraints need not always result in lower risk-adjusted returns.

We develop our main framework in Section 2. In Section 3, we incorporate estimation risk into our framework using Bayesian portfolio analysis. We consider two common examples of portfolio constraints in Section 4 and provide an illustrative application to ESG investing in Section 5. We conclude in Section 6.

2. A Framework for Constraint Attribution

We consider a universe of N assets whose returns are given by the random vector $\mathbf{r}_t = [r_{1,t} \cdots r_{N,t}]'$.[3] Because we principally consider the static portfolio-selection problem in this article, we omit the time subscript t and simply write $\mathbf{r}$ in most cases. We denote by $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ the expected value and covariance matrix of $\mathbf{r}$, respectively. Investors solve the following mean-variance portfolio:

$$\begin{aligned} \max_{\boldsymbol{\omega}} \quad & \boldsymbol{\omega}'\boldsymbol{\mu} - \frac{\gamma}{2}\boldsymbol{\omega}'\boldsymbol{\Sigma}\boldsymbol{\omega} \\ \text{s.t.} \quad & \mathbf{A}\boldsymbol{\omega} = \mathbf{b}, \end{aligned} \tag{1}$$

where $\boldsymbol{\omega} \equiv [\omega_1 \cdots \omega_N]'$ is an N-dimensional vector representing portfolio weights, γ is the relative risk-aversion parameter, $\mathbf{b} \equiv [b_1 \ \cdots \ b_J]'$ is a J-dimensional vector, and

$$\mathbf{A} \equiv \begin{pmatrix} \mathbf{A}_1' \\ \cdots \\ \mathbf{A}_J' \end{pmatrix}$$

is a $J \times N$ full-rank matrix. Together, $\mathbf{b}$ and $\mathbf{A}$ describe J constraints. In particular, $\mathbf{A}_j'$ is the jth row of $\mathbf{A}$ and b_j is the jth element of $\mathbf{b}$, which together describe the jth constraint.

We consider the case of equality constraints in (1) for expositional simplicity. It is also easy to derive a parallel set of results under inequality constraints, $\mathbf{A}\boldsymbol{\omega} \le \mathbf{b}$, and we describe ways to generalize our results throughout our exposition. Common examples of portfolio constraints that can be described by (1) include $\boldsymbol{\omega}'\mathbf{1} = 1$ representing a full investment constraint, $\omega_i = 0$ representing the exclusion of asset i, and $\boldsymbol{\omega}'\mathbf{A}_1 = b_1$ representing a certain level of factor exposure.

A critical implicit assumption in the existing literature on constraint attribution is that the constraints, $\mathbf{A}$, are treated as constants and are, therefore, independent of returns, $\mathbf{r}$. Under this setting, the solution to the optimization problem in (1) without constraints, which we refer to as the unconstrained MVO portfolio, yields the best portfolio in terms of the objective value, and imposing constraints can only decrease the objective value.[4] The following result summarizes the optimal portfolio weights and the decomposition of portfolio holdings, expected return, and expected utility attributable to each constraint.[5] We provide proofs of all propositions in Online Appendix A.

Proposition 1 (Static Constraints). *The optimal portfolio weight*, $\boldsymbol{\omega}^*$, *of Problem* (1) *is given by*

$$\boldsymbol{\omega}^* = \frac{1}{\gamma}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \mathbf{A}'\boldsymbol{\lambda}^*), \tag{2}$$

where the Lagrange multipliers are given by $\boldsymbol{\lambda}^* = (\mathbf{A}\boldsymbol{\Sigma}^{-1}\mathbf{A}')^{-1}(\mathbf{A}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \gamma\mathbf{b})$ *provided that the feasible region of the constrained optimization problem is nonempty. Here,* $\boldsymbol{\lambda}^*$ *is a measure of the shadow cost of the portfolio's expected utility with respect to each constraint.*[6]

Equation (2) *leads to a series of decompositions:*

1. *Portfolio holdings decomposition:*

$$\boldsymbol{\omega}^* = \frac{1}{\gamma}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{1}{\gamma}\boldsymbol{\Sigma}^{-1}\mathbf{A}'\boldsymbol{\lambda}^*. \tag{3}$$

- $\boldsymbol{\omega}_{\text{MVO}} \equiv \frac{1}{\gamma}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$: *holdings of the unconstrained MVO portfolio.*
- $\boldsymbol{\omega}_{\text{CSTR}} \equiv -\frac{1}{\gamma}\boldsymbol{\Sigma}^{-1}\mathbf{A}'\boldsymbol{\lambda}^*$: *components attributable to each constraint.*

2. *Expected return decomposition:*

$$\boldsymbol{\mu}'\boldsymbol{\omega}^* = \frac{1}{\gamma}\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{1}{\gamma}\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{A}'\boldsymbol{\lambda}^*. \tag{4}$$

- $\frac{1}{\gamma}\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$: *expected return of the unconstrained MVO portfolio.*
- $-\frac{1}{\gamma}\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{A}'\boldsymbol{\lambda}^*$: *components attributable to each constraint.*

3. *Expected utility decomposition:*

$$\boldsymbol{\mu}'\boldsymbol{\omega}^* - \frac{\gamma}{2}\boldsymbol{\omega}^{*\prime}\boldsymbol{\Sigma}\boldsymbol{\omega}^* = \frac{1}{2\gamma}\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{1}{2\gamma}\boldsymbol{\lambda}^{*\prime}\mathbf{A}\boldsymbol{\Sigma}^{-1}\mathbf{A}'\boldsymbol{\lambda}^*. \tag{5}$$

- $\frac{1}{2\gamma}\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$: *expected utility of the unconstrained MVO portfolio.*
- $-\frac{1}{2\gamma}\boldsymbol{\lambda}^{*\prime}\mathbf{A}\boldsymbol{\Sigma}^{-1}\mathbf{A}'\boldsymbol{\lambda}^*$: *components attributable to all constraints combined together. This term can be equivalently written as* $-\frac{\gamma}{2}\boldsymbol{\omega}_{\text{CSTR}}'\boldsymbol{\Sigma}\boldsymbol{\omega}_{\text{CSTR}}$.

A few observations regarding the intuition behind these decompositions are in order. First, constraints may change the portfolio weights in either direction because the last term in (3) can lead to both positive and negative entries.

Second, in the expected return decomposition in (4), if the Lagrange multiplier, or the shadow cost, for the ith constraint $\lambda_i > 0$, the sign of the marginal contribution of that constraint is determined by the sign of $-\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{A}_j$, which may be positive. To see that, we observe $-\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{A}_j = -|\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}| \cdot |\mathbf{A}_j| \cdot \cos\theta$, where θ is the angle between $\boldsymbol{\omega}_{\text{MVO}} = \frac{1}{\gamma}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ (holdings of the MVO portfolio) and $\mathbf{A}_j$ (constraint coefficients). When these two vectors are negatively correlated, $\cos\theta$ is negative, which implies that the ith constraint increases expected returns.

Third, when constraints are static, they always decrease expected utility relative to the unconstrained MVO portfolio because the last term in (5) is always negative. In addition, this last term provides an attribution of expected utility to all constraints combined together. Unlike the decomposition of the portfolio holdings and expected return, which both provide a simple linear additive attribution to individual constraints, the expected utility cannot be decomposed into a linear combination of each constraint because of the risk term. Nonetheless, there is a clear interpretation of the portfolio holdings decomposition, $\boldsymbol{\omega}_{\text{CSTR}}$, operating on the covariance matrix $\boldsymbol{\Sigma}$.

Finally, the portfolio holdings decomposition in (3) provides a way to derive investor views on $\boldsymbol{\mu}$ implied by their holdings. For example, if a portfolio deviates from the benchmark portfolio, $\boldsymbol{\omega}_{\text{MVO}}$, by $\boldsymbol{\omega}_{\text{CSTR}}$, it is equivalent to solving an unconstrained mean-variance problem using $\boldsymbol{\mu} - \mathbf{A}'\boldsymbol{\lambda}^*$ as the expected value of assets. Therefore, the deviation, $-\mathbf{A}'\boldsymbol{\lambda}^*$, can be treated as the investor's view implied by the portfolio holdings. This idea is explored extensively by, for example, Tu and Zhou (2010) and Ardia and Boudt (2015). We follow Tu and Zhou (2010) and the performance analysis literature mentioned in Section 1 to consider a partial equilibrium analysis of the impact of constraints from the investment perspective. The equilibrium analysis that accounts for the impact of investors on asset prices is an important problem but beyond the scope of this paper.

2.1. Constraints with Information

To extend the framework beyond static constraints, we consider constraints that are potentially correlated with returns. Let the random vector $\mathbf{x}_t \equiv [x_{1,t} \cdots x_{N,t}]'$ be a characteristic or score associated with each asset at time t. The randomness in $\mathbf{x}$ comes from the fact that these scores may change period by period and, more importantly, may be correlated with $\mathbf{r}_t$. For example, $\mathbf{x}_t$ can represent a firm's ESG score, price-to-earnings ratio, momentum measure, or even a proprietary alpha signal at time t. We again omit the subscript t and simply write $\mathbf{x}$ in most cases.

Investors form constraints based on the value of $\mathbf{x}$, and we denote the jth constraint by $\mathbf{A}_j(\mathbf{x})$. For example, $\mathbf{A}_j(\mathbf{x}) = \mathbf{x}$ corresponds to the factor exposure constraint, $\mathbf{x}'\boldsymbol{\omega} = b_j$. More generally, investors can impose J constraints based on J different characteristics denoted by the vector

$$\mathbf{X} \equiv [\mathbf{x}_1' \, \mathbf{x}_2' \cdots \mathbf{x}_J']',$$

where $\mathbf{x}_j$ represents the jth characteristic that forms the jth constraint, $\mathbf{A}_j(\mathbf{x}_j)$. We denote

$$\mathbf{A}(\mathbf{X}) = \begin{pmatrix} \mathbf{A}_1'(\mathbf{x}_1) \\ \vdots \\ \mathbf{A}_J'(\mathbf{x}_J) \end{pmatrix}$$

to be the $J \times N$ coefficient matrix of J constraints that each depend on one characteristic. Investors observe the characteristics $\mathbf{X}$ at the time of portfolio construction but not the returns $\mathbf{r}$. We assume the following about the distribution of asset characteristics, $\mathbf{X}$.

Assumption 1. *The characteristics* $\mathbf{X}$ *have finite moments up to order two.*

2.2. Attribution with Information

We use $\mathbf{r}|X$ to denote the distribution of returns conditioned on information in $\mathbf{X}$ and $\boldsymbol{\mu}_{\mathbf{X}}$ and $\boldsymbol{\Sigma}_{\mathbf{X}}$ to denote its conditional mean and covariance matrix. We first attribute portfolio performance metrics conditioned on $\mathbf{X}$. This can be interpreted as a per-period attribution because it is a function of the realizations of asset characteristics $\mathbf{X}$ in each period. After that, the overall attribution is simply the expectation with respect to $\mathbf{X}$, which can be interpreted as the long-run average of the per-period attribution.

Portfolio weights depend on the constraints and, therefore, $\mathbf{X}$. We still use $\boldsymbol{\omega}^*$ and $\boldsymbol{\omega}_{\text{CSTR}}$ to represent the weights of the constrained portfolio and components attributable to constraints, respectively, but it is worth noting that they are functions of $\mathbf{X}$ in the context of random characteristics. The following result summarizes the attribution of expected return and utility because of information in each constraint, conditioned on $\mathbf{X}$.

Proposition 2 (Conditional Attribution with Information). *Under Assumption 1 and conditioned on information in* $\mathbf{X}$ *that is used to form constraints* $\mathbf{A}(X)$*, the following decompositions hold for the optimal portfolio* $\boldsymbol{\omega}^*$.

1. *Expected return decomposition:*

$$\mathbb{E}[\boldsymbol{\omega}^{*\prime}\mathbf{r}|\mathbf{X}] = \boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}^* = \boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}_{\text{MVO}} + \boldsymbol{\mu}'\boldsymbol{\omega}_{\text{CSTR}} + (\boldsymbol{\mu}_{\mathbf{X}}' - \boldsymbol{\mu}')\boldsymbol{\omega}_{\text{CSTR}}, \tag{6}$$

where

- $\boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}_{\text{MVO}} = \frac{1}{\gamma}\boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$*: expected return of the unconstrained MVO portfolio.*
- $\boldsymbol{\mu}'\boldsymbol{\omega}_{\text{CSTR}} = -\frac{1}{\gamma}\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{A}(\mathbf{X})'\boldsymbol{\lambda}^*$*: components attributable to each constraint treated as static.*
- $(\boldsymbol{\mu}_{\mathbf{X}}' - \boldsymbol{\mu}')\boldsymbol{\omega}_{\text{CSTR}} = -\frac{1}{\gamma}(\boldsymbol{\mu}_{\mathbf{X}}' - \boldsymbol{\mu}')\boldsymbol{\Sigma}^{-1}\mathbf{A}(\mathbf{X})'\boldsymbol{\lambda}^*$*: components attributable to information in constraints.*

Here, the Lagrange multipliers are given by $\boldsymbol{\lambda}^* = (\mathbf{A}(\mathbf{X})\boldsymbol{\Sigma}^{-1}\mathbf{A}(\mathbf{X})')^{-1}(\mathbf{A}(\mathbf{X})\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \gamma\mathbf{b})$ *provided that the feasible region of the constrained optimization problem is nonempty.*

2. *Expected utility decomposition:*

$$\begin{aligned} &\boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}^* - \frac{\gamma}{2}\boldsymbol{\omega}^{*\prime}\boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{\omega}^* \\ &= \boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}_{\text{MVO}} - \frac{\gamma}{2}\boldsymbol{\omega}_{\text{MVO}}'\boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{\omega}_{\text{MVO}} \\ &\quad - \frac{\gamma}{2}\boldsymbol{\omega}_{\text{CSTR}}'\boldsymbol{\Sigma}\boldsymbol{\omega}_{\text{CSTR}} + (\boldsymbol{\mu}_{\mathbf{X}}' - \boldsymbol{\mu}')\boldsymbol{\omega}_{\text{CSTR}} \\ &\quad - \gamma\boldsymbol{\omega}_{\text{SHR}}'(\boldsymbol{\Sigma}_{\mathbf{X}} - \boldsymbol{\Sigma})\boldsymbol{\omega}_{\text{CSTR}}. \end{aligned} \tag{7}$$

- $\boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}_{\text{MVO}} - \frac{\gamma}{2}\boldsymbol{\omega}_{\text{MVO}}'\boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{\omega}_{\text{MVO}} = \frac{1}{\gamma}\boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{1}{2\gamma}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})'\boldsymbol{\Sigma}_{\mathbf{X}}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})$*: optimal expected utility of the unconstrained MVO portfolio.*

- $-\frac{\gamma}{2}\boldsymbol{\omega}'_{\text{CSTR}}\boldsymbol{\Sigma}\boldsymbol{\omega}_{\text{CSTR}} = -\frac{1}{2\gamma}\boldsymbol{\lambda}^{*\prime}\mathbf{A}(\mathbf{X})\boldsymbol{\Sigma}^{-1}\mathbf{A}(\mathbf{X})'\boldsymbol{\lambda}^{*}$: *components attributable to all constraints combined together, treated as static.*
- $(\boldsymbol{\mu}'_{\mathbf{X}} - \boldsymbol{\mu}')\boldsymbol{\omega}_{\text{CSTR}} - \gamma\boldsymbol{\omega}'_{\text{SHR}}(\boldsymbol{\Sigma}_{\mathbf{X}} - \boldsymbol{\Sigma})\boldsymbol{\omega}_{\text{CSTR}} = -\frac{1}{\gamma}(\boldsymbol{\mu}'_{\mathbf{X}} - \boldsymbol{\mu}')\boldsymbol{\Sigma}^{-1}\mathbf{A}(\mathbf{X})'\boldsymbol{\lambda}^{*} - \frac{1}{\gamma}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{\Sigma}^{-1}\mathbf{A}(\mathbf{X})'\boldsymbol{\lambda}^{*})'(\boldsymbol{\Sigma}_{\mathbf{X}} - \boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\mathbf{A}(\mathbf{X})'\boldsymbol{\lambda}^{*}$: *component attributable to information in constraints.*

Here, $\boldsymbol{\omega}_{\text{SHR}}$ *is a shrinkage portfolio defined as*

$$\boldsymbol{\omega}_{\text{SHR}} \equiv \boldsymbol{\omega}_{\text{MVO}} + \frac{1}{2}\boldsymbol{\omega}_{\text{CSTR}} = \boldsymbol{\omega}^{*} - \frac{1}{2}\boldsymbol{\omega}_{\text{CSTR}}. \quad (8)$$

Proposition 2 provides a decomposition of the expected return and utility into components attributable to the unconstrained MVO portfolio, static constraints, and information.

We provide a few remarks regarding the choice of the benchmark portfolio in this decomposition. First, the unconstrained MVO portfolio here is with respect to Problem (1), using the unconditional expected return $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. This is consistent with how performance attribution is typically carried out in the investment management industry; investors pick a widely accepted benchmark portfolio, such as the market portfolio or a passive index, and compare the performance of the actual portfolio, which may contain additional constraints and information, against the chosen benchmark. The decomposition in Proposition 2 provides investors with a quantitative measure of information contained in portfolio constraints relative to any benchmark portfolios they may use.[7]

Second, benchmarking toward the unconstrained MVO portfolio without information in $\mathbf{X}$ provides a way to quantify the value-added performance that investors receive by paying for professional portfolio management. Investors delegate the process of incorporating information through constraints either because of limited cognitive resources or time budgets or because they are not able to fully incorporate such information in $\mathbf{X}$ into return forecasts. The large literature documenting limited attention (Corwin and Coughenour 2008, Hirshleifer et al. 2011) and bounded rationality (Simon 1955, Hirshleifer et al. 2006) in investor behavior provide support for this view. This is particularly relevant for uninformed and less sophisticated investors who lack the conditioning information available only to informed investors as in the frameworks of Dybvig and Ross (1985) and Ferson and Siegel (2001).

We emphasize that the decomposition in Proposition 2 is fundamentally different from the traditional constraint attribution given in Proposition 1, in which the coefficients that form constraints are assumed to be constant. Once the constraints depend on asset characteristics $\mathbf{X}$ that are potentially correlated with returns, they provide information. Proposition 2 quantifies this effect explicitly by showing how information contributes to the expected return and utility of a portfolio.

The information component of expected return is $(\boldsymbol{\mu}'_{\mathbf{X}} - \boldsymbol{\mu}')\boldsymbol{\omega}_{\text{CSTR}}$, which implies that the information contributes positively when portfolio holdings attributable to constraints are positively correlated with the excess return vector of all assets, $\boldsymbol{\mu}'_{\mathbf{X}} - \boldsymbol{\mu}'$. This term can be further decomposed into components attributable to each individual constraint under certain distributional assumptions of $\mathbf{X}$. We discuss this in Section 2.3.

The information component of expected utility is $(\boldsymbol{\mu}'_{\mathbf{X}} - \boldsymbol{\mu}')\boldsymbol{\omega}_{\text{CSTR}} - \gamma\boldsymbol{\omega}'_{\text{SHR}}(\boldsymbol{\Sigma}_{\mathbf{X}} - \boldsymbol{\Sigma})\boldsymbol{\omega}_{\text{CSTR}}$. The first part is the same as the information component of the expected return, and the second part corresponds to the information component from the variance, which itself consists of two terms:

$$\begin{aligned} -\gamma\boldsymbol{\omega}_{\text{SHR}}(\boldsymbol{\Sigma}_{\mathbf{X}} - \boldsymbol{\Sigma})\boldsymbol{\omega}_{\text{CSTR}} &= -\gamma\left(\boldsymbol{\omega}'_{\text{MVO}} + \frac{1}{2}\boldsymbol{\omega}'_{\text{CSTR}}\right)(\boldsymbol{\Sigma}_{\mathbf{X}} - \boldsymbol{\Sigma})\boldsymbol{\omega}_{\text{CSTR}} \\ &= -\frac{\gamma}{2}\boldsymbol{\omega}'_{\text{CSTR}}(\boldsymbol{\Sigma}_{\mathbf{X}} - \boldsymbol{\Sigma})\boldsymbol{\omega}_{\text{CSTR}} - \gamma\boldsymbol{\omega}'_{\text{MVO}}(\boldsymbol{\Sigma}_{\mathbf{X}} - \boldsymbol{\Sigma})\boldsymbol{\omega}_{\text{CSTR}}. \end{aligned} \quad (9)$$

The first term corresponds to the excess variance, $\boldsymbol{\Sigma}_{\mathbf{X}} - \boldsymbol{\Sigma}$, from portfolio holdings attributable to constraints, $\boldsymbol{\omega}_{\text{CSTR}}$. The second term corresponds to an interaction effect between the unconstrained MVO portfolio, $\boldsymbol{\omega}_{\text{MVO}}$, and the component attributable to constraints, $\boldsymbol{\omega}_{\text{CSTR}}$.

Taken together, (9) can be interpreted as the covariance between the shrinkage portfolio, $\boldsymbol{\omega}_{\text{SHR}}$, and the portfolio attributable to constraints, $\boldsymbol{\omega}_{\text{CSTR}}$. However, this covariance is not with respect to the returns of the original assets, but with respect to a set of hypothetical assets whose covariance is determined by the negative excess covariance matrix, $-(\boldsymbol{\Sigma}_{\mathbf{X}} - \boldsymbol{\Sigma})$.[8] Figure 1 shows two examples of the shrinkage portfolio, which shrinks the optimal constrained portfolio, $\boldsymbol{\omega}^{*}$, toward the unconstrained MVO portfolio, $\boldsymbol{\omega}_{MVO}$.

Finally, Online Appendix B.3 provides the unconditional version of Proposition 2, which can be interpreted as the average decomposition over multiple time periods.

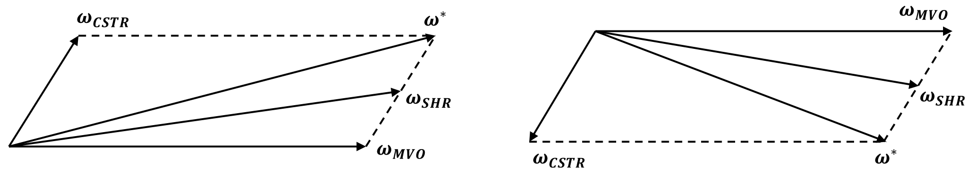
2.3. Decomposing Information with Specific Distributions

The attribution in Proposition 2 depends critically on two terms, $(\boldsymbol{\mu}_{\mathbf{X}} - \boldsymbol{\mu})$ and $(\boldsymbol{\Sigma}_{\mathbf{X}} - \boldsymbol{\Sigma})$, which capture the excess return and covariance because of information in $\mathbf{X}$, respectively. They cannot be simplified further for general distributions of $\mathbf{r}$ and $\mathbf{X}$. However, for certain classes of specific distributions, we can decompose these terms further and obtain considerable intuition about their contribution.

We consider multivariate normal and multivariate Student's t (MVT) distributions in this section. The normal distribution is a common choice in the portfolio

Figure 1. Illustration of Shrinkage Portfolios Defined in (8)

(a) Acute angle between $\boldsymbol{\omega}_{\mathrm{SHR}}$ and $\boldsymbol{\omega}_{\mathrm{CSTR}}$ (b) Obtuse angle between $\boldsymbol{\omega}_{\mathrm{SHR}}$ and $\boldsymbol{\omega}_{\mathrm{CSTR}}$



literature, and the Student t distribution is widely adopted to model heavy-tail returns. Both appear in the posterior predictive distribution in Bayesian portfolio analysis, which we use to account for estimation risk in Section 3.

We make the following two parallel assumptions.

Assumption 2. *The returns and characteristics,* $(\mathbf{r}', \mathbf{x}_1', \ldots, \mathbf{x}_J')$*, are jointly normally distributed.*

Assumption 2′. *The returns and characteristics,* $(\mathbf{r}', \mathbf{x}_1', \ldots, \mathbf{x}_J')$*, follow a multivariate Student t distribution:*

$$\begin{pmatrix} \mathbf{r} \\ \mathbf{X} \end{pmatrix} \sim \mathrm{MVT}\left(\begin{pmatrix} \boldsymbol{\mu} \\ \bar{\mathbf{X}} \end{pmatrix}, \begin{pmatrix} \mathbf{V} & \mathbf{V}_{\mathbf{r},\mathbf{X}} \\ \mathbf{V}_{\mathbf{X},\mathbf{r}} & \mathbf{V}_{\mathbf{X},\mathbf{X}} \end{pmatrix}, \nu \right).$$

Recall that $\mathbf{X}$ represents the $(N \times J)$-dimensional vector $[\mathbf{x}_1' \cdots \mathbf{x}_J']'$, and we use $\bar{\mathbf{X}} \equiv [\bar{\mathbf{x}}_1' \cdots \bar{\mathbf{x}}_J']'$ to denote the expected value of $\mathbf{X}$. Assumption 2′ implies that the returns $\mathbf{r} \sim \mathrm{MVT}(\boldsymbol{\mu}, \mathbf{V}, \nu)$. The parameters $\boldsymbol{\mu}$ and $\mathbf{V}$ are usually referred to as the location vector and scale matrix, respectively, and ν is the degree of freedom. The first two moments of an MVT distribution are given by

$$\mathbb{E}[\mathbf{r}] = \boldsymbol{\mu} \quad \text{and} \quad \boldsymbol{\Sigma} = \mathrm{Cov}(\mathbf{r}) = \frac{\nu}{\nu - 2}\mathbf{V}. \tag{10}$$

The expected value is the location vector, but the covariance matrix is not the scale matrix. Hence, we use $\mathbf{V}$ to denote the scale matrix to emphasize its difference from the covariance matrix, $\boldsymbol{\Sigma}$.

The next assumption describes the dependence between characteristics and returns.

Assumption 3. *The joint distribution of the return vector,* $\mathbf{r}$*, and characteristics* $\mathbf{X} = [\mathbf{x}_1' \, \mathbf{x}_2' \cdots \mathbf{x}_J']'$ *satisfies the following conditions:*

1. *The characteristic values are independent both across different assets and between the J different constraints.*
2. *For the jth constraint, the correlation between the return and characteristic value of each asset is* ρ_j*, and there is no cross-correlation between the return and characteristic value of different assets. In other words, the covariance between returns* $\mathbf{r}$ *and characteristics* $\mathbf{x}_j$ *is given by* $\mathrm{Cov}(\mathbf{r}, \mathbf{x}_j) = \rho_j \sigma_{\mathbf{r}} \sigma_{\mathbf{x}_j} \mathbf{I}$ *with* $\sigma_{\mathbf{r}}$ *the cross-sectional standard deviation of returns,* $\sigma_{\mathbf{x}_j}$ *the cross-sectional standard deviation of the jth characteristic, and* $\mathbf{I}$ *the identity matrix.*

Assumption 3(1) asserts that multiple asset characteristics are independent of each other, which makes it possible to decompose the contribution from each constraint in mathematically intuitive forms (see Propositions 3 and 4). More generally, when characteristics are dependent, a similar decomposition is also possible with additional interaction terms between different constraints, and we provide this generalization in Online Appendix B.1.

Assumption 3(2) simplifies the dependence between returns and characteristics, still allowing the returns, $\mathbf{r}$, to be arbitrarily dependent cross-sectionally. This assumption is first used in Lo and MacKinlay (1990) to describe cross-sectional estimation errors of intercepts in capital asset pricing model regressions, and later in Lo and Zhang (2024) to describe the dependence structure between returns and an impact factor, such as ESG. Online Appendix B.1 provides a generalization that relaxes this assumption as well.

Given Assumption 3, the covariance matrix of $[\mathbf{r}' \; \mathbf{x}_1' \cdots \mathbf{x}_J']$ can be written as

$$\begin{pmatrix} \boldsymbol{\Sigma} & \rho_1 \sigma_{\mathbf{r}} \sigma_{\mathbf{x}_1} \mathbf{I} \cdots \rho_J \sigma_{\mathbf{r}} \sigma_{\mathbf{x}_J} \mathbf{I} \\ \rho_1 \sigma_{\mathbf{r}} \sigma_{\mathbf{x}_1} \mathbf{I} & \\ \vdots & \begin{pmatrix} \sigma_{\mathbf{x}_1}^2 \mathbf{I} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_{\mathbf{x}_J}^2 \mathbf{I} \end{pmatrix} \\ \rho_J \sigma_{\mathbf{r}} \sigma_{\mathbf{x}_J} \mathbf{I} & \end{pmatrix}. \tag{11}$$

The following result characterizes the excess return and covariance because of the information in $\mathbf{X}$.

Proposition 3 (Information Decomposition). *Under Assumptions 1 and 3,*

- *if Assumption 2 holds,* $\mathbf{r}|\mathbf{X}$ *is normally distributed with an expected value given by*

$$\boldsymbol{\mu}_{\mathbf{X}} = \mathbb{E}[\mathbf{r}|\mathbf{X}] = \boldsymbol{\mu} + \sum_{j=1}^{J} \rho_j \sigma_{\mathbf{r}} \frac{(\mathbf{x}_j - \bar{\mathbf{x}}_j)}{\sigma_{\mathbf{x}_j}}, \tag{12}$$

and a covariance matrix given by

$$\boldsymbol{\Sigma}_{\mathbf{X}} = \mathrm{Cov}(\mathbf{r}|\mathbf{X}) = \boldsymbol{\Sigma} - \sum_{j=1}^{J} \rho_j^2 \sigma_{\mathbf{r}}^2 \mathbf{I}; \tag{13}$$

- *if Assumption 2′ holds,* $\mathbf{r}|\mathbf{X}$ *follows a multivariate Student t distribution:*

$$\mathrm{MVT}(\boldsymbol{\mu} + \mathbf{V}_{\mathbf{r},\mathbf{X}} \mathbf{V}_{\mathbf{X},\mathbf{X}}^{-1} (\mathbf{X} - \bar{\mathbf{X}}),\ s(\mathbf{V} - \mathbf{V}_{\mathbf{r},\mathbf{X}} \mathbf{V}_{\mathbf{X},\mathbf{X}}^{-1} \mathbf{V}_{\mathbf{X},\mathbf{r}}),\ \nu + NJ), \tag{14}$$

where $s = \frac{\nu + (\mathbf{X}-\bar{\mathbf{X}})'\mathbf{V}_{\mathbf{X},\mathbf{X}}^{-1}(\mathbf{X}-\bar{\mathbf{X}})}{\nu + NJ}$ *is a scaling parameter that approaches one as the number of assets, N, increases without bound. In particular, its expected value is given by*

$$\boldsymbol{\mu}_{\mathbf{X}} = \mathbb{E}[\mathbf{r}|\mathbf{X}] = \boldsymbol{\mu} + \sum_{j=1}^{J} \rho_j \sigma_{\mathbf{r}} \frac{(\mathbf{x}_j - \bar{\mathbf{x}}_j)}{\sigma_{\mathbf{x}_j}}, \tag{15}$$

for all N, and its covariance matrix is given by

$$\boldsymbol{\Sigma}_{\mathbf{X}} = \mathrm{Cov}(\mathbf{r}|\mathbf{X}) \overset{p}{=} (1 - 2/\nu)\left(\boldsymbol{\Sigma} - \sum_{j=1}^{J} \rho_j^2 \sigma_{\mathbf{r}}^2 \mathbf{I}\right), \tag{16}$$

where $\overset{p}{=}$ *denotes equality in probability as N increases without bound.*

Both the normal distribution and the multivariate Student t distribution are special cases of the elliptical distribution family (Fang et al. 1990). Therefore, it is not surprising that their conditional moments conform to similar analytical forms.

Proposition 3 allows for more explicit decompositions of the expected return and utility of the constrained portfolio by substituting (12)–(16) into Proposition 2, which we summarize in Proposition 4.

Proposition 4 (Attribution with Normal and MVT Returns). *Under Assumptions 1, 2 (or 2′), and 3 and conditioned on information in* $\mathbf{X}$ *that is used to form constraints* $\mathbf{A}(\mathbf{X})$*, the following decompositions hold for the optimal portfolio,* $\boldsymbol{\omega}^*$*:*

1. *Expected return decomposition:*

$$\mathbb{E}[\boldsymbol{\omega}^{*\prime}\mathbf{r}|\mathbf{X}] = \boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}^* = \boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}_{\mathrm{MVO}} + \boldsymbol{\mu}'\boldsymbol{\omega}_{\mathrm{CSTR}} + \sum_{j=1}^{J} \rho_j \sigma_{\mathbf{r}} \frac{(\mathbf{x}_j' - \bar{\mathbf{x}}_j')\boldsymbol{\omega}_{\mathrm{CSTR}}}{\sigma_{\mathbf{x}_j}}, \tag{17}$$

where the Lagrange multipliers are the same as in Proposition 2 provided that the feasible region of the constrained optimization problem is nonempty.

- $\boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}_{\mathrm{MVO}}$: *expected return of the unconstrained MVO portfolio.*
- $\boldsymbol{\mu}'\boldsymbol{\omega}_{\mathrm{CSTR}}$: *components attributable to each constraint treated as static.*
- $\rho_j \sigma_{\mathbf{r}} \frac{(\mathbf{x}_j' - \bar{\mathbf{x}}'_j)\boldsymbol{\omega}_{\mathrm{CSTR}}}{\sigma_{\mathbf{x}_j}}$: *component attributable to information in the jth constraint.*

2. *Expected utility decomposition:*

$$\begin{aligned}\boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}^* &- \frac{\gamma}{2}\boldsymbol{\omega}^{*\prime}\boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{\omega}^* \\ &= \boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}_{\mathrm{MVO}} - \frac{\gamma}{2}\boldsymbol{\omega}_{\mathrm{MVO}}'\boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{\omega}_{\mathrm{MVO}} - \frac{\gamma}{2}\boldsymbol{\omega}_{\mathrm{CSTR}}'\boldsymbol{\Sigma}\boldsymbol{\omega}_{\mathrm{CSTR}} \\ &+ \sum_{j=1}^{J}\left(\rho_j \sigma_{\mathbf{r}} \frac{(\mathbf{x}_j' - \bar{\mathbf{x}}_j')\boldsymbol{\omega}_{\mathrm{CSTR}}}{\sigma_{\mathbf{x}_j}} + \gamma \rho_j^2 \sigma_{\mathbf{r}}^2 \boldsymbol{\omega}_{\mathrm{SHR}}'\boldsymbol{\omega}_{\mathrm{CSTR}}\right).\end{aligned} \tag{18}$$

- $\boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}_{\mathrm{MVO}} - \frac{\gamma}{2}\boldsymbol{\omega}_{\mathrm{MVO}}'\boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{\omega}_{\mathrm{MVO}}$: *optimal expected utility of the unconstrained MVO portfolio.*
- $-\frac{\gamma}{2}\boldsymbol{\omega}_{\mathrm{CSTR}}'\boldsymbol{\Sigma}\boldsymbol{\omega}_{\mathrm{CSTR}}$: *components attributable to all constraints combined together, treated as static.*
- $\rho_j \sigma_{\mathbf{r}} \frac{(\mathbf{x}_j' - \bar{\mathbf{x}}_j')\boldsymbol{\omega}_{\mathrm{CSTR}}}{\sigma_{\mathbf{x}_j}} + \gamma \rho_j^2 \sigma_{\mathbf{r}}^2 \boldsymbol{\omega}_{\mathrm{SHR}}'\boldsymbol{\omega}_{\mathrm{CSTR}}$: *component attributable to information in the jth constraint.*

When Assumption 2′ is true, (18) holds when N and ν increase without bound.[9]

The last term of (17) (see also (12)) shows that the excess expected return, $\boldsymbol{\mu}_{\mathbf{X}} - \boldsymbol{\mu}$, is linear in $\mathbf{x}$. More importantly, it is determined by three terms: the first term, ρ_j, determines the correlation between asset characteristics and returns; the second term, $\sigma_{\mathbf{r}}$, measures the standard deviation of returns; and the third term, $(\mathbf{x}_j - \bar{\mathbf{x}}_j)/\sigma_{\mathbf{x}_j}$, determines whether each asset's characteristic value is above or below the average characteristic value, much as does a z-score. When the asset characteristics are positively correlated with returns, those assets with above-average characteristic values have positive excess returns. When the asset characteristics are negatively correlated with returns, those assets with below-average characteristic values have positive excess returns.

Proposition 4 also highlights a connection between the information component of our decomposition and traditional regression-based performance attribution (Fama 1972, Sharpe 1992). The last term of (17) can be written equivalently as the product of two terms. The first term, $\rho_j \sigma_{\mathbf{r}}/\sigma_{\mathbf{x}_j}$, is the beta of the asset returns regressed on each of the jth characteristics, which is a standard procedure in classic style analysis and performance attribution. The second term, $(\mathbf{x}_j' - \bar{\mathbf{x}}_j')\boldsymbol{\omega}_{\mathrm{CSTR}}$, is the average loading on the jth characteristic of the portfolio holdings directly attributable to constraints. Our results show that an interaction effect between the two determines the information contribution from constraints.

The last term of (18) (see also (13)) shows that the excess covariance, $(\boldsymbol{\Sigma}_{\mathbf{X}} - \boldsymbol{\Sigma})$, is always negative. In other words, when $\boldsymbol{\omega}_{\mathrm{SHR}}'\boldsymbol{\omega}_{\mathrm{CSTR}} > 0$, that is, the shrinkage portfolio holdings and the holdings attributable to constraints are positively correlated, incorporating information in $\mathbf{X}$ always reduces the variance of a portfolio. In addition, the magnitude of reduction in variance because of the jth constraint depends on ρ_j^2, the squared correlation between the jth asset characteristics and returns. The constraints with a higher magnitude of correlation lead to larger reductions in variance. On the other hand, when $\boldsymbol{\omega}_{\mathrm{SHR}}'\boldsymbol{\omega}_{\mathrm{CSTR}} < 0$, incorporating information in $\mathbf{X}$ increases the variance and reduces the expected utility of a portfolio. These two scenarios correspond to the two cases in Figure 1, respectively.

3. Accounting for Estimation Risk via Bayesian Portfolio Analysis

Our framework so far relies on knowing the population value of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ but does not account for the estimation risk when these parameters are estimated using a sample. In this section, we apply Bayesian portfolio analysis to accommodate estimation risk and analyze out-of-sample returns. We emphasize that we present our main framework in the previous section in the population model because it highlights the fundamental role of information in constraints in determining the performance of a portfolio. This holds true even if investors know the true $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ without any estimation risk. The goal of this section is not to propose a new robust portfolio rule or a better way to deal with estimation risk, but rather to generalize our framework to out-of-sample returns of existing robust portfolio rules.

3.1. Bayesian Portfolio Analysis

Bayesian portfolio analysis is a classic and natural framework to account for estimation risk (Avramov and Zhou 2010, Fabozzi et al. 2010, Jacquier and Polson 2011). We follow the classic literature to assume that returns $\mathbf{r}_t$, $t = 1, 2, \ldots$ are independently and identically distributed (IID) over time with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Investors at time T have observed returns $\Phi_T = \{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_T\}$, and the predictive density of the next return is

$$\begin{aligned}\mathbb{P}(\mathbf{r}_{T+1} \mid \Phi_T) &= \int_{\boldsymbol{\mu}}\int_{\boldsymbol{\Sigma}} \mathbb{P}(\mathbf{r}_{T+1}, \boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \Phi_T)\, d\boldsymbol{\mu}\, d\boldsymbol{\Sigma} \\ &= \int_{\boldsymbol{\mu}}\int_{\boldsymbol{\Sigma}} \mathbb{P}(\mathbf{r}_{T+1} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})\mathbb{P}(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \Phi_T)\, d\boldsymbol{\mu}\, d\boldsymbol{\Sigma}, \quad (19)\end{aligned}$$

where $\mathbb{P}(\mathbf{r}_{T+1} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$ describes the data-generating process, that is, the distribution of returns conditioned on known parameters, and $\mathbb{P}(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \Phi_T)$ is the posterior density of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. Given an appropriate prior, $\mathbb{P}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the posterior density is $\mathbb{P}(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \Phi_T) \propto \mathbb{P}(\Phi_T \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) \times \mathbb{P}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Unlike the conditional distribution, the Bayesian predictive distribution (19) accounts for estimation error by integrating over the unknown parameter space. The optimal portfolio is then constructed by maximizing the expected utility with respect to the predictive distribution. For notational convenience, we denote the $(N \times 1)$-vector, $\tilde{\mathbf{r}} \equiv \mathbf{r}_{T+1} \mid \Phi_T$, as the predictive return vector that follows (19), and $\tilde{\boldsymbol{\mu}} \equiv \mathbb{E}[\tilde{\mathbf{r}}]$ and $\tilde{\boldsymbol{\Sigma}} \equiv \mathrm{Cov}(\tilde{\mathbf{r}})$ are the expected value and covariance matrix of $\tilde{\mathbf{r}}$, respectively. By convention, we always consider predictive returns conditioned on the previous T returns in the Bayesian framework, and we omit the time subscript $T+1$ in $\tilde{\mathbf{r}}$ and its moments for notational convenience without causing any confusion.

In the case of mean-variance utility, the optimal portfolio with respect to the predictive distribution solves the optimization problem in (1) with $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ replaced by $\tilde{\boldsymbol{\mu}}$ and $\tilde{\boldsymbol{\Sigma}}$:

$$\begin{aligned}\max_{\boldsymbol{\omega}}\ & \boldsymbol{\omega}'\tilde{\boldsymbol{\mu}} - \frac{\gamma}{2}\boldsymbol{\omega}'\tilde{\boldsymbol{\Sigma}}\boldsymbol{\omega} \\ \text{s.t.}\ & \mathbf{A}(\mathbf{X}_T)\boldsymbol{\omega} = \mathbf{b},\end{aligned} \quad (20)$$

where $\mathbf{X}_T$ is the asset characteristics at time T. We use $\tilde{\boldsymbol{\omega}}^*$ to denote the weights for the optimal Bayesian portfolio by solving the optimization problem (20).

3.2. Attribution with Information Using Predictive Distribution

We assume that the asset characteristics, $\mathbf{X}_t$, $t = 1, 2, \ldots$, are identically distributed over time, and $\mathbf{X}_t$ may be correlated with $\mathbf{r}_{t+1}$ in general. Therefore, at time T, $\mathbf{X}_T$ is not independent of the predictive returns, $\tilde{\mathbf{r}}$, that follow (19). Following our notations so far and again omitting the time subscript T in asset characteristics without causing any confusion, we use $\tilde{\mathbf{r}} \mid \mathbf{X}$ to denote the predictive returns conditioned on additional information in $\mathbf{X}_T$ and $\tilde{\boldsymbol{\mu}}_{\mathbf{X}} \equiv \mathbb{E}[\tilde{\mathbf{r}} \mid \mathbf{X}]$ and $\tilde{\boldsymbol{\Sigma}}_{\mathbf{X}} \equiv \mathrm{Cov}[\tilde{\mathbf{r}} \mid \mathbf{X}]$ to denote the conditional mean and covariance matrix.

With these notations, the performance attribution in Proposition 2 can be generalized to accommodate the Bayesian predictive distribution by replacing the population moments ($\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, $\boldsymbol{\mu}_{\mathbf{X}}$, $\boldsymbol{\Sigma}_{\mathbf{X}}$) with the corresponding moments from the predictive distribution ($\tilde{\boldsymbol{\mu}}$, $\tilde{\boldsymbol{\Sigma}}$, $\tilde{\boldsymbol{\mu}}_{\mathbf{X}}$, $\tilde{\boldsymbol{\Sigma}}_{\mathbf{X}}$). We provide the formal statement of this result in Proposition B.3 in Online Appendix B.2.1.

We emphasize that so far we have not made any assumptions about the specific form of the prior $\mathbb{P}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ nor the distribution of returns conditioned on known parameters $\mathbb{P}(\mathbf{r}_{T+1} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$. Therefore, the decomposition in Online Proposition B.3 holds for any Bayesian portfolios constructed using the predictive distribution. In general, the predictive distribution in (19) and its moments, $\tilde{\boldsymbol{\mu}}$, $\tilde{\boldsymbol{\Sigma}}$, $\tilde{\boldsymbol{\mu}}_{\mathbf{X}}$, and $\tilde{\boldsymbol{\Sigma}}_{\mathbf{X}}$, can be obtained via Monte Carlo simulation in the Bayesian statistics literature.[10] Therefore, as long as a portfolio rule can be expressed as a Bayesian portfolio, our framework can be applied to derive a decomposition of the performance into the different components outlined in Propositions 2 and B.3, at least numerically.

For certain Bayesian portfolios in the literature with specific distributions of the prior, we can use our results in Section 2.3 to derive more explicit decompositions, which we discuss next. We follow the majority of this literature to assume that the return distribution conditioned on $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ is normal to allow for tractability of the posterior predictive distribution.

3.2.1. Normal Predictive Returns with Unknown $\boldsymbol{\mu}$ and Known $\boldsymbol{\Sigma}$. It is known that the expected return vector $\boldsymbol{\mu}$ is much more challenging to estimate than the covariance matrix $\boldsymbol{\Sigma}$, partially because using higher sampling frequencies improves the accuracy of the covariance estimator but not the mean estimator (Merton 1980). Several studies document that the estimation risk is mainly due to errors in the estimated expected return (Broadie 1993, Chopra and Ziemba 1993, Ceria and Stubbs 2006), which suggests that uncertainty in the mean is the first order effect to address in portfolio optimization (Jacquier and Polson 2011).

Therefore, we first consider uncertainties in $\boldsymbol{\mu}$ and treat $\boldsymbol{\Sigma}$ as known. In this case, the posterior predictive distribution is usually normal. Proposition B.2 in Online Appendix B.2.2 derives an explicit expression of the information component for predictive returns, which is similar to Proposition 3 except that the population moments ($\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, $\boldsymbol{\mu}_{\mathbf{X}}$, $\boldsymbol{\Sigma}_{\mathbf{X}}$) are replaced by the predictive moments ($\tilde{\boldsymbol{\mu}}$, $\tilde{\boldsymbol{\Sigma}}$, $\tilde{\boldsymbol{\mu}}_{\mathbf{X}}$, $\tilde{\boldsymbol{\Sigma}}_{\mathbf{X}}$).

This result can be applied to several classic models such as Klein and Bawa (1976), Jorion (1986), and Black and Litterman (1992), and Online Appendix B.2.2 provides more details. In each model, they adopt different priors that lead to normal predictive distributions with different parameters. In our framework, these portfolios without constraints correspond to different benchmark portfolios $\tilde{\boldsymbol{\omega}}_{\text{MVO}}$.

3.2.2. Student *t* Predictive Returns with Unknown $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

When $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are both unknown and modeled as random parameters in the Bayesian framework, the predictive density of returns is typically a multivariate Student *t* distribution, $\tilde{\mathbf{r}} \sim \text{MVT}(\tilde{\boldsymbol{\mu}}, \tilde{\mathbf{V}}, \nu)$.

Similar to the case of normal predictive returns, the result of Proposition B.4 in Online Appendix B.2.2 can be applied to a broader class of Bayesian portfolios and shrinkage portfolios with certain equivalence to Bayesian portfolios, including the noninformative diffuse prior $\mathbb{P}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-(N+1)/2}$ as in Klein and Bawa (1976) and Brown (1978) and the normal–inverse–Wishart conjugate prior for $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ as in Frost and Savarino (1986), Stambaugh (1997), Pástor (2000), Pástor and Stambaugh (2000), Zhou (2009), and Lai et al. (2011). In these examples, the predictive distribution is MVT with different parameters.

In addition, Tu and Zhou (2010) establish a certain equivalence between priors on portfolio weights, $\boldsymbol{\omega}$, and priors on $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. DeMiguel et al. (2009b) show that certain norm-constrained portfolios can be equivalently interpreted as Bayesian portfolios with prior beliefs on portfolio weights. The insights of these two seminal articles provide an important link between the literature on robust portfolios with priors on portfolio weights[11] and the literature on Bayesian portfolio analysis. This implies that our attribution framework can, in principle, be applied to robust portfolio rules that directly impose priors or shrinkage on portfolio weights through their Bayesian equivalents. However, deriving the specific Bayesian formulation for each of them is not the focus and beyond the scope of this article. Online Appendix B.2.2 discusses this literature in more detail.

3.3. Interpreting Constraints as Bayesian Priors

In this section, we formally establish that, in the context of Bayesian portfolios, solving a portfolio with constraints can be equivalently interpreted as solving an unconstrained portfolio with a certain prior on $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

In particular, following Section 3.2, we consider normally distributed returns conditioned on $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$: $\mathbf{r}|(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and two priors, π_1 and π_2, on the unknown parameters, which lead to posterior predictive returns with moments $(\tilde{\boldsymbol{\mu}}_1, \tilde{\boldsymbol{\Sigma}}_1)$ and $(\tilde{\boldsymbol{\mu}}_2, \tilde{\boldsymbol{\Sigma}}_2)$, respectively. We consider a constrained Bayesian portfolio (I) with prior π_1 that solves the problem in (20):

$$\tilde{\boldsymbol{\omega}}_1^* = \frac{1}{\gamma}\tilde{\boldsymbol{\Sigma}}_1^{-1}(\tilde{\boldsymbol{\mu}}_1 - \mathbf{A}'\tilde{\boldsymbol{\lambda}}) \quad \text{where}$$
$$\tilde{\boldsymbol{\lambda}} = (\mathbf{A}\tilde{\boldsymbol{\Sigma}}_1^{-1}\mathbf{A}')^{-1}(\mathbf{A}\tilde{\boldsymbol{\Sigma}}_1^{-1}\tilde{\boldsymbol{\mu}}_1 - \gamma\mathbf{b}), \tag{21}$$

and an unconstrained Bayesian portfolio (II) with prior π_2 that maximizes the same predictive mean-variance utility in (20) but without any constraint:

$$\tilde{\boldsymbol{\omega}}_2^* = \frac{1}{\gamma}\tilde{\boldsymbol{\Sigma}}_2^{-1}\tilde{\boldsymbol{\mu}}_2. \tag{22}$$

The following result establishes the equivalence between the two portfolios.

Proposition 5 (Constraints as Priors). *With portfolios I and II in (21) and (22), respectively,*

1. *When $\boldsymbol{\mu}$ is unknown and $\boldsymbol{\Sigma}$ is known, consider two priors on $\boldsymbol{\mu}$:*

$$\pi_1 : \boldsymbol{\mu} \sim N\left(\boldsymbol{\mu}_1, \frac{\boldsymbol{\Sigma}}{\tau}\right),$$
$$\pi_2 : \boldsymbol{\mu} \sim N\left(\boldsymbol{\mu}_2, \frac{\boldsymbol{\Sigma}}{\tau}\right),$$

where $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are prior means, and τ is a precision hyperparameter. Portfolios I and II are identical ($\tilde{\boldsymbol{\omega}}_1^* = \tilde{\boldsymbol{\omega}}_2^*$) if the following condition holds:

$$\boldsymbol{\mu}_2 = \boldsymbol{\mu}_1 - \mathbf{A}'\tilde{\boldsymbol{\lambda}}_{1,\text{PR}} - \frac{T}{\tau}\mathbf{A}'\tilde{\boldsymbol{\lambda}}_{1,\text{DT}}, \tag{23}$$

where

$$\tilde{\boldsymbol{\lambda}}_{1,\text{PR}} \equiv (\mathbf{A}\boldsymbol{\Sigma}^{-1}\mathbf{A}')^{-1}(\mathbf{A}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1 - (1 + 1/\tau)\gamma\mathbf{b}),$$
$$\tilde{\boldsymbol{\lambda}}_{1,\text{DT}} \equiv (\mathbf{A}\boldsymbol{\Sigma}^{-1}\mathbf{A}')^{-1}(\mathbf{A}\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\mu}} - \gamma\mathbf{b})$$

are Lagrange multipliers of two hypothetical constrained portfolios (20) with predictive distributions

$$\tilde{\mathbf{r}} \sim N\left(\boldsymbol{\mu}_1, \left(1 + \frac{1}{\tau}\right)\boldsymbol{\Sigma}\right) \quad \text{and} \quad \tilde{\mathbf{r}} \sim N(\hat{\boldsymbol{\mu}}, \boldsymbol{\Sigma}),$$

respectively. Here $\hat{\boldsymbol{\mu}}$ is the sample mean of the returns.

2. *When $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are both unknown, consider two priors on $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$:*

$$\pi_1 : \boldsymbol{\mu}|\boldsymbol{\Sigma} \sim N\left(\boldsymbol{\mu}_1, \frac{\boldsymbol{\Sigma}}{\tau}\right), \quad \boldsymbol{\Sigma} \sim IW(\boldsymbol{\Sigma}_1, v_0),$$
$$\pi_2 : \boldsymbol{\mu}|\boldsymbol{\Sigma} \sim N\left(\boldsymbol{\mu}_2, \frac{\boldsymbol{\Sigma}}{\tau}\right), \quad \boldsymbol{\Sigma} \sim IW(\boldsymbol{\Sigma}_2, v_0),$$

where $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are prior means, $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are prior covariance matrices, τ and v_0 are hyperparameters, and IW *stands for the inversed-Wishart distribution. Portfolios I and II are identical ($\tilde{\boldsymbol{\omega}}_1^* = \tilde{\boldsymbol{\omega}}_2^*$) if the following conditions hold:*

$$\boldsymbol{\mu}_2 = \boldsymbol{\mu}_1 - \mathbf{A}'\tilde{\boldsymbol{\lambda}}_{2,\mathrm{PR}} - \frac{T}{\tau}\mathbf{A}'\tilde{\boldsymbol{\lambda}}_{2,\mathrm{DT}}, \tag{24}$$

$$\boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_1 + \frac{T\tau}{T+\tau}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2 - 2\hat{\boldsymbol{\mu}})', \tag{25}$$

where

$$\tilde{\boldsymbol{\lambda}}_{2,\mathrm{PR}} \equiv (\mathbf{A}\tilde{\boldsymbol{\Sigma}}_1^{-1}\mathbf{A}')^{-1}(\mathbf{A}\tilde{\boldsymbol{\Sigma}}_1^{-1}\boldsymbol{\mu}_1 - \gamma\mathbf{b}),$$
$$\tilde{\boldsymbol{\lambda}}_{2,\mathrm{DT}} \equiv (\mathbf{A}\tilde{\boldsymbol{\Sigma}}_1^{-1}\mathbf{A}')^{-1}(\mathbf{A}\tilde{\boldsymbol{\Sigma}}_1^{-1}\hat{\boldsymbol{\mu}} - \gamma\mathbf{b})$$

are Lagrange multipliers of two hypothetical constrained portfolios (20) with predictive distributions

$$\tilde{\mathbf{r}} \sim MVT\left(\boldsymbol{\mu}_1, \frac{v_0 + T - N - 1}{v_0 + T - N + 1}\tilde{\boldsymbol{\Sigma}}_1, v_0 + T - N + 1\right),$$
$$\tilde{\mathbf{r}} \sim MVT\left(\hat{\boldsymbol{\mu}}, \frac{v_0 + T - N - 1}{v_0 + T - N + 1}\tilde{\boldsymbol{\Sigma}}_1, v_0 + T - N + 1\right),$$

respectively, and

$$\tilde{\boldsymbol{\Sigma}}_1 \equiv \left(1 + \frac{1}{T+\tau}\right)\left(\frac{1}{v_0 + T - N - 1}\right)\left(\boldsymbol{\Sigma}_1 + T\hat{\boldsymbol{\Sigma}} + \frac{T\tau}{T+\tau}(\boldsymbol{\mu}_1 - \hat{\boldsymbol{\mu}})(\boldsymbol{\mu}_1 - \hat{\boldsymbol{\mu}})'\right)$$

is the predictive covariance matrix of returns under prior π_1.

Proposition 5 shows that Bayesian portfolio I with constraints is equivalent to Bayesian portfolio II with a different prior and without constraints. The relationships between their priors are characterized by (23) when $\boldsymbol{\Sigma}$ is known and (24) and (25) when $\boldsymbol{\Sigma}$ is unknown. In both cases, imposing a constraint is equivalent to moving the prior mean on $\boldsymbol{\mu}$ by two terms, both of which yield clear financial intuitions. The first term reflects the effect of prior π_1 (see $\tilde{\boldsymbol{\lambda}}_{1,\mathrm{PR}}$ and $\tilde{\boldsymbol{\lambda}}_{2,\mathrm{PR}}$), and the second term reflects the effect of the observed data (see $\tilde{\boldsymbol{\lambda}}_{1,\mathrm{DT}}$ and $\tilde{\boldsymbol{\lambda}}_{2,\mathrm{DT}}$).[12]

The connection between portfolio constraints and investor views is closely related to the seminal work of Tu and Zhou (2010), DeMiguel et al. (2009b), and Ardia and Boudt (2015). Tu and Zhou (2010) establish the equivalence between (objective-based) priors on $\boldsymbol{\omega}$ and priors on $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ but do not explicitly consider the effect of portfolio constraints. DeMiguel et al. (2009b) consider minimum variance portfolios and show that certain norm constraints on $\boldsymbol{\omega}$ are equivalent to Bayesian portfolios with a prior on $\boldsymbol{\omega}$. We consider mean-variance portfolios and priors on $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, and our constraints are different from theirs. Ardia and Boudt (2015) point out the connection between portfolio constraints and investor views but do not explore the equivalence in the context of Bayesian portfolios.

3.4. Summary

We summarize the procedure of applying our framework to Bayesian portfolios that account for estimation risk. This allows for performance attribution and constraint selection based on out-of-sample returns.

- Step 1: Pick a Bayesian method to derive the predictive distribution of asset returns that accounts for estimation risk. Different methods correspond to different benchmark portfolios $\tilde{\boldsymbol{\omega}}_{\mathrm{MVO}}$.
- Step 2: Performance attribution using Online Proposition B.3 either directly with Monte Carlo methods or using the analytical expressions given in Online Proposition B.4. For the latter, one estimates the correlation, ρ_j, between the characteristics of the jth constraint, $\mathbf{x}_j$, and the predictive returns, $\tilde{\mathbf{r}}$, using a rolling window of panel data in the past T periods. This is a low-dimensional problem because there is only one parameter to estimate for each constraint, ρ_j, which can be achieved with a fairly high degree of accuracy (Lo and Zhang 2024, Lo et al. 2024).
- Step 3: The results can be used either for purposes of performance attribution and disclosure or to choose the constraint that provides the highest contribution to performance. Because portfolio constraints can be interpreted as Bayesian priors as shown in Section 3.3, the attribution result also reflects the influence of investors' views on returns (expressed as Bayesian priors) on the portfolio's performance.

Different Bayesian portfolios correspond to different predictive distributions of returns and different benchmark portfolios. Whereas our framework can be applied to Bayesian portfolios, we emphasize that it should not be treated only as a way to deal with estimation risk. In fact, even without estimation error, the same decomposition holds. Constraints can serve as a mechanism for incorporating information in both population and out-of-sample returns. Our methodology provides a way to quantify this effect.

We consider two common types of portfolio constraints and demonstrate how to select constraints that improve portfolio performance using simulations in Section 4. Our empirical analysis in Section 5 applies the framework to portfolios that account for estimation risk using two robust portfolio rules: Jorion's (1986) rule and the $1/N$ rule, which is shown to be equivalent to Bayesian portfolios with a particular prior (DeMiguel et al. 2009a, b).

4. Common Examples of Portfolio Constraints

In this section, we consider two common examples of portfolio constraints: factor exposures and exclusions. We derive additional analytical results for these constraints and conduct simulations to illustrate the attribution of expected returns and utility. All results in this section can be derived from either the unconditional $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ or the posterior predictive moments, $\tilde{\boldsymbol{\mu}}$ and $\tilde{\boldsymbol{\Sigma}}$. The latter can be used to select constraints ex ante.

4.1. Factor Exposure

A common constraint in portfolio construction arises when investors wish to control the average value of a characteristic or the exposure to a certain factor, such as the average ESG score, market capitalization, beta, or book-to-market values of the portfolio.

We consider the case of a single constraint, $\mathbf{A}(\mathbf{x}) = \mathbf{x}'$, for simplicity. In this case, the Lagrange multiplier is a scalar, which leads to the following result.

Proposition 6 (Factor Exposure). *Under Assumptions 1, 2 (or 2′), and 3, and assuming without loss of generality that the cross-sectional average factor value $\bar{\mathbf{X}} = 0$, the expected return of the optimal portfolio with a factor exposure constraint can be decomposed into*

$$\mathbb{E}[\boldsymbol{\omega}^{*\prime}\mathbf{r}|\mathbf{x}] = \frac{1}{\gamma}\boldsymbol{\mu}_X'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \frac{\mathbf{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\mathbf{x}'\boldsymbol{\Sigma}^{-1}\mathbf{x}}(b - \mathbf{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}/\gamma) + \frac{\rho\sigma_{\mathbf{r}}}{\sigma_{\mathbf{x}}}(b - \mathbf{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}/\gamma), \quad (26)$$

if the constraint is binding. In the case of a nonbinding inequality constraint, the Lagrange multiplier $\lambda^* = 0$ and (26) reduces to $\mathbb{E}[\boldsymbol{\omega}^{*\prime}\mathbf{r}|\mathbf{x}] = \boldsymbol{\mu}_X'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}/\gamma$.

If the moments of returns in Proposition 6 are replaced by their counterparts from the predictive distribution, we derive the same decomposition for the expected return of the predictive distribution.

There is no information contribution when the constraint is not binding in the case of an inequality constraint. When the constraint is binding, the information component depends on the correlation, ρ, between returns and asset characteristics. In addition, $\mathbf{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ is the characteristic value of the unconstrained MVO portfolio, so $b - \mathbf{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ measures the constrained characteristic value b relative to the unconstrained MVO portfolio. If a particular asset characteristic, for example, ESG, is positively correlated with returns ($\rho > 0$), a positive desired ESG level relative to the unconstrained MVO portfolio ($b - \mathbf{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} > 0$) adds value to expected returns. However, if ESG is negatively correlated with returns ($\rho < 0$), the same constraint hurts expected returns. This simple intuition informs the selection of constraints using our framework: pick constraints with a positive correlation, ρ, with the predictive returns. We demonstrate this further with the following simulation.

Simulation. We consider a world with 10 assets in which investors adopt a certain Bayesian method to account for estimation risk and derive the predictive density of returns. The expected predictive returns are generated from a normal distribution $\tilde{\mu}_i \sim N(0.05, 0.05^2)$ for $i = 1, 2, \ldots, 10$. The covariance matrix, $\tilde{\boldsymbol{\Sigma}} = (\sigma_{i,j})_{10\times 10}$, is generated by using the algorithm of Davies and Higham (2000)—implemented by `scipy.stats.random_correlation` in Python—which generates a random correlation matrix given a set of eigenvalues. We set the 10 eigenvalues to range from 0.9 to 1.1 and then multiply the correlation matrix by 0.1 to get $\tilde{\boldsymbol{\Sigma}}$.[13]

Investors solve the following optimization problem with two constraints on factor exposures:

$$\max_{\boldsymbol{\omega}} \ \boldsymbol{\omega}'\tilde{\boldsymbol{\mu}} - \frac{\gamma}{2}\boldsymbol{\omega}'\tilde{\boldsymbol{\Sigma}}\boldsymbol{\omega} \quad \text{s.t.} \ \boldsymbol{\omega}'\mathbf{x} \geq 0.5 \ \text{ and } \ \boldsymbol{\omega}'\mathbf{y} \geq 0.5. \quad (27)$$

Here, $\mathbf{x}$ and $\mathbf{y}$ represent two characteristics such as an ESG score and return momentum. They are both 10-dimensional $N(0, 1)$ random vectors that are IID over time. We denote the correlations between asset returns and these two characteristics by $\rho_1 \equiv \text{Corr}(\mathbf{x}, \tilde{\mathbf{r}})$ and $\rho_2 \equiv \text{Corr}(\mathbf{y}, \tilde{\mathbf{r}})$.

Figure 2 shows the attribution of expected returns following Proposition 6. Figure 2(a) shows the expected return of the constrained portfolio, which ranges between 1.6 and 3.0 as ρ_1 and ρ_2 vary between –0.8 and 0.8. Figure 2(b) shows the expected return of the unconstrained MVO portfolio, which is a constant value around 2.5 regardless of the values of ρ_1 and ρ_2.

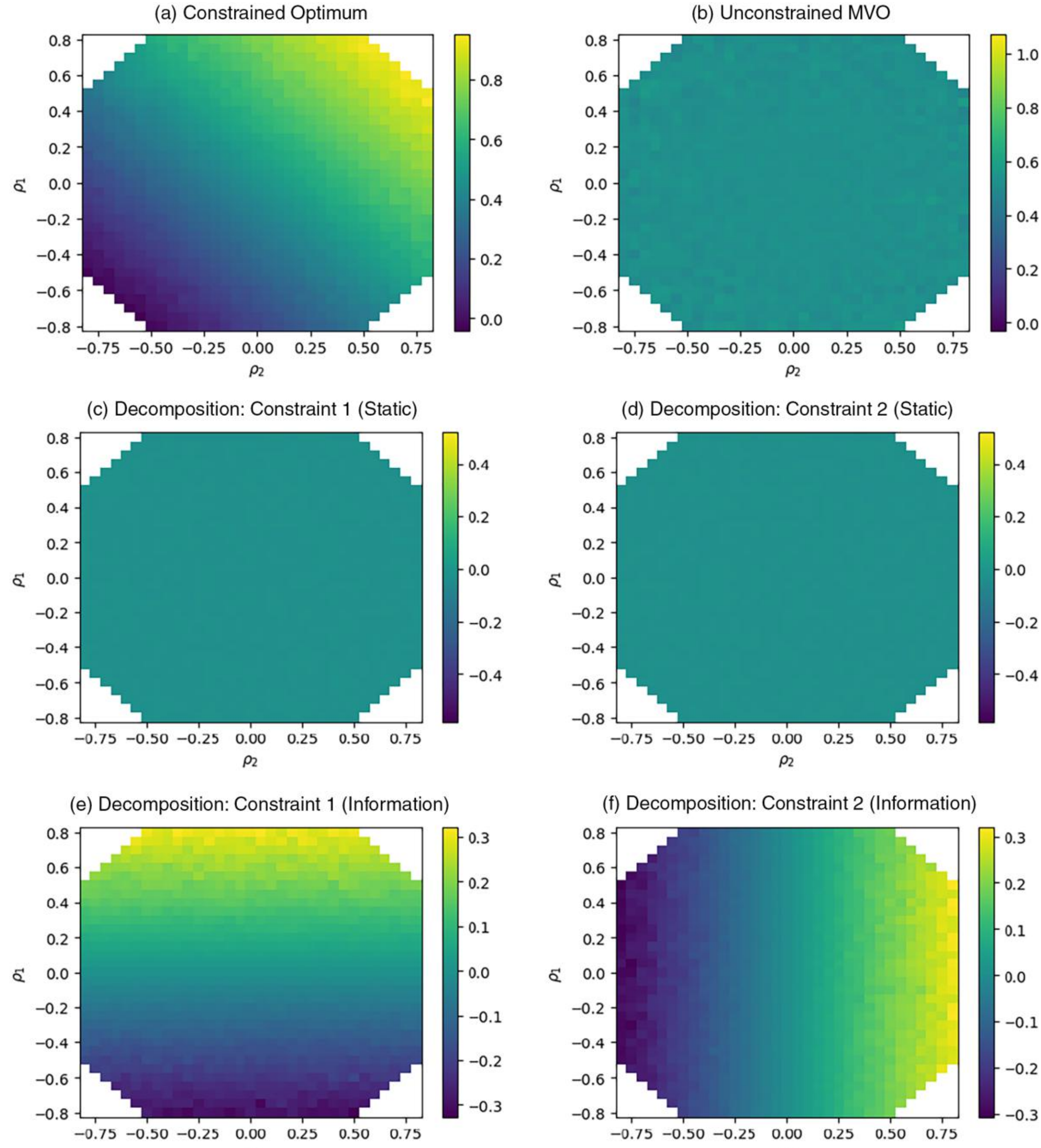
The source of the difference in expected returns between the unconstrained MVO portfolio and the constrained portfolio becomes clear in Figure 2, (c)–(f). Figure 2, (c) and (d), shows the expected returns attributable to the two constraints, respectively, as if they are static. They each contribute to the expected returns with a negative constant value of around –0.1. Figure 2(e) shows the expected returns attributable to information in the first constraint, which increase as ρ_1 increases but remains constant as ρ_2 varies. Figure 2(f) shows the expected returns attributable to information in the second constraint, which increase as ρ_2 increases but remains constant as ρ_1 varies. Similar patterns are seen in decomposing expected utility as shown in Figure B.1 in Online Appendix B.5.

Overall, these results demonstrate how to understand the expected return and utility of a constrained portfolio by decomposing them into an unconstrained MVO portfolio, static constraints, and information in each constraint. In particular, when the information in constraints is sufficiently positively correlated with returns, they can lead to higher expected returns and utilities for the constrained portfolio.

4.2. Exclusionary Investing

Another common form of constraint in portfolio construction is the exclusionary constraint, in which assets are excluded from the portfolio based on certain criteria, such as a minimum ESG score or the particular industry in which a firm operates.

Figure 2. (Color online) Decomposition of Expected Return for the Optimization Problem in (27) with Two Constraints That Depend on Random Characteristics as Correlations (ρ_1 and ρ_2) Between Random Characteristics and Asset Returns Vary



Note. The expected return of the constrained portfolio (a) is decomposed into components corresponding to the unconstrained MVO portfolio (b), static constraints (c and d), and information in the constraints (e and f).

For convenience, we introduce some new notation. Suppose that assets are excluded based on the ranking of a characteristic **x**. We use the notation $x_{1:N} < x_{2:N} < x_{N:N}$ to denote the ranked values of **x** or their order statistics. We then denote by $r_{[i:N]}$ the return associated with the ith order statistic, $x_{i:N}$. This return is called the ith induced order statistic in the literature to emphasize the fact that its ranking is determined not by its own value, but by the value of **x** (David 1973). For simplicity, we also use the subscript $[N]$ to denote a vector or a matrix that is reordered based on values of **x**. For example, $\boldsymbol{\omega}_{[N]}$ represents the vector of weights for assets that are reordered based on values of **x**.

Investors solve the following optimization problem in which the top N_0 assets ranked by **x** are allowed to enter the portfolio, whereas the bottom $N - N_0$ assets are excluded:

$$\begin{aligned}\min_{\boldsymbol{\omega}_{[N]}} \quad & \boldsymbol{\omega}'_{[N]}\boldsymbol{\mu}_{[N]} - \frac{\gamma}{2}\boldsymbol{\omega}'_{[N]}\boldsymbol{\Sigma}_{[N]}\boldsymbol{\omega}_{[N]} \\ \text{s.t.} \quad & \omega_{[i:N]} = 0 \quad \text{for } i \leq N - N_0.\end{aligned} \tag{28}$$

The optimal portfolio for (28) is simply the optimal portfolio restricted to N_0 assets. Therefore, the optimal portfolio weights are given by $\boldsymbol{\omega}^*_{N_0} = \boldsymbol{\Sigma}^{-1}_{N_0}\boldsymbol{\mu}_{N_0}$.

If $\mathbf{x}$ is a vector of continuous random variables, the expected return of the optimal portfolio with an exclusionary constraint can be decomposed by (17) in Proposition 4. If $\mathbf{x}$ is a vector of binary random variables, as is the case for exclusion based on industry or sin stock labels, the following result provides a more intuitive form of attribution for the expected return of this portfolio.

Proposition 7 (Exclusion Based on Binary Characteristic). *Under Assumptions 1, 2 (or 2′), and 3, if $\mathbf{x}$ is a vector of binary random variables and x_i follows a Bernoulli distribution, the expected return of the optimal portfolio with an exclusionary constraint in (28) can be decomposed into*

$$\begin{aligned}\mathbb{E}[\boldsymbol{\omega}^{*\prime}\mathbf{r}|\mathbf{x}] = \boldsymbol{\mu}'_{\mathbf{x},N_0}\boldsymbol{\omega}^*_{N_0} &= \boldsymbol{\mu}'_{\mathbf{x}}\boldsymbol{\omega}_{\text{MVO}} + \boldsymbol{\mu}'\boldsymbol{\omega}_{\text{CSTR}} \\ &\quad + \rho\sigma_{\mathbf{r}}(\mathbf{x}\odot\mathbf{u} - (1-\mathbf{x})\odot\mathbf{v})'\boldsymbol{\omega}_{\text{CSTR}},\end{aligned} \tag{29}$$

where

$$\mathbf{u} = \left(\sqrt{\frac{\psi_{x_1=0}}{\psi_{x_1=1}}}, \ldots, \sqrt{\frac{\psi_{x_N=0}}{\psi_{x_N=1}}}\right)' \quad \text{and}$$

$$\mathbf{v} = \left(\sqrt{\frac{\psi_{x_1=1}}{\psi_{x_1=0}}}, \ldots, \sqrt{\frac{\psi_{x_N=1}}{\psi_{x_N=0}}}\right)'$$

are two vectors of the odds ratio (that is, the relative chance) of each asset being excluded from the portfolio, $\psi_{x_i=0} \equiv \mathbb{P}(x_i = 0)$, $\psi_{x_i=1} \equiv \mathbb{P}(x_i = 1)$, and $\odot$ represents element-wise multiplication of two vectors.

If the return moments in Proposition 7 are replaced by their counterparts from the predictive distribution, we derive the same decomposition for the expected return of the predictive distribution. Proposition 7 shows that the component attributable to information depends on the correlation, ρ, between returns and characteristics $\mathbf{x}$, and the chance of being excluded from the portfolio, $\frac{\psi_{x_i=0}}{\psi_{x_i=1}}$.

Simulation. We consider a world with 10 assets, and as before, investors adopt a certain Bayesian method to account for estimation risk and derive the predictive density of returns. The expected predictive return, $\tilde{\boldsymbol{\mu}}$, and covariance matrix, $\tilde{\boldsymbol{\Sigma}}$, are generated in the same way as in Section 4.1.

Investors solve the following problem:

$$\begin{aligned}\max_{\boldsymbol{\omega}} \quad & \boldsymbol{\omega}'\tilde{\boldsymbol{\mu}} - \frac{\gamma}{2}\boldsymbol{\omega}'\tilde{\boldsymbol{\Sigma}}\boldsymbol{\omega} \\ \text{s.t.} \quad & \omega_{[i:10]} = 0, \text{ for } i \leq N - N_0 \text{ assets} \\ & \qquad \text{ordered by } \mathbf{x}.\end{aligned} \tag{30}$$

Here, $\mathbf{x}$ represents the asset characteristic that is used to exclude assets (e.g., minimum ESG score), a 10-dimensional $N(0, 1)$ random vector that is IID over time. We denote the correlations of the asset characteristics with returns by $\rho \equiv \text{Corr}(\mathbf{x}, \tilde{\mathbf{r}})$.

Figure 3 shows the attribution of expected returns following Proposition 7 with ρ varying between –0.8 and 0.8 and the number of excluded assets varying between one and nine. Figure 3(a) shows the expected return of the constrained problem, which varies from a low of –0.2 to a high of 3.0 as ρ and the number of excluded assets vary. Figure 3(b) shows the expected return of the unconstrained MVO portfolio, which has a constant value of around 2.5.

The source of the difference in expected returns between the unconstrained MVO and the constrained portfolio becomes clear in Figure 3, (c) and (d). Figure 3(c) shows the expected returns attributable to the constraints as if they are static. As more assets are excluded, the contributions from static constraints also increase. However, this component remains unchanged as the correlation, ρ, varies. Figure 3(d) shows the expected returns attributable to the information in the constraint, which increase as ρ increases. This result highlights a trade-off when a greater number of assets are excluded: when the correlation, ρ, is nonzero, excluding more assets implies that only assets with positive or negative returns are included in the portfolio, but excluding too many assets allows the portfolio too little choice in the universe of available assets. As a result, the highest returns are achieved when an intermediate number of assets are excluded, given a positive correlation ρ. Similarly, the lowest returns are achieved with an intermediate number of assets given a negative correlation. These results are consistent with Proposition 7.

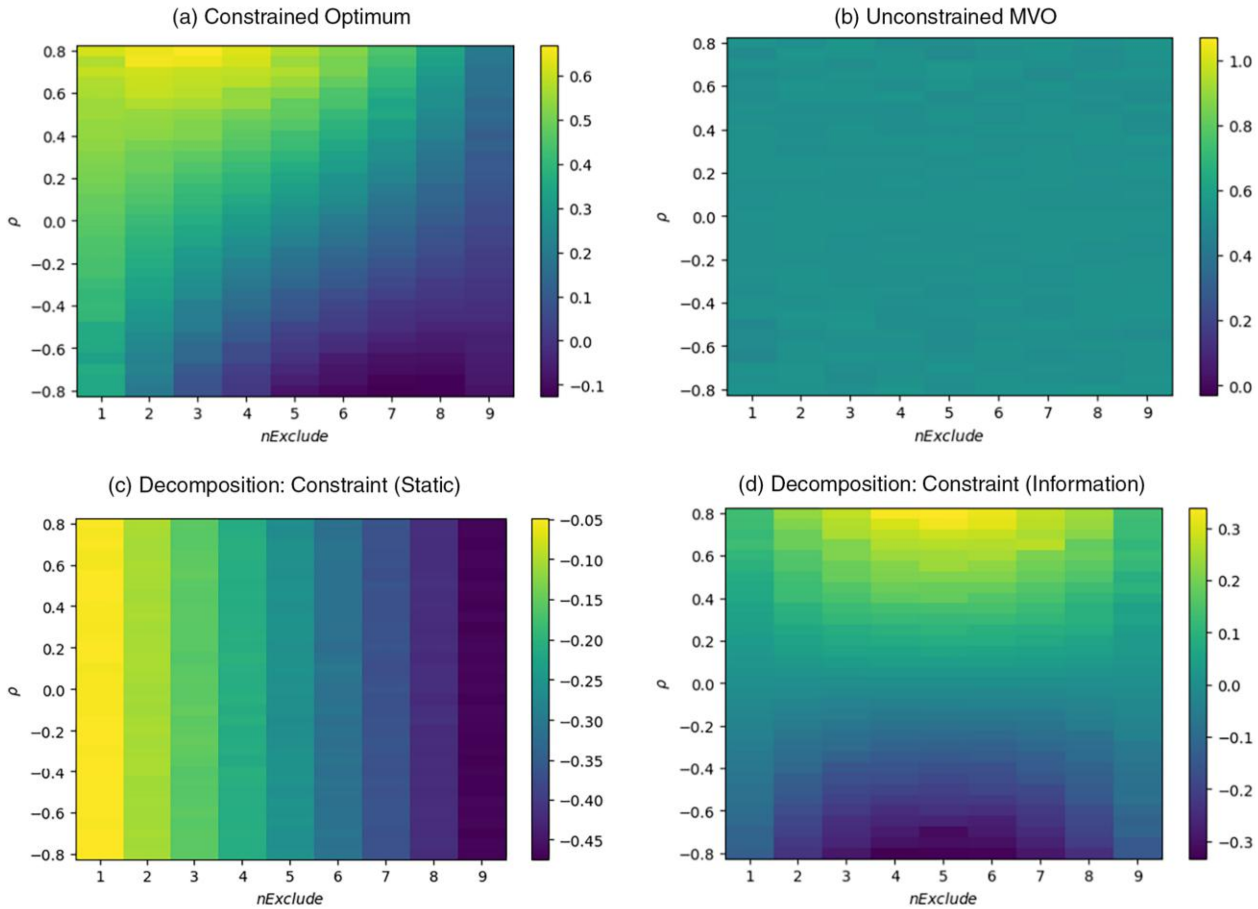
Similarly, Figure B.2 in Online Appendix B.5 demonstrates the attribution of expected utility. The patterns are similar to those in Figure 3.

Overall, these results demonstrate how to understand the expected return and utility of an exclusionary portfolio by decomposing them into an unconstrained MVO portfolio, static constraints, and information in each constraint.

4.3. Selection of Constraints

These examples highlight how investors can use our framework to select constraints that help improve performance ex ante. If investors can select one or several constraints from a collection of factor exposure constraints, they should select the binding constraint(s)

Figure 3. (Color online) Decomposition of Expected Return for the Problem in (30) with One Exclusionary Constraint That Depends on Random Characteristics as the Number of Excluded Assets (*nExclude*) and the Correlation (ρ) Between the Random Characteristic and Asset Returns Vary



Note. The expected return of the constrained portfolio (a) is decomposed into components corresponding to the unconstrained MVO portfolio (b), static constraints (c), and information in the constraints (d).

with information that is sufficiently positively correlated with returns because they can lead to higher expected returns and utilities for the constrained portfolio as demonstrated by the simulation in Section 4.1.

The analysis in Section 4.2 demonstrates that, with exclusionary investing, to decide which characteristic to use and how many assets to include, investors should select the characteristic with the highest correlation with returns and the optimal number of assets to exclude. The latter depends on the correlation of the specific characteristic and may lie between zero and N, reflecting a trade-off between information and diversification. As a result, the best performance is achieved when an intermediate number of assets is excluded.

Our empirical analysis in Section 5.4 demonstrates an example with real data in which different forms of ESG constraints lead to different impact on performance.

5. Empirical Analysis

In this section, we apply our framework to real-world data sets and consider an example of ESG investing in which the average portfolio ESG score is required to be above a certain threshold. In Online Appendix B.6.4, we also consider exclusionary investing, excluding sin stocks and stranded assets.

5.1. Data

Returns. We obtain daily return data for all U.S. stocks from 2001 to 2020 from the Center for Research in Security Prices (CRSP) available through Wharton Research Data Services. The CRSP data set also contains basic firm characteristics, such as market capitalization. We obtain the daily Fama–French factor data from Kenneth French's website.[14]

Because the ESG data are updated annually, we require that a stock has at least 10 years of valid return data to be included in our analysis. For each stock, we estimate a Fama–French five-factor model (Fama and French 2015) based on daily returns:

$$R_{i,t} = \alpha_i + \beta_{i,1}(R_{M,t} - R_{f,t}) + \beta_{i,2}SMB_t + \beta_{i,3}HML_t + \beta_{i,5}RMW_t + \beta_{i,6}CMA_t + \epsilon_{i,t}. \quad (31)$$

We follow the common practice in the empirical asset pricing literature of winsorizing the raw returns used in the regression at 2.5% on both sides for each stock in each year in order to mitigate the impact of extreme outliers on the regression results. We then use the residual returns $r_{i,t} \equiv \alpha_i + \epsilon_{i,t}$ as the main target of interest in our analysis. We summarize the residual returns annually to match the frequency of the ESG data.

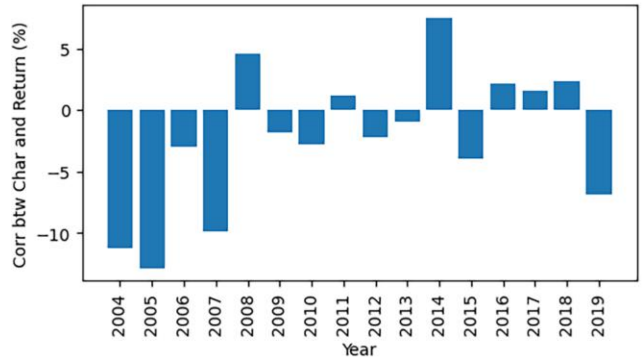
ESG. We merge the CRSP data set with the MSCI KLD ESG data set, which contains yearly environmental, social, and governance ratings of roughly 3,000 large U.S. publicly traded companies from 2003 to 2018. It is used in numerous studies examining the effect of ESG ratings on firm performance.[15] We follow Lins et al. (2017) to aggregate the raw data into an ESG score, and details are provided in Online Appendix B.6.1.

5.2. Descriptive Statistics

Table 1 shows, for each year, the number of firms in our data set, the summary statistics of its annualized residual returns, and the summary statistics of the aggregate ESG score we construct lagged by one year.

Figure 4 shows the year-over-year cross-sectional correlations between the residual returns and lag-1 ESG scores.[16] The correlations are generally negative before 2007, implying that high ESG stocks, on average, delivered lower excess returns, consistent with equilibrium theories of ESG returns (Pástor et al. 2021, Pedersen et al. 2021). After 2008, the correlations fluctuate around zero and are positive in certain years, reflecting increasing attention toward ESG and climate-related issues, an effect consistent with the Pástor et al. (2022) and Lo et al. (2022) findings.

Figure 4. (Color online) Cross-Sectional Correlations Between Asset Returns and Lag-1 ESG Scores Each Year



Although the magnitude of the correlations—between 2% and 10%—may not seem large, they can still have a nonnegligible impact on asset prices. In fact, as shown by Lo and MacKinlay (1990) and Lo and Zhang (2024), the correlation between stock alphas and firm characteristics such as beta, size, or the ESG score of a firm also ranges from 2% to 20% depending on the year. Yet they can have significant implications for tests of asset-pricing models or yield portfolios with significant alphas. In fact, our empirical analysis that follows shows that even a seemingly low level of correlation can still contribute to the performance of a portfolio by a significant amount. This is precisely what we hope to highlight using our framework.

We also emphasize that our intention is not to find the best ESG score or to provide a comprehensive study of whether ESG delivers positive or negative excess returns. Whereas Berg et al. (2022) show that there exists substantial noise in ESG measures and ESG

Table 1. Summary Statistics of the Annualized Residual Returns (in Percentage) from the Fama–French Five-Factor Model and the Aggregate ESG Score Lagged by One Year

		Annualized residual return, %							ESG score (lag one year)						
Year	#firms	Mean	Std Dev	Min	25%	50%	75%	Max	Mean	Std Dev	Min	25%	50%	75%	Max
2004	1,246	4.9	30.0	−69.1	−12.6	1.9	16.9	330.3	−0.1	0.5	−3.4	−0.3	0.0	0.1	2.9
2005	1,344	4.0	29.0	−64.0	−14.1	−0.3	17.6	156.5	−0.1	0.7	−3.0	−0.6	−0.2	0.2	2.2
2006	1,305	3.3	27.2	−74.4	−13.1	−0.1	15.5	234.1	−0.3	0.6	−3.3	−0.5	−0.2	0.0	2.5
2007	1,383	8.1	43.1	−81.3	−17.0	1.4	23.0	511.0	−0.3	0.6	−3.7	−0.7	−0.2	0.0	3.0
2008	1,487	−0.7	51.3	−95.3	−32.7	−7.5	24.2	647.1	−0.3	0.6	−3.5	−0.7	−0.3	0.0	3.4
2009	1,614	13.1	75.4	−98.0	−20.5	1.5	27.9	1,768.7	−0.3	0.6	−3.6	−0.7	−0.3	0.0	2.8
2010	1,611	5.5	35.6	−86.0	−13.9	0.8	18.3	549.0	−0.3	0.6	−3.5	−0.7	−0.3	0.0	2.8
2011	1,710	2.3	30.4	−96.5	−15.5	1.7	18.4	202.9	−0.4	0.7	−2.8	−0.7	−0.6	0.0	3.9
2012	1,643	2.3	31.0	−81.6	−13.0	−1.0	13.0	407.6	−0.4	0.8	−2.7	−0.9	−0.6	−0.2	4.2
2013	1,573	3.7	31.0	−92.1	−13.5	0.5	16.0	456.9	0.1	0.7	−2.3	−0.3	0.0	0.5	3.8
2014	1,621	1.1	23.9	−93.4	−12.4	1.0	14.2	142.6	0.1	0.8	−2.4	−0.3	0.0	0.3	3.2
2015	1,330	2.7	28.3	−86.4	−14.2	3.9	19.8	194.9	0.1	0.5	−3.7	0.0	0.0	0.3	3.2
2016	1,374	2.7	27.6	−82.2	−12.3	0.4	14.7	245.4	0.2	0.6	−2.6	−0.1	0.1	0.5	3.3
2017	1,320	4.0	27.3	−83.0	−10.7	2.0	16.0	216.5	0.2	0.7	−2.4	0.0	0.1	0.5	3.1
2018	1,440	1.0	26.0	−71.9	−15.6	−0.5	15.2	130.0	0.2	0.7	−2.9	0.0	0.2	0.6	4.1
2019	1,476	8.0	26.5	73.8	−6.8	7.8	20.9	259.8	0.7	0.8	−2.3	0.1	0.5	1.0	4.7

Note. #firms, number of firms; Std Dev, standard deviation; Min, minimum; Max, maximum.

scores from different data providers may lead to very different correlations, Berg et al. (2024) further show that aggregating individual ESG ratings improves portfolio performance, and there exists a significant signal in ESG rating scores that can be used for portfolio construction despite their noisy nature. What we hope to demonstrate in this article is that, given any ESG score, it is possible to attribute portfolio performance metrics to different constraints and to the information in those constraints, and we use the MSCI KLD ESG data set as an illustrative example.

5.3. Performance Attribution

5.3.1. Portfolio Construction. We consider investors who construct portfolios each year by solving the following problem:

$$\max_{\boldsymbol{\omega}} \ \boldsymbol{\omega}'\boldsymbol{\mu} - \frac{\gamma}{2}\boldsymbol{\omega}'\boldsymbol{\Sigma}\boldsymbol{\omega} \quad \text{s.t.} \ \boldsymbol{\omega}'\mathbf{1} = 1 \quad \text{and} \quad \boldsymbol{\omega}'\mathbf{x}_{\text{ESG}} \geq b. \tag{32}$$

The first constraint implies that the portfolio is fully invested, allowing for both long and short positions in individual assets. The second constraint imposes a minimum level of portfolio ESG score. In our example, we set $b = 1$ and $\gamma = 5$.

To construct these portfolios, investors need to estimate the expected residual return, $\boldsymbol{\mu}$, and the covariance matrix of the residual returns, $\boldsymbol{\Sigma}$, each year. We consider the two sets of estimators discussed in Section 3, both of which account for estimation risk. The goal of our analysis is not to exhaustively compare different portfolio rules and find the best one in terms of its ability to account for estimation risk, but rather to demonstrate how to attribute performance to constraints given any such portfolio rules.

In our empirical analysis, we implement two portfolio rules. The first is Jorion's (1986) portfolio rule, which is also summarized in Avramov and Zhou (2010). In each period t, the predictive mean and covariance matrix of returns are given by

$$\hat{\boldsymbol{\mu}}_t^{\text{Jorion}} = (1-\xi_1)\hat{\boldsymbol{\mu}}_s + \xi_1\hat{\boldsymbol{\mu}}_g\mathbf{1} \quad \text{and}$$
$$\hat{\boldsymbol{\Sigma}}_t^{\text{Jorion}} = \left(1 + \frac{1}{T+\xi_2}\right)\bar{\boldsymbol{\Sigma}} + \frac{\xi_2}{T(T+1+\xi_2)\mathbf{1}'\bar{\boldsymbol{\Sigma}}^{-1}\mathbf{1}}\mathbf{1}\mathbf{1}', \tag{33}$$

where

$$\xi_1 = \frac{N+2}{(N+2) + T(\hat{\boldsymbol{\mu}}_s - \hat{\boldsymbol{\mu}}_g\mathbf{1})'\bar{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{\mu}}_s - \hat{\boldsymbol{\mu}}_g\mathbf{1})},$$
$$\xi_2 = \frac{N+2}{(\hat{\boldsymbol{\mu}}_s - \hat{\boldsymbol{\mu}}_g\mathbf{1})'\bar{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{\mu}}_s - \hat{\boldsymbol{\mu}}_g\mathbf{1})}, \quad \bar{\boldsymbol{\Sigma}} = \frac{T}{T-N-2}\hat{\boldsymbol{\Sigma}}_s,$$

$\hat{\boldsymbol{\mu}}_s$, and $\hat{\boldsymbol{\Sigma}}_s$ are the sample mean and sample covariance matrix of returns over the last T periods, respectively, and $\hat{\boldsymbol{\mu}}_g = \mathbf{1}'\bar{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}_s/\mathbf{1}'\bar{\boldsymbol{\Sigma}}^{-1}\mathbf{1}$ is the estimated return of the global minimum variance portfolio.[17]

The second is the $1/N$ rule, which is extensively studied by DeMiguel et al. (2009a). As discussed in Section 3, the $1/N$ rule can be interpreted as a Bayesian approach with an appropriate prior. We specify the predictive mean and covariance matrix of returns so that they are consistent with an unconstrained optimal portfolio $\boldsymbol{\omega}_{\text{MVO}} = \mathbf{1}/N$:

$$\hat{\boldsymbol{\mu}}_t^{\text{Equal}} = \frac{\gamma}{N}\mathbf{1} \quad \text{and} \quad \hat{\boldsymbol{\Sigma}}_t^{\text{Equal}} = \mathbf{I}. \tag{34}$$

Combining (32) and the estimators in (33) or (34), we can solve for the optimal constrained portfolios. Online Appendix B.6.2 shows the decomposition of portfolio holdings based on Proposition 1.

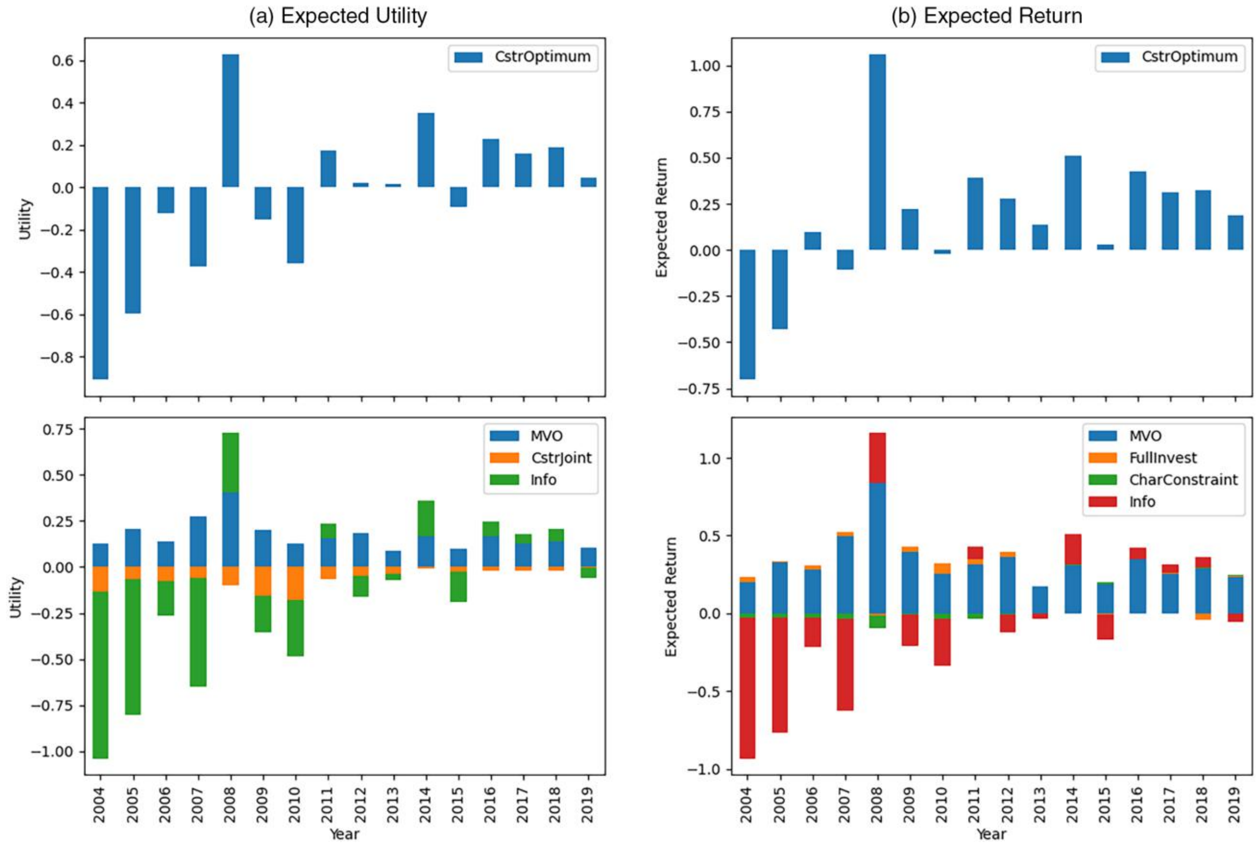
5.3.2. Expected Return and Utility Decomposition. Figure 5 demonstrates the decomposition of the expected return and utility of the portfolio into different components for Jorion's (1986) rule as an example. The upper panel of Figure 5(a) shows that the expected utility of the optimal portfolio is generally negative in the first half of our sample period and starts to turn positive toward the second half. In the lower panel, this expected utility is decomposed into three components using (18) in Proposition 4 and its predictive return version in Online Appendix B.2.2.[18] The expected utility of the unconstrained MVO portfolio is positive over the 16 years in our sample. As the conventional wisdom of constrained optimization suggests, the expected utility contribution of the two constraints (CstrJoint), treated as static, is indeed negative.

However, the expected utility contribution from the information contained in the constraints (Info) varies over time. During the first four years in our period, the expected utility contribution from information is negative. After 2008, it alternates in sign, with 2008, 2011, 2014, 2016, 2017, and 2018 being positive years in information. This pattern coincides with the signs of correlations between asset returns and ESG scores in Figure 4. The magnitude of the information contribution, however, differs from the patterns of correlations because the former is jointly determined by several terms as shown in Propositions 3 and 4.

This example vividly demonstrates that, whereas constraints must decrease the overall expected utility of a portfolio when treated as static, they can sometimes increase the expected utility relative to a passive benchmark, depending on the information contained in the constraints.

Figure 5(b) shows the expected return of the optimal portfolio and its decomposition based on (17) in Proposition 4 and its predictive return version in Online Appendix B.2.2. Whereas the two constraints (FullInvest and CharConstraint) can contribute either positively or negatively to the expected returns,[19] the main source of contribution in terms of expected

Figure 5. (Color online) Expected Return and Utility and Their Decomposition for the Portfolio in (32) with a Constraint on the Average Portfolio Characteristic Value ($\boldsymbol{\omega}'\mathbf{x}_{\text{ESG}} \geq 1.0$) and Jorion's (1986) Estimates of Predictive Moments



Notes. In (a), the top panel shows the expected utility of the constrained portfolio and the bottom panel shows its decomposition into components corresponding to the unconstrained MVO portfolio (MVO), all constraints treated as static (CstrJoint), and the information from the ESG constraint (Info). In (b), the top panel shows the expected return in excess of the Fama–French five-factor model of the constrained portfolio and the bottom panel shows its decomposition into components corresponding to the unconstrained MVO portfolio (MVO), the full investment constraint (FullInvest), the ESG constraint ($\boldsymbol{\omega}'\mathbf{x}_{\text{ESG}} \geq 1.0$) treated as static (CharConstraint), and the information from the ESG constraint (Info).

return is from the information in the constraints (Info). As is the expected utility decomposition, the expected return contribution from information is strongly negative between 2004 and 2007 and is strongly positive in 2008 and 2014.

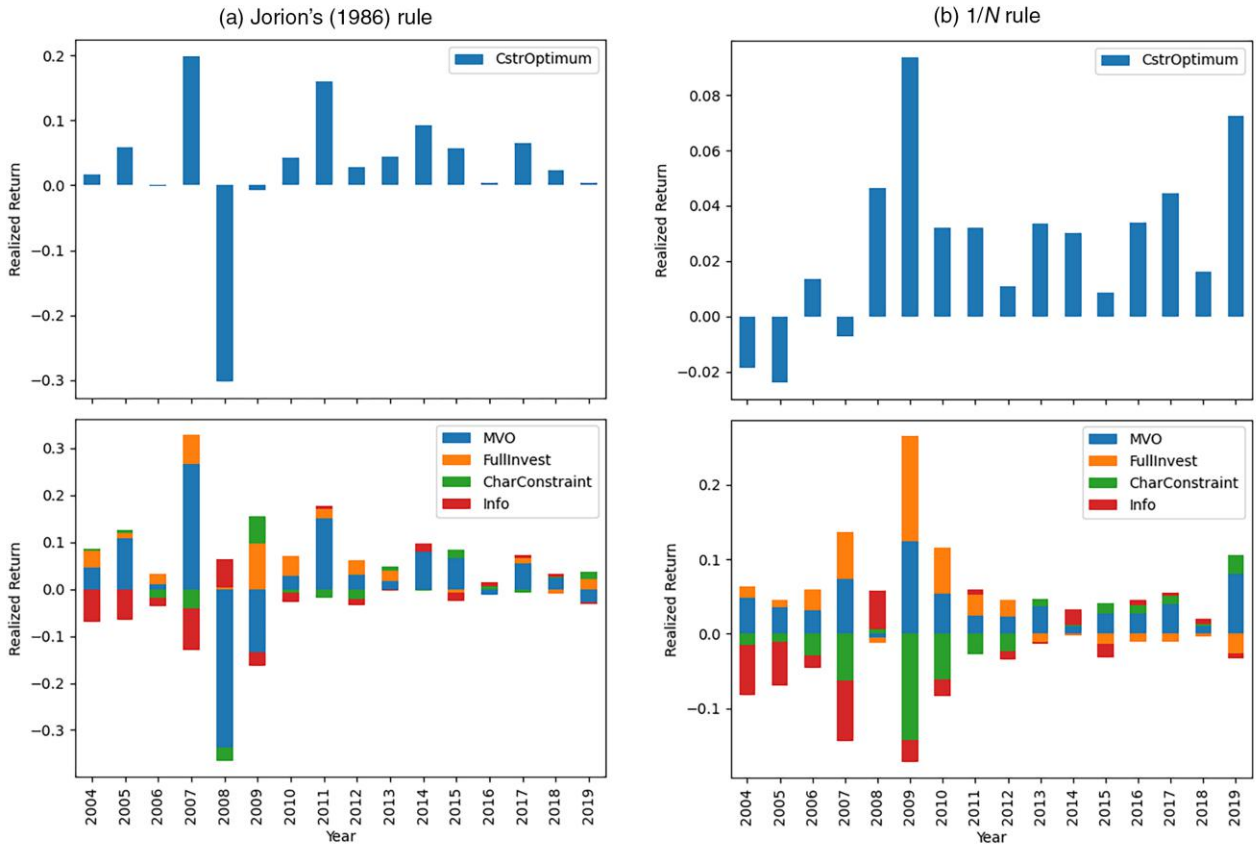
5.3.3. Realized Return Decomposition. Figure 6 shows the realized returns of the optimal portfolio for both Jorion's (1986) rule and the $1/N$ rule as the unconstrained MVO portfolio.

The upper panel of Figure 6(a) shows that, for Jorion's (1986) rule, the realized residual returns in excess of the Fama–French five-factor model of the constrained portfolio is generally positive over the 16 years in our sample except in 2008.

The lower panel decomposes the realized return of the constrained portfolio based on Proposition B.6 in Online Appendix B.4. The full investment constraint generally contributes positively to the returns. This is consistent with the fact that the average residual returns during these years are positive as shown in Table 1. The contribution of the ESG constraint, treated as static, also varies over time. In contrast, the information component contributes negatively to realized returns before 2007 and positively in certain years after 2008. Notably, in 2008, the information component from the ESG constraint contributed positively to the large negative returns from the unconstrained portfolio. These patterns are consistent with results in Online Proposition B.6 and the correlations between asset returns and ESG scores shown in Figure 4. Overall, these components explain the difference in residual returns between the unconstrained MVO and the constrained portfolio.

Figure 6(b) shows parallel results, this time with the $1/N$ rule as the unconstrained MVO portfolio. The realized residual returns for both the unconstrained and the constrained portfolio are much more stable over time. The decomposition is also similar to the case of Jorion's (1986) rule except that the contributions from the information in the constraints, in relative terms, are bigger.

Figure 6. (Color online) Realized Return for the Portfolio Defined in (32) with a Constraint on the Average Portfolio Characteristic Value ($\boldsymbol{\omega}'\mathbf{x}_{\mathrm{ESG}} \geq 1.0$)



Notes. Panel (a) corresponds to Jorion's (1986) estimates of predictive moments in (33), and (b) corresponds to predictive moments consistent with the $1/N$ rule in (34). In each subfigure, the top panel shows the realized return in excess of the Fama–French five-factor model of the constrained portfolio, and the bottom panel shows its decomposition into components corresponding to the unconstrained MVO portfolio (MVO), the full investment constraint (FullInvest), the ESG constraint ($\boldsymbol{\omega}'\mathbf{x}_{\mathrm{ESG}} \geq 1.0$) treated as static (CharConstraint), and the information from the ESG constraint (Info).

For example, the negative residual returns before 2007 were primarily driven by the negative contribution from the ESG constraints.

5.4. Other Portfolios and Selection of Constraints

The portfolio given by (32) is one way to construct ESG portfolios. In practice, investors need to determine the threshold, b. In addition, they may face long-only constraints, and we demonstrate how to decompose the performance of these portfolios using our framework in Online Appendix B.6.3.

Whereas the choice of the specific form of constraints depends on many factors such as the regulatory requirements and other business considerations, here, we show how to use our framework to compare the impact of different forms of ESG constraints purely from the perspective of financial performance.

We consider 32 portfolios with different constraints, including the factor exposure constraint in (32) with $b = 0, 0.1, \ldots, 2$ and the factor exposure constraint and long-only constraint in (B.45) in Online Appendix B.6.3 with $b = 0, 0.1, \ldots, 1$. We use a rolling window of three years to estimate the financial performance of these portfolios and select constraints for the next year based on their expected utilities with and without the information component.

Table 2 summarizes the key out-of-sample performance metrics and their decomposition averaged over 13 years in our sample. For portfolios that allow short positions (first two columns) and long-only portfolios (last two columns), we compare the performance of portfolios whose constraints are selected based on their estimated expected utilities with and without the information component. For both Jorion's (1986) rule and the $1/N$ rule, if the information component is ignored, investors achieve a lower average ESG score, lower expected utility, and lower realized return. In other words, if the information in constraints is not properly accounted for, investors may select constraints that are more detrimental to performance, leading to suboptimal portfolios.

Overall, these comparisons demonstrate how to use our framework to select constraints and the importance of accounting for information in the constraints. In

Table 2. Out-of-Sample Performance Metrics of ESG Portfolios with Different Criteria to Select Constraints for Jorion's (1986) Rule and the $1/N$ Rule

		Allowing short positions		Long-only	
Selection criteria		Exp util	Exp util w/o info	Exp util	Exp util w/o info
Panel A: Jorion's (1986) rule					
ESG	$\tilde{\boldsymbol{\omega}}^*$	**0.33**	0.13	**0.15**	0.11
Expected utility	$\tilde{\boldsymbol{\omega}}_{\text{MVO}}$	17.16	17.16	17.16	17.16
	$\tilde{\boldsymbol{\omega}}_{\text{CSTR}}$ (static)	−1.45	**−1.15**	−9.60	**−9.51**
	$\tilde{\boldsymbol{\omega}}_{\text{CSTR}}$ (info)	**0.72**	−1.39	**−1.04**	−1.33
	$\tilde{\boldsymbol{\omega}}^*$	**16.43**	14.62	**6.49**	6.29
Realized return	$\tilde{\boldsymbol{\omega}}_{\text{MVO}}$	1.59	1.59	1.59	1.59
	$\tilde{\boldsymbol{\omega}}_{\text{CSTR}}$ (static)	1.52	**1.68**	1.67	1.67
	$\tilde{\boldsymbol{\omega}}_{\text{CSTR}}$ (info)	**0.07**	−0.14	**−0.11**	−0.14
	$\tilde{\boldsymbol{\omega}}^*$	**3.17**	3.13	**3.16**	3.12
Panel B: $1/N$ rule					
ESG	$\tilde{\boldsymbol{\omega}}^*$	**0.52**	0.13	**0.33**	0.13
Expected utility	$\tilde{\boldsymbol{\omega}}_{\text{MVO}}$	0.17	0.17	0.17	0.17
	$\tilde{\boldsymbol{\omega}}_{\text{CSTR}}$ (static)	−0.23	**−0.02**	−0.09	**−0.02**
	$\tilde{\boldsymbol{\omega}}_{\text{CSTR}}$ (info)	**1.64**	−0.36	**0.63**	−0.36
	$\tilde{\boldsymbol{\omega}}^*$	**1.58**	−0.21	**0.71**	−0.21
Realized return	$\tilde{\boldsymbol{\omega}}_{\text{MVO}}$	4.03	4.03	4.03	4.03
	$\tilde{\boldsymbol{\omega}}_{\text{CSTR}}$ (static)	0.00	0.00	−0.07	**−0.03**
	$\tilde{\boldsymbol{\omega}}_{\text{CSTR}}$ (info)	**0.36**	−0.14	**0.11**	−0.14
	$\tilde{\boldsymbol{\omega}}^*$	**4.39**	3.89	**4.07**	3.86

Notes. The portfolio with a better performance metric in each comparison is bolded. $\tilde{\boldsymbol{\omega}}^*$, the optimal constrained portfolio; $\tilde{\boldsymbol{\omega}}_{\text{MVO}}$, the unconstrained MVO portfolio; $\tilde{\boldsymbol{\omega}}_{\text{CSTR}}$ (static), the component attributable to constraints treated as static; $\tilde{\boldsymbol{\omega}}_{\text{CSTR}}$ (info), the information component in constraints.

this example, the ESG constraints are detrimental (beneficial) to the financial performance of the portfolio because the ESG scores are negatively (positively) correlated with residual returns in 7 (6) out of the 13 years in our out-of-sample period (see Figure 4). Because there is substantial noise in the ESG measures (Berg et al. 2022), the specific numbers of our empirical example should be taken with a grain of salt. Nonetheless, this example demonstrates that there exists a significant signal in ESG scores that may be used for portfolio construction despite their noisy nature, which is consistent with the Berg et al. (2024) findings, and adequately accounting for information in the constraints is important for the performance attribution of a portfolio.

More generally, in other applications in which constraints contain information that is positively correlated with returns, investors miss out on gains from information if they are not properly accounted for. Examples include other company sustainability measures that may provide material information to the operations and risks of the company (Khan et al. 2016), and proprietary constraints that may contain private information about future returns.

6. Conclusion

Constraints are an integral part of the portfolio-construction process, and they have become particularly relevant as investors and regulators debate whether investing with ESG constraints or excluding stranded assets is to the benefit or detriment of investors. We propose a framework for constraint attribution that decomposes portfolio holdings, expected returns, variance, expected utilities, and realized returns into components attributable to each constraint and the information contained in each constraint.

Our framework provides a quantifiable and easily implementable measure of the information content from constraints when they are stochastic and potentially correlated with asset returns. The correlation between characteristics that form constraints and the individual asset returns plays a key role in determining the sign and magnitude of the excess return and variance from information. Constraints serve as an indirect mechanism for incorporating this information into the portfolio that is otherwise unavailable to investors.

We demonstrate that our methodology can accommodate estimation risk in parameter values of the portfolio construction process using Bayesian portfolio analysis, which provides the same decomposition of performance out-of-sample and a method to select constraints that improve—or are least detrimental to—portfolio performance ex ante.

Our framework can be applied to common examples of constraints, including the level of a characteristic, such as ESG scores, and exclusion constraints, such as divesting from sin stocks and energy stocks. Our results show that these constraints can contribute positively (or negatively) to portfolio performance compared with a passive benchmark when the information contained in the constraints is not accounted for in the passive benchmark but is sufficiently positively (or negatively) correlated with asset returns.

Acknowledgments
The authors thank Kay Giesecke (editor), Lisa Goldberg, Umit Gurun, Nabil Kahale, Alok Kumar, Alberto Martin-Utrera, Tong Yao, an anonymous associate editor, and two anonymous reviewers as well as seminar and conference participants at the 15th Annual Hedge Fund Research Conference, the 2023 INFORMS Annual Meeting, the 2023 China Financial Review International (CFRI) and China International Risk Forum (CIRF) Joint Conference, and the Seventh PKU-NUS Annual International Conference on Quantitative Finance and Economics for very helpful comments and discussion. They also thank Manish Singh and Chaoyi Zhao for assistance in obtaining ESG data, and Yumin Xu and Ziyue Yu for excellent research assistance. This article is the recipient of the 2023 CFRI & CIRF–China Finance Review International Research Excellence Award (Asset Management) and the 2024 Commodity and Energy Markets Association (CEMA) Questrom–CEMA Best Paper Prize.

Endnotes

[1] This is a key distinction from the literature that shows constraints may improve out-of-sample performance because of the parameter estimation risk in portfolio construction. We provide more details in the related literature.

[2] For example, on August 4, 2022, a letter signed by 19 attorneys general was sent to BlackRock expressing concern over its ESG policies, stating in part, "BlackRock's actions on a variety of governance objectives may violate multiple state laws. … BlackRock has a private motivation that differs from its public commitments and statements. This is likely insufficient to satisfy state laws requiring a sole focus on financial return." See https://www.texasattorneygeneral.gov/sites/default/files/images/executive-management/BlackRock\%20Letter.pdf (accessed December 15, 2022).

[3] We follow the common convention that all vectors are column vectors, and all vectors and matrices are boldface.

[4] We are comparing unconstrained and constrained portfolios in the same context, that is, with or without a risk-free asset for both portfolios.

[5] This decomposition is known in the literature; see, for example, Stubbs and Vandenbussche (2010), Menchero and Davis (2011), and Kan et al. (2022).

[6] When the optimization problem in (1) contains inequality constraints, the Lagrange multipliers $\boldsymbol{\lambda}^* \geq 0$ satisfy the complementary slackness condition: $(b_i - \mathbf{A}_i'\boldsymbol{\omega}^*)\lambda_i^* = 0$ for $i = 1, 2, \ldots, J$.

[7] If one compares the performance of the constrained portfolio with the mean-variance optimal portfolio using the conditional moments, $\boldsymbol{\mu}_\mathbf{X}$ and $\boldsymbol{\Sigma}_\mathbf{X}$, one recovers the classic result that constraints must decrease the expected utility of the optimal portfolio with respect to $\boldsymbol{\mu}_\mathbf{X}$ and $\boldsymbol{\Sigma}_\mathbf{X}$. However, this comparison is not very realistic because, in practice, investors often need to understand the impact of a managed portfolio relative to simple and passive benchmarks to which they have easy access rather than to a hypothetical portfolio that is optimal with respect to the conditional mean and variance with full information in $\mathbf{X}$.

[8] To see this, imagine a set of N hypothetical assets whose returns, $\mathbf{s}$, have a covariance matrix $\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_\mathbf{X}$. We have $\text{Cov}(\boldsymbol{\omega}_{\text{SHR}}'\mathbf{s}, \boldsymbol{\omega}_{\text{CSTR}}'\mathbf{s}) = \boldsymbol{\omega}_{\text{SHR}}'\text{Cov}(\mathbf{s},\mathbf{s})\boldsymbol{\omega}_{\text{CSTR}} = \boldsymbol{\omega}_{\text{SHR}}'(\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_\mathbf{X})\boldsymbol{\omega}_{\text{CSTR}}$. We show in Section 2.3 and Online Appendix B.1 that $\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_\mathbf{X}$ is always positive semidefinite under certain distributional assumptions of $\mathbf{r}$ and $\mathbf{X}$.

[9] This is true for Bayesian portfolios with common choices of priors as shown in Section 3.2.

[10] Pástor (2000) and Harvey et al. (2010) provide examples of using Monte Carlo simulation to obtain Bayesian portfolio weights. Tu and Zhou (2010) suggest using Monte Carlo simulation to derive the relationship between expected return and optimal Bayesian portfolio weights for non–mean-variance utility functions.

[11] In addition to the studies mentioned above, see, for example, Brodie et al. (2009) and Fan et al. (2012a, b) as well as combinations of different portfolio rules, such as Kan and Zhou (2007), Tu and Zhou (2011), Kan et al. (2022), and Kan and Wang (2023).

[12] For example, as a sanity check, $\tilde{\boldsymbol{\lambda}}_{1,\text{PR}}$ vanishes if the unconstrained optimal portfolio under $\tilde{\mathbf{r}} \sim N\left(\boldsymbol{\mu}_1, \left(1 + \frac{1}{\tau}\right)\boldsymbol{\Sigma}\right)$ already satisfies the constraint: $\mathbf{A}\left(\frac{1}{\gamma}\left(1 + \frac{1}{\tau}\right)\boldsymbol{\Sigma}\right)^{-1}\boldsymbol{\mu}_1 = \mathbf{b}$. Similarly, $\tilde{\boldsymbol{\lambda}}_{1,\text{DT}}$ vanishes if the optimal portfolio under the plug-in mean, $\tilde{\mathbf{r}} \sim N(\hat{\boldsymbol{\mu}}, \boldsymbol{\Sigma})$, already satisfies the constraint $\mathbf{A}\left(\frac{1}{\gamma}\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\mu}}\right) = \mathbf{b}$.

[13] Results are not sensitive to the specific parameterization of $\tilde{\boldsymbol{\Sigma}}$ as long as it is positive definite.

[14] See https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. Accessed July 20, 2022.

[15] See, for example, Hong and Kostovetsky (2012), Lins et al. (2017), and Berg et al. (2022). Makridis and Simaan (2023) use Refinitiv ESG data to compare the performance of different shrinkage estimators of the covariance matrix and different ESG-based rules.

[16] The ESG scores are available from 2003 to 2018, so the correlations are computed from 2004 to 2019.

[17] We encounter the high-dimensional problem in our analysis whereby the total number of assets, N, in the portfolio—typically between 1,000 to 4,000—is larger than the number of daily, T, in a year. Therefore, we use the shrinkage covariance estimator of Chen et al. (2010), implemented by `sklearn.covariance.OAS` in Python, multiplied by a constant factor (100) to replace $\bar{\Sigma}$ in order to ensure numerical stability.

[18] We estimate the correlations, ρ, between lagged ESG values and residual returns rather than predictive returns because the $1/N$ rule implies constant predictive returns for all assets which lead to a trivial decomposition.

[19] The sign is determined by the correlation between the holdings of the MVO portfolio and the coefficients of the constraint. See the remarks after Proposition 1.

References

Ardia D, Boudt K (2015) Implied expected returns and the choice of a mean–variance efficient portfolio proxy. *J. Portfolio Management* 41(4):68–81.

Avramov D, Zhou G (2010) Bayesian portfolio analysis. *Annual Rev. Financial Econom.* 2:25–47.

Berg F, Koelbel JF, Rigobon R (2022) Aggregate confusion: The divergence of ESG ratings. *Rev. Finance* 26(6):1315–1344.

Berg F, Lo AW, Rigobon R, Singh M, Zhang R (2024) Quantifying the returns of ESG investing: An empirical analysis with six ESG metrics. *J. Portfolio Management* 50(8):216–238.

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

Bonami P, Lejeune MA (2009) An exact solution approach for portfolio optimization problems under stochastic and integer constraints. *Oper. Res.* 57(3):650–670.

Brandt MW, Santa-Clara P, Valkanov R (2009) Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns. *Rev. Financial Stud.* 22(9):3411–3447.

Brinson GP, Hood LR, Beebower GL (1986) Determinants of portfolio performance. *Financial Anal. J.* 42(4):39–44.

Broadie M (1993) Computing efficient frontiers using estimated parameters. *Ann. Oper. Res.* 45:21–58.

Brodie J, Daubechies I, De Mol C, Giannone D, Loris I (2009) Sparse and stable Markowitz portfolios. *Proc. Natl. Acad. Sci. USA* 106(30):12267–12272.

Brown SJ (1978) The portfolio choice problem: Comparison of certainty equivalence and optimal Bayes portfolios. *Comm. Statist. Simulation Comput.* 7(4):321–334.

Ceria S, Stubbs RA (2006) Incorporating estimation errors into portfolio selection: Robust portfolio construction. *J. Asset Management* 7(2):109–127.

Chen Y, Wiesel A, Eldar YC, Hero AO (2010) Shrinkage algorithms for MMSE covariance estimation. *IEEE Trans. Signal Processing* 58(10):5016–5029.

Chopra VK, Ziemba WT (1993) The effect of errors in means, variances, and covariances on optimal portfolio choice. *J. Portfolio Management* 19(2):6–11.

Clarke R, de Silva H, Thorley S (2002) Portfolio constraints and the fundamental law of active management. *Financial Anal. J.* 58(5): 48–66.

Corwin SA, Coughenour JF (2008) Limited attention and the allocation of effort in securities trading. *J. Finance* 63(6):3031–3067.

David H (1973) Concomitants of order statistics. *Bull. Internat. Statist. Inst.* 45:295–300.

Davies PI, Higham NJ (2000) Numerically stable generation of correlation matrices and their factors. *BIT* 40:640–651.

DeMiguel V, Garlappi L, Uppal R (2009a) Optimal vs. naive diversification: How inefficient is the 1/N portfolio strategy? *Rev. Financial Stud.* 22(5):1915–1953.

DeMiguel V, Garlappi L, Nogales FJ, Uppal R (2009b) A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Sci.* 55(5):798–812.

Dybvig PH, Ross SA (1985) Differential information and performance measurement using a security market line. *J. Finance* 40(2):383–399.

Fabozzi FJ, Huang D, Zhou G (2010) Robust portfolios: Contributions from operations research and finance. *Ann. Oper. Res.* 176:191–220.

Fama EF (1972) Components of investment performance. *J. Finance* 27(3):551–567.

Fama EF, French KR (2015) A five-factor asset pricing model. *J. Financial Econom.* 116(1):1–22.

Fan J, Li Y, Yu K (2012a) Vast volatility matrix estimation using high-frequency data for portfolio selection. *J. Amer. Statist. Assoc.* 107:412–428.

Fan J, Zhang J, Yu K (2012b) Vast portfolio selection with gross-exposure constraints. *J. Amer. Statist. Assoc.* 107(497):592–606.

Fang K-T, Kotz S, Ng KW (1990) *Symmetric Multivariate and Related Distributions* (Chapman and Hall, New York).

Ferson WE, Siegel AF (2001) The efficient use of conditioning information in portfolios. *J. Finance* 56(3):967–982.

Frost PA, Savarino JE (1986) An empirical Bayes approach to efficient portfolio selection. *J. Financial Quant. Anal.* 21(3):293–305.

Goldberg LR (2021) Constraint implied returns and stress in mean-variance optimized portfolios. Working paper, University of California, Berkeley, CA.

Grinold R (2005) Implementation efficiency. *Financial Anal. J.* 61(5):52–64.

Harvey CR, Liechty JC, Liechty MW, Müller P (2010) Portfolio selection with higher moments. *Quant. Finance* 10(5):469–485.

Hirshleifer D, Lim SS, Teoh SH (2011) Limited investor attention and stock market misreactions to accounting information. *Rev. Asset Pricing Stud.* 1(1):35–73.

Hirshleifer D, Subrahmanyam A, Titman S (2006) Feedback and the success of irrational investors. *J. Financial Econom.* 81(2):311–338.

Hjalmarsson E, Manchev P (2012) Characteristic-based mean-variance portfolio choice. *J. Banking Finance* 36(5):1392–1401.

Hong H, Kostovetsky L (2012) Red and blue investing: Values and finance. *J. Financial Econom.* 103(1):1–19.

Jacquier E, Polson N (2011) Bayesian methods in finance. Geweke J, Koop G, van Dijk H, eds. *The Oxford Handbook of Bayesian Econometrics* (Oxford University Press, Oxford, UK), 439–512.

Jagannathan R, Ma T (2003) Risk reduction in large portfolios: Why imposing the wrong constraints helps. *J. Finance* 58(4): 1651–1683.

Jin H, Xu Z-Q, Zhou XY (2008) A convex stochastic optimization problem arising from portfolio selection. *Math. Finance* 18(1): 171–183.

Jorion P (1986) Bayes-Stein estimation for portfolio analysis. *J. Financial Quant. Anal.* 21(3):279–292.

Kan R, Wang X (2023) Optimal portfolio choice with unknown benchmark efficiency. *Management Sci.* 70(9):6117–6138.

Kan R, Zhou G (2007) Optimal portfolio choice with parameter uncertainty. *J. Financial Quant. Anal.* 42(3):621–656.

Kan R, Wang X, Zhou G (2022) Optimal portfolio choice with estimation risk: No risk-free asset case. *Management Sci.* 68(3): 2047–2068.

Khan M, Serafeim G, Yoon A (2016) Corporate sustainability: First evidence on materiality. *Accounting Rev.* 91(6):1697–1724.

Klein RW, Bawa VS (1976) The effect of estimation risk on optimal portfolio choice. *J. Financial Econom.* 3(3):215–231.

Lai TL, Xing H, Chen Z (2011) Mean–variance portfolio optimization when means and covariances are unknown. *Ann. Appl. Statist.* 5(2):798–823.

Levy H, Levy M (2014) The benefits of differential variance-based constraints in portfolio optimization. *Eur. J. Oper. Res.* 234(2): 372–381.

Lins KV, Servaes H, Tamayo A (2017) Social capital, trust, and firm performance: The value of corporate social responsibility during the financial crisis. *J. Finance* 72(4):1785–1824.

Lo AW, MacKinlay AC (1990) Data-snooping biases in tests of financial asset pricing models. *Rev. Financial Stud.* 3(3):431–467.

Lo AW, Zhang R (2024) Quantifying the impact of impact investing. *Management Sci.* 70(10):7161–7186.

Lo AW, Zhang R, Zhao C (2022) Measuring and optimizing the risk and reward of green portfolios. *J. Impact ESG Investing* 3(2): 55–93.

Lo AW, Wu L, Zhang R, Zhao C (2024) Optimal impact portfolios with general dependence and marginals. *Oper. Res.* 72(5): 1775–1789.

Makridis C, Simaan M (2023) Balancing returns and responsibility: Evidence from shrinkage-based portfolios, Preprint, submitted November 7, https://dx.doi.org/10.2139/ssrn.4597152.

Menchero J, Davis B (2011) Risk contribution is exposure times volatility times correlation: Decomposing risk using the x-sigma-rho formula. *J. Portfolio Management* 37(2):97–106.

Merton RC (1980) On estimating the expected return on the market: An exploratory investigation. *J. Financial Econom.* 8(4):323–361.

Pástor L (2000) Portfolio selection and asset pricing models. *J. Finance* 55(1):179–223.

Pástor L, Stambaugh RF (2000) Comparing asset pricing models: An investment perspective. *J. Financial Econom.* 56(3):335–381.

Pástor L, Stambaugh RF, Taylor LA (2021) Sustainable investing in equilibrium. *J. Financial Econom.* 142(2):550–571.

Pástor L, Stambaugh RF, Taylor LA (2022) Dissecting green returns. *J. Financial Econom.* 146(2):403–424.

Pedersen LH, Fitzgibbons S, Pomorski L (2021) Responsible investing: The ESG-efficient frontier. *J. Financial Econom.* 142(2):572–597.

Powell WB (2019) A unified framework for stochastic optimization. *Eur. J. Oper. Res.* 275(3):795–821.

Sharpe WF (1992) Asset allocation: Management style and performance measurement. *J. Portfolio Management* 18(2):7–19.

Simon HA (1955) A behavioral model of rational choice. *Quart. J. Econom.* 69(1):99–118.

Stambaugh RF (1997) Analyzing investments whose histories differ in length. *J. Financial Econom.* 45(3):285–331.

Stubbs RA, Vandenbussche D (2010) Constraint attribution. *J. Portfolio Management* 36(4):48–59.

Tu J, Zhou G (2010) Incorporating economic objectives into Bayesian priors: Portfolio choice under parameter uncertainty. *J. Financial Quant. Anal.* 45(4):959–986.

Tu J, Zhou G (2011) Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies. *J. Financial Econom.* 99(1):204–215.

Zhou G (2009) Beyond Black–Litterman: Letting the data speak. *J. Portfolio Management* 36(1):36–45.