Performance Attribution for Portfolio Constraints

Andrew W. Lo* and Ruixun Zhang†

3 Aug 2023

Abstract

We propose a new performance attribution framework that decomposes a constrained portfolio's holdings, expected utility, expected returns, variance, 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. The excess return from information is positive (negative) when this correlation is positive (negative) and the constraint is binding. The excess variance of a portfolio is negative when the holdings of a *shrinkage portfolio* and the holdings attributable to constraints are positively correlated, and the degree to which variance is reduced depends on the squared correlation between returns and constraints. We provide simulated and empirical examples involving constraints on ESG portfolios. Contrary to conventional wisdom, constraints may improve portfolio performance under certain scenarios.

Keywords: Constraints, Performance Attribution, Portfolio Theory, Information, ESG Investing, Socially Responsible Investing.

JEL Classification: C10, G11, G12, Q56

*Charles E. and Susan T. Harris Professor, MIT Sloan School of Management; Director, MIT Laboratory for Financial Engineering; Principal Investigator, MIT Computer Science and Artificial Intelligence Laboratory; External Faculty, Santa Fe Institute. alo-admin@mit.edu (email).

†Assistant Professor and Boya Young Fellow, Peking University 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. zhangruixun@pku.edu.cn (email).

Contents

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 (Clarke, de Silva, and Thorley, 2002; Grinold, 2005).

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. The construction of these portfolios typically involves constraints based on a firm's characteristics, such as its ESG score, its amount of carbon emissions, its prospect of developing a disease-curing drug, or the industry to which it belongs.[1] Popular methods include negative screening, which imposes filters so that certain companies are excluded from the investable universe; positive screening, where companies are selected for high values of certain attributes; and factor integration, which imposes constraints on the average level of a portfolio's characteristics, such as its ESG score or other fundamental and technical factors of a company.[2]

The growth in popularity and assets under management of SRI and ESG has also triggered a backlash. For example, on 4 August 2022, a letter signed by the attorneys general of nineteen states was sent to BlackRock's CEO, Laurence Fink, expressing concern over its asset manager's ESG policies and how those policies may affect their holdings of fossil fuel energy companies.[3] These are not minor concerns, given that the legal penalty for violating one's fiduciary duty involves personally making up any losses suffered by the client and restoring to the client any profits made by the fiduciary's service provision to said client.[4] In addition, it is also unclear from the empirical literature whether SRI and ESG investing are adding or removing value from an investor's point of view.

[1] For examples on the way constraints are imposed to construct SRI and ESG portfolios, and recent theories on the asset-pricing implications, see Pástor, Stambaugh, and Taylor (2021, 2022), Pedersen, Fitzgibbons, and Pomorski (2021), Idzorek, Kaplan, and Ibbotson (2021), Zerbib (2022), and Lo and Zhang (2023).

[2] For a discussion of these different portfolio construction methodologies, see Roselle (2016), Eccles, Kastrapeli, and Potter (2017), Amel-Zadeh and Serafeim (2018), and Cappucci (2018).

[3] "BlackRock's actions on a variety of governance objectives may violate multiple state laws. Mr. McCombe's letter asserts compliance with our fiduciary laws because 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. Our states will not idly stand for our pensioners' retirements to be sacrificed for BlackRock's climate agenda." See `https://www.texasattorneygeneral.gov/sites/default/files/images/executive-management/BlackRock%20Letter.pdf`, accessed 15 December 2022.

[4] See 9 U.S. Code §1109 - Liability for breach of fiduciary duty.

How can we reconcile SRI and ESG investing with fiduciary duty? The answer lies in developing a framework in which the financial impact of constraints can be measured, the subject of this article.

We develop a general framework to attribute the performance of portfolios to contributions from individual constraints. Conventional wisdom typically maintains that a constrained portfolio must have a non-superior risk/reward profile compared to the unconstrained case, because the former contains a proper subset of securities of the unconstrained version, and mathematical logic suggests that the constrained optimum is at best equal to the unconstrained optimum or, more likely, inferior.

However, the non-superiority of constrained optima relies on a key assumption that is almost never explicitly stated: the constraint does not provide additional information regarding asset returns. This implies that either investors have full information about asset returns when constructing portfolios *ex post*, or when 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 CUSIP identifiers that contain prime numbers. Clearly such a constraint has no relation to the returns of any security, hence imposing such a constraint can only reduce the risk-adjusted return of the optimized portfolio.

But what if investors do not have full information about returns *ex ante* and the constraint is *not* independent of the returns? For example, consider the constraint, "invest only in those companies for which their stock 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, the answer to the question of 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 whose objective is to obtain a portfolio that makes an optimal trade-off between return and risk using the standard mean-variance utility. We denote the optimal portfolio with respect to this objective under no constraints as the mean-variance optimal (MVO) portfolio. However, the portfolio obtained while imposing all constraints will likely differ from the MVO portfolio due to the effect of the constraints. Therefore, we develop a methodology to decompose the constrained portfolio's holdings, expected utility, expected returns, and realized returns into different components: those attributable to the MVO portfolio, those to the individual constraints treated as static, and those to 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. For a universe of N assets and portfolio weights $\boldsymbol{\omega} \equiv [\omega_1 \ \omega_2 \ \cdots \ \omega_N]'$, 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 i-th asset, such as its ESG score or a label representing its industry. A constraint is denoted by $\mathbf{A}(\mathbf{x})'\boldsymbol{\omega} = a_1(x_1)\omega_1 + \cdots + a_N(x_N)\omega_N = b$ (or $\geq b$), where b represents the constraint threshold and $a_i(x_i)$ symbolizes that the coefficient of the i-th asset depends on its characteristic x_i. In our framework, investors do not always have perfect knowledge of—nor are they necessarily fully rational in exploiting—information in these constraints in terms of incorporating it into return forecasts.[5]

By modeling the firm characteristic $\mathbf{x}$ 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, under the special case of normally distributed returns and independent characteristics, 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 18) 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 is able to contribute to the performance of a portfolio. While 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.

We apply our framework to two common classes of portfolio constraints. The first is when investors constrain the 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 a minimum ESG score or whether the firm belongs to an industry associated with "sin" stocks. We derive additional analytical results for these special types of constraints and provide simulated examples to illustrate the attribution of expected returns and utilities in these scenarios. In particular, constraints may contribute positively to both the expected return

[5]There is a large literature documenting limited attention (Corwin and Coughenour, 2008; Hirshleifer, Lim, and Teoh, 2011) and bounded rationality (Simon, 1955; Hirshleifer, Subrahmanyam, and Titman, 2006; Kogan et al., 2006) in investor behavior.

and utility if the information contained in the constraint is sufficiently positively correlated with asset returns, because they serve as an indirect mechanism to use the information.

Finally, using real-world datasets, including MSCI KLD ESG ratings, the Compustat Historical Segment compilation, daily returns and industry classification from the CRSP dataset, and the well-known Fama-French factor data, we illustrate how our framework can be applied to quantify the impact of ESG investing 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. While 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 2010, implying that high ESG stocks and non-stranded assets delivered lower excess returns relative to the Fama-French five-factor model on average, which is consistent with equilibrium theories of ESG returns (Pástor, Stambaugh, and Taylor, 2021; Pedersen, Fitzgibbons, and Pomorski, 2021). However, after 2011, the information in the constraints contributes positively to portfolio performance, reflecting the increasing attention toward SRI and ESG-related issues, an effect consistent with Pástor, Stambaugh, and Taylor's (2022) and Lo, Zhang, and Zhao's (2022) findings.

We emphasize that our intention in this article is not to provide the best measure of whether SRI and ESG deliver positive or negative excess returns.[6] Instead, our primary objective is to illustrate how our framework can be used to attribute performance to any portfolio constraints and—given that any firm characteristics may be correlated to returns—to the information contained in those constraints.

2 Related Literature

Our article contributes to the literature on portfolio theory, and in particular, to performance attribution in portfolio optimization. The major breakthrough dates back to 1952, when Harry Markowitz launched the field of modern portfolio theory (Markowitz, 1952). The classical literature on performance attribution include Fama (1972), Daniel et al. (1997), and Brinson, Hood, and Beebower (1986). Surveys in this area include Grinold and Kahn (1999, 2019), Steinbach (2001), Rubinstein (2002), Kolm, Tütüncü, and Fabozzi (2014), Markowitz (2014), and Bacon (2019).

One of the pioneering contributions for understanding and quantifying the impact of

[6]In fact, there is substantial divergence among ESG measures, even when they purport to capture the same concepts (Gibson, Krueger, and Schmidt, 2021; Berg, Koelbel, and Rigobon, 2022). In particular, Khan, Serafeim, and Yoon (2016) find that only firms with good ratings on *material* sustainability issues significantly outperform firms with poor ratings on these issues.

constraints is the transfer coefficient of Clarke, de Silva, and Thorley (2002, 2005). However, one limitation of the transfer coefficient is that it only provides an aggregate measure of the impact of constraints, and does not offer a scalable solution to decompose the effects of individual constraints. As a result, several studies rely on the notion of shadow cost (or Lagrange multipliers) to measure the first-order marginal cost of each constraint (Grinold, 2005; Scherer and Xu, 2007; Stubbs and Vandenbussche, 2010; Menchero and Davis, 2011; Davis and Menchero, 2012; Tütüncü, 2012).

In the standard framework, constraints are believed to never improve the expected utility of the optimized portfolio on an *ex ante* basis. However, through *ex post* constraint attribution analysis, certain constraints may actually improve performance by preventing the portfolio from taking bets that turn out to be harmful out of sample. This is often referred to as the *model insurance* function of constraints.

However, as demonstrated in our framework, there is actually a more fundamental reason why constraints may contribute positively to the expected utility and other portfolio performance metrics: the information contained in constraints. In our framework, investors do not always have perfect knowledge of, nor are they assumed to be fully rational in exploiting, such information. As a result, constraints serve as an indirect mechanism to use that information, and they may contribute positively to the expected utility even on an *ex ante* basis if the information contained in the constraint is sufficiently positively correlated with returns. Our methodology provides a way to quantify this effect.

Our constraint attribution framework also contributes to the literature on the impact of SRI, ESG, and other non-financial objectives on investment returns. Theories on the asset pricing implications of sustainable investing date back to Merton's (1987) model of neglected stocks and segmented markets, Fama and French's (2007) taste model, and, more recently, Pástor, Stambaugh, and Taylor (2021) and Pedersen, Fitzgibbons, and Pomorski (2021).[7]

There is also a vast empirical literature focused on measuring the returns of SRI and ESG investing. On the one hand, studies suggest that these investments may sacrifice returns in markets including stocks (Fabozzi, Ma, and Oliphant, 2008; Hong and Kacperczyk, 2009; Statman and Glushkov, 2009; Fauver and McDonald IV, 2014; Alessandrini and Jondeau, 2020), bonds (Baker et al., 2022), venture capital funds (Barber, Morse, and Yasuda, 2021), and mutual funds (Geczy, Stambaugh, and Levin, 2021). Pástor, Stambaugh, and Taylor (2022) show that the high returns for green assets in recent years reflect unexpectedly strong increases in environmental concerns, not high expected returns.

On the other hand, recent empirical evidence (Madhavan, Sobczyk, and Ang, 2021;

[7]See also Albuquerque, Koskinen, and Zhang (2019), Berk and van Binsbergen (2021), Goldstein et al. (2021), Idzorek, Kaplan, and Ibbotson (2021), Zerbib (2022), and Lo and Zhang (2023).

Bansal, Wu, and Yaron, 2022; Lo, Zhang, and Zhao, 2022; Berg et al., 2023) suggests that ESG measures are associated with higher returns, at least under certain market conditions. Lindsey, Pruitt, and Schiller (2021) find that modifying optimal portfolio weights to achieve an ESG-investing tilt negligibly affects portfolio performance, because ESG measures do not provide information in addition to other observable firm characteristics.

Our findings show 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. Our framework provides a methodology for quantifying the financial implications of a wide range of portfolio constraints, not merely SRI or ESG measures.

3 A Framework for Constraint Attribution

We consider a universe of N assets whose returns are given by the random vector $\mathbf{r}_t = (r_{1,t}, \ldots, r_{N,t})'$.[8] 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 construct portfolios based on the following mean-variance optimization:

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

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

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

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

We choose $\frac{1}{2}$ to be the coefficient of the variance term in the utility function, and consider the case of equality constraints in (1), both for expositional simplicity. All of our results carry out for general coefficients of the variance term other than $\frac{1}{2}$. It is also easy to derive a parallel set of results under inequality constraints, $\mathbf{A}\boldsymbol{\omega} \leq \mathbf{b}$, and we describe ways to

[8] We follow the common convention that all vectors are assumed to be column vectors unless stated otherwise, and all vectors and matrices are boldface.

generalize our results throughout our exposition.

Common examples of portfolio constraints include:

$$\boldsymbol{\omega}'\mathbf{1} = 1: \text{ full investment} \tag{2}$$

$$\omega_i = 0: \text{ exclusion of asset } i \tag{3}$$

$$\boldsymbol{\omega}'\mathbf{A}_1 = b_1: \text{maintain average firm characteristic, or factor exposure.} \tag{4}$$

Equation (3) represents a common constraint in exclusionary investing that removes certain assets based on a particular criterion, such as divesting from sin stocks. Equation (4) represents a wide class of constraints that restricts the average value of a particular asset characteristic, such as ESG, value, size, or momentum measures.

There also exist variations of the objective function in (1), or the expected utility of the portfolio. The returns $\mathbf{r}$ can be interpreted either as raw returns or residual returns in excess of some factor model. For example, Grinold and Eaton (1998), Grinold (2005), and Stubbs and Vandenbussche (2010) consider active returns or alphas with respect to a benchmark portfolio instead of raw returns. We also consider residual returns in excess of the Fama-French five-factor model in our empirical analysis in Section 5. Goldberg (2021) considers the minimization of the tracking error as the objective. Our framework can also be easily generalized to account for these objectives.

A critical implicit assumption in the existing literature of constraint attribution is that the constraints, $\mathbf{A}$, are treated as constants, and are therefore independent of returns, $\mathbf{r}$. Under this assumption, the solution of 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 function, and imposing constraints can only decrease the objective function. The following result summarizes the optimal portfolio weights and the decomposition of portfolio holdings, expected return, and expected utility attributable to each constraint.[9] We provide proofs of all propositions in the Appendix.

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

$$\boldsymbol{\Sigma}\boldsymbol{\omega}^* = \boldsymbol{\mu} - \mathbf{A}'\boldsymbol{\lambda}^* \tag{5}$$

where the Lagrange multipliers, $\boldsymbol{\lambda}^*$, are given by:

$$\boldsymbol{\lambda}^* = \left(\mathbf{A}\boldsymbol{\Sigma}^{-1}\mathbf{A}'\right)^{-1}\left(\mathbf{A}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \mathbf{b}\right), \tag{6}$$

provided that the feasible region of the constrained optimization problem is nonempty. Here

[9] See also Stubbs and Vandenbussche (2010) and Menchero and Davis (2011).

$\boldsymbol{\lambda}^*$ *measure the shadow cost of the portfolio's expected utility with respect to each constraint.*[10]

Equation (5) leads to a series of decompositions.

1. *Portfolio holdings decomposition.*

$$\boldsymbol{\omega}^* = \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \boldsymbol{\Sigma}^{-1}\mathbf{A}'\boldsymbol{\lambda}^*. \tag{7}$$

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

Because we refer to this decomposition repeatedly in subsequent sections, we denote:

$$\boldsymbol{\omega}_{\text{MVO}} \equiv \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}, \tag{8}$$

$$\boldsymbol{\omega}_{\text{CSTR}} \equiv -\boldsymbol{\Sigma}^{-1}\mathbf{A}'\boldsymbol{\lambda}^*. \tag{9}$$

2. *Expected return decomposition.*

$$\boldsymbol{\mu}'\boldsymbol{\omega}^* = \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{A}'\boldsymbol{\lambda}^*. \tag{10}$$

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

3. *Expected utility decomposition.*

$$\boldsymbol{\mu}'\boldsymbol{\omega}^* - \frac{1}{2}\boldsymbol{\omega}^{*\prime}\boldsymbol{\Sigma}\boldsymbol{\omega}^* = \frac{1}{2}\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{\lambda}^{*\prime}\mathbf{A}\boldsymbol{\Sigma}^{-1}\mathbf{A}'\boldsymbol{\lambda}^*. \tag{11}$$

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

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 (7) can lead to both positive and negative entries.

Second, in the expected return decomposition in (10), if the Lagrange multiplier, or the shadow cost, for the i-th constraint $\lambda_i > 0$,[11] the sign of the marginal contribution of that

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

[11] This is always true for constraints that are binding.

constraint is determined by the sign of $-\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{A}_j$, which can be positive in certain cases. To see that, observe that:

$$-\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{A}_j = -|\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}| \cdot |\mathbf{A}_j| \cdot \cos\theta \tag{12}$$

where θ is the angle between $\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ (holdings of the MVO portfolio) and $\mathbf{A}_j$ (constraint coefficients). When these two vectors have a negative inner product (that is, they are negatively correlated), $\cos\theta$ is negative, which implies that the i-th constraint *increases* expected returns.

Finally, when constraints are static, they always decrease expected utility relative to the unconstrained MVO portfolio, as shown by the fact that the last term in (11) 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 due to the risk term. Nonetheless, there is a clear interpretation of the portfolio holdings decomposition $\boldsymbol{\omega}_{\text{CSTR}}$ operating on the covariance matrix $\boldsymbol{\Sigma}$.

3.1 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, P/E 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 j-th constraint by $\mathbf{A}_j(\mathbf{x})$. For example, $\mathbf{A}_j(\mathbf{x}) = \mathbf{x}$ corresponds to the factor exposure constraint in (4):

$$\mathbf{A}_j(\mathbf{x})'\boldsymbol{\omega} = b_j \implies \mathbf{x}'\boldsymbol{\omega} = b_j.^{12} \tag{13}$$

As another example, we consider exclusionary investing based on values of $\mathbf{x}$. If we rank assets by their values of $\mathbf{x}$, denote by $\mathbf{e}_{[k:N]}$ the unit vector that takes the value 1 only in the entry corresponding to the k-th ranked asset and zero otherwise, and $\omega_{[k:N]}$ the portfolio weight of the k-th ranked asset. With this notation, $\mathbf{A}_j(\mathbf{x}) = \begin{pmatrix} \mathbf{e}_{[1:N]} & \cdots & \mathbf{e}_{[N-N_0:N]} \end{pmatrix}'$

[12] In the case of inequality constraints, this becomes $\mathbf{x}'\boldsymbol{\omega} \geq b_j$.

corresponds to exclusionary investing in (3) that only keeps the top N_0 assets ranked by $\mathbf{x}$:

$$\mathbf{A}_j(\mathbf{x})'\boldsymbol{\omega} = 0 \implies \mathbf{e}'_{[1:N]}\boldsymbol{\omega} = 0, \cdots, \mathbf{e}'_{[N-N_0:N]}\boldsymbol{\omega} = 0 \implies \omega_{[1:N]} = 0, \cdots, \omega_{[N-N_0:N]} = 0.$$

More generally, investors can impose J constraints based on J different characteristics, which we denote by the vector:

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

where $\mathbf{x}_j$ represents the j-th characteristic that forms the j-th constraint, $\mathbf{A}_j(\mathbf{x}_j)$. We use the notation:

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

to denote the $J \times N$ coefficient matrix of J constraints that each depend on one characteristic. We assume the following about the distribution of asset characteristics, $\mathbf{X}$:

Assumption 1. *The characteristics X_t are identically distributed over time $t = 1, 2, \cdots$, and described by the probability distribution function $\Phi(X)$. In addition, X_t have finite moments up to order 2.*

Assumption 1 simply requires that X_t are drawn from some well-behaved distribution, and still allows X_t to be dependent over time.

Investors observe the characteristics $\mathbf{X}$ at the time of portfolio construction, but not the returns $\mathbf{r}$. If constraints are formed from values of $\mathbf{X}$, which may contain information of $\mathbf{r}$, how can we then attribute portfolio holdings and performance metrics to constraints?

3.2 Attribution with Information

We use $\mathbf{r}|\mathbf{X}$ 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, 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 due to 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* $\mathbf{A}(\mathbf{X})$*, the following decompositions hold for the optimal portfolio* $\boldsymbol{\omega}^*$*.*

1. *Expected return decomposition.*

$$\mathbb{E}\left[\boldsymbol{\omega}^{*\prime}\mathbf{r}|\mathbf{X}\right] = \boldsymbol{\mu}_{\mathbf{X}}^{\prime}\boldsymbol{\omega}^* = \boldsymbol{\mu}_{\mathbf{X}}^{\prime}\boldsymbol{\omega}_{\mathrm{MVO}} + \boldsymbol{\mu}^{\prime}\boldsymbol{\omega}_{\mathrm{CSTR}} + (\boldsymbol{\mu}_{\mathbf{X}}^{\prime} - \boldsymbol{\mu}^{\prime})\boldsymbol{\omega}_{\mathrm{CSTR}}, \tag{14}$$

where

- $\boldsymbol{\mu}_{\mathbf{X}}^{\prime}\boldsymbol{\omega}_{\mathrm{MVO}} = \boldsymbol{\mu}_{\mathbf{X}}^{\prime}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$*: expected return of the unconstrained MVO portfolio.*
- $\boldsymbol{\mu}^{\prime}\boldsymbol{\omega}_{\mathrm{CSTR}} = -\boldsymbol{\mu}^{\prime}\boldsymbol{\Sigma}^{-1}\mathbf{A}(\mathbf{X})^{\prime}\boldsymbol{\lambda}^*$*: components attributable to each constraint treated as static.*
- $(\boldsymbol{\mu}_{\mathbf{X}}^{\prime} - \boldsymbol{\mu}^{\prime})\boldsymbol{\omega}_{\mathrm{CSTR}} = -(\boldsymbol{\mu}_{\mathbf{X}}^{\prime} - \boldsymbol{\mu}^{\prime})\boldsymbol{\Sigma}^{-1}\mathbf{A}(\mathbf{X})^{\prime}\boldsymbol{\lambda}^*$*: components attributable to information in constraints.*[13]

Here the Lagrange multipliers are given by:

$$\boldsymbol{\lambda}^* = \left(\mathbf{A}(\mathbf{X})\boldsymbol{\Sigma}^{-1}\mathbf{A}(\mathbf{X})^{\prime}\right)^{-1}\left(\mathbf{A}(\mathbf{X})\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \mathbf{b}\right) \tag{15}$$

provided that the feasible region of the constrained optimization problem is nonempty.[14]

2. *Expected utility decomposition.*

$$\begin{aligned}\boldsymbol{\mu}_{\mathbf{X}}^{\prime}\boldsymbol{\omega}^* - \frac{1}{2}\boldsymbol{\omega}^{*\prime}\boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{\omega}^* = \ & \boldsymbol{\mu}_{\mathbf{X}}^{\prime}\boldsymbol{\omega}_{\mathrm{MVO}} - \frac{1}{2}\boldsymbol{\omega}_{\mathrm{MVO}}^{\prime}\boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{\omega}_{\mathrm{MVO}} \\ & - \frac{1}{2}\boldsymbol{\omega}_{\mathrm{CSTR}}^{\prime}\boldsymbol{\Sigma}\boldsymbol{\omega}_{\mathrm{CSTR}} \\ & + (\boldsymbol{\mu}_{\mathbf{X}}^{\prime} - \boldsymbol{\mu}^{\prime})\boldsymbol{\omega}_{\mathrm{CSTR}} - \boldsymbol{\omega}_{\mathrm{SHR}}^{\prime}(\boldsymbol{\Sigma}_{\mathbf{X}} - \boldsymbol{\Sigma})\boldsymbol{\omega}_{\mathrm{CSTR}}.\end{aligned} \tag{17}$$

[13]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 3.3.

[14]In the special case of only one constraint, the Lagrange multiplier reduces to:

$$\lambda^* = \frac{\mathbf{A}(\mathbf{x})\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - b}{\mathbf{A}(\mathbf{x})\boldsymbol{\Sigma}^{-1}\mathbf{A}(\mathbf{x})^{\prime}}. \tag{16}$$

The sign of λ^* is determined by whether the constraint is binding, i.e., the average characteristic value (e.g. ESG score) of the unconstrained MVO portfolio subtracted by the constraint threshold b.

- $\boldsymbol{\mu}'_{\mathbf{X}}\boldsymbol{\omega}_{\text{MVO}} - \frac{1}{2}\boldsymbol{\omega}'_{\text{MVO}}\boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{\omega}_{\text{MVO}} = \boldsymbol{\mu}'_{\mathbf{X}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{1}{2}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right)'\boldsymbol{\Sigma}_{\mathbf{X}}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right)$: *optimal expected utility of the unconstrained MVO portfolio.*
- $-\frac{1}{2}\boldsymbol{\omega}'_{\text{CSTR}}\boldsymbol{\Sigma}\boldsymbol{\omega}_{\text{CSTR}} = -\frac{1}{2}\boldsymbol{\lambda}^{*'}\mathbf{A}(\mathbf{X})\boldsymbol{\Sigma}^{-1}\mathbf{A}(\mathbf{X})'\boldsymbol{\lambda}^*$: *components attributable to all constraints combined together, treated as static.*
- $(\boldsymbol{\mu}'_{\mathbf{X}} - \boldsymbol{\mu}')\boldsymbol{\omega}_{\text{CSTR}} - \boldsymbol{\omega}'_{\text{SHR}}(\boldsymbol{\Sigma}_{\mathbf{X}} - \boldsymbol{\Sigma})\boldsymbol{\omega}_{\text{CSTR}} = -(\boldsymbol{\mu}'_{\mathbf{X}} - \boldsymbol{\mu}')\boldsymbol{\Sigma}^{-1}\mathbf{A}(\mathbf{X})'\boldsymbol{\lambda}^* - \left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{\Sigma}^{-1}\mathbf{A}(\mathbf{X})'\boldsymbol{\lambda}^*\right)'(\boldsymbol{\Sigma}_{\mathbf{X}} - \boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\mathbf{A}(\mathbf{X})'\boldsymbol{\lambda}^*$: *component attributable to information in constraints.*[15]

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

$$\boldsymbol{\omega}_{\text{SHR}} \equiv \boldsymbol{\omega}_{\text{MVO}} + \frac{1}{2}\boldsymbol{\omega}_{\text{CSTR}} = \boldsymbol{\omega}^* - \frac{1}{2}\boldsymbol{\omega}_{\text{CSTR}}. \tag{18}$$

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

$$\begin{aligned}\text{Expected Return or Utility} = &\ \text{Unconstrained MVO Portfolio} \\ &+ \text{ Static Constraint } + \text{ Information }.\end{aligned} \tag{19}$$

This result 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 the 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, $\boldsymbol{\mu}'_{\mathbf{X}} - \boldsymbol{\mu}'$.

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

$$\begin{aligned}-\boldsymbol{\omega}_{\text{SHR}}(\boldsymbol{\Sigma}_{\mathbf{X}} - \boldsymbol{\Sigma})\boldsymbol{\omega}_{\text{CSTR}} &= -\left(\boldsymbol{\omega}'_{\text{MVO}} + \frac{1}{2}\boldsymbol{\omega}'_{\text{CSTR}}\right)(\boldsymbol{\Sigma}_{\mathbf{X}} - \boldsymbol{\Sigma})\boldsymbol{\omega}_{\text{CSTR}} \\ &= -\frac{1}{2}\boldsymbol{\omega}'_{\text{CSTR}}(\boldsymbol{\Sigma}_{\mathbf{X}} - \boldsymbol{\Sigma})\boldsymbol{\omega}_{\text{CSTR}} - \boldsymbol{\omega}'_{\text{MVO}}(\boldsymbol{\Sigma}_{\mathbf{X}} - \boldsymbol{\Sigma})\boldsymbol{\omega}_{\text{CSTR}}.\end{aligned} \tag{20}$$

[15] 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 3.3.

The first term corresponds to the excess variance, $\mathbf{\Sigma}_{\mathbf{X}} - \mathbf{\Sigma}$, 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, (20) 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, $-(\mathbf{\Sigma}_{\mathbf{X}} - \mathbf{\Sigma})$.[16] Figure 1 demonstrates two representative examples of the shrinkage portfolio, which shrinks the optimal constrained portfolio $\boldsymbol{\omega}^*$ towards the unconstrained MVO portfolio $\boldsymbol{\omega}_{MVO}$. We emphasize that the coefficient of the variance term is 1/2 in our expected utility which corresponds to a specific level of risk aversion. More generally, different levels of risk aversion correspond to different shrinkage portfolios in (18).

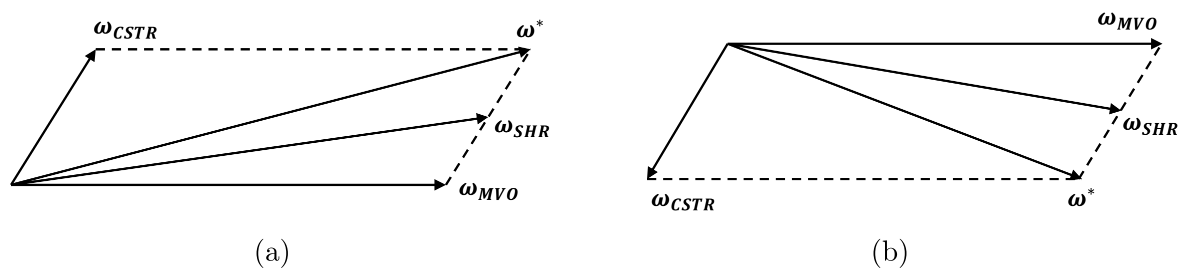


Figure 1: Illustration of shrinkage portfolios defined in (18). (a) shows an example with an acute angle between the shrinkage portfolio and the portfolio attributable to constraints. (b) shows an example with an obtuse angle between the shrinkage portfolio and the portfolio attributable to constraints.

It is important to note that the results in Proposition 2 are conditional on $\mathbf{X}$. To obtain an unconditional decomposition over multiple time periods, we must compute the expectation of (14) and (17) with respect to $\mathbf{X}$. This does not change the decompositions in Proposition 2 because they are linear, and the unconditional expected return and utility are simple generalizations of the decompositions conditioned on $\mathbf{X}$. Nonetheless, we summarize the unconditional results formally.

Proposition 3 (Attribution with Information)**.** *Under Assumption 1, the unconditional expected return and utility can be decomposed into components that are attributable to the unconstrained MVO portfolio, static constraints, and information, respectively.*

[16]To see this, imagine a set of N hypothetical assets whose returns, $\mathbf{s}$, have a covariance matrix $\mathbf{\Sigma} - \mathbf{\Sigma}_{\mathbf{X}}$. We have $\text{Cov}\left(\boldsymbol{\omega}_{\text{SHR}}'\mathbf{s}, \boldsymbol{\omega}_{\text{CSTR}}'\mathbf{s}\right) = \boldsymbol{\omega}_{\text{SHR}}'\text{Cov}(\mathbf{s},\mathbf{s})\boldsymbol{\omega}_{\text{CSTR}} = \boldsymbol{\omega}_{\text{SHR}}'\left(\mathbf{\Sigma} - \mathbf{\Sigma}_{\mathbf{X}}\right)\boldsymbol{\omega}_{\text{CSTR}}$. We show in Section 3.3 and Appendix B.1 that $\mathbf{\Sigma} - \mathbf{\Sigma}_{\mathbf{X}}$ is always positive semi-definite under certain distributional assumptions of $\mathbf{r}$ and $\mathbf{X}$.

1. *Expected return decomposition.*

$$\mathbb{E}\left[\boldsymbol{\omega}^{*\prime}\mathbf{r}\right] = \boldsymbol{\mu}'\boldsymbol{\omega}_{\text{MVO}} + \boldsymbol{\mu}'\mathbb{E}\left[\boldsymbol{\omega}_{\text{CSTR}}\right] + \mathbb{E}\left[(\boldsymbol{\mu}'_{\mathbf{X}} - \boldsymbol{\mu}')\boldsymbol{\omega}_{\text{CSTR}}\right]. \tag{21}$$

2. *Expected utility decomposition.*

$$\begin{aligned}\mathbb{E}\left[\boldsymbol{\omega}^{*\prime}\mathbf{r}\right] - \frac{1}{2}\text{Var}\left(\boldsymbol{\omega}^{*\prime}\mathbf{r}\right) = \boldsymbol{\mu}'\boldsymbol{\omega}_{\text{MVO}} &- \frac{1}{2}\boldsymbol{\omega}'_{\text{MVO}}\boldsymbol{\Sigma}\boldsymbol{\omega}_{\text{MVO}} \\ &- \frac{1}{2}\mathbb{E}\left[\boldsymbol{\omega}'_{\text{CSTR}}\boldsymbol{\Sigma}\boldsymbol{\omega}_{\text{CSTR}}\right] \\ &+ \mathbb{E}\left[(\boldsymbol{\mu}'_{\mathbf{X}} - \boldsymbol{\mu}')\boldsymbol{\omega}_{\text{CSTR}} - \boldsymbol{\omega}'_{\text{SHR}}(\boldsymbol{\Sigma}_{\mathbf{X}} - \boldsymbol{\Sigma})\boldsymbol{\omega}_{\text{CSTR}}\right] \\ &- \frac{1}{2}\left(\text{Var}\left(\boldsymbol{\omega}'_{\text{CSTR}}\boldsymbol{\mu}_{\mathbf{X}}\right) + 2\text{Cov}\left(\boldsymbol{\omega}'_{\text{MVO}}\boldsymbol{\mu}_{\mathbf{X}}, \boldsymbol{\omega}'_{\text{CSTR}}\boldsymbol{\mu}_{\mathbf{X}}\right)\right).\end{aligned} \tag{22}$$

3.3 Decomposing Information with Normally Distributed Returns

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 due to information in $\mathbf{X}$. They cannot be simplified further for general distributions of $\mathbf{r}$ and $\mathbf{X}$. However, in the special case of conditional normality and independent characteristics, we can decompose these terms further and obtain considerable intuition about their contribution.

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

2. *The return vector is homoskedastic with variance* $\sigma_{\mathbf{r}}$*.*

3. *Each characteristic is homoskedastic with variance* $\sigma_{\mathbf{x}_j}$ *for* $j = 1, 2, \ldots, J$*.*

4. *The characteristic values are independent both across different assets and between the* J *different constraints.*

5. *For the* j*-th constraint, there is a homogeneous correlation between the return and characteristic value of each asset, 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*

$$\text{Cov}(\mathbf{r}, \mathbf{x}_j) = \rho_j \sigma_{\mathbf{r}} \sigma_{\mathbf{x}_j} \mathbf{I}$$

where $\mathbf{I}$ *is the identity matrix.*

We make a few remarks about Assumption 2. First, although results in this section are derived under joint normality of Assumption 2.1, we provide a generalization in Section 4.2 when asset characteristics $\mathbf{x}$ are binary-valued Bernoulli random variables, which is useful for capturing exclusionary investing based on binary labels, such as the industry of a company.

Second, Assumption 2.2 specifies homoskedastic asset returns, which are more suitable for residual returns in excess of a particular asset-pricing model. Our empirical analysis in Section 5 analyzes the residual returns in excess of the Fama-French five-factor model.

Third, Assumption 2.4 asserts that multiple asset characteristics are independent of each other, which makes it possible to decompose the contribution from each constraint in mathematically simple forms (see Propositions 4–5). More generally, when characteristics are dependent, a similar decomposition is still possible in a rotated space that orthogonalizes $\mathbf{X}$, rather than the original space of $\mathbf{X}$.

Finally, Assumption 2.5 specifies that, given a particular constraint, the correlation between the return and characteristic value of the i-th asset, $\text{Corr}(r_i, x_i) = \rho$, is a constant for $i = 1, 2, \ldots, N$, and there are no cross-correlations between r_i and x_j when $i \neq j$. It is worth noting, though, that the returns $\mathbf{r}$ are allowed to be dependent cross-sectionally. This assumption was first used in Lo and MacKinlay (1990) to describe cross-sectional estimation errors of intercepts in CAPM regressions, and later in Lo and Zhang (2023) to describe the dependence structure between returns and an impact factor such as ESG.

We emphasize that Assumptions 2.2, 2.3, and 2.5 are imposed to allow for a more intuitive explanation of the attribution results below. All our results remain valid, albeit with more complicated mathematical expressions, when returns and characteristics are allowed to be dependent in more general forms. We choose to present more intuitive results using stronger assumptions, and provide the general results in the Appendix. Nevertheless, none of our empirical results in Section 5 relies on these assumptions and our attribution framework can easily be carried out numerically under general distributional assumptions.

Given Assumption 2, the covariance matrix of $[\,\mathbf{r}' \ \mathbf{x}_1' \ \cdots \ \mathbf{x}_J'\,]$ can be written as:

$$\begin{pmatrix} \boldsymbol{\Sigma} & \rho_1\sigma_{\mathbf{r}}\sigma_{\mathbf{x}_1}\mathbf{I} \ \cdots \ \rho_J\sigma_{\mathbf{r}}\sigma_{\mathbf{x}_J}\mathbf{I} \\ & \\ \rho_1\sigma_{\mathbf{r}}\sigma_{\mathbf{x}_1}\mathbf{I} & \begin{pmatrix} \sigma_{\mathbf{x}_1}^2\mathbf{I} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_{\mathbf{x}_J}^2\mathbf{I} \end{pmatrix} \\ \vdots & \\ \rho_J\sigma_{\mathbf{r}}\sigma_{\mathbf{x}_J}\mathbf{I} & \end{pmatrix}. \tag{23}$$

Recall that $\mathbf{X}$ represents the $(NJ \times 1)$-dimensional vector $[\mathbf{x}_1' \ \cdots \ \mathbf{x}_J']'$, and we use $\boldsymbol{\nu} \equiv [\,\boldsymbol{\nu}_1' \ \cdots \ \boldsymbol{\nu}_J'\,]'$ to denote the expected value of $\mathbf{X}$. The following result characterizes the *excess* return and covariance due to the information in $\mathbf{X}$.

Proposition 4 (Information Decomposition)**.** *Under Assumptions 1 and 2,* $\mathbf{r}|\mathbf{X}$ *is normally distributed, with an expected value given by:*

$$\boldsymbol{\mu}_{\mathbf{X}} = \mathbb{E}[\mathbf{r}|\mathbf{X}] = \boldsymbol{\mu} + \sum_{j=1}^{J} \rho_j \sigma_{\mathbf{r}} \frac{(\mathbf{x}_j - \boldsymbol{\nu}_j)}{\sigma_{\mathbf{x}_j}}, \tag{24}$$

and a covariance matrix given by:

$$\boldsymbol{\Sigma}_{\mathbf{X}} = \mathrm{Cov}(\mathbf{r}|\mathbf{X}) = \boldsymbol{\Sigma} - \sum_{j=1}^{J} \rho_j^2 \sigma_{\mathbf{r}}^2 \mathbf{I}. \tag{25}$$

Proposition 4 provides a decomposition of the excess return and excess covariance due to information into contributions from each of the J constraints. It also allows for more explicit decompositions of the expected return and utility of the constrained portfolio by substituting (24)–(25) into Proposition 2, which we summarize below.

Proposition 5 (Attribution with Normally Distributed Returns)**.** *Under Assumptions 1 and 2 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.*

$$\mathbb{E}\left[\boldsymbol{\omega}^{*\prime}\mathbf{r}|\mathbf{X}\right] = \boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}^* = \boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}_{\mathrm{MVO}} + \boldsymbol{\mu}'\boldsymbol{\omega}_{\mathrm{CSTR}} + \sum_{j=1}^{J} \rho_j \sigma_{\mathbf{r}} \frac{(\mathbf{x}_j' - \boldsymbol{\nu}_j')\boldsymbol{\omega}_{\mathrm{CSTR}}}{\sigma_{\mathbf{x}_j}}, \tag{26}$$

where the Lagrange multipliers are given by (15) 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.*
- $\rho_j \sigma_{\mathbf{r}} \frac{(\mathbf{x}_j' - \boldsymbol{\nu}_j')\boldsymbol{\omega}^*}{\sigma_{\mathbf{x}_j}}$*: component attributable to information in the* j*-th constraint.*

2. *Expected utility decomposition.*

$$\begin{aligned}\boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}^* - \frac{1}{2}\boldsymbol{\omega}^{*\prime}\boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{\omega}^* = \boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}_{\mathrm{MVO}} &- \frac{1}{2}\boldsymbol{\omega}_{\mathrm{MVO}}'\boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{\omega}_{\mathrm{MVO}} \\ &- \frac{1}{2}\boldsymbol{\omega}_{\mathrm{CSTR}}'\boldsymbol{\Sigma}\boldsymbol{\omega}_{\mathrm{CSTR}} \\ &+ \sum_{j=1}^{J}\left(\rho_j \sigma_{\mathbf{r}} \frac{(\mathbf{x}_j' - \boldsymbol{\nu}_j')\boldsymbol{\omega}_{\mathrm{CSTR}}}{\sigma_{\mathbf{x}_j}} + \rho_j^2 \sigma_{\mathbf{r}}^2 \boldsymbol{\omega}_{\mathrm{SHR}}'\boldsymbol{\omega}_{\mathrm{CSTR}}\right).\end{aligned} \tag{27}$$

- $\boldsymbol{\mu}'_{\mathbf{X}}\boldsymbol{\omega}_{\mathrm{MVO}} - \frac{1}{2}\boldsymbol{\omega}'_{\mathrm{MVO}}\boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{\omega}_{\mathrm{MVO}}$: *optimal expected utility of the unconstrained MVO portfolio.*
- $-\frac{1}{2}\boldsymbol{\omega}'_{\mathrm{CSTR}}\boldsymbol{\Sigma}\boldsymbol{\omega}_{\mathrm{CSTR}}$: *components attributable to all constraints combined together, treated as static.*
- $\rho_j\sigma_{\mathbf{r}}\frac{(\mathbf{x}'_j-\boldsymbol{\nu}'_j)\boldsymbol{\omega}_{\mathrm{CSTR}}}{\sigma_{\mathbf{x}_j}} + \rho_j^2\sigma_{\mathbf{r}}^2\boldsymbol{\omega}'_{\mathrm{SHR}}\boldsymbol{\omega}_{\mathrm{CSTR}}$: *component attributable to information in the j-th constraint.*

Furthermore, the unconditional expected return and utility can be decomposed into components that are attributable to the unconstrained MVO portfolio, static constraints, and information, respectively, by following Proposition 3.

The last term of (26) (see also (24)) 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, i.e., opportunity in the market. The third term, $(\mathbf{x}_j - \boldsymbol{\nu}_j)/\sigma_{\mathbf{x}_j}$, determines whether each asset's characteristic value is above or below the average characteristic value, much like 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.

The last term of (27) (see also (25)) shows that the excess covariance, $(\boldsymbol{\Sigma}_{\mathbf{X}} - \boldsymbol{\Sigma})$, is always negative. In other words, when $\boldsymbol{\omega}'_{\mathrm{SHR}}\boldsymbol{\omega}_{\mathrm{CSTR}} > 0$, i.e., the shrinkage portfolio holdings and the holdings attributable to constraints are positively correlated, incorporating information in $\mathbf{X}$ always reduces the variance and improves the expected utility of a portfolio. In addition, the magnitude of reduction in variance due to the j-th constraint depends on ρ_j^2, the squared correlation between the j-th asset characteristics and returns. The constraints with a larger magnitude of correlation lead to larger reductions in variance. On the other hand, when the shrinkage portfolio holdings and the holdings attributable to constraints are negatively correlated, 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.

3.4 Ex Post Return Attribution

Propositions 2–5 provide a theoretical framework to decompose expected returns. In practice, investors can also use this framework to decompose realized returns *ex post*, as we show in this section.

We use an $(N \times 1)$-vector $\tilde{\mathbf{r}}$ to represent the realized returns of all assets. The goal for ex post attribution is to decompose the realized portfolio return, $\tilde{\mathbf{r}}'\boldsymbol{\omega}^*$, into components attributable to the unconstrained MVO portfolio and each constraint.

If we treat constraints as static, (10) in Proposition 1 already provides such a decomposition as long as the *ex ante* expected returns are replaced by *ex post* realized returns:

$$\tilde{\mathbf{r}}'\boldsymbol{\omega}^* = \tilde{\mathbf{r}}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \tilde{\mathbf{r}}'\boldsymbol{\Sigma}^{-1}\mathbf{A}'\boldsymbol{\lambda}^*. \tag{28}$$

- $\tilde{\mathbf{r}}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$: realized return that the unconstrained MVO portfolio would have achieved.
- $-\tilde{\mathbf{r}}'\boldsymbol{\Sigma}^{-1}\mathbf{A}'\boldsymbol{\lambda}^*$: realized return attributable to constraints.

However, this decomposition does not account for the information contained in each constraint. Equation (24) in Proposition 4 quantifies the excess return due to information, and we use the sample version of this decomposition to quantify realized returns attributable to information:

$$\tilde{\mathbf{r}}_{\text{Info}} = \sum_{j=1}^{J} \frac{\rho(\tilde{\mathbf{r}}, \tilde{\mathbf{x}}_j)\tilde{\sigma}_{\mathbf{r}}(\tilde{\mathbf{x}}_j - \tilde{\boldsymbol{\nu}}_j)}{\tilde{\sigma}_{\mathbf{x}_j}}. \tag{29}$$

We can therefore define the static returns as $\tilde{\mathbf{r}}_{\text{Static}} \equiv \tilde{\mathbf{r}} - \tilde{\mathbf{r}}_{\text{Info}}$. This leads to the following decomposition of realized portfolio returns.

Proposition 6 (Ex Post Return Attribution)**.** *Under Assumptions 1 and 2, realized portfolio returns can be decomposed into:*

$$\tilde{\mathbf{r}}'\boldsymbol{\omega}^* = \tilde{\mathbf{r}}'\boldsymbol{\omega}_{\text{MVO}} + \tilde{\mathbf{r}}'_{\text{Static}}\boldsymbol{\omega}_{\text{CSTR}} + \sum_{j=1}^{J} \rho(\tilde{\mathbf{r}}, \tilde{\mathbf{x}}_j)\tilde{\sigma}_{\mathbf{r}} \frac{(\tilde{\mathbf{x}}'_j - \tilde{\boldsymbol{\nu}}'_j)\boldsymbol{\omega}_{\text{CSTR}}}{\tilde{\sigma}_{\mathbf{x}_j}}. \tag{30}$$

- $\tilde{\mathbf{r}}'\boldsymbol{\omega}_{\text{MVO}} = \tilde{\mathbf{r}}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$*: realized return of the unconstrained MVO portfolio.*
- $\tilde{\mathbf{r}}'_{\text{Static}}\boldsymbol{\omega}_{\text{CSTR}} = -\tilde{\mathbf{r}}'_{\text{Static}}\boldsymbol{\Sigma}^{-1}\mathbf{A}'\boldsymbol{\lambda}^*$*: realized return attributable to constraints treated as static.*
- $\rho(\tilde{\mathbf{r}}, \tilde{\mathbf{x}}_j)\tilde{\sigma}_{\mathbf{r}}\frac{(\tilde{\mathbf{x}}'_j - \tilde{\boldsymbol{\nu}}'_j)\boldsymbol{\omega}_{\text{CSTR}}}{\tilde{\sigma}_{\mathbf{x}_j}} = -\rho(\tilde{\mathbf{r}}, \tilde{\mathbf{x}}_j)\tilde{\sigma}_{\mathbf{r}}\frac{(\tilde{\mathbf{x}}'_j - \tilde{\boldsymbol{\nu}}'_j)\boldsymbol{\Sigma}^{-1}\mathbf{A}'\boldsymbol{\lambda}^*}{\tilde{\sigma}_{\mathbf{x}_j}}$*: realized return attributable to information in the j-th constraint.*

It is worth noting that in the last term of (30), the information component contains two terms. The first term reflects the correlation of each characteristic with returns, which captures the information content of each characteristic. The second term reflects the portfolio holdings attributable to each constraint, which captures how information is realized into actual returns. As a result, there are interactions from the information contained in each constraint with the portfolio holdings attributable to other constraints. Together, they determine the information contribution to the realized returns.

4 Common Examples of Portfolio Constraints

In this section, we apply the results in Section 3 and consider two common examples of portfolio constraints: factor exposures and exclusions. We derive additional analytical results for these special types of constraints and provide simulations to illustrate the attribution of expected returns.

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. This corresponds to $\mathbf{A}(\mathbf{x}) = \mathbf{x}'$ in (13).

We consider the case of one single constraint $\mathbf{A}(\mathbf{x}) = \mathbf{x}'$ for simplicity, and focus on the last term in (26) of Proposition 5 that corresponds to the information component. Given a scalar Lagrange multiplier as in (16), we have the following result.

Proposition 7 (Factor Exposure). *Under Assumptions 1 and 2, and assuming without loss of generality that the cross-sectional average factor value* $\boldsymbol{\nu} = 0$*, the expected return of the optimal portfolio with a factor exposure constraint can be decomposed into:*

$$\mathbb{E}\left[\boldsymbol{\omega}^{*\prime}\mathbf{r}|\mathbf{x}\right] = \boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \frac{\mathbf{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\mathbf{x}'\boldsymbol{\Sigma}^{-1}\mathbf{x}}\left(b - \mathbf{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right) + \frac{\rho\sigma_{\mathbf{r}}}{\sigma_{\mathbf{x}}}\left(b - \mathbf{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right), \tag{31}$$

if the constraint is binding. In the case of a non-binding inequality constraint, the Lagrange multiplier $\lambda^* = 0$*, and (31) reduces to:* $\mathbb{E}\left[\boldsymbol{\omega}^{*\prime}\mathbf{r}|\mathbf{x}\right] = \boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$.

Proposition 7 provides the same decomposition of expected returns into three components as in Proposition 5, except that the last term attributable to information has a much simpler form. First of all, 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, note that $\mathbf{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ is the characteristic value of the unconstrained MVO portfolio. Therefore $b - \mathbf{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ measures the constrained characteristic value b *relative to* the unconstrained MVO portfolio. For example, if a particular asset characteristic, such as ESG, is positively correlated with returns ($\rho > 0$), a positive desired ESG level relative to the unconstrained MVO portfolio ($b - \mathbf{x}'\mathbf{\Sigma}^{-1}\boldsymbol{\mu} > 0$) adds value to expected returns. On the other hand, if ESG is negatively correlated with returns ($\rho < 0$), the same constraint will hurt expected returns. This intuition holds true for any asset characteristic such as value, size, or measures of momentum, as well as denizens of the "factor zoo" described in the recent literature (Harvey, Liu, and Zhu, 2016; Feng, Giglio, and Xiu, 2020; Hou, Xue, and Zhang, 2020).

Simulation. We consider a world with 10 assets whose expected return $\boldsymbol{\mu} = [\,\mu_1 \ \cdots \ \mu_{10}\,]'$ and covariance matrix $\mathbf{\Sigma} = (\sigma_{i,j})_{10\times 10}$ are randomly generated in the following way:

$$\mu_i \sim N(0.05, 0.05^2), \quad \text{for } i = 1, 2, \ldots, 10 \tag{32a}$$

$$\sigma_{i,i} \sim U[0, 0.01], \quad \text{for } i = 1, 2, \ldots, 10 \tag{32b}$$

$$\sigma_{i,j} \sim U[0, 0.001], \quad \text{for } i, j = 1, 2, \ldots, 10 \text{ and } i \neq j \tag{32c}$$

where N and U denote the normal and uniform distributions, respectively. Investors solve the following optimization problem with two constraints on factor exposures:

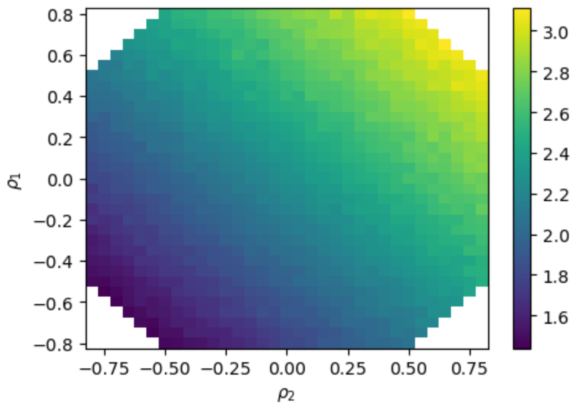
$$\begin{aligned} \max_{\boldsymbol{\omega}} \ & \boldsymbol{\omega}'\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{\omega}'\mathbf{\Sigma}\boldsymbol{\omega} \\ \text{s.t.} \ & \boldsymbol{\omega}'\mathbf{x} \geq 0.5 \\ & \boldsymbol{\omega}'\mathbf{y} \geq 0.5. \end{aligned} \tag{33}$$

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

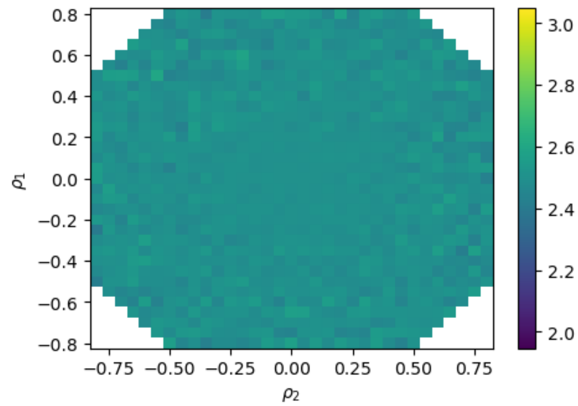
$$\rho_1 \equiv \text{Corr}(x_i, r_i) \quad \text{and} \quad \rho_2 \equiv \text{Corr}(y_i, r_i), \quad \text{for } i = 1, 2, \ldots, 10.$$

As ρ_1 and ρ_2 vary between -0.8 and 0.8, Figure 2 demonstrates the attribution of expected returns following Proposition 7. Figure 2a shows the expected return of the constrained portfolio, which varies between 1.6 to as high as 3.0 as ρ_1 and ρ_2 vary between -0.8 and 0.8. Figure 2b shows the expected return of the unconstrained MVO problem, which is a constant value around 2.5 regardless of values of ρ_1 and ρ_2.

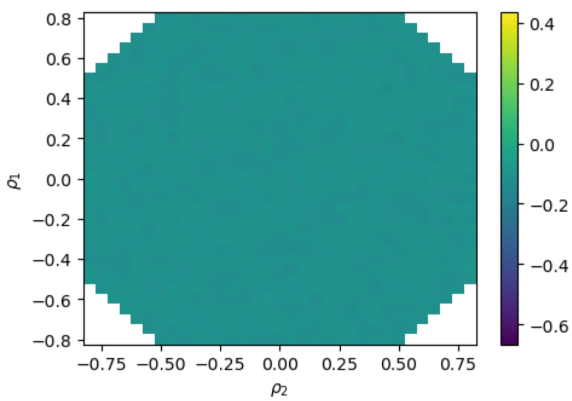
The source of the difference in expected returns between the unconstrained MVO and the constrained portfolio becomes clear in Figures 2c–2f. Figures 2c and 2d show the expected returns attributable to the two constraints, respectively, as if they are static. They each



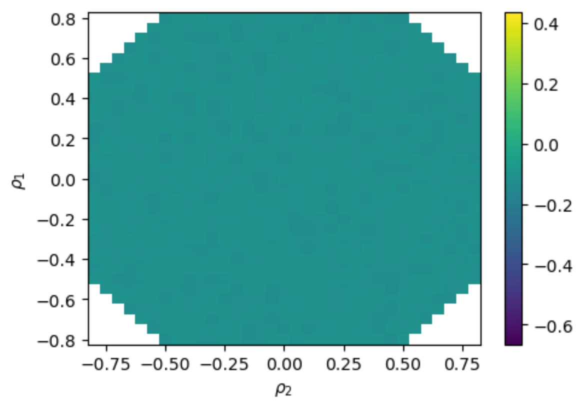
(a) Constrained Optimum



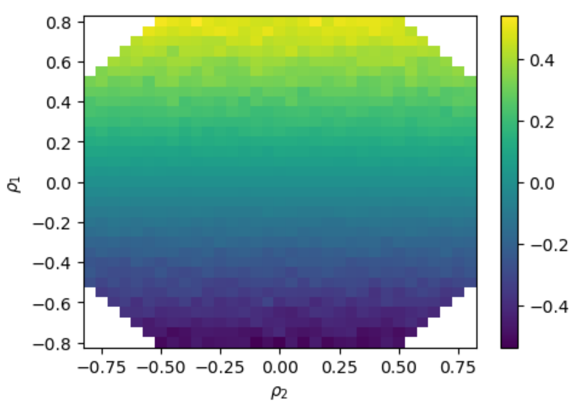
(b) Unconstrained MVO



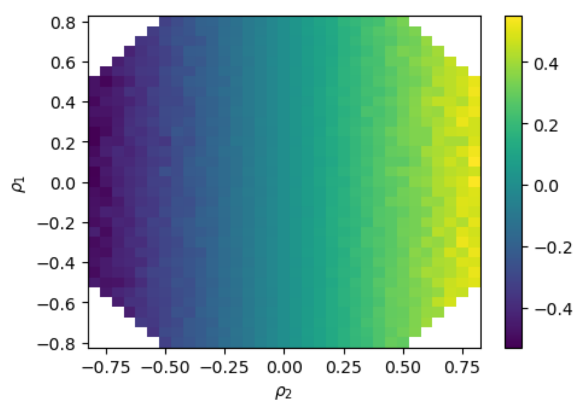
(c) Decomposition: Constraint 1 (Static)



(d) Decomposition: Constraint 2 (Static)



(e) Decomposition: Constraint 1 (Information)

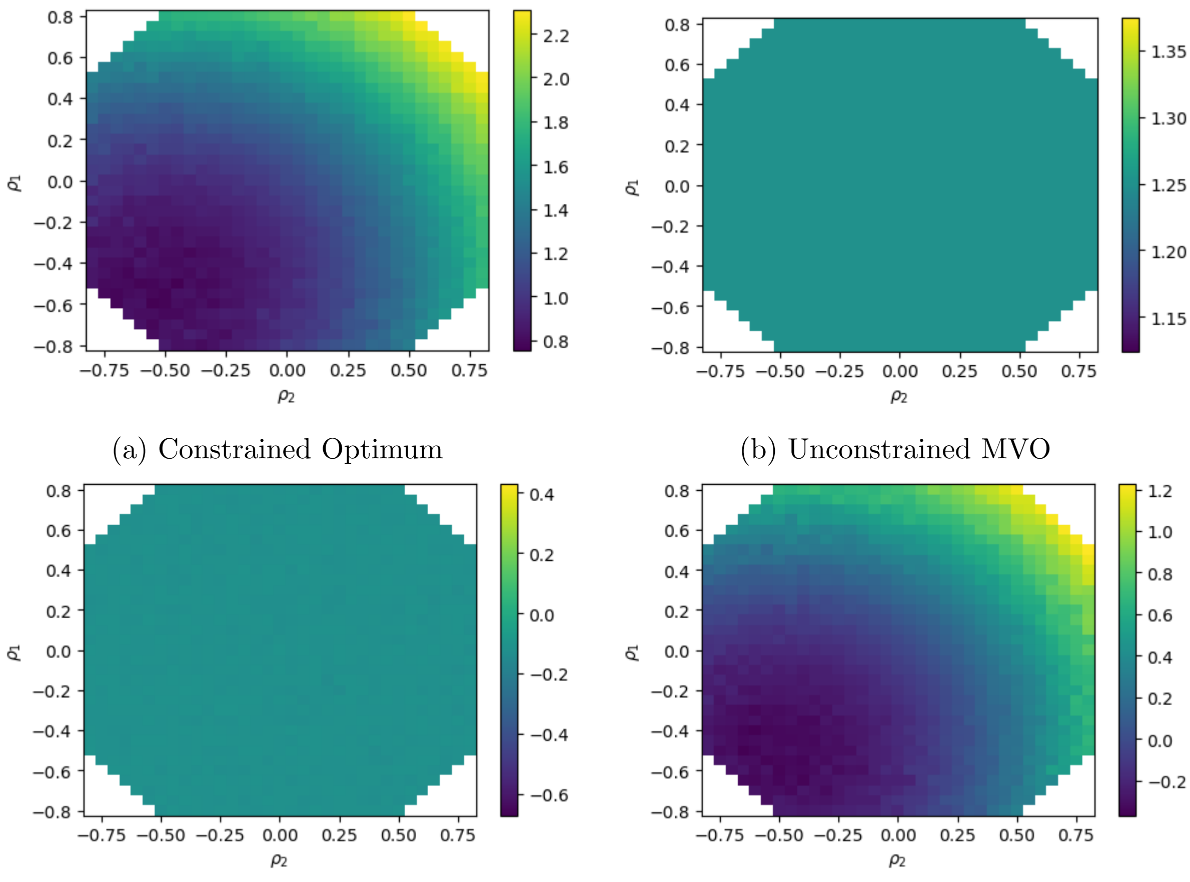


(f) Decomposition: Constraint 2 (Information)

Figure 2: Decomposition of expected return for (33) with two constraints which depend on random characteristics, as correlations (ρ_1 and ρ_2) between random characteristics and asset returns vary. The expected return of the constrained portfolio (a) is decomposed into components corresponding to the unconstrained MVO portfolio (b), static constraints (c–d), and information in the constraints (e–f).

contribute to the expected returns with a negative constant value of around -0.1. Figure 2e shows the expected returns attributable to information in the first constraint, which increase as ρ_1 increases but remain constant as ρ_2 varies. Similarly, Figure 2f shows the expected returns attributable to information in the second constraint, which increase as ρ_2 increases but remain constant as ρ_1 varies.

Similarly, Figure 3 demonstrates the attribution for expected utility. Figure 3a shows the expected utility of the constrained portfolio, which varies between 0.8 to as high as 2.2 as ρ_1 and ρ_2 vary between -0.8 and 0.8. Figure 3b shows the expected utility of the unconstrained MVO problem, which is around 1.25 regardless of values of ρ_1 and ρ_2. When ρ_1 and ρ_2 are high, the expected utility of the constrained portfolio can actually be higher than that of the unconstrained portfolio.



(a) Constrained Optimum

(b) Unconstrained MVO

(c) Decomposition: Constraint (Static)

(d) Decomposition: Constraint (Information)

Figure 3: Decomposition of expected utility for (33) with two constraints which depend on random characteristics, as correlations (ρ_1 and ρ_2) between random characteristics and asset returns vary. The expected utility 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).

The difference in expected utility between the unconstrained MVO and the constrained portfolio is decomposed into the static and information components, respectively. Figure 3c shows the expected utility attributable to the two constraints as if they are static, which contributes to the expected utility with a negative constant value of around -0.1. Figure 3d shows the expected utility attributable to information, which is negative in most regions marked by dark blue. As both ρ_1 and ρ_2 increase, the expected utility contribution from information increases.

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, where certain assets are excluded from the portfolio based on certain criteria such as a minimum level of ESG scores, or whether the firm belongs to a particular industry. Without loss of generality, suppose the j-th asset is excluded based on a characteristic $\mathbf{x}$.

For convenience, we introduce some new notation. We rank securities by $\mathbf{x}$, and use the notation $x_{1:N} < x_{2:N} < x_{N:N}$ to denote the ranked values of $\mathbf{x}$, also known as "order statistics" in the statistics literature. We then denote by $r_{[i:N]}$ the return associated with the i-th order statistic $x_{i:N}$. This return is called the i-th *induced order statistic* to emphasize the fact that its ranking is determined not by its own value but by the value of $\mathbf{x}$.[17] For simplicity, we also use the subscript "$[N]$" to denote a vector or a matrix that is reordered based on values of $\mathbf{x}$. For example, $\boldsymbol{\omega}_{[N]}$ represents the vector of weights for assets that are reordered based on values of $\mathbf{x}$.

Investors solve the following optimization problem in which the top N_0 assets ranked by $\mathbf{x}$ are allowed to enter the portfolio, while the bottom $N - N_0$ assets are excluded:

$$\begin{aligned} \min_{\boldsymbol{\omega}_{[N]}} \quad & \boldsymbol{\omega}'_{[N]}\boldsymbol{\mu}_{[N]} - \frac{1}{2}\boldsymbol{\omega}'_{[N]}\boldsymbol{\Sigma}_{[N]}\boldsymbol{\omega}_{[N]} \\ \text{s.t.} \quad & \omega_{[i:N]} = 0 \quad \text{for } i \leq N - N_0. \end{aligned} \tag{34}$$

The optimal portfolio for (34) is simply the optimal portfolio restricted to N_0 assets. There-

[17]The term was coined by Bhattacharya (1974). These indirectly ranked statistics are also referred to as *concomitants* of the order statistic $x_{i:N}$ (David, 1973).

fore, 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 (26) in Proposition 5. 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 8 (Exclusion based on Binary Characteristic)**.** *Under Assumptions 1 and 2, 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 found in (34) can be decomposed into:*

$$\mathbb{E}\left[\boldsymbol{\omega}^{*\prime}\mathbf{r}|\mathbf{x}\right] = \boldsymbol{\mu}'_{\mathbf{x},N_0}\boldsymbol{\omega}^*_{N_0} = \boldsymbol{\mu}'_{\mathbf{x}}\boldsymbol{\omega}_{\text{MVO}} + \boldsymbol{\mu}'\boldsymbol{\omega}_{\text{CSTR}} + \rho\sigma_{\mathbf{r}}\left(\mathbf{x}\odot\mathbf{u} - (1-\mathbf{x})\odot\mathbf{v}\right)'\boldsymbol{\omega}_{\text{CSTR}} \tag{35}$$

where

$$\mathbf{u} = \left(\sqrt{\frac{\pi_{x_1=0}}{\pi_{x_1=1}}},\ldots,\sqrt{\frac{\pi_{x_N=0}}{\pi_{x_N=1}}}\right)' \quad \textit{and} \quad \mathbf{v} = \left(\sqrt{\frac{\pi_{x_1=1}}{\pi_{x_1=0}}},\ldots,\sqrt{\frac{\pi_{x_N=1}}{\pi_{x_N=0}}}\right)'.$$

are two vectors of the odds ratio (that is, the relative chance) of each asset being excluded from the portfolio.

Proposition 8 shows that the component attributable to information depends on the correlation ρ between asset returns and characteristics $\mathbf{x}$, and the chance of being excluded from the portfolio, $\frac{\pi_{x_i=0}}{\pi_{x_i=1}}$. It is worth emphasizing that, in broad terms, the excess return attributable to information has two different interpretations. On the one hand, when $\mathbf{x}$ corresponds to characteristics such as ESG, size, or the book-to-market of a firm, the excess return represents the conditional return premium from the information in these characteristics. On the other hand, the excess return can also represent an investor's ex ante expectation of asset returns going forward.

A concrete example is so-called "stranded" assets such as fossil fuels. These assets may have had acceptable returns in the past, but "green" investors may believe that they will be subject to lower returns in the future which will be reflected in a decomposition of portfolio performance according to their beliefs about their future expected returns.

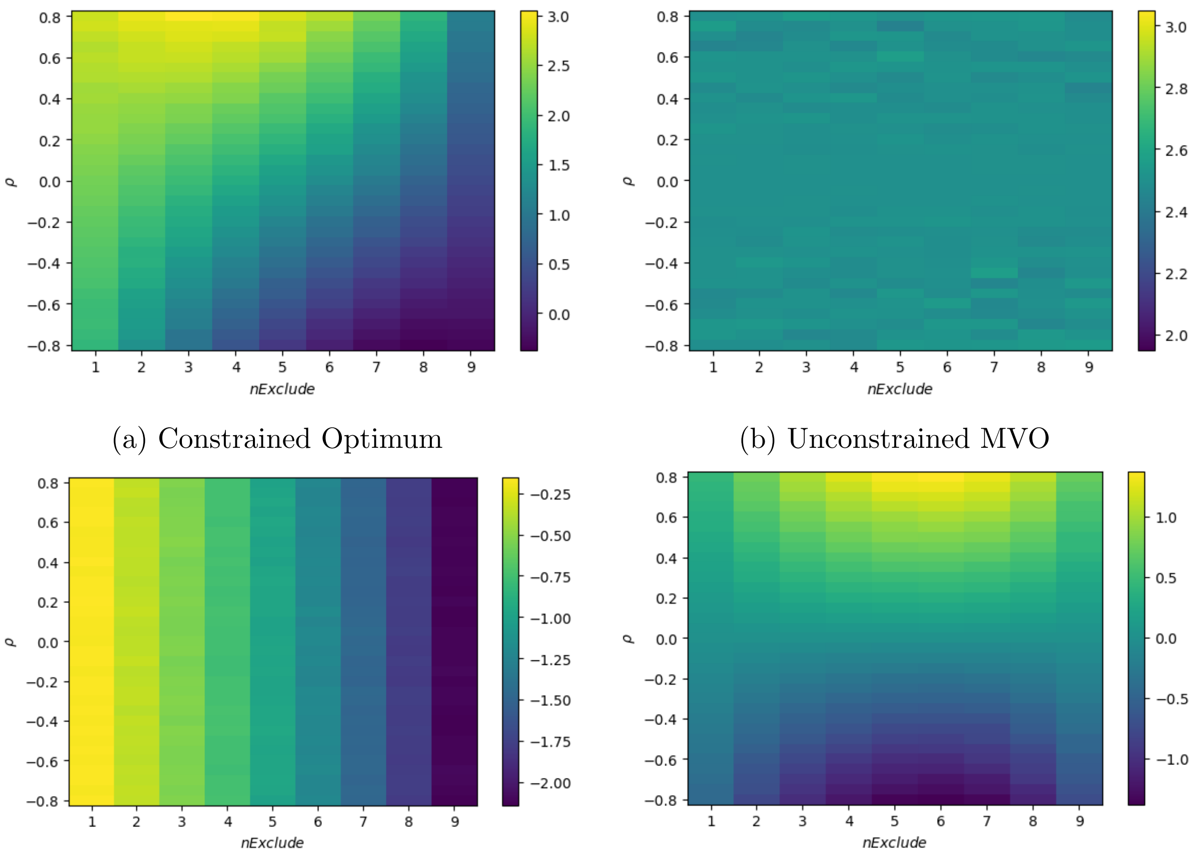
Simulation. We consider a world with 10 assets whose expected return $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ are given in the same way as in (32)–(32). Investors solve the following problem:

$$\begin{aligned} \max_{\boldsymbol{\omega}} \quad & \boldsymbol{\omega}'\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{\omega}'\boldsymbol{\Sigma}\boldsymbol{\omega} \\ \text{s.t.} \quad & \omega_{[i:10]} = 0, \text{ for } i \leq N - N_0 \text{ assets ordered by } \mathbf{x}. \end{aligned} \tag{36}$$

Here, $\mathbf{x}$ represents the asset characteristic (such as ESG score) that is used to exclude assets, a 10-dimensional $N(0, 1)$ random vector that is IID over time. We denote the correlations of the asset characteristics with returns by

$$\rho \equiv \mathrm{Corr}(x_i, r_i), \quad \text{for } i = 1, 2, \ldots, 10.$$

As ρ varies between -0.8 and 0.8 and the number of excluded assets varies between 1 and 9, Figure 4 demonstrates the attribution of expected returns following Proposition 8. Figure 4a shows the expected return of the constrained problem, which varies between -0.2 to as high as 3.0 as ρ and the number of excluded assets vary. Figure 4b shows the expected return of the unconstrained MVO portfolio, which has a constant value of around 2.5.



(a) Constrained Optimum

(b) Unconstrained MVO

(c) Decomposition: Constraint (Static)

(d) Decomposition: Constraint (Information)

Figure 4: Decomposition of expected return for problem (36) with one exclusionary constraint which depends on random characteristics, as the number of excluded assets (*nExclude*) and the correlation (ρ) between the random characteristic and asset returns vary. 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).

The source of the difference in expected returns between the unconstrained MVO and the constrained portfolio becomes clear in Figures 4c–4d. Figure 4c shows the expected returns attributable to the constraints as if they are static. As more assets are excluded, the contributions from the static constraints also increase. However, this component remains unchanged as the correlation ρ varies.

Figure 4d shows the expected returns attributable to the information in the constraint, which increase as ρ increases. In addition, this result highlights a trade-off when a greater number of assets are excluded. On the one hand, when the correlation ρ is nonzero, excluding more assets implies that only assets with positive or negative returns are included in the portfolio. On the other hand, 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, and similarly, the lowest returns are achieved with an intermediate number of assets given a negative correlation. These results are consistent with Proposition 8.

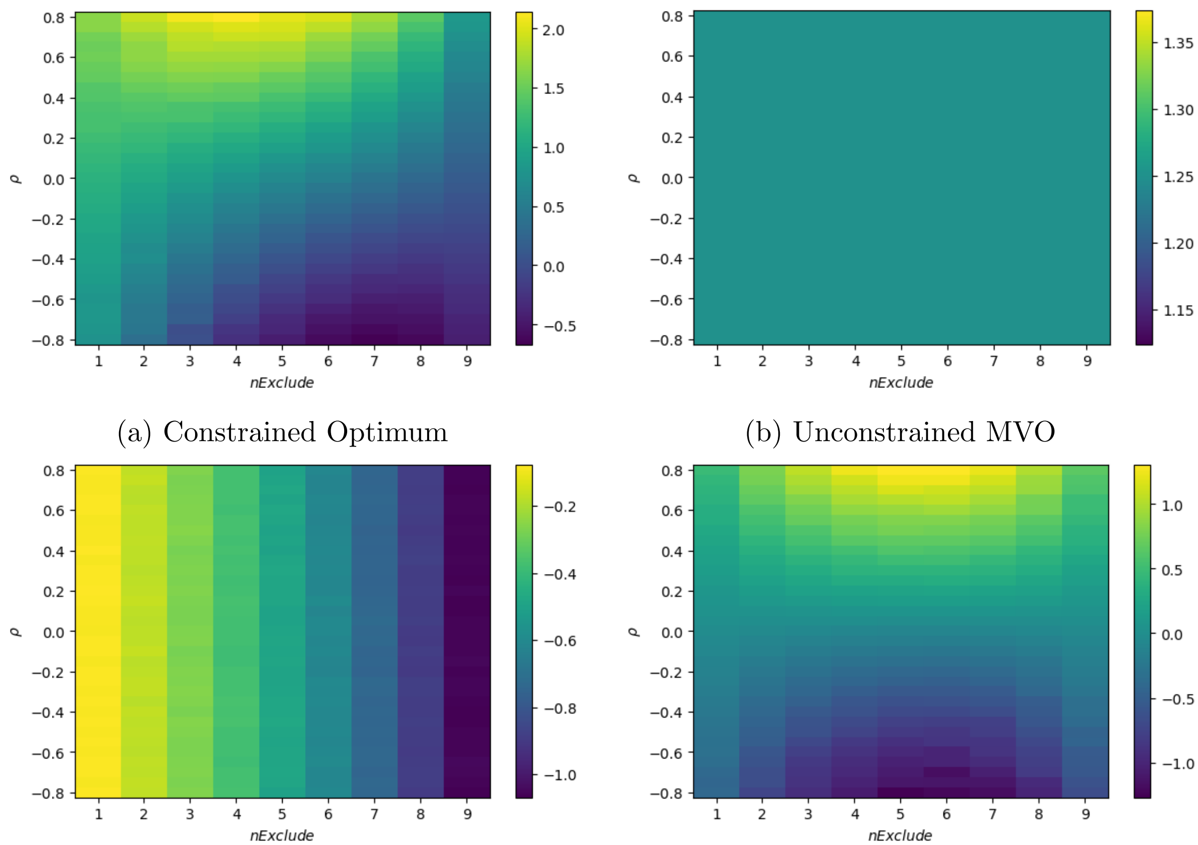
Similarly, Figure 5 demonstrates the attribution for expected utility. Figure 5a shows the expected utility of the constrained portfolio, which varies between -0.5 to 2.0 as ρ_1 and ρ_2 vary between -0.8 and 0.8. Figure 5b shows the expected utility of the unconstrained MVO problem, which is again around 1.25 regardless of values of ρ_1 and ρ_2.

The difference in expected utility between the unconstrained MVO and the constrained portfolio is decomposed into the static and information components, respectively. Figure 5c shows the expected utility attributable to the exclusionary constraints as if they are static, which contributes to the expected utility negatively, ranging from -1.0 to -0.2. Figure 5d shows the expected utility attributable to information, which increases as ρ increases, and has a similar pattern to Figure 4d.

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.

5 Empirical Analysis

In this section, we apply our attribution framework to real-world datasets, and consider two examples where constraints are applied in portfolio construction: ESG investing in which the average portfolio ESG score is required to be above a certain threshold, and exclusionary investing of sin stocks and stranded assets.



(a) Constrained Optimum

(b) Unconstrained MVO

(c) Decomposition: Constraint (Static)

(d) Decomposition: Constraint (Information)

Figure 5: Decomposition of expected utility for problem (36) with one exclusionary constraint which depends on random characteristics, as the number of excluded assets ($nExclude$) and the correlation (ρ) between the random characteristic and asset returns vary. The expected utility 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).

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) at the Wharton Research Data Services (WRDS). The CRSP dataset also contains basic firm characteristics such as market capitalization. We obtain the daily Fama-French factor data from Kenneth French's website.[18]

Because the ESG data is updated annually, we require that a stock has at least five 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:

$$R_{i,t} = \alpha_i + \beta_{i,1}(R_{M,t} - R_{f,t}) + \beta_{i,2}SMB_t + \beta_{i,3}HML_t + \beta_{i,5}RMW_t + \beta_{i,6}CMA_t + \epsilon_{i,t}. \quad (37)$$

We use the residual returns:

$$r_{i,t} \equiv \alpha_i + \epsilon_{i,t}, \quad (38)$$

winsorized at 2.5% on both sides, 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 use the MSCI KLD ESG dataset which contains environmental, social, and governance ratings of large publicly traded companies from 2003 to 2018. This database contains yearly ratings on roughly the 3,000 largest U.S. companies, and has been used in numerous studies examining the effect of ESG ratings on firm performance.[19] The raw ESG data classifies environmental, social, and governance performance into 13 different categories, including seven qualitative categories (community, diversity, employee relations, environment, human rights, product) and six controversial-business categories (alcohol, gambling, firearms, military, nuclear, tobacco, and corporate governance). The raw data rates each firm in terms of both strength and concern in the seven qualitative categories, and only in terms of concern in the six controversial-business categories.

We follow Lins, Servaes, and Tamayo (2017) in aggregating the raw data into an ESG score. First, we mark all missing ratings as zero. As the maximum number of strengths and concerns for any given category will vary over time, we scale them for each category by dividing the number of strengths or concerns for each firm-year by the maximum number of strengths or concerns possible for that category in that year. This procedure yields strength and concern indices that range from zero to one for each category-year. Our measure in

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

[19] See, for example, Hong and Kostovetsky (2012), Deng, Kang, and Low (2013), Krüger (2015), and Borisov, Goldman, and Gupta (2016), Lins, Servaes, and Tamayo (2017), and Berg, Koelbel, and Rigobon (2022).

each category-year is then obtained by subtracting the concerns index from the strengths index. The net score per category therefore ranges from -1 to $+1$. Finally, to obtain the aggregated ESG score of a firm, we combine the net score for seven qualitative categories, which leads to a final score that ranges from -7 to $+7$.[20]

We provide basic descriptive statistics of the ESG score in Section 5.2.

Sin Stocks and Energy Stocks. The CRSP data contains several basic firm characteristics, including the industry classification of the firm. We complement the CRSP data with the Compustat Historical Segment data, which also contains industry classification information for the different segments of a company in the U.S.

We follow Hong and Kacperczyk (2009) in identifying sin stocks as those with SIC codes 2100–2199 belong to the alcohol group, and those with SIC codes 2080–2085 are in the tobacco group. In addition, we identify gaming stocks as those with the following NAICS codes: 7132, 71312, 713210, 71329, 713290, 72112, and 721120. We then augment this list by searching across companies at the company segment level using the Compustat Segments data, identifying a company as a sin stock if any of its segments has an SIC code in either the alcohol or the tobacco group, or an NAICS code in the gaming group, as defined above. Accordingly, our final list of sin stocks is the union of these two screening procedures.

In addition, there is a growing literature on the effects of excluding stranded assets such as energy stocks (Bohn, Goldberg, and Ulucam, 2022). Therefore, we add energy stocks to the list of assets excluded in portfolio construction, and follow Bohn, Goldberg, and Ulucam (2022) in identifying energy stocks as those with SIC codes 1000–1519.

We provide basic descriptive statistics of the final list of excluded firms in Section 5.3.

5.2 ESG Constraints

We merge the CRSP dataset with the MSCI KLD ESG dataset. Table 1 shows, for each year, the number of firms in our dataset, the summary statistics of its annualized residual returns, and the summary statistics of the aggregate ESG score we construct lagged by one year. In general, we have around 1,800 stocks each year in our sample period.

Table 2 shows the cross-sectional correlations between the residual returns and ESG scores lagged by zero to four years, averaged over all years in our sample. When ESG scores are compared with residual returns in the same year (lag 0), there is a -2.98% correlation.

[20]We follow Lins, Servaes, and Tamayo (2017) in excluding the six controversial-business categories. Lins, Servaes, and Tamayo (2017) use only the first five qualitative categories because they consider the other categories irrelevant for their purposes of corporate social responsibility. However, we choose to include those ratings that correspond to human rights and product.

However, this correlation is not realizable in actual portfolios, since the ESG scores are not released at the start of the year. When ESG scores lag by one to four years, there is a slight increase in the correlation. We use the ESG scores lagged by one year as our main measure to construct annually rebalanced portfolios.

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.

Year	#Firms	Annualized Residual Return (%)							ESG (lag one year)						
		mean	std	min	25%	50%	75%	max	mean	std	min	25%	50%	75%	max
2004	1876	0.7	25.9	-87.6	-15.8	-0.7	13.9	172.5	-0.1	0.5	-3.4	-0.3	0.0	0.1	2.9
2005	1916	1.3	28.2	-93.6	-16.3	-1.5	15.6	164.9	-0.1	0.7	-3.0	-0.6	-0.2	0.2	2.2
2006	1734	0.9	26.7	-78.3	-15.1	-2.0	13.3	171.1	-0.3	0.6	-3.3	-0.5	-0.2	0.0	2.5
2007	1682	5.6	38.1	-80.7	-17.5	1.3	22.1	321.5	-0.3	0.6	-3.7	-0.7	-0.2	0.0	3.0
2008	1721	3.3	44.9	-92.2	-25.7	-1.3	27.8	497.2	-0.3	0.6	-3.5	-0.7	-0.3	0.0	3.4
2009	1785	1.5	38.2	-88.8	-21.7	-2.9	19.4	290.5	-0.3	0.6	-3.6	-0.7	-0.3	0.0	2.8
2010	1776	0.7	28.4	-87.1	-17.1	-2.2	14.6	174.1	-0.3	0.6	-3.5	-0.7	-0.3	0.0	2.8
2011	1896	-0.4	28.3	-92.3	-17.6	0.4	16.2	165.8	-0.4	0.6	-2.8	-0.7	-0.6	-0.1	3.9
2012	1809	-0.5	27.7	-82.3	-15.1	-2.9	10.8	297.0	-0.5	0.8	-2.7	-0.9	-0.6	-0.2	4.2
2013	1785	-0.9	27.7	-89.9	-17.4	-3.6	12.0	302.6	0.1	0.7	-2.3	-0.3	0.0	0.5	3.8
2014	1845	-0.1	24.0	-81.9	-13.7	-0.7	13.6	142.1	0.1	0.7	-2.4	-0.3	0.0	0.3	3.2
2015	1603	2.0	27.4	-91.4	-15.1	2.9	18.7	148.9	0.1	0.5	-3.7	0.0	0.0	0.3	3.2
2016	1708	-1.3	25.1	-82.7	-16.4	-2.5	11.5	201.5	0.2	0.6	-2.6	-0.1	0.0	0.5	3.3
2017	1687	1.0	24.7	-79.5	-13.5	-0.6	13.0	153.4	0.1	0.7	-2.4	-0.1	0.1	0.5	3.1
2018	1838	3.6	27.8	-71.1	-14.2	1.8	18.0	158.7	0.2	0.7	-2.9	0.0	0.2	0.5	4.1
2019	1898	5.4	29.4	-69.8	-9.8	4.4	17.9	585.6	0.6	0.8	-2.3	0.1	0.5	1.0	4.7

Table 2: Average cross-sectional correlation between residual returns and lagged values of ESG scores.

Lag (ESG Score)	4	3	2	1	0
Correlation	0.97%	0.01%	-0.89%	-1.95%	-2.98%

Figure 6 shows the year-over-year cross-sectional correlations between the residual returns and lag-1 ESG scores.[21] The correlations are generally negative before 2010, implying that high ESG stocks on average delivered lower excess returns relative to the Fama-French five-factor model, consistent with equilibrium theories of ESG returns (Pástor, Stambaugh, and Taylor, 2021; Pedersen, Fitzgibbons, and Pomorski, 2021). After 2011, the correlations fluctuate around zero, and are positive in certain years, reflecting increasing attention toward ESG and climate-related issues, an effect consistent with Pástor, Stambaugh, and Taylor's (2022) and Lo, Zhang, and Zhao's (2022) findings.

[21]The ESG scores are available from 2003 to 2018, and the correlations are necessarily computed from 2004 to 2019.

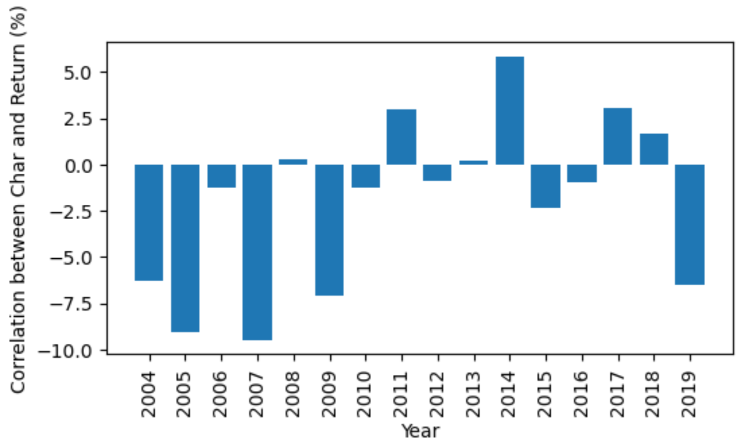


Figure 6: Cross-sectional correlations between asset returns and lag-1 ESG scores each year.

We 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. In fact, Berg, Koelbel, and Rigobon (2022) and Berg et al. (2023) show that there exists substantial noise in ESG measures, and ESG scores from different data providers may lead to very different correlations. What we hope to demonstrate is that, given any ESG score, it is possible to attribute portfolio performance metrics to different constraints and to the information in those constraints.

Portfolio Construction. We consider investors who construct long/short portfolios each year by solving the following problem.

$$\begin{aligned} \max_{\boldsymbol{\omega}} \quad & \boldsymbol{\omega}'\boldsymbol{\mu} - \frac{\eta}{2}\boldsymbol{\omega}'\boldsymbol{\Sigma}\boldsymbol{\omega} \\ \text{s.t.} \quad & \boldsymbol{\omega}'\mathbf{1} = 1 \\ & \boldsymbol{\omega}'\mathbf{x}_{\text{ESG}} \geq b. \end{aligned} \tag{39}$$

The first constraint guarantees that the portfolio is fully invested and can therefore be compared across different configurations of the constraint. The second constraint imposes a minimum level of portfolio ESG score, and we set $b = 1$ to use as an example in our analysis. We choose η such that the unconstrained MVO portfolio has a realistic leverage.

To construct these portfolios, investors need an estimate of the expected residual return $\boldsymbol{\mu}$ and the covariance matrix of the residual returns $\boldsymbol{\Sigma}$ each year. We assume that investors have access to the following estimator:

$$\begin{aligned}\hat{\mu}_{i,t} &= pr_{i,t} + (1-p)z_{i,t}, \quad i = 1, 2, \ldots, N \\ \hat{\mathbf{\Sigma}}_t &= p\mathbf{\Sigma}_t + (1-p)\bar{\sigma}^2\mathbf{I}\end{aligned} \tag{40}$$

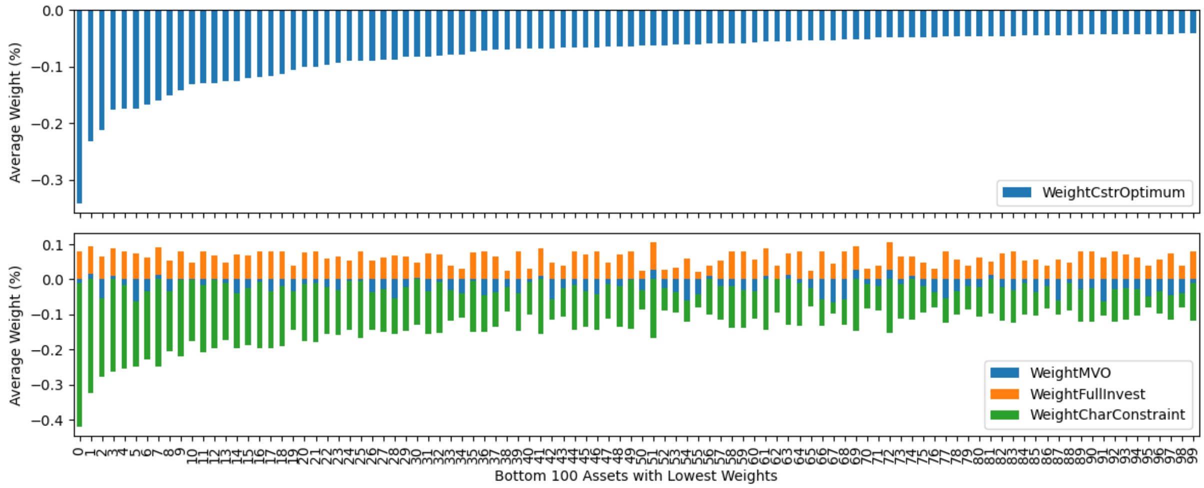
where $r_{i,t}$ is the average residual return from the Fama-French five-factor model of the i-th stock,[22] $z_{i,t} \sim N\left(0, (0.1 \times \bar{\sigma}_t)^2\right)$ is white noise, $\mathbf{\Sigma}_t$ is the sample covariance matrix estimated from the residuals in year t, $\bar{\sigma}_t^2$ is the cross-sectional average residual variance in year t, and $\bar{\sigma}$ is the annualized cross-sectional average residual variance over all years. The key parameter $p \in [0, 1]$ controls the level of look-ahead information the investor has. For most of our analysis below, we set $p = 0$, which corresponds to an investor with no information about expected returns in the future. We also consider an example of $p > 0$, where the estimators can be interpreted as a certain degree of return predictability.

Portfolio Holdings Decomposition. Combining (39) and (40), we can solve for the optimal constrained portfolio. Figures 7a and 7b show the bottom and top 100 assets with the lowest and highest portfolio weights for the optimal portfolio, respectively, averaged over all years. We decompose the portfolio weights into components corresponding to the unconstrained MVO portfolio and two constraints, based on (7) in Proposition 1. In both the top and bottom assets, the full investment constraint (orange) unanimously gives positive weights.

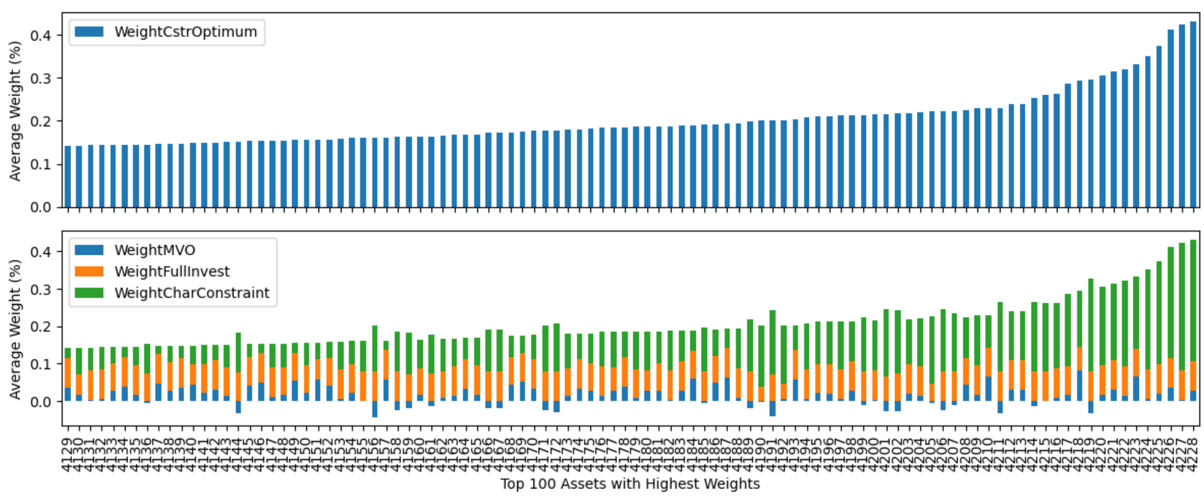
For the bottom 100 assets in Figure 7a, the unconstrained MVO portfolio (blue) generally leads to negative weights, and the ESG constraint (green) further adds to the negative portfolio holdings. Overall, these assets tend to have low ESG scores, and are therefore assigned the lowest weights in the portfolio. In contrast, for the top 100 assets in Figure 7b, the unconstrained MVO portfolio (blue) generally gives positive weights and the ESG constraint (green) leads to additional positive weights.

Expected Return and Utility Decomposition. Figure 8 demonstrates the decomposition of the expected return and utility of the portfolio into different components. The upper panel of Figure 8a 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 towards the second half. This expected utility is decomposed into three components in the lower panel using (27) in Proposition 5. The expected utility of the unconstrained MVO portfolio (blue) is positive over the 16 years in our sample. As the conventional wisdom of constrained optimization

[22] We estimate this by taking the average of the actual residuals (38) over year t



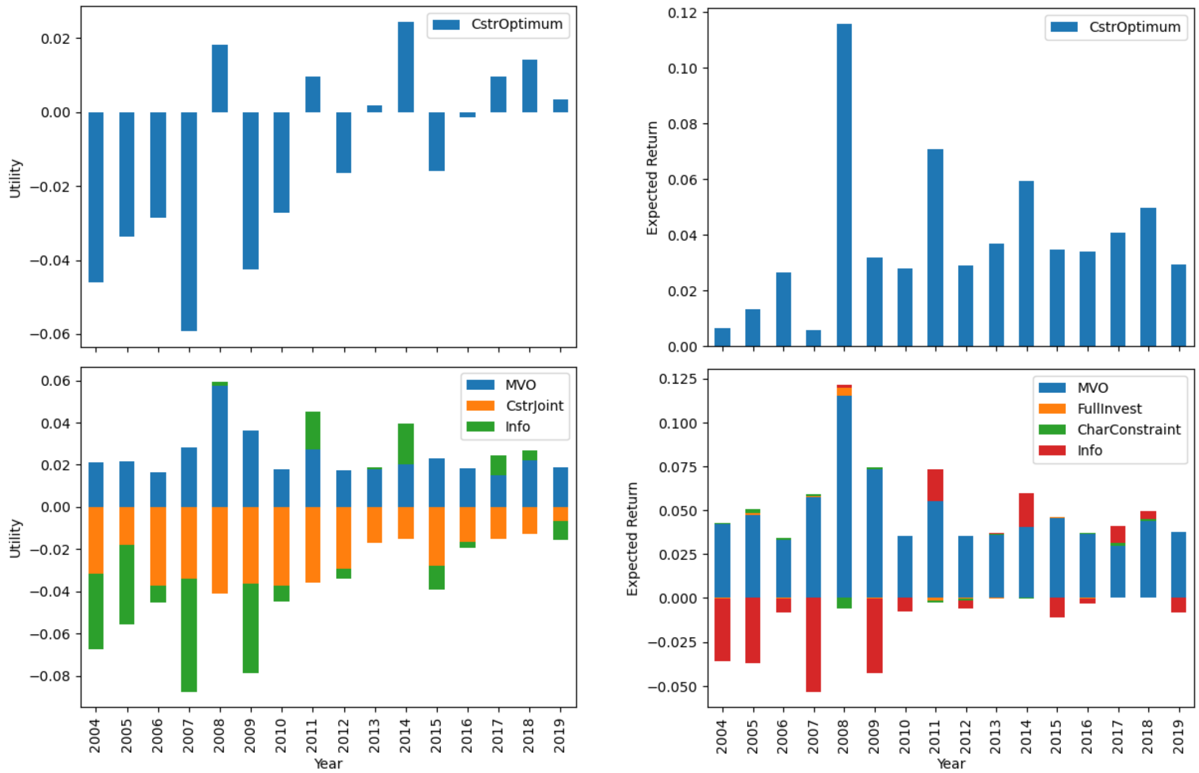
(a) Bottom assets with the lowest weights.



(b) Top assets with the highest weights.

Figure 7: Average portfolio weights over all years and their decomposition, for the long/short portfolio defined in (39) with a constraint on the average portfolio characteristic value ($\boldsymbol{\omega}'\mathbf{x}_{\text{ESG}} \geq 1.0$). (a) shows the bottom assets with the lowest weights and (b) shows the top assets with the highest weights. In each subfigure, the top panel shows the portfolio weights (%) of the constrained portfolio. The bottom panel shows the decomposition of the weights into components corresponding to the unconstrained MVO portfolio (blue), the full investment constraint ($\boldsymbol{\omega}'\mathbf{1} = 1$, orange), and the ESG constraint ($\boldsymbol{\omega}'\mathbf{x}_{\text{ESG}} \geq 1.0$, green).

would suggest, the expected utility contribution of the two constraints (orange), treated as static, is indeed negative.[23]



(a) Expected Utility

(b) Expected Return

Figure 8: Expected return and utility and their decomposition, for the long/short portfolio (39) with a constraint on the average portfolio characteristic value ($\boldsymbol{\omega}'\mathbf{x}_{\mathrm{ESG}} \geq 1.0$). 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 (blue), all constraints treated as static (orange), and the information from the ESG constraint (green). 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 (blue), the full investment constraint (orange), the ESG constraint treated as static ($\boldsymbol{\omega}'\mathbf{x}_{\mathrm{ESG}} \geq 1.0$, green), and the information from the ESG constraint (red).

However, the expected utility contribution from the information contained in the constraints (green) 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 2009

[23]We use decomposition conditioned on $\mathbf{X}$ because we decompose expected utility year over year. This is different from the decomposition of the overall unconditional expected utility in our simulated examples in Section 4. In addition, the attribution of expected utility to static constraints is only possible when they are combined together because of the risk term in expected utility. See the earlier remarks after Proposition 1.

being a strong negative year and 2011, 2014, and 2017 being strong positive years in information. This pattern is consistent with the intuition derived from Propositions 4–5, and the correlations between asset returns and ESG scores in Figure 6. This example vividly demonstrates that while constraints must decrease the overall expected utility of a portfolio when treated as static, they can sometimes increase the expected utility depending on the information contained in the constraints.

Figure 8b shows the expected return of the optimal portfolio and its decomposition based on (26) in Proposition 5. While the two constraints (orange and green) can contribute either positively or negatively to the expected returns,[24] the main source of contribution in terms of expected return is from the information in the constraints (red). Like the case of expected utility decomposition in Figure 8a, the expected return contribution from information is strongly negative in 2004, 2005, 2007, and 2009, and is strongly positive in 2011, 2014, and 2017.

Realized Return Decomposition. Figure 9 shows the realized returns of the optimal portfolio and *ex post* attribution of returns. We compare a portfolio constructed without a return forecast ($p = 0$) in Figure 9a with one based on a return forecast ($p = 0.1$) in Figure 9b.

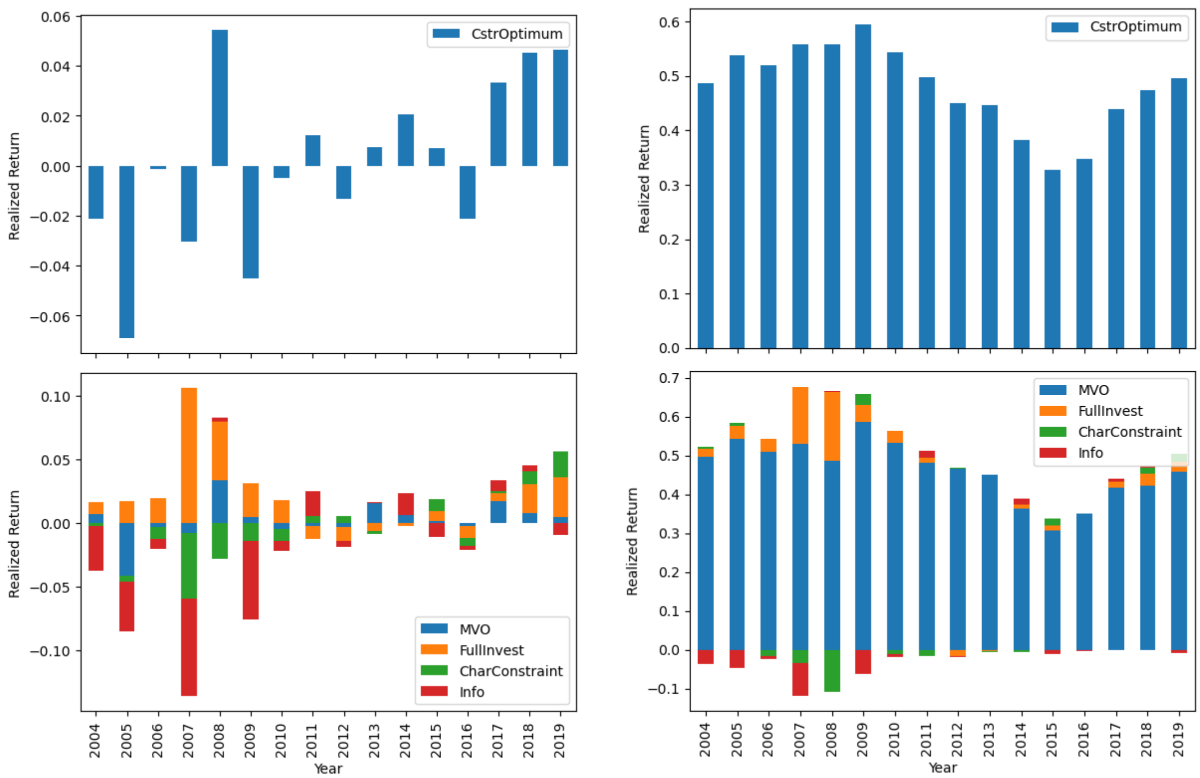
The upper panel of Figure 9a shows that, without any return forecast, the realized residual returns in excess of the Fama-French five-factor model of the constrained portfolio fluctuate around zero over the 16 years in our sample. The bottom panel shows the realized residual returns of the unconstrained MVO portfolio.

To understand the difference between these returns, the lower panel decomposes the realized return of the constrained portfolio based on Proposition 6. The full investment constraint generally contributes to the returns positively, especially before 2010 and after 2017. 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. On the other hand, the information component contributes negatively to realized returns before 2010, and positively in 2011, 2014, 2017, and 2018, consistent with results in Propositions 6 and the pattern of correlations between asset returns and ESG scores in Figure 6. Overall, these components explain the difference in residual returns between the unconstrained MVO and the constrained portfolio.

In addition, Figure 9b shows results parallel to those in Figure 9a, but with a return forecast ($p = 0.1$) in portfolio construction. As expected, the realized residual returns for both

[24]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.

the unconstrained and the constrained portfolio dramatically increase. The decomposition is also similar to the case without a return forecast, except that the contributions from the information in the constraints, in relative terms, are much smaller.



(a) No Return Predictability

(b) Return Predictability $p = 0.1$

Figure 9: Realized return and their *ex post* decomposition, for the long/short portfolio defined in (39) with a constraint on the average portfolio characteristic value ($\boldsymbol{\omega}'\mathbf{x}_{\text{ESG}} \geq 1.0$). (a) corresponds to a return estimator in (40) with no ability to forecast future expected returns ($p = 0$), and (b) corresponds to a return estimator with some level of predictability ($p = 0.1$). 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 (blue), the full investment constraint (orange), the ESG constraint treated as static ($\boldsymbol{\omega}'\mathbf{x}_{\text{ESG}} \geq 1.0$, green), and the information from the ESG constraint (red).

Other Portfolios. The long/short portfolio given by (39) is one way to construct ESG portfolios. In practice, investors may face long-only constraints or use exclusionary constraints to construct ESG portfolios. We demonstrate how to decompose the performance of these portfolios using our framework in Appendices B.2 and B.3, respectively.

5.3 Excluding Sin Stocks and Energy Stocks

Our universe of stocks contains those with valid CRSP returns and industry labels as described in Section 5.1. We also require a firm to have a market capitalization of at least 100 million USD in a particular year to be included in the universe for next year. Table 3 shows, for each year, the number of firms available in our dataset, the number of excluded firms based on sin stock and energy stock classification at the end of the last year, and the summary statistics of the annualized residual returns. In general, we have around 4,000 stocks each year, of which 4.8% to 7.5% firms are excluded each year because they are labeled as either sin stocks or energy stocks as of the previous year.

Table 3: Summary statistics of the annualized residual returns (in percentage) from the Fama-French five-factor model and the number of excluded firms based on sin stock and energy stock classification at the end of last year (as a percentage of the total number of firms in the sample).

Year	#Firms	Excluded Firms (%)	Annualized Residual Return (%)						
			mean	std	min	25%	50%	75%	max
2001	3265	4.9	6.8	45.5	-89.2	-18.4	1.8	22.6	796.7
2002	3505	4.8	2.5	36.0	-94.3	-18.1	3.1	19.3	356.4
2003	3440	5.2	3.3	29.1	-77.3	-14.2	0.5	14.9	276.3
2004	4224	5.4	0.9	28.3	-88.7	-15.1	0.0	12.9	320.3
2005	4484	5.8	0.7	30.1	-87.3	-16.1	-1.9	14.2	256.0
2006	4567	6.1	3.8	28.5	-81.1	-12.1	1.8	15.8	236.9
2007	4659	6.8	4.2	40.1	-88.9	-17.9	-0.6	19.9	487.2
2008	4669	7.0	-8.0	40.9	-93.4	-34.2	-11.4	11.3	515.8
2009	3775	7.0	7.1	38.0	-93.0	-16.5	3.3	26.9	233.4
2010	4195	7.2	1.1	27.7	-87.5	-14.0	0.1	13.4	296.2
2011	4439	7.3	-3.2	28.3	-93.2	-18.7	-1.1	13.3	177.5
2012	4300	7.3	-0.5	25.6	-95.3	-13.1	-0.4	10.5	303.8
2013	4437	7.5	-4.4	28.1	-96.0	-19.2	-6.4	6.9	320.5
2014	4744	6.8	-2.1	26.0	-90.5	-15.3	-1.2	12.0	270.6
2015	4890	5.9	-2.7	28.4	-92.9	-17.6	-0.5	11.7	189.9
2016	4771	5.4	-2.6	27.1	-96.8	-15.5	-2.5	9.1	258.8
2017	4669	5.7	1.1	24.7	-96.7	-10.2	1.2	11.1	223.8
2018	4625	5.8	-1.5	26.9	-96.8	-13.5	-2.1	9.8	198.5
2019	4317	5.3	2.8	25.5	-92.0	-8.4	3.5	14.6	252.4
2020	4215	5.0	-5.4	30.8	-92.1	-20.9	-5.6	8.0	349.3

We define a binary variable to represent whether an asset can be included in the portfolio:

$$x_i = \begin{cases} 0, & \text{if stock } i \text{ belongs to sin stocks or energy stocks} \\ 1, & \text{otherwise.} \end{cases} \tag{41}$$

Table 4 shows the cross-sectional correlations between the residual returns and the inclusion variable defined in (41) lagged by zero to four years, averaged over all years in our sample. When the inclusion variable is compared with the residual returns in the same year (lag 0), there is a 3.59% correlation. When the inclusion variable lags by one to four years, the correlation increases slightly to around 5%. We use the inclusion variable lagged by one year as our main measure to construct annually rebalanced portfolios.

Table 4: Average cross-sectional correlation between residual returns and lagged values of the inclusion variable defined in (41).

Lag (Inclusion Variable (41))	4	3	2	1	0
Correlation	5.20%	5.17%	5.09%	4.81%	3.59%

Figure 10 shows the year-over-year cross-sectional correlations between the residual returns and the inclusion variable (41), which is generally negative before 2010 and positive after 2011. This implies that sin stocks and energy stocks have higher excess returns relative to the Fama-French five-factor model compared to other stocks before 2010, which is consistent with Hong and Kacperczyk's (2009) results. After 2011, as the attention to SRI investing increases, sin stocks and energy stocks tend to deliver lower excess returns.

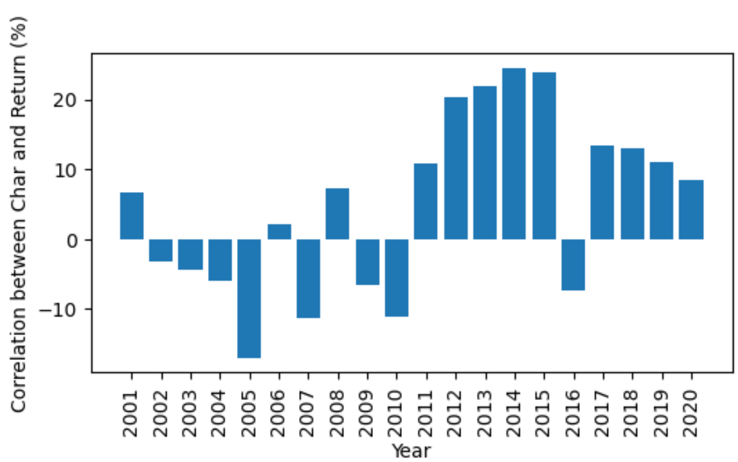


Figure 10: Cross-sectional correlations between asset returns and the inclusion variable defined in (41) each year.

Portfolio Construction. Exclusionary investing usually does not consider short positions, because otherwise excluded assets can arguably be shorted. Therefore, we consider long-only portfolios by solving the following problem each year.

$$\begin{aligned} \max_{\boldsymbol{\omega}} \quad & \boldsymbol{\omega}'\boldsymbol{\mu} - \frac{\eta}{2}\boldsymbol{\omega}'\boldsymbol{\Sigma}\boldsymbol{\omega} \\ \text{s.t.} \quad & \boldsymbol{\omega}'\mathbf{1} = 1 \\ & \omega_i = 0 \quad \text{if } x_i = 0 \\ & \boldsymbol{\omega} \geq \mathbf{0}. \end{aligned} \tag{42}$$

We again use the return forecast in (40) to estimate $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. We choose η such that the unconstrained MVO portfolio has a realistic leverage.

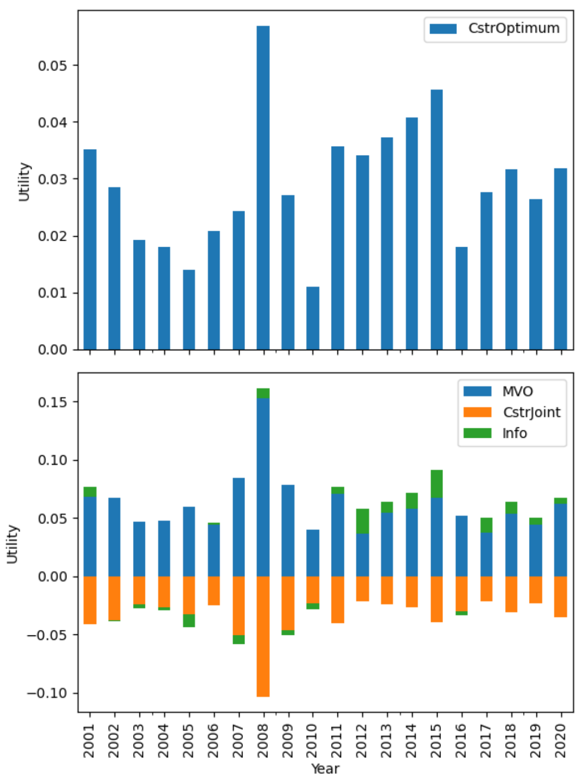
Expected Return and Utility Decomposition. Figure 11 demonstrates the decomposition of the expected utility and expected return of the portfolio into different components.

The upper panel of Figure 11a shows that the expected utility of the optimal portfolio is positive through our 20-year sample. This utility is decomposed into three components in the lower panel using (27) in Proposition 5. The expected utility of the unconstrained MVO portfolio (blue) is positive, while the expected utility contribution of the three constraints (orange), treated as static, is negative. The expected utility contribution from information contained in the constraints (green), however, varies over time. It is generally negative before 2010 and positive after 2011, consistent with the pattern of correlations between asset returns and the inclusion variable in Figure 10.

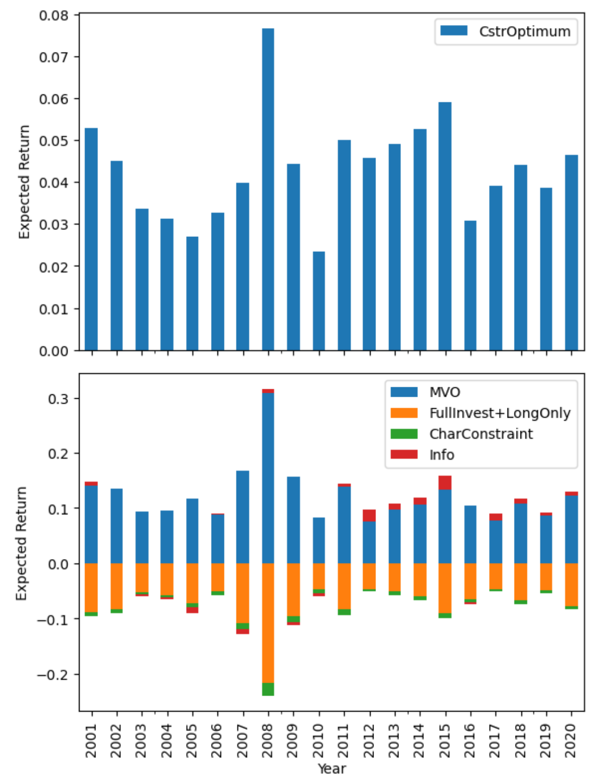
Figure 11b shows the expected utility of the optimal portfolio and its decomposition based on (26) in Proposition 5. The two constraints (orange and green) contribute negatively to expected returns. The expected return contribution from information (read) is generally negative before 2010 and generally positive after 2011. Together, the expected return of the constrained portfolio is lower than that of the unconstrained MVO portfolio primarily driven by the full investment and long-only constraints.

Realized Return Decomposition. Finally, we show the realized returns of the optimal portfolio and *ex post* attribution of returns in Figure 12, in which we compare a portfolio constructed without a return forecast ($p = 0$) in Figure 12a with one based on a return forecast ($p = 0.1$) in Figure 12b.

The upper panel of Figure 12a shows the realized residual returns in excess of the Fama-French five-factor model for the constrained portfolio, which is decomposed into several components based on Proposition 6 in the lower panel. The contribution from the full

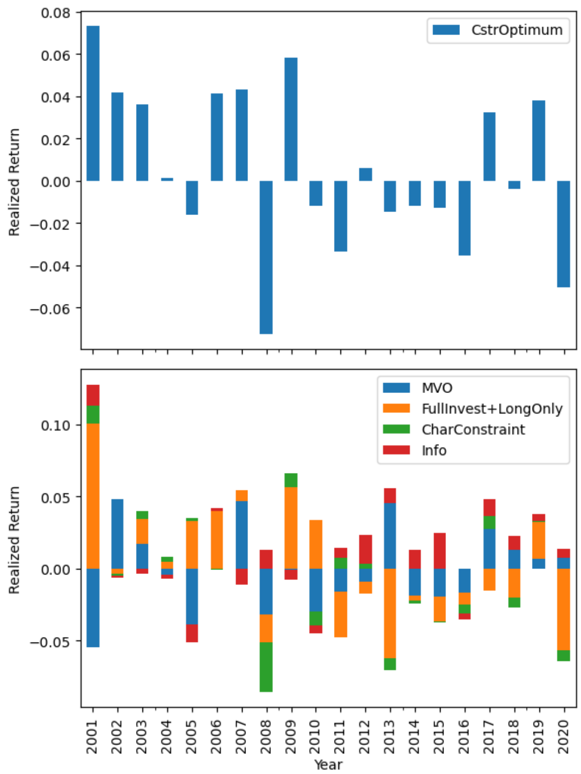


(a) Expected Utility

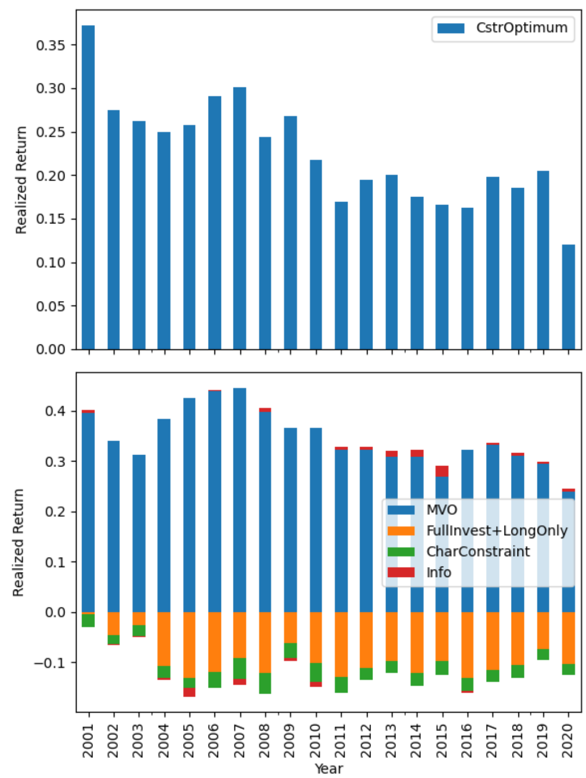


(b) Expected Return

Figure 11: Expected return and utility and their decomposition, for the long-only portfolio defined in (42) with an exclusionary constraint based on the inclusion variable defined in (41). 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 (blue), all constraints treated as static (orange), and the information from the exclusionary constraint (green). 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 (blue), the full investment and long-only constraints combined together (orange), the exclusionary constraint treated as static (green), and the information from the exclusionary constraint (red).



(a) No Return Predictability



(b) Return Predictability $p = 0.1$

Figure 12: Realized return and their *ex post* decomposition, for the long-only portfolio defined in (42) with an exclusionary constraint based on the inclusion variable defined in (41). (a) corresponds to a return estimator in (40) with no ability to forecast future expected returns ($p = 0$), and (b) corresponds to a return estimator with some level of predictability ($p = 0.1$). 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 a component corresponding to the unconstrained MVO portfolio (blue), a component attributable to the full investment and long-only constraints combined together (orange), a component attributable to the exclusionary constraint treated as static (green), and a component attributable to information from the exclusionary constraint (red).

investment and the long-only constraints (orange), is generally positive before 2010 except for 2008, and generally negative after 2011. This pattern is consistent with the average residual returns shown in Table 3. The exclusionary investing constraint (green), treated as static, also contributes either positively or negatively over the 20-year period.

In addition, Figure 12b shows results parallel to those in Figure 12a, but with a return forecast ($p = 0.1$) in portfolio construction. As expected, the realized residual returns for both the unconstrained and the constrained portfolio dramatically increase. However, the constraints greatly limit the ability of the portfolio to take advantage of the return predictability.

In both cases, the contribution attributable to information from the exclusionary investing constraint (red) is generally negative before 2010 and positive after 2011, matching the patterns of the correlation between the returns and the inclusion variable (41) in Figure 10.

Overall, these results demonstrate that our performance attribution framework works not only for constraints that restrict the level of ESG scores, but also for constraints that directly exclude assets from a portfolio.

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, utilities, and both expected and realized returns into components attributable to each constraint treated as static and the information from each constraint. While it is commonly believed that constraints can only decrease the expected utility of a portfolio, we show that this is only true when they are treated as static. We quantify the information content from constraints when they are stochastic and potentially correlated with asset returns.

We demonstrate that our methodology can be applied to common examples of constraints including the level of a particular characteristic, such as ESG scores, and exclusion constraints, such as divesting from sin stocks and energy stocks. Our results show that these constraints do not necessarily decrease the expected utility and returns of the portfolio, and can even contribute positively to portfolio performance when information contained in the constraints correlates positively to asset returns.

References

Albuquerque, R., Y. Koskinen, and C. Zhang, 2019, Corporate social responsibility and firm risk: Theory and empirical evidence, *Management Science* 65, 4451–4469.

Alessandrini, F., and E. Jondeau, 2020, ESG investing: From sin stocks to smart beta, *The Journal of Portfolio Management* 46, 75–94.

Amel-Zadeh, A., and G. Serafeim, 2018, Why and how investors use ESG information: Evidence from a global survey, *Financial Analysts Journal* 74, 87–103.

Bacon, C. R., 2019, *Performance Attribution: History and Progress* (CFA Institute Research Foundation).

Baker, M., D. Bergstresser, G. Serafeim, and J. Wurgler, 2022, The pricing and ownership of US green bonds, *Annual Review of Financial Economics* 14, 415–437.

Bansal, R., D. Wu, and A. Yaron, 2022, Socially responsible investing in good and bad times, *The Review of Financial Studies* 35, 2067–2099.

Barber, B. M., A. Morse, and A. Yasuda, 2021, Impact investing, *Journal of Financial Economics* 139, 162–185.

Berg, F., J. F. Koelbel, and R. Rigobon, 2022, Aggregate confusion: The divergence of ESG ratings, *Review of Finance* 26, 1315–1344.

Berg, F., A. W. Lo, R. Rigobon, M. Singh, and R. Zhang, 2023, Quantifying the returns of ESG investing: An empirical analysis with six ESG metrics, Available at SSRN.

Berk, J., and J. H. van Binsbergen, 2021, The impact of impact investing, Available at SSRN 3909166.

Bhattacharya, P., 1974, Convergence of sample paths of normalized sums of induced order statistics, *The Annals of Statistics* 2, 1034–1039.

Bohn, J., L. R. Goldberg, and S. Ulucam, 2022, Sustainable investing from a practitioner's viewpoint: What's in your ESG portfolio, *Journal of Investment Management* 20, 14–29.

Borisov, A., E. Goldman, and N. Gupta, 2016, The corporate value of (corrupt) lobbying, *The Review of Financial Studies* 29, 1039–1071.

Brinson, G. P., L. R. Hood, and G. L. Beebower, 1986, Determinants of portfolio performance, *Financial Analysts Journal* 42, 39–44.

Cappucci, M., 2018, The ESG integration paradox, *Journal of Applied Corporate Finance* 30, 22–28.

Clarke, R., H. de Silva, and S. Thorley, 2002, Portfolio constraints and the fundamental law of active management, *Financial Analysts Journal* 58, 48–66.

Clarke, R., H. de Silva, and S. Thorley, 2005, Performance attribution and the fundamental law, *Financial Analysts Journal* 61, 70–83.

Corwin, S. A., and J. F. Coughenour, 2008, Limited attention and the allocation of effort in securities trading, *The Journal of Finance* 63, 3031–3067.

Daniel, K., M. Grinblatt, S. Titman, and R. Wermers, 1997, Measuring mutual fund performance with characteristic-based benchmarks, *The Journal of finance* 52, 1035–1058.

David, H., 1973, Concomitants of order statistics, *Bulletin of the International Statistical Institute* 45, 295–300.

Davis, B., and J. Menchero, 2012, The alpha and beta of risk attribution, *The Journal of Portfolio Management* 38, 99–107.

Deng, X., J.-k. Kang, and B. S. Low, 2013, Corporate social responsibility and stakeholder value maximization: Evidence from mergers, *Journal of Financial Economics* 110, 87–109.

Eccles, R. G., M. D. Kastrapeli, and S. J. Potter, 2017, How to integrate ESG into investment decision-making: Results of a global survey of institutional investors, *Journal of Applied Corporate Finance* 29, 125–133.

Fabozzi, F. J., K. Ma, and B. J. Oliphant, 2008, Sin stock returns, *The Journal of Portfolio Management* 35, 82–94.

Fama, E. F., 1972, Components of investment performance, *The Journal of finance* 27, 551–567.

Fama, E. F., and K. R. French, 2007, Disagreement, tastes, and asset prices, *Journal of Financial Economics* 83, 667–689.

Fama, E. F., and K. R. French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.

Fauver, L., and M. B. McDonald IV, 2014, International variation in sin stocks and its effects on equity valuation, *Journal of Corporate Finance* 25, 173–187.

Feng, G., S. Giglio, and D. Xiu, 2020, Taming the factor zoo: A test of new factors, *The Journal of Finance* 75, 1327–1370.

Geczy, C. C., R. F. Stambaugh, and D. Levin, 2021, Investing in socially responsible mutual funds, *The Review of Asset Pricing Studies* 11, 309–351.

Gibson, R., P. Krueger, and P. S. Schmidt, 2021, ESG rating disagreement and stock returns, *Financial Analysts Journal* 1–24.

Goldberg, L. R., 2021, Constraint implied returns and stress in mean-variance optimized portfolios, Working paper.

Goldstein, I., A. Kopytov, L. Shen, and H. Xiang, 2021, On ESG investing: Heterogeneous preferences, information, and asset prices, Available at SSRN 3823042.

Grinold, R., 2005, Implementation efficiency, *Financial Analysts Journal* 61, 52–64.

Grinold, R. C., and K. Eaton, 1998, Attribution of performance and holdings, in J. Mulvey, and W. Ziemba, eds., *Worldwide Asset and Liability Modeling*, 87–113 (Cambridge University Press).

Grinold, R. C., and R. N. Kahn, 1999, *Active Portfolio Management: A Quantitative Approach for Producing Superior Returns and Controlling Risk*, second edition (McGraw-Hill Education, New York).

Grinold, R. C., and R. N. Kahn, 2019, *Advances in Active Portfolio Management: New Developments in Quantitative Investing* (McGraw Hill Professional).

Harvey, C. R., Y. Liu, and H. Zhu, 2016, ... and the cross-section of expected returns, *The Review of Financial Studies* 29, 5–68.

Hirshleifer, D., S. S. Lim, and S. H. Teoh, 2011, Limited investor attention and stock market misreactions to accounting information, *The Review of Asset Pricing Studies* 1, 35–73.

Hirshleifer, D., A. Subrahmanyam, and S. Titman, 2006, Feedback and the success of irrational investors, *Journal of Financial Economics* 81, 311–338.

Hong, H., and M. Kacperczyk, 2009, The price of sin: The effects of social norms on markets, *Journal of Financial Economics* 93, 15–36.

Hong, H., and L. Kostovetsky, 2012, Red and blue investing: Values and finance, *Journal of Financial Economics* 103, 1–19.

Hou, K., C. Xue, and L. Zhang, 2020, Replicating anomalies, *The Review of Financial Studies* 33, 2019–2133.

Idzorek, T. M., P. D. Kaplan, and R. G. Ibbotson, 2021, The popularity asset pricing model, *Available at SSRN 3451554* .

Khan, M., G. Serafeim, and A. Yoon, 2016, Corporate sustainability: First evidence on materiality, *The Accounting Review* 91, 1697–1724.

Kogan, L., S. A. Ross, J. Wang, and M. M. Westerfield, 2006, The price impact and survival of irrational traders, *The Journal of Finance* 61, 195–229.

Kolm, P. N., R. Tütüncü, and F. J. Fabozzi, 2014, 60 years of portfolio optimization: Practical challenges and current trends, *European Journal of Operational Research* 234, 356–371.

Krüger, P., 2015, Corporate goodness and shareholder wealth, *Journal of Financial Economics* 115, 304–329.

Lindsey, L. A., S. Pruitt, and C. Schiller, 2021, The cost of esg investing, *Available at SSRN 3975077* .

Lins, K. V., H. Servaes, and A. Tamayo, 2017, Social capital, trust, and firm performance: The value of corporate social responsibility during the financial crisis, *the Journal of Finance* 72, 1785–1824.

Lo, A. W., and A. C. MacKinlay, 1990, Data-snooping biases in tests of financial asset pricing models, *The Review of Financial Studies* 3, 431–467.

Lo, A. W., and R. Zhang, 2023, Quantifying the impact of impact investing, *Management Science* forthcoming.

Lo, A. W., R. Zhang, and C. Zhao, 2022, Measuring and optimizing the risk and reward of green portfolios, *The Journal of Impact and ESG Investing* .

Madhavan, A., A. Sobczyk, and A. Ang, 2021, Toward ESG alpha: Analyzing ESG exposures through a factor lens, *Financial Analysts Journal* 77, 69–88.

Markowitz, H., 1952, Portfolio selection, *The Journal of Finance* 7, 77–91.

Markowitz, H., 2014, Mean–variance approximations to expected utility, *European Journal of Operational Research* 234, 346–355.

Menchero, J., and B. Davis, 2011, Risk contribution is exposure times volatility times correlation: decomposing risk using the x-sigma-rho formula, *The Journal of Portfolio Management* 37, 97–106.

Merton, R. C., 1987, A simple model of capital market equilibrium with incomplete information, *The Journal of Finance* 42, 483–510.

Pástor, L., R. F. Stambaugh, and L. A. Taylor, 2021, Sustainable investing in equilibrium, *Journal of Financial Economics* 142, 550–571.

Pástor, L., R. F. Stambaugh, and L. A. Taylor, 2022, Dissecting green returns, *Journal of Financial Economics* 146, 403–424.

Pedersen, L. H., S. Fitzgibbons, and L. Pomorski, 2021, Responsible investing: The ESG-efficient frontier, *Journal of Financial Economics* .

Roselle, P., 2016, The evolution of integrating ESG analysis into wealth management decisions, *Journal of Applied Corporate Finance* 28, 75–79.

Rubinstein, M., 2002, Markowitz's" portfolio selection": A fifty-year retrospective, *The Journal of finance* 57, 1041–1045.

Scherer, B., and X. Xu, 2007, The impact of constraints on value-added, *The Journal of Portfolio Management* 33, 45–54.

Simon, H. A., 1955, A behavioral model of rational choice, *The Quarterly Journal of Economics* 69, 99–118.

Statman, M., and D. Glushkov, 2009, The wages of social responsibility, *Financial Analysts Journal* 65, 33–46.

Steinbach, M. C., 2001, Markowitz revisited: Mean-variance models in financial portfolio analysis, *SIAM review* 43, 31–85.

Stubbs, R. A., and D. Vandenbussche, 2010, Constraint attribution, *The Journal of Portfolio Management* 36, 48–59.

Tütüncü, R., 2012, Recent advances in portfolio optimization, in B. Scherer, and K. Winston, eds., *The Oxford Handbook of Quantitative Asset Management* (OUP Oxford).

Zerbib, O. D., 2022, A sustainable capital asset pricing model (S-CAPM): Evidence from environmental integration and sin stock exclusion, *Review of Finance* 26, 1345–1388.

A Appendix

In this Appendix, we provide proofs for all the propositions.

A.1 Proof of Proposition 1

Problem (1) can be solved by considering the Lagrangian:

$$L(\boldsymbol{\omega}, \boldsymbol{\lambda}) = \boldsymbol{\omega}'\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{\omega}'\boldsymbol{\Sigma}\boldsymbol{\omega} - \boldsymbol{\lambda}'(\mathbf{A}\boldsymbol{\omega} - \mathbf{b}) = \boldsymbol{\omega}'\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{\omega}'\boldsymbol{\Sigma}\boldsymbol{\omega} - \sum_{j=1}^{J} \lambda_j(\mathbf{A}_j'\boldsymbol{\omega} - b_j) \tag{A.1}$$

where $\mathbf{A}_j$ represents the j-th row (constraint) of the matrix $\mathbf{A}$. The first-order conditions:

$$\begin{aligned} \frac{\partial L(\boldsymbol{\omega}, \boldsymbol{\lambda})}{\partial \boldsymbol{\omega}} &= 0, \\ \frac{\partial L(\boldsymbol{\omega}, \boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} &= 0, \end{aligned} \tag{A.2}$$

lead to:

$$\begin{aligned} &\boldsymbol{\mu} - \boldsymbol{\Sigma}\boldsymbol{\omega} - \mathbf{A}'\boldsymbol{\lambda} = 0 \implies \boldsymbol{\Sigma}\boldsymbol{\omega} = \boldsymbol{\mu} - \mathbf{A}'\boldsymbol{\lambda} \implies \boldsymbol{\omega} = \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu} - \mathbf{A}'\boldsymbol{\lambda}\right), \\ &\mathbf{A}\boldsymbol{\omega} - \mathbf{b} = 0. \end{aligned} \tag{A.3}$$

The first equation proves (5). Combining the two equations leads to the system of equations that the optimal Lagrange multipliers, $\boldsymbol{\lambda}^*$, should satisfy:

$$\mathbf{A}\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu} - \mathbf{A}'\boldsymbol{\lambda}\right) - \mathbf{b} = 0 \implies \mathbf{A}\boldsymbol{\Sigma}^{-1}\mathbf{A}'\boldsymbol{\lambda} = \mathbf{A}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \mathbf{b} = 0. \tag{A.4}$$

In particular, when the feasible region of the constrained optimization problem is nonempty, $\mathbf{A}\boldsymbol{\Sigma}^{-1}\mathbf{A}'$ is invertible, which implies that:

$$\boldsymbol{\lambda}^* = \left(\mathbf{A}\boldsymbol{\Sigma}^{-1}\mathbf{A}'\right)^{-1}\left(\mathbf{A}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \mathbf{b}\right). \tag{A.5}$$

This completes the proof of (6) and (7).

To derive the expected return decomposition of (10), multiplying the expected return $\boldsymbol{\mu}$ by the portfolio holdings of (7) leads directly to:

$$\boldsymbol{\mu}'\boldsymbol{\omega}^* = \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \boldsymbol{\lambda}^{*'}\mathbf{A}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}. \tag{A.6}$$

To derive the expected utility decomposition of (11), we have

$$\begin{aligned}\boldsymbol{\mu}'\boldsymbol{\omega}^* - \frac{1}{2}\boldsymbol{\omega}^{*\prime}\boldsymbol{\Sigma}\boldsymbol{\omega}^* &= \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{A}'\boldsymbol{\lambda}^* - \frac{1}{2}\left(\boldsymbol{\mu}' - \boldsymbol{\lambda}^{*\prime}\mathbf{A}\right)\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu} - \mathbf{A}'\boldsymbol{\lambda}^*\right) \\ &= \frac{1}{2}\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{A}'\boldsymbol{\lambda}^* + \frac{1}{2}\left(2\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{A}'\boldsymbol{\lambda}^* - \boldsymbol{\lambda}^{*\prime}\mathbf{A}\boldsymbol{\Sigma}^{-1}\mathbf{A}'\boldsymbol{\lambda}^*\right) \\ &= \frac{1}{2}\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{\lambda}^{*\prime}\mathbf{A}\boldsymbol{\Sigma}^{-1}\mathbf{A}'\boldsymbol{\lambda}^*.\end{aligned} \tag{A.7}$$

The second term can be equivalently written as $-\frac{1}{2}\left(\boldsymbol{\Sigma}^{-1}\mathbf{A}'\boldsymbol{\lambda}^*\right)'\boldsymbol{\Sigma}\left(\boldsymbol{\Sigma}^{-1}\mathbf{A}'\boldsymbol{\lambda}^*\right)$, in which $\boldsymbol{\Sigma}^{-1}\mathbf{A}'\boldsymbol{\lambda}^*$ is the portfolio holdings attributable to constraints.

A.2 Proof of Proposition 2

We first observe that, conditioned on $\mathbf{X}$, the investor's optimization problem in (1) remains the same and, therefore, the portfolio holdings and Lagrange multipliers are given by (5)–(6) with static constraints $\mathbf{A}$ replaced by constraints that depend on characteristics $\mathbf{A}(\mathbf{X})$:

$$\boldsymbol{\omega}^* = \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu} - \mathbf{A}(\mathbf{X})'\boldsymbol{\lambda}^*\right) \tag{A.8}$$

$$\boldsymbol{\lambda}^* = \left(\mathbf{A}(\mathbf{X})\boldsymbol{\Sigma}^{-1}\mathbf{A}(\mathbf{X})'\right)^{-1}\left(\mathbf{A}(\mathbf{X})\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \mathbf{b}\right). \tag{A.9}$$

Therefore, the conditional expected return is given by:

$$\begin{aligned}\mathbb{E}\left[\boldsymbol{\omega}^{*\prime}\mathbf{r}|\mathbf{X}\right] &= \boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}^* = \boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}_{\mathrm{MVO}} + \boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}_{\mathrm{CSTR}} \\ &= \boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}_{\mathrm{MVO}} + \boldsymbol{\mu}'\boldsymbol{\omega}_{\mathrm{CSTR}} + (\boldsymbol{\mu}_{\mathbf{X}}' - \boldsymbol{\mu}')\boldsymbol{\omega}_{\mathrm{CSTR}},\end{aligned} \tag{A.10}$$

which proves (14). To get explicit expressions for each term, we have:

$$\boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}_{\mathrm{MVO}} = \boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \tag{A.11}$$

$$\boldsymbol{\mu}'\boldsymbol{\omega}_{\mathrm{CSTR}} = -\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{A}(\mathbf{X})'\boldsymbol{\lambda}^* \tag{A.12}$$

$$(\boldsymbol{\mu}_{\mathbf{X}}' - \boldsymbol{\mu}')\boldsymbol{\omega}_{\mathrm{CSTR}} = -(\boldsymbol{\mu}_{\mathbf{X}}' - \boldsymbol{\mu}')\boldsymbol{\Sigma}^{-1}\mathbf{A}(\mathbf{X})'\boldsymbol{\lambda}^*. \tag{A.13}$$

The conditional expected utility consists of two parts. The decomposition of the expected return is given by (A.10). Therefore, the conditional expected utility can be decomposed by:

$$
\begin{aligned}
\boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}^* - \frac{1}{2}\boldsymbol{\omega}^{*\prime}\boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{\omega}^* &= \boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}_{\mathrm{MVO}} + \boldsymbol{\mu}'\boldsymbol{\omega}_{\mathrm{CSTR}} + (\boldsymbol{\mu}_{\mathbf{X}}' - \boldsymbol{\mu}')\boldsymbol{\omega}_{\mathrm{CSTR}} \\
&\quad - \frac{1}{2}\left(\boldsymbol{\omega}_{\mathrm{MVO}} + \boldsymbol{\omega}_{\mathrm{CSTR}}\right)'\boldsymbol{\Sigma}_{\mathbf{X}}\left(\boldsymbol{\omega}_{\mathrm{MVO}} + \boldsymbol{\omega}_{\mathrm{CSTR}}\right) \\
&= \boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}_{\mathrm{MVO}} + \boldsymbol{\mu}'\boldsymbol{\omega}_{\mathrm{CSTR}} + (\boldsymbol{\mu}_{\mathbf{X}}' - \boldsymbol{\mu}')\boldsymbol{\omega}_{\mathrm{CSTR}} \\
&\quad - \frac{1}{2}\boldsymbol{\omega}_{\mathrm{MVO}}'\boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{\omega}_{\mathrm{MVO}} - \frac{1}{2}\boldsymbol{\omega}_{\mathrm{CSTR}}'\boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{\omega}_{\mathrm{CSTR}} - \boldsymbol{\omega}_{\mathrm{MVO}}'\boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{\omega}_{\mathrm{CSTR}} \\
&= \boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}_{\mathrm{MVO}} - \frac{1}{2}\boldsymbol{\omega}_{\mathrm{MVO}}'\boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{\omega}_{\mathrm{MVO}} \\
&\quad + \boldsymbol{\mu}'\boldsymbol{\omega}_{\mathrm{CSTR}} - \frac{1}{2}\boldsymbol{\omega}_{\mathrm{CSTR}}'\boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{\omega}_{\mathrm{CSTR}} \\
&\quad + (\boldsymbol{\mu}_{\mathbf{X}}' - \boldsymbol{\mu}')\boldsymbol{\omega}_{\mathrm{CSTR}} - \boldsymbol{\omega}_{\mathrm{MVO}}'\boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{\omega}_{\mathrm{CSTR}} \\
&= \boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}_{\mathrm{MVO}} - \frac{1}{2}\boldsymbol{\omega}_{\mathrm{MVO}}'\boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{\omega}_{\mathrm{MVO}} \\
&\quad + \boldsymbol{\mu}'\boldsymbol{\omega}_{\mathrm{CSTR}} - \frac{1}{2}\boldsymbol{\omega}_{\mathrm{CSTR}}'\boldsymbol{\Sigma}\boldsymbol{\omega}_{\mathrm{CSTR}} - \frac{1}{2}\boldsymbol{\omega}_{\mathrm{CSTR}}'\left(\boldsymbol{\Sigma}_{\mathbf{X}} - \boldsymbol{\Sigma}\right)\boldsymbol{\omega}_{\mathrm{CSTR}} \\
&\quad + (\boldsymbol{\mu}_{\mathbf{X}}' - \boldsymbol{\mu}')\boldsymbol{\omega}_{\mathrm{CSTR}} - \boldsymbol{\omega}_{\mathrm{MVO}}'\boldsymbol{\Sigma}\boldsymbol{\omega}_{\mathrm{CSTR}} - \boldsymbol{\omega}_{\mathrm{MVO}}'\left(\boldsymbol{\Sigma}_{\mathbf{X}} - \boldsymbol{\Sigma}\right)\boldsymbol{\omega}_{\mathrm{CSTR}} \\
&\overset{(1)}{=} \boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}_{\mathrm{MVO}} - \frac{1}{2}\boldsymbol{\omega}_{\mathrm{MVO}}'\boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{\omega}_{\mathrm{MVO}} \\
&\quad - \frac{1}{2}\boldsymbol{\omega}_{\mathrm{CSTR}}'\boldsymbol{\Sigma}\boldsymbol{\omega}_{\mathrm{CSTR}} \\
&\quad + (\boldsymbol{\mu}_{\mathbf{X}}' - \boldsymbol{\mu}')\boldsymbol{\omega}_{\mathrm{CSTR}} - \frac{1}{2}\boldsymbol{\omega}_{\mathrm{CSTR}}'\left(\boldsymbol{\Sigma}_{\mathbf{X}} - \boldsymbol{\Sigma}\right)\boldsymbol{\omega}_{\mathrm{CSTR}} - \boldsymbol{\omega}_{\mathrm{MVO}}'\left(\boldsymbol{\Sigma}_{\mathbf{X}} - \boldsymbol{\Sigma}\right)\boldsymbol{\omega}_{\mathrm{CSTR}} \\
&= \boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}_{\mathrm{MVO}} - \frac{1}{2}\boldsymbol{\omega}_{\mathrm{MVO}}'\boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{\omega}_{\mathrm{MVO}} \\
&\quad - \frac{1}{2}\boldsymbol{\omega}_{\mathrm{CSTR}}'\boldsymbol{\Sigma}\boldsymbol{\omega}_{\mathrm{CSTR}} \\
&\quad + (\boldsymbol{\mu}_{\mathbf{X}}' - \boldsymbol{\mu}')\boldsymbol{\omega}_{\mathrm{CSTR}} - \left(\frac{1}{2}\boldsymbol{\omega}_{\mathrm{CSTR}}' + \boldsymbol{\omega}_{\mathrm{MVO}}'\right)(\boldsymbol{\Sigma}_{\mathbf{X}} - \boldsymbol{\Sigma})\boldsymbol{\omega}_{\mathrm{CSTR}}.
\end{aligned}
\tag{A.14}
$$

Here step (1) follows from the fact that

$$\boldsymbol{\omega}_{\mathrm{MVO}}'\boldsymbol{\Sigma}\boldsymbol{\omega}_{\mathrm{CSTR}} = \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}\boldsymbol{\omega}_{\mathrm{CSTR}} = \boldsymbol{\mu}'\boldsymbol{\omega}_{\mathrm{CSTR}}.$$

This proves (17). To get explicit expressions for each term, we have:

$$\boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}_{\mathrm{MVO}} - \frac{1}{2}\boldsymbol{\omega}_{\mathrm{MVO}}'\boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{\omega}_{\mathrm{MVO}} = \boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{1}{2}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right)'\boldsymbol{\Sigma}_{\mathbf{X}}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right) \tag{A.15}$$

$$-\frac{1}{2}\boldsymbol{\omega}'_{\text{CSTR}}\boldsymbol{\Sigma}\boldsymbol{\omega}_{\text{CSTR}} = -\frac{1}{2}\left(\boldsymbol{\lambda}^{*\prime}\mathbf{A}(\mathbf{X})\boldsymbol{\Sigma}^{-1}\right)\boldsymbol{\Sigma}\left(\boldsymbol{\Sigma}^{-1}\mathbf{A}(\mathbf{X})'\boldsymbol{\lambda}^{*}\right) = -\frac{1}{2}\boldsymbol{\lambda}^{*\prime}\mathbf{A}(\mathbf{X})\boldsymbol{\Sigma}^{-1}\mathbf{A}(\mathbf{X})'\boldsymbol{\lambda}^{*} \tag{A.16}$$

$$\begin{aligned}(\boldsymbol{\mu}'_{\mathbf{X}} - \boldsymbol{\mu}')\boldsymbol{\omega}_{\text{CSTR}} &- \left(\frac{1}{2}\boldsymbol{\omega}'_{\text{CSTR}} + \boldsymbol{\omega}'_{\text{MVO}}\right)(\boldsymbol{\Sigma}_{\mathbf{X}} - \boldsymbol{\Sigma})\boldsymbol{\omega}_{\text{CSTR}} \\ &= -(\boldsymbol{\mu}'_{\mathbf{X}} - \boldsymbol{\mu}')\boldsymbol{\Sigma}^{-1}\mathbf{A}(\mathbf{X})'\boldsymbol{\lambda}^{*} - \left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{\Sigma}^{-1}\mathbf{A}(\mathbf{X})'\boldsymbol{\lambda}^{*}\right)'(\boldsymbol{\Sigma}_{\mathbf{X}} - \boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\mathbf{A}(\mathbf{X})'\boldsymbol{\lambda}^{*}.\end{aligned} \tag{A.17}$$

A.3 Proof of Proposition 3

Because the linear decomposition in Proposition 2 is conditioned on characteristics $\mathbf{X}$, the unconditional decomposition of the expected return follows from the linearity of expected value with respect to the distribution of $\mathbf{X}$.

For the unconditional decomposition of the expected utility, we observe that

$$\text{Var}\left(\boldsymbol{\omega}^{*\prime}\mathbf{r}\right) = \mathbb{E}\left[\text{Var}\left(\boldsymbol{\omega}^{*\prime}\mathbf{r}|\mathbf{X}\right)\right] + \text{Var}\left(\mathbb{E}\left[\boldsymbol{\omega}^{*\prime}\mathbf{r}|\mathbf{X}\right]\right). \tag{A.18}$$

Therefore,

$$\begin{aligned}\mathbb{E}\left[\boldsymbol{\omega}^{*\prime}\mathbf{r}\right] - \frac{1}{2}\text{Var}\left(\boldsymbol{\omega}^{*\prime}\mathbf{r}\right) &= \mathbb{E}\left[\mathbb{E}\left[\boldsymbol{\omega}^{*\prime}\mathbf{r}|\mathbf{X}\right]\right] - \frac{1}{2}\mathbb{E}\left[\text{Var}\left(\boldsymbol{\omega}^{*\prime}\mathbf{r}|\mathbf{X}\right)\right] - \frac{1}{2}\text{Var}\left(\mathbb{E}\left[\boldsymbol{\omega}^{*\prime}\mathbf{r}|\mathbf{X}\right]\right) \\ &= \mathbb{E}\left[\mathbb{E}\left[\boldsymbol{\omega}^{*\prime}\mathbf{r}|\mathbf{X}\right] - \frac{1}{2}\text{Var}\left(\boldsymbol{\omega}^{*\prime}\mathbf{r}|\mathbf{X}\right)\right] - \frac{1}{2}\text{Var}\left(\mathbb{E}\left[\boldsymbol{\omega}^{*\prime}\mathbf{r}|\mathbf{X}\right]\right) \\ &= \mathbb{E}\left[\boldsymbol{\mu}'_{\mathbf{X}}\boldsymbol{\omega}^{*} - \frac{1}{2}\boldsymbol{\omega}^{*\prime}\boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{\omega}^{*}\right] - \frac{1}{2}\text{Var}\left(\boldsymbol{\mu}'_{\mathbf{X}}\boldsymbol{\omega}^{*}\right).\end{aligned} \tag{A.19}$$

The first term follows from (17) in Proposition 2. The variance in the second term can be further decomposed into:

$$\begin{aligned}\text{Var}\left(\boldsymbol{\mu}'_{\mathbf{X}}\boldsymbol{\omega}^{*}\right) &= \text{Var}\left(\boldsymbol{\mu}'_{\mathbf{X}}\boldsymbol{\omega}_{\text{MVO}} + \boldsymbol{\mu}'_{\mathbf{X}}\boldsymbol{\omega}_{\text{CSTR}}\right) \\ &= \boldsymbol{\omega}'_{\text{MVO}}\text{Var}\left(\boldsymbol{\mu}'_{\mathbf{X}}\right)\boldsymbol{\omega}_{\text{MVO}} + \text{Var}\left(\boldsymbol{\mu}'_{\mathbf{X}}\boldsymbol{\omega}_{\text{CSTR}}\right) + 2\text{Cov}\left(\boldsymbol{\mu}'_{\mathbf{X}}\boldsymbol{\omega}_{\text{MVO}}, \boldsymbol{\mu}'_{\mathbf{X}}\boldsymbol{\omega}_{\text{CSTR}}\right).\end{aligned} \tag{A.20}$$

Substituting both terms back to (A.19) leads to:

$$\begin{aligned}
\mathbb{E}\left[\boldsymbol{\omega}^{*\prime}\mathbf{r}\right] - \frac{1}{2}\mathrm{Var}\left(\boldsymbol{\omega}^{*\prime}\mathbf{r}\right) &= \mathbb{E}[\boldsymbol{\mu}_{\mathbf{X}}^{\prime}]\boldsymbol{\omega}_{\mathrm{MVO}} - \frac{1}{2}\boldsymbol{\omega}_{\mathrm{MVO}}^{\prime}\mathbb{E}[\boldsymbol{\Sigma}_{\mathbf{X}}]\boldsymbol{\omega}_{\mathrm{MVO}} - \frac{1}{2}\boldsymbol{\omega}_{\mathrm{MVO}}^{\prime}\mathrm{Var}\left(\boldsymbol{\mu}_{\mathbf{X}}^{\prime}\right)\boldsymbol{\omega}_{\mathrm{MVO}} \\
&\quad - \frac{1}{2}\mathbb{E}[\boldsymbol{\omega}_{\mathrm{CSTR}}^{\prime}\boldsymbol{\Sigma}\boldsymbol{\omega}_{\mathrm{CSTR}}] \\
&\quad + \mathbb{E}\left[(\boldsymbol{\mu}_{\mathbf{X}}^{\prime} - \boldsymbol{\mu}^{\prime})\boldsymbol{\omega}_{\mathrm{CSTR}} - \boldsymbol{\omega}_{\mathrm{SHR}}^{\prime}(\boldsymbol{\Sigma}_{\mathbf{X}} - \boldsymbol{\Sigma})\boldsymbol{\omega}_{\mathrm{CSTR}}\right] \\
&\quad - \frac{1}{2}\mathrm{Var}\left(\boldsymbol{\mu}_{\mathbf{X}}^{\prime}\boldsymbol{\omega}_{\mathrm{CSTR}}\right) - \frac{1}{2}\mathrm{Cov}\left(\boldsymbol{\mu}_{\mathbf{X}}^{\prime}\boldsymbol{\omega}_{\mathrm{MVO}}, \boldsymbol{\mu}_{\mathbf{X}}^{\prime}\boldsymbol{\omega}_{\mathrm{CSTR}}\right) \\
&= \boldsymbol{\mu}^{\prime}\boldsymbol{\omega}_{\mathrm{MVO}} - \frac{1}{2}\boldsymbol{\omega}_{\mathrm{MVO}}^{\prime}\boldsymbol{\Sigma}\boldsymbol{\omega}_{\mathrm{MVO}} \\
&\quad - \frac{1}{2}\mathbb{E}\left[\boldsymbol{\omega}_{\mathrm{CSTR}}^{\prime}\boldsymbol{\Sigma}\boldsymbol{\omega}_{\mathrm{CSTR}}\right] \\
&\quad + \mathbb{E}\left[(\boldsymbol{\mu}_{\mathbf{X}}^{\prime} - \boldsymbol{\mu}^{\prime})\boldsymbol{\omega}_{\mathrm{CSTR}} - \boldsymbol{\omega}_{\mathrm{SHR}}^{\prime}(\boldsymbol{\Sigma}_{\mathbf{X}} - \boldsymbol{\Sigma})\boldsymbol{\omega}_{\mathrm{CSTR}}\right] \\
&\quad - \frac{1}{2}\left(\mathrm{Var}\left(\boldsymbol{\omega}_{\mathrm{CSTR}}^{\prime}\boldsymbol{\mu}_{\mathbf{X}}\right) + 2\mathrm{Cov}\left(\boldsymbol{\omega}_{\mathrm{MVO}}^{\prime}\boldsymbol{\mu}_{\mathbf{X}}, \boldsymbol{\omega}_{\mathrm{CSTR}}^{\prime}\boldsymbol{\mu}_{\mathbf{X}}\right)\right),
\end{aligned} \tag{A.21}$$

which completes the proof.

A.4 Proof of Proposition 4

Under Assumption 2, item 1, the conditional distribution $\mathbf{r}|\mathbf{X}$ is normal. To compute its conditional expected value, we first find a constant matrix $\mathbf{C}$ such that $\mathbf{Z} \equiv \mathbf{r} - \mathbf{C}\mathbf{X}$ is uncorrelated with $\mathbf{X}$. For this to be true, we require

$$0 = \mathrm{Cov}\left(\mathbf{Z}, \mathbf{X}\right) = \mathrm{Cov}\left(\mathbf{r} - \mathbf{C}\mathbf{X}, \mathbf{X}\right) = \mathrm{Cov}(\mathbf{r}, \mathbf{X}) - \mathbf{C} \cdot \mathrm{Cov}(\mathbf{X}, \mathbf{X}), \tag{A.22}$$

which yields:

$$\mathbf{C} = \mathrm{Cov}(\mathbf{r}, \mathbf{X})\mathrm{Cov}(\mathbf{X}, \mathbf{X})^{-1}. \tag{A.23}$$

Therefore,

$$\begin{aligned}
\boldsymbol{\mu}_{\mathbf{X}} &= \mathbb{E}[\mathbf{r}|\mathbf{X}] = \mathbb{E}[\mathbf{Z} + \mathbf{C}\mathbf{X}|\mathbf{X}] = \mathbb{E}[\mathbf{Z}|\mathbf{X}] + \mathbf{C}\mathbf{X} \\
&\overset{(1)}{=} \mathbb{E}[\mathbf{Z}] + \mathbf{C}\mathbf{X} = \mathbb{E}[\mathbf{r}] + \mathbf{C}(\mathbf{X} - \mathbb{E}[\mathbf{X}]) \\
&= \boldsymbol{\mu} + \mathrm{Cov}(\mathbf{r}, \mathbf{X})\mathrm{Cov}(\mathbf{X}, \mathbf{X})^{-1}(\mathbf{X} - \boldsymbol{\nu}), \\
&\overset{(2)}{=} \boldsymbol{\mu} + \begin{pmatrix} \mathrm{Cov}(\mathbf{r}, \mathbf{x}_1) & \cdots & \mathrm{Cov}(\mathbf{r}, \mathbf{x}_J) \end{pmatrix} \begin{pmatrix} \frac{1}{\sigma^2_{\mathbf{x}_1}}\mathbf{I} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \frac{1}{\sigma^2_{\mathbf{x}_J}}\mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{x}_1 - \boldsymbol{\nu}_1 \\ \vdots \\ \mathbf{x}_J - \boldsymbol{\nu}_J \end{pmatrix}, \\
&= \boldsymbol{\mu} + \sum_{j=1}^{J} \frac{\mathrm{Cov}(\mathbf{r}, \mathbf{x}_j)(\mathbf{x}_j - \boldsymbol{\nu}_j)}{\sigma^2_{\mathbf{x}_j}}.
\end{aligned} \tag{A.24}$$

Here step (1) follows from the fact that $\mathbf{Z}$ and $\mathbf{X}$ are uncorrelated multivariate Gaussian random vectors and therefore are independent. Step (2) follows from Assumption 2, items 3 and 4 that each characteristic is homoskedastic, and the characteristic values are independent both across different assets and between the J different constraints.

This proves (B.2) in Proposition B.1 of the Appendix. Note that we have only used items 1, 3, and 4 of Assumption 2, but not items 2 and 5.

To further prove (24) of Proposition 4, we observe that items 2 and 5 of Assumption 2 yield that:

$$\mathrm{Cov}(\mathbf{r}, \mathbf{x}_j) = \rho_j \sigma_{\mathbf{r}} \sigma_{\mathbf{x}_j} \mathbf{I}. \tag{A.25}$$

Combining (A.24) with (A.25) completes the proof of (24):

$$\boldsymbol{\mu}_{\mathbf{X}} = \boldsymbol{\mu} + \sum_{j=1}^{J} \frac{\rho_j \sigma_{\mathbf{r}} (\mathbf{x}_j - \boldsymbol{\nu}_j)}{\sigma_{\mathbf{x}_j}}. \tag{A.26}$$

Next, we compute the conditional covariance matrix of $\mathbf{r}|\mathbf{X}$:

$$
\begin{aligned}
\boldsymbol{\Sigma}_{\mathbf{X}} = \mathrm{Cov}(\mathbf{r}|\mathbf{X}) &= \mathrm{Cov}(\mathbf{Z}+\mathbf{C}\mathbf{X}|\mathbf{X}) = \mathrm{Cov}(\mathbf{Z}|\mathbf{X}) \\
&\overset{(1)}{=} \mathrm{Cov}(\mathbf{Z}) = \mathrm{Cov}(\mathbf{r}-\mathbf{C}\mathbf{X}) = \mathrm{Cov}(\mathbf{r}) - \mathbf{C}\mathrm{Cov}(\mathbf{X},\mathbf{X})\mathbf{C}' \\
&= \boldsymbol{\Sigma} - \mathrm{Cov}(\mathbf{r},\mathbf{X})\mathrm{Cov}(\mathbf{X},\mathbf{X})^{-1}\mathrm{Cov}(\mathbf{r},\mathbf{X})' \\
&\overset{(2)}{=} \boldsymbol{\Sigma} - \begin{pmatrix}\mathrm{Cov}(\mathbf{r},\mathbf{x}_1) & \cdots & \mathrm{Cov}(\mathbf{r},\mathbf{x}_J)\end{pmatrix} \begin{pmatrix} \frac{1}{\sigma^2_{\mathbf{x}_1}}\mathbf{I} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \frac{1}{\sigma^2_{\mathbf{x}_J}}\mathbf{I} \end{pmatrix} \begin{pmatrix}\mathrm{Cov}(\mathbf{r},\mathbf{x}_1)' \\ \vdots \\ \mathrm{Cov}(\mathbf{r},\mathbf{x}_J)'\end{pmatrix}, \\
&= \boldsymbol{\Sigma} - \sum_{j=1}^{J} \frac{\mathrm{Cov}(\mathbf{r},\mathbf{x}_j)\mathrm{Cov}(\mathbf{r},\mathbf{x}_j)'}{\sigma^2_{\mathbf{x}_j}}.
\end{aligned}
\tag{A.27}
$$

Again, step (1) follows from the fact that $\mathbf{Z}$ and $\mathbf{X}$ are uncorrelated multivariate Gaussian random vectors and therefore are independent. Step (2) follows from Assumption 2, items 3 and 4 that each characteristic is homoskedastic, and the characteristic values are independent both across different assets and between the J different constraints.

This proves (B.3) in Proposition B.1 of the Appendix. Note that we have only used items 1, 3, and 4 of Assumption 2, but not items 2 and 5.

To further prove (25) of Proposition 4, we again combine items 2 and 5 of Assumption 2 with (A.27),

$$
\boldsymbol{\Sigma}_{\mathbf{X}} = \boldsymbol{\Sigma} - \sum_{j=1}^{J} \frac{\rho_j^2 \sigma_{\mathbf{r}}^2 \sigma_{\mathbf{x}_j}^2 \mathbf{I}}{\sigma^2_{\mathbf{x}_j}} = \boldsymbol{\Sigma} - \sum_{j=1}^{J} \rho_j^2 \sigma_{\mathbf{r}}^2 \mathbf{I}, \tag{A.28}
$$

which completes the proof of (25).

A.5 Proof of Proposition 5

Substituting the excess return from information in Proposition 4 into the expected return decomposition of Proposition 2, we have:

$$
\begin{aligned}
\mathbb{E}\left[\boldsymbol{\omega}^{*\prime}\mathbf{r}|\mathbf{X}\right] = \boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}^* &= \boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}_{\mathrm{MVO}} + \boldsymbol{\mu}'\boldsymbol{\omega}_{\mathrm{CSTR}} + (\boldsymbol{\mu}_{\mathbf{X}}' - \boldsymbol{\mu}')\boldsymbol{\omega}_{\mathrm{CSTR}} \\
&= \boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}_{\mathrm{MVO}} + \boldsymbol{\mu}'\boldsymbol{\omega}_{\mathrm{CSTR}} + \sum_{j=1}^{J} \rho_j \sigma_{\mathbf{r}} \frac{(\mathbf{x}_j - \boldsymbol{\nu}_j)\boldsymbol{\omega}_{\mathrm{CSTR}}}{\sigma_{\mathbf{x}_j}},
\end{aligned}
\tag{A.29}
$$

which completes the proof of (26).

Substituting the excess return and excess covariance from information in Proposition 4 into the expected utility decomposition of Proposition 2, we have:

$$
\begin{aligned}
\boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}^* - \frac{1}{2}\boldsymbol{\omega}^{*\prime}\boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{\omega}^* &= \boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}_{\mathrm{MVO}} - \frac{1}{2}\boldsymbol{\omega}_{\mathrm{MVO}}'\boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{\omega}_{\mathrm{MVO}} \\
&\quad - \frac{1}{2}\boldsymbol{\omega}_{\mathrm{CSTR}}'\boldsymbol{\Sigma}\boldsymbol{\omega}_{\mathrm{CSTR}} \\
&\quad + (\boldsymbol{\mu}_{\mathbf{X}}' - \boldsymbol{\mu}')\boldsymbol{\omega}_{\mathrm{CSTR}} - \boldsymbol{\omega}_{\mathrm{SHR}}'(\boldsymbol{\Sigma}_{\mathbf{X}} - \boldsymbol{\Sigma})\boldsymbol{\omega}_{\mathrm{CSTR}} \\
&= \boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}_{\mathrm{MVO}} - \frac{1}{2}\boldsymbol{\omega}_{\mathrm{MVO}}'\boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{\omega}_{\mathrm{MVO}} \\
&\quad - \frac{1}{2}\boldsymbol{\omega}_{\mathrm{CSTR}}'\boldsymbol{\Sigma}\boldsymbol{\omega}_{\mathrm{CSTR}} \\
&\quad + \left(\sum_{j=1}^{J}\rho_j\sigma_{\mathbf{r}}\frac{(\mathbf{x}_j - \boldsymbol{\nu}_j)\boldsymbol{\omega}_{\mathrm{CSTR}}}{\sigma_{\mathbf{x}_j}} - \boldsymbol{\omega}_{\mathrm{SHR}}'\left(-\sum_{j=1}^{J}\rho_j^2\sigma_{\mathbf{r}}^2\mathbf{I}\right)\boldsymbol{\omega}_{\mathrm{CSTR}}\right) \\
&= \boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}_{\mathrm{MVO}} - \frac{1}{2}\boldsymbol{\omega}_{\mathrm{MVO}}'\boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{\omega}_{\mathrm{MVO}} \\
&\quad - \frac{1}{2}\boldsymbol{\omega}_{\mathrm{CSTR}}'\boldsymbol{\Sigma}\boldsymbol{\omega}_{\mathrm{CSTR}} \\
&\quad + \sum_{j=1}^{J}\left(\rho_j\sigma_{\mathbf{r}}\frac{(\mathbf{x}_j - \boldsymbol{\nu}_j)\boldsymbol{\omega}_{\mathrm{CSTR}}}{\sigma_{\mathbf{x}_j}} + \rho_j^2\sigma_{\mathbf{r}}^2\boldsymbol{\omega}_{\mathrm{SHR}}'\boldsymbol{\omega}_{\mathrm{CSTR}}\right),
\end{aligned}
\tag{A.30}
$$

which completes the proof of (27).

Similarly, substituting the excess return and excess covariance from the information in Proposition B.1 into the decomposition of Proposition 2 yields more general results in Proposition B.2 under items 1, 3, and 4 of Assumption 2, but not requiring items 2 and 5.

A.6 Proof of Proposition 6

Given the decomposition of portfolio holdings in (7), we have:

$$
\begin{aligned}
\tilde{\mathbf{r}}'\boldsymbol{\omega}^* &= \tilde{\mathbf{r}}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \tilde{\mathbf{r}}'\boldsymbol{\Sigma}^{-1}\mathbf{A}'\boldsymbol{\lambda}^* \\
&= \tilde{\mathbf{r}}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \tilde{\mathbf{r}}_{\mathrm{Static}}'\boldsymbol{\Sigma}^{-1}\mathbf{A}'\boldsymbol{\lambda}^* - \tilde{\mathbf{r}}_{\mathrm{Info}}'\boldsymbol{\Sigma}^{-1}\mathbf{A}'\boldsymbol{\lambda}^* \\
&\overset{(1)}{=} \tilde{\mathbf{r}}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \tilde{\mathbf{r}}_{\mathrm{Static}}'\boldsymbol{\Sigma}^{-1}\mathbf{A}'\boldsymbol{\lambda}^* - \left(\sum_{j=1}^{J}\frac{\rho(\tilde{\mathbf{r}}, \tilde{\mathbf{x}}_j)\tilde{\sigma}_{\mathbf{r}}(\tilde{\mathbf{x}}_j' - \tilde{\boldsymbol{\nu}}_j')}{\tilde{\sigma}_{\mathbf{x}_j}}\right)\boldsymbol{\Sigma}^{-1}\mathbf{A}'\boldsymbol{\lambda}^* \\
&= \tilde{\mathbf{r}}'\boldsymbol{\omega}_{\mathrm{MVO}} + \tilde{\mathbf{r}}_{\mathrm{Static}}'\boldsymbol{\omega}_{\mathrm{CSTR}} + \sum_{j=1}^{J}\rho(\tilde{\mathbf{r}}, \tilde{\mathbf{x}}_j)\tilde{\sigma}_{\mathbf{r}}\frac{(\tilde{\mathbf{x}}_j' - \tilde{\boldsymbol{\nu}}_j')\boldsymbol{\omega}_{\mathrm{CSTR}}}{\tilde{\sigma}_{\mathbf{x}_j}},
\end{aligned}
\tag{A.31}
$$

where step (1) follows from the definition of $\tilde{\mathbf{r}}_{\mathrm{Info}}$ in (29).

A.7 Proof of Proposition 7

When there is a single constraint $\mathbf{A}(\mathbf{x}) = \mathbf{x}'$, the Lagrange multiplier λ^* is a scalar. From (16), we can further simplify λ^* to:

$$\lambda^* = \frac{\mathbf{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - b}{\mathbf{x}'\boldsymbol{\Sigma}^{-1}\mathbf{x}}. \tag{A.32}$$

Substituting this Lagrange multiplier into the decomposition of portfolio holdings in (7) yields:

$$\boldsymbol{\omega}^* = \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \lambda^*\boldsymbol{\Sigma}^{-1}\mathbf{x} = \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{\mathbf{x}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - b}{\mathbf{x}'\boldsymbol{\Sigma}^{-1}\mathbf{x}}\boldsymbol{\Sigma}^{-1}\mathbf{x}. \tag{A.33}$$

Substituting (A.33) into the last term (the information component) of the expected return decomposition in (26) and assuming the expected value $\boldsymbol{\nu} = \mathbb{E}[\mathbf{x}] = 0$ yields:

$$\begin{aligned}
\frac{\rho\sigma_{\mathbf{r}}}{\sigma_{\mathbf{x}}}\mathbf{x}'\boldsymbol{\omega}_{\text{CSTR}} &= \frac{\rho\sigma_{\mathbf{r}}}{\sigma_{\mathbf{x}}}\mathbf{x}'\left(-\lambda^*\boldsymbol{\Sigma}^{-1}\mathbf{x}\right) \\
&= \frac{\rho\sigma_{\mathbf{r}}}{\sigma_{\mathbf{x}}}\mathbf{x}'\left(-\frac{\mathbf{x}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - b}{\mathbf{x}'\boldsymbol{\Sigma}^{-1}\mathbf{x}}\boldsymbol{\Sigma}^{-1}\mathbf{x}\right) \\
&= \frac{\rho\sigma_{\mathbf{r}}}{\sigma_{\mathbf{x}}}\left(-\frac{\mathbf{x}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - b}{\mathbf{x}'\boldsymbol{\Sigma}^{-1}\mathbf{x}}\mathbf{x}'\boldsymbol{\Sigma}^{-1}\mathbf{x}\right) \\
&= \frac{\rho\sigma_{\mathbf{r}}}{\sigma_{\mathbf{x}}}\left(b - \mathbf{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right).
\end{aligned} \tag{A.34}$$

Therefore, the full decomposition of expected return in (26) reduces to:

$$\begin{aligned}
\mathbb{E}\left[\boldsymbol{\omega}^{*\prime}\mathbf{r}|\mathbf{x}\right] &= \boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}_{\text{MVO}} + \boldsymbol{\mu}'\boldsymbol{\omega}_{\text{CSTR}} + \frac{\rho\sigma_{\mathbf{r}}}{\sigma_{\mathbf{x}}}\mathbf{x}'\boldsymbol{\omega}_{\text{CSTR}} \\
&= \boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \lambda^*\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{x} + \frac{\rho\sigma_{\mathbf{r}}}{\sigma_{\mathbf{x}}}\left(b - \mathbf{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right) \\
&= \boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \frac{\mathbf{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\mathbf{x}'\boldsymbol{\Sigma}^{-1}\mathbf{x}}\left(b - \mathbf{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right) + \frac{\rho\sigma_{\mathbf{r}}}{\sigma_{\mathbf{x}}}\left(b - \mathbf{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right).
\end{aligned} \tag{A.35}$$

Note that both (A.34) and (A.35) assumes that the Lagrange multiplier $\lambda^* \neq 0$. When the constraint is not binding for inequality constraints, (A.34) becomes zero and the last two terms of (A.35) vanishes.

A.8 Proof of Proposition 8

When $\mathbf{x}$ is a vector of binary random variables following the Bernoulli distribution, we need to quantify the excess expected return $\boldsymbol{\mu}_{\mathbf{x}} - \boldsymbol{\mu}$ in (14) of Proposition 2. The i-the element of $\boldsymbol{\mu}_{\mathbf{x}} - \boldsymbol{\mu}$ is given by:

$$\begin{aligned} \mu_{i,\mathbf{x}} - \mu_i &= \mathbb{E}[r_i|\mathbf{x}] - \mathbb{E}[r_i] \\ &\overset{(1)}{=} \mathbb{E}[r_i|x_i] - \mathbb{E}[r_i] \\ &\overset{(2)}{=} x_i\left(\mathbb{E}[r_i|x_i=1] - \mathbb{E}[r_i]\right) + (1-x_i)\left(\mathbb{E}[r_i|x_i=0] - \mathbb{E}[r_i]\right), \end{aligned} \tag{A.36}$$

for $i = 1, 2, \ldots, N$. Here step (1) follows from item 5 of Assumption 2 which guarantees that there is no cross-correlation between the return and characteristic value of different assets. Step (2) uses the fact that x_i is either 1 or 0.

To compute $\mathbb{E}[r_i|x_i = 1]$, we consider the correlation between the characteristic value and return of the i-th asset:

$$\begin{aligned} \rho \equiv \mathrm{Corr}(x_i, r_i) &= \frac{\mathrm{Cov}(x_i, r_i)}{\sqrt{\mathrm{Var}(x_i)\mathrm{Var}(r_i)}} \\ &= \frac{\mathbb{E}[x_i r_i] - \mathbb{E}[x_i]\mathbb{E}[r_i]}{\sqrt{\mathrm{Var}(x_i)\mathrm{Var}(r_i)}} \\ &\overset{(1)}{=} \frac{\mathbb{E}[r_i|x_i=1]\mathbb{P}(x_i=1) - \mathbb{P}(x_i=1)\mathbb{E}[r_i]}{\sqrt{\mathbb{P}(x_i=1)(1-\mathbb{P}(x_i=1))\mathrm{Var}(r_i)}} \\ &\overset{(2)}{=} \frac{\left(\mathbb{E}[r_i|x_i=1] - \mathbb{E}[r_i]\right)\pi_{x_i=1}}{\sqrt{\pi_{x_i=1}(1-\pi_{x_i=1})\sigma_{\mathbf{r}}^2}} \\ &= \frac{\mathbb{E}[r_i|x_i=1] - \mathbb{E}[r_i]}{\sigma_{\mathbf{r}}}\sqrt{\frac{\pi_{x_i=1}}{\pi_{x_i=0}}}. \end{aligned} \tag{A.37}$$

Here step (1) follows from the fact that x_i follows the Bernoulli distribution, whose variance is given by $\mathbb{P}(x_i = 1)(1 - \mathbb{P}(x_i = 1))$. Step (2) assumes homoskedastic returns in item #2 of Assumption 2. We use $\pi_{x_i=1} = \mathbb{P}(x_i = 1)$ and $\pi_{x_i=0} = \mathbb{P}(x_i = 0)$ to denote the marginal probability of the i-th asset being included or excluded from the portfolio.

Equation (A.37) implies that

$$\mathbb{E}[r_i|x_i=1] - \mathbb{E}[r_i] = \rho\sigma_{\mathbf{r}}\sqrt{\frac{\pi_{x_i=0}}{\pi_{x_i=1}}}. \tag{A.38}$$

To compute $\mathbb{E}[r_i|x_i = 0]$, we observe that:

$$\mathbb{E}[r_i] = \mathbb{E}[r_i|x_i=1]\pi_{x_i=1} + \mathbb{E}[r_i|x_i=0]\pi_{x_i=0}, \tag{A.39}$$

which yields:

$$\mathbb{E}[r_i|x_i=0]-\mathbb{E}[r_i]=-\left(\mathbb{E}[r_i|x_i=1]-\mathbb{E}[r_i]\right)\frac{\pi_{x_i=1}}{\pi_{x_i=0}}=-\rho\sigma_{\mathbf{r}}\sqrt{\frac{\pi_{x_i=1}}{\pi_{x_i=0}}}, \tag{A.40}$$

Substituting (A.38) and (A.40) into (A.36) yields:

$$\mu_{i,\mathbf{x}}-\mu_i=\rho\sigma_{\mathbf{r}}\left(x_i\sqrt{\frac{\pi_{x_i=0}}{\pi_{x_i=1}}}-(1-x_i)\sqrt{\frac{\pi_{x_i=1}}{\pi_{x_i=0}}}\right) \tag{A.41}$$

for $i=1,2,\ldots,N$. Therefore,

$$\boldsymbol{\mu}_{\mathbf{x}}-\boldsymbol{\mu}=\rho\sigma_{\mathbf{r}}\left(\mathbf{x}\odot\mathbf{u}-(1-\mathbf{x})\odot\mathbf{v}\right), \tag{A.42}$$

where

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

Substituting this into (14) of Proposition 2 yields:

$$\begin{aligned}\mathbb{E}\left[\boldsymbol{\omega}^{*\prime}\mathbf{r}|\mathbf{x}\right]&=\boldsymbol{\mu}_{\mathbf{x}}'\boldsymbol{\omega}_{\mathrm{MVO}}+\boldsymbol{\mu}'\boldsymbol{\omega}_{\mathrm{CSTR}}+(\boldsymbol{\mu}_{\mathbf{x}}'-\boldsymbol{\mu}')\boldsymbol{\omega}_{\mathrm{CSTR}}\\&=\boldsymbol{\mu}_{\mathbf{x}}'\boldsymbol{\omega}_{\mathrm{MVO}}+\boldsymbol{\mu}'\boldsymbol{\omega}_{\mathrm{CSTR}}+\rho\sigma_{\mathbf{r}}\left(\mathbf{x}\odot\mathbf{u}-(1-\mathbf{x})\odot\mathbf{v}\right)'\boldsymbol{\omega}_{\mathrm{CSTR}},\end{aligned} \tag{A.43}$$

which completes the proof of (35).

B Internet Appendix

In this Internet Appendix, we provide additional technical results.

B.1 General Dependence between Returns and Characteristics

In this section, we revisit Assumption 2 and results in Section 3.3. In particular, we derive a version of Propositions 4–5 under items 1, 3, and 4 of Assumption 2, but not requiring items 2 and 5.

When returns and asset characteristics have a general dependence, the covariance matrix of $(\mathbf{r}', \mathbf{x}_1', \ldots, \mathbf{x}_J')$ can be written as:

$$\begin{pmatrix} \boldsymbol{\Sigma} & \mathrm{Cov}\left(\mathbf{r}, (\mathbf{x}_1', \ldots, \mathbf{x}_J')\right) \\ \mathrm{Cov}\left((\mathbf{x}_1', \ldots, \mathbf{x}_J'), \mathbf{r}\right) & \begin{pmatrix} \sigma^2_{\mathbf{x}_1}\mathbf{I} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma^2_{\mathbf{x}_J}\mathbf{I} \end{pmatrix} \end{pmatrix}. \tag{B.1}$$

Recall that $\mathbf{X}$ represents the $(N \times J)$-dimensional vector $(\mathbf{x}_1', \ldots, \mathbf{x}_J')'$, and we use $\boldsymbol{\nu} \equiv (\boldsymbol{\nu}_1', \ldots, \boldsymbol{\nu}_J')'$ to denote the expected value of $\mathbf{X}$. The following result generalizes Proposition 4 under the above weaker assumptions. We provide the proof together with the proof of Proposition 4 in Section A.4.

Proposition B.1. *Under the covariance structure in (B.1),* $\mathbf{r}|\mathbf{X}$ *is normally distributed with an expected value given by:*

$$\boldsymbol{\mu}_{\mathbf{X}} = \mathbb{E}[\mathbf{r}|\mathbf{X}] = \boldsymbol{\mu} + \sum_{j=1}^{J} \frac{\mathrm{Cov}(\mathbf{r}, \mathbf{x}_j)}{\sigma_{\mathbf{x}_j}} \frac{(\mathbf{x}_j - \boldsymbol{\nu}_j)}{\sigma_{\mathbf{x}_j}}, \tag{B.2}$$

and a covariance matrix given by:

$$\boldsymbol{\Sigma}_{\mathbf{X}} = \mathrm{Cov}(\mathbf{r}|\mathbf{X}) = \boldsymbol{\Sigma} - \sum_{j=1}^{J} \frac{\mathrm{Cov}(\mathbf{r}, \mathbf{x}_j)\mathrm{Cov}(\mathbf{r}, \mathbf{x}_j)'}{\sigma^2_{\mathbf{x}_j}}, \tag{B.3}$$

where $\mathrm{Cov}(\mathbf{r}, \mathbf{x}_j)$ *is the* $N \times N$ *covariance matrix between two* N*-dimensional vector* $\mathbf{r}$ *and* $\mathbf{x}_j$*.*

Proposition B.1 also allows for more explicit decompositions of the expected return and utility of the portfolio by substituting (B.2)–(B.3) into Proposition 2, which gives the following generalized result of Proposition 5 under the above weaker assumptions. We provide the proof together with the proof of Proposition 5 in Section A.5.

Proposition B.2. *Under the covariance structure in (B.1) 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.*

$$\mathbb{E}\left[\boldsymbol{\omega}^{*\prime}\mathbf{r}|\mathbf{X}\right] = \boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}^* = \boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}_{\text{MVO}} + \boldsymbol{\mu}'\boldsymbol{\omega}_{\text{CSTR}} + \sum_{j=1}^{J}\boldsymbol{\omega}_{\text{CSTR}}'\frac{\text{Cov}(\mathbf{r},\mathbf{x}_j)}{\sigma_{\mathbf{x}_j}}\frac{(\mathbf{x}_j-\boldsymbol{\nu}_j)}{\sigma_{\mathbf{x}_j}}. \quad \text{(B.4)}$$

2. *Expected utility decomposition.*

$$\begin{aligned}\boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}^* - \frac{1}{2}\boldsymbol{\omega}^{*\prime}\boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{\omega}^* &= \boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}_{\text{MVO}} - \frac{1}{2}\boldsymbol{\omega}_{\text{MVO}}'\boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{\omega}_{\text{MVO}} - \frac{1}{2}\boldsymbol{\omega}_{\text{CSTR}}'\boldsymbol{\Sigma}\boldsymbol{\omega}_{\text{CSTR}} \\ &+ \sum_{j=1}^{J}\left(\boldsymbol{\omega}_{\text{CSTR}}'\frac{\text{Cov}(\mathbf{r},\mathbf{x}_j)}{\sigma_{\mathbf{x}_j}}\frac{(\mathbf{x}_j-\boldsymbol{\nu}_j)}{\sigma_{\mathbf{x}_j}} + \frac{1}{2}\boldsymbol{\omega}_{\text{SHR}}'\frac{\text{Cov}(\mathbf{r},\mathbf{x}_j)\text{Cov}(\mathbf{r},\mathbf{x}_j)'}{\sigma_{\mathbf{x}_j}^2}\boldsymbol{\omega}_{\text{CSTR}}\right).\end{aligned} \quad \text{(B.5)}$$

Furthermore, the unconditional expected return and utility can be decomposed into components that are attributable to the unconstrained MVO portfolio, static constraints, and information, respectively, by following Proposition 3.

B.2 Long-Only ESG Portfolios

Portfolio Construction. In this section, we consider investors who construct long-only portfolios each year by solving the following problem.

$$\begin{aligned}\max_{\boldsymbol{\omega}} \quad & \boldsymbol{\omega}'\boldsymbol{\mu} - \frac{\eta}{2}\boldsymbol{\omega}'\boldsymbol{\Sigma}\boldsymbol{\omega} \\ \text{s.t.} \quad & \boldsymbol{\omega}'\mathbf{1} = 1 \\ & \boldsymbol{\omega}'\mathbf{x}_{\text{ESG}} \geq b \\ & \boldsymbol{\omega} \geq \mathbf{0}.\end{aligned} \quad \text{(B.6)}$$

In contrast to (39), we have an additional constraint that all portfolio weights must be non-negative. We again set $b = 1$ as an example in our analysis, and use the return forecast in (40) to estimate $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. We choose η such that the unconstrained MVO portfolio has a realistic leverage.

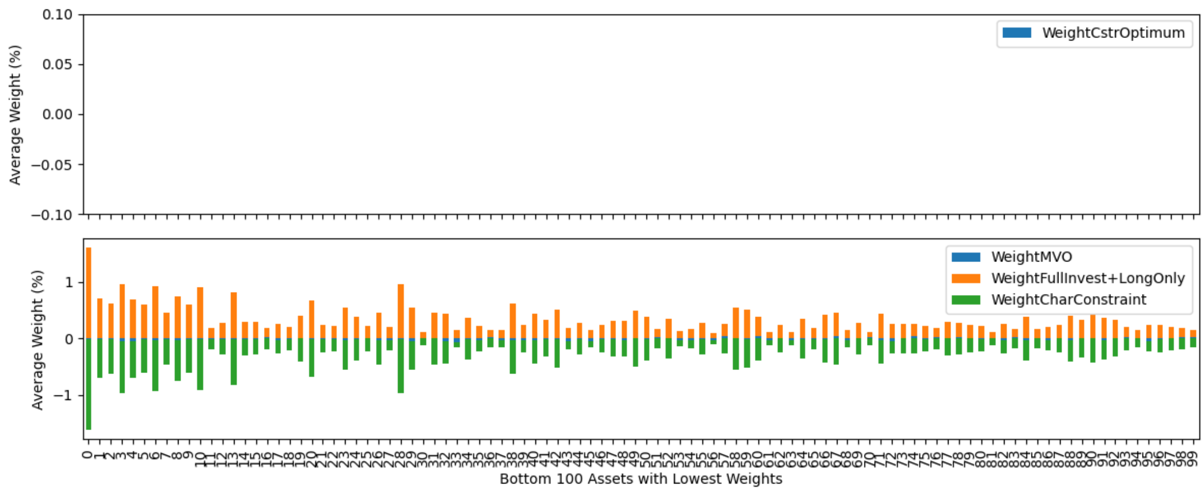
Portfolio Holdings Decomposition. Figure B.1 shows the bottom and top 100 assets with the lowest and highest portfolio weights for the optimal portfolio, averaged over all years, respectively. We decompose the portfolio weights into components corresponding to the unconstrained MVO portfolio and constraints, respectively, based on (7) in Proposition 1. For performance attributions of long-only portfolios, we always combined the full investment constraint and the long-only constraint for simplicity.

In Figure B.1a, the bottom assets, by design, have zero weights. This is a result of the negative contribution from the ESG constraint (green) combined with the positive contribution from the full investment and long-only constraints (orange). In other words, these assets tend to have low ESG scores, but the long-only constraint forces them to have zero weights instead of negative weights. On the other hand, for the top 100 assets in Figure B.1b, the most significant contribution comes from the ESG constraint (green). The full investment and long-only constraints (orange) generally add negative contributions.

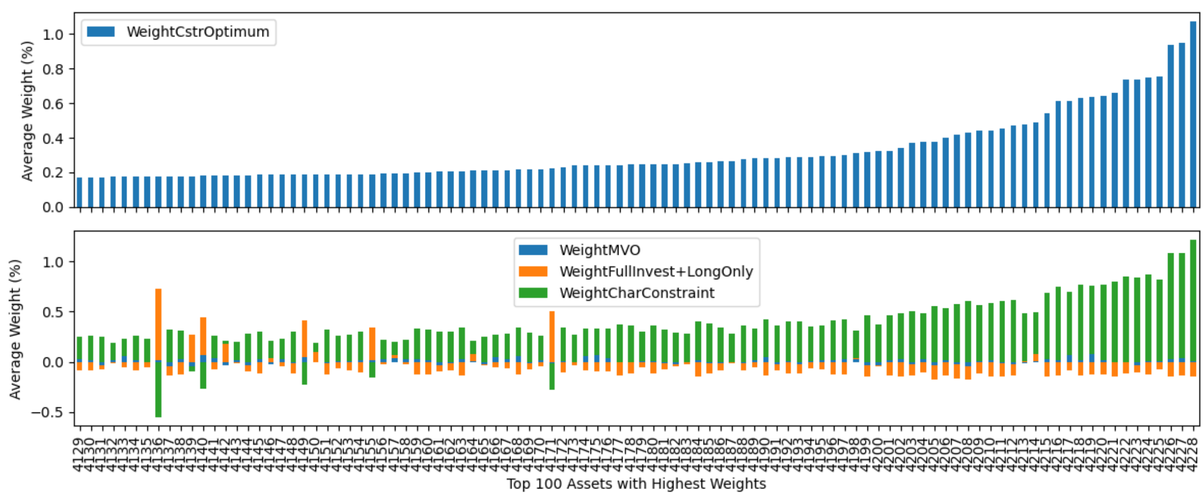
Expected Return and Utility Decomposition. Figure B.2 demonstrates the decomposition of the expected utility and expected return of the long-only portfolio into different components.

The upper panel of Figure B.2a shows that the expected utility of the optimal portfolio is negative in most years in our sample. This utility is decomposed into three components in the lower panel using (27) in Proposition 5. The expected utility of the unconstrained MVO portfolio (blue) is positive over all years. Compared with the long/short portfolio in Figure 8, the expected utility contribution of the three constraints (orange), treated as static, is now much more negative due to the addition of the long-only constraint. Like the long/short portfolio, the expected utility contribution from the information contained in the constraints (green) varies over time. The pattern is again consistent with the pattern of correlations between asset returns and ESG scores in Figure 6.

Figure B.2b shows the expected utility of the optimal portfolio and its decomposition based on (26) in Proposition 5. The full investment and long-only constraints (orange) contribute negatively to expected returns. The ESG constraint (green) can contribute either positively or negatively to the expected returns, but generally on a very small scale relative to other components. The expected return contribution from the information is very significant, which is strongly negative in 2004, 2005, 2007, and 2009, and is strongly positive in 2011, 2014, and 2017. However, the negative contributions from the full investment and long-only constraints are so strong that the expected return of the constrained portfolio is lower than that of the unconstrained MVO portfolio in most years, except for 2014.

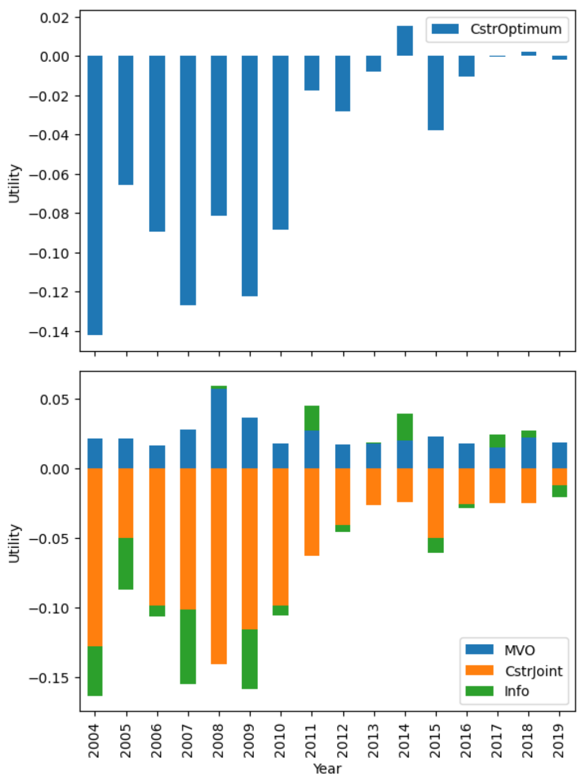


(a) Bottom assets with lowest weights.

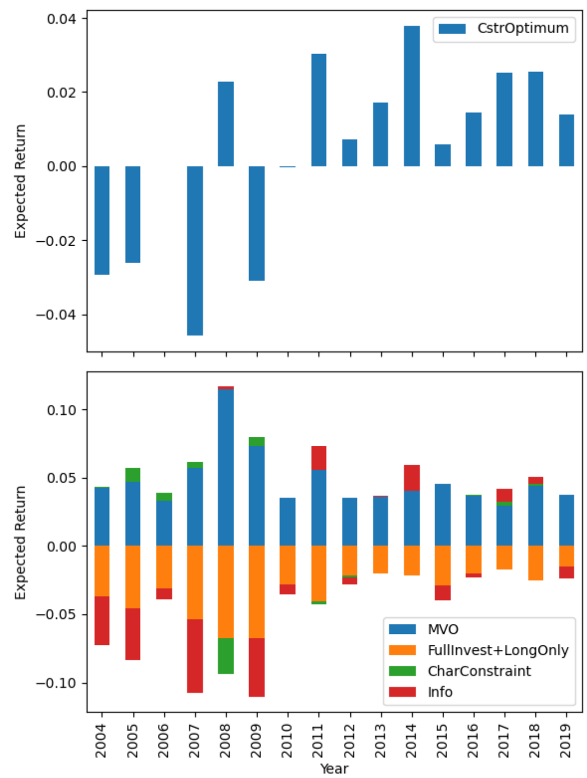


(b) Top assets with highest weights.

Figure B.1: Average portfolio weights over all years and their decomposition, for the long-only portfolio defined in (B.6) with a constraint on the average portfolio characteristic value ($\boldsymbol{\omega}'\mathbf{x}_{\text{ESG}} \geq 1.0$). (a) shows the bottom assets with the lowest weights and (b) shows the top assets with the highest weights. In each subfigure, the top panel shows the portfolio weights (%) of the constrained portfolio. The bottom panel shows the decomposition of the weights into components corresponding to the unconstrained MVO portfolio (blue), the full investment and long-only constraints combined together ($\boldsymbol{\omega}'\mathbf{1} = 1$ and $\boldsymbol{\omega} \geq \mathbf{0}$, orange), and the ESG constraint ($\boldsymbol{\omega}'\mathbf{x}_{\text{ESG}} \geq 1.0$, green).



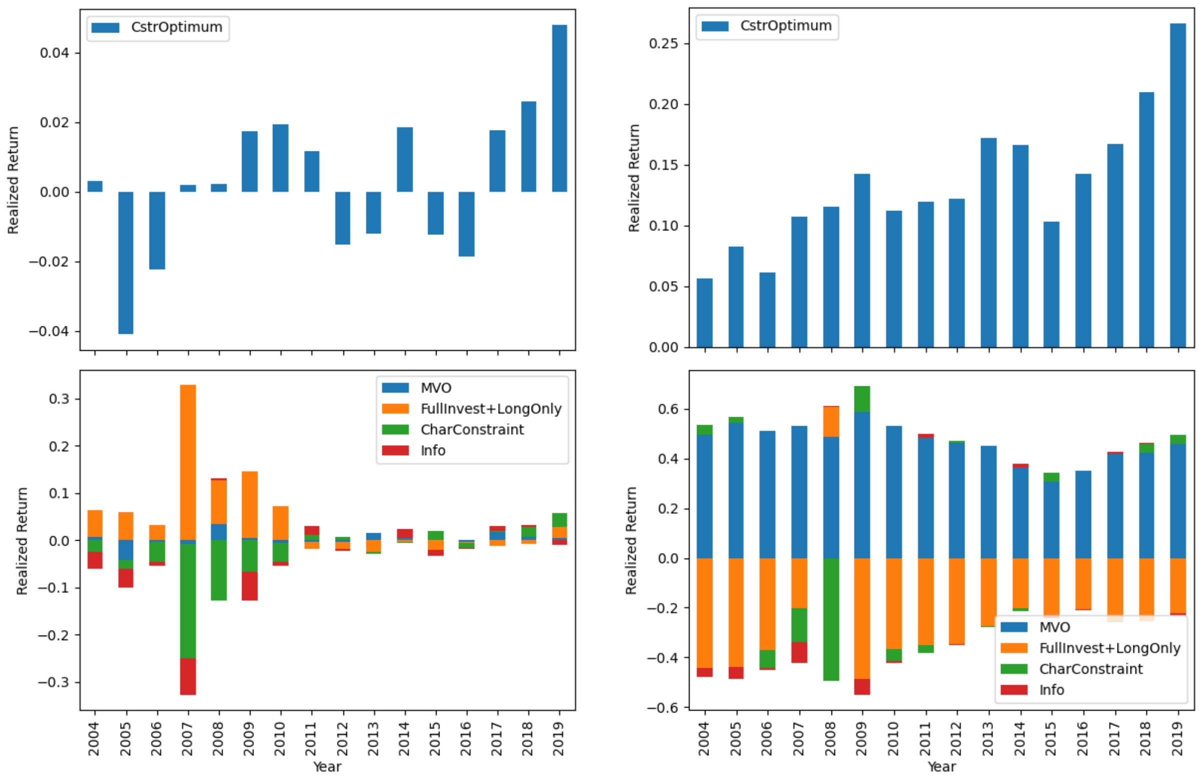
(a) Expected Utility



(b) Expected Return

Figure B.2: Expected return and utility and their decomposition, for the long-only portfolio defined in (B.6) with a constraint on the average portfolio characteristic value ($\boldsymbol{\omega}'\mathbf{x}_{\text{ESG}} \geq 1.0$). 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 (blue), all constraints treated as static (orange), and the information from the ESG constraint (green). 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 (blue), the full investment and long-only constraints combined together (orange), the ESG constraint treated as static ($\boldsymbol{\omega}'\mathbf{x}_{\text{ESG}} \geq 1.0$, green), and the information from the ESG constraint (red).

Realized Return Decomposition. Figure B.3 shows the realized returns of the optimal portfolio and *ex post* attribution of returns. Here we again compare a portfolio constructed without a return forecast ($p = 0$) in Figure B.3a with one based on a return forecast ($p = 0.1$) in Figure B.3b.



(a) No Return Predictability

(b) Return Predictability $p = 0.1$

Figure B.3: Realized return and their *ex post* decomposition, for the long-only portfolio defined in (B.6) with a constraint on the average portfolio characteristic value ($\boldsymbol{\omega}'\mathbf{x}_{\text{ESG}} \geq 1.0$). (a) corresponds to a return estimator in (40) with no ability to forecast future expected returns ($p = 0$), and (b) corresponds to a return estimator with some level of predictability ($p = 0.1$). 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 (blue), the full investment and long-only constraints combined together (orange), the ESG constraint treated as static ($\boldsymbol{\omega}'\mathbf{x}_{\text{ESG}} \geq 1.0$, green), and the information from the ESG constraint (red).

The upper panel of Figure B.3a shows that, without any return forecast, the realized residual returns in excess of the Fama-French five-factor model of the constrained portfolio fluctuate around zero over the 16 years in our sample.

The lower panel decomposes the realized return of the constrained portfolio based on Proposition 6. The full investment and long-only constraints (orange) generally contribute to the returns positively, especially before 2010. This is again because the average residual returns during these years are positive, as shown in Table 1, and the long-only constraint now plays the role of forcing the portfolio to take advantage of these positive returns. The ESG constraint, though, generally contributes negatively to realized returns in these years.

In addition, Figure B.3b shows results parallel to those in Figure B.3a, but with a return forecast ($p = 0.1$) in portfolio construction. As expected, the realized residual returns for both the unconstrained and the constrained portfolio dramatically increase. However, the constraints limit the ability of the portfolio to take advantage of the return predictability.

In both cases, the information component contributes negatively to realized returns before 2010, and positively in 2011, 2014, 2017, and 2018, which is consistent with results in Propositions 6 and the pattern of correlations between asset returns and ESG scores in Figure 6. Overall, these components explain the difference in residual returns between the unconstrained MVO and the constrained portfolio.

B.3 Exclusionary ESG Investing

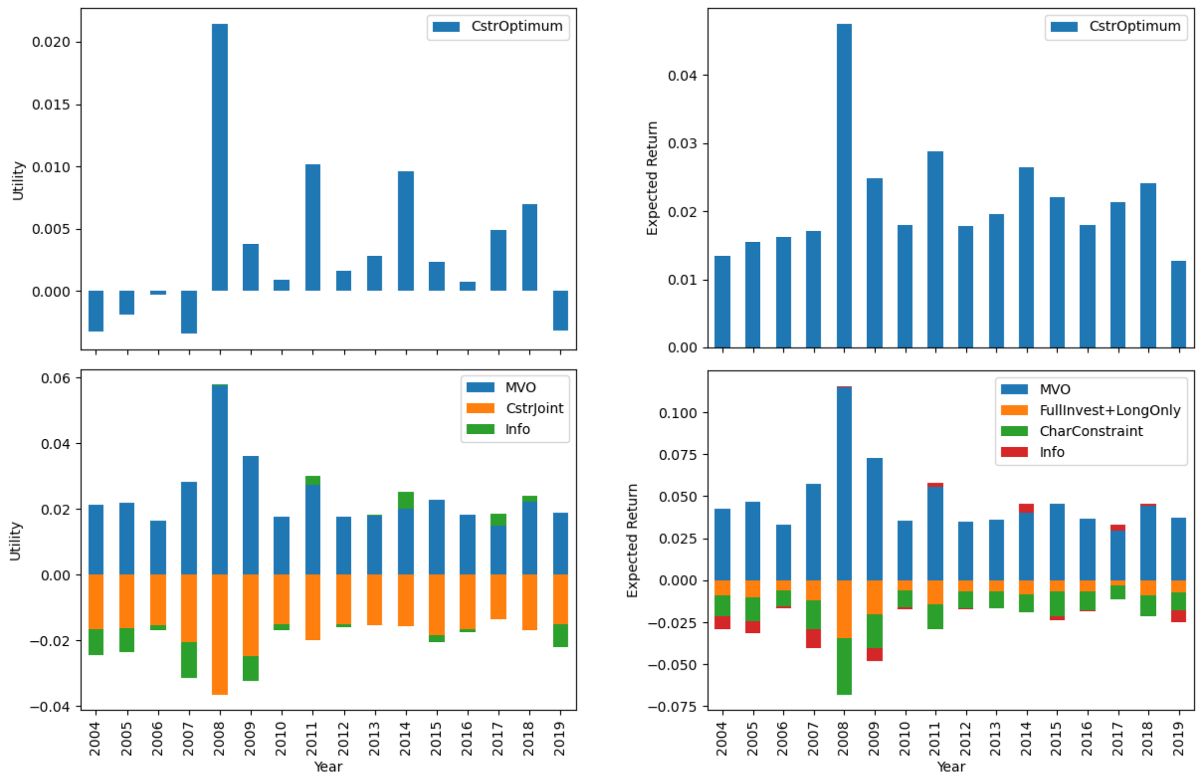
Portfolio Construction. Another common heuristic to implement ESG investing is to directly exclude assets with low ESG scores. We consider investors who construct long-only portfolios each year by solving the following problem.

$$\begin{aligned} \max_{\boldsymbol{\omega}} \quad & \boldsymbol{\omega}'\boldsymbol{\mu} - \frac{\eta}{2}\boldsymbol{\omega}'\boldsymbol{\Sigma}\boldsymbol{\omega} \\ \text{s.t.} \quad & \boldsymbol{\omega}'\mathbf{1} = 1 \\ & \omega_{[i]} = 0 \quad \text{for } i \leq 500 \text{ assets ordered by } \mathbf{x}_{\text{ESG}} \\ & \boldsymbol{\omega} \geq \mathbf{0}. \end{aligned} \tag{B.7}$$

In contrast to (B.6), we replace the constraint on the average ESG score of the portfolio with an alternative constraint that excludes the bottom 500 assets ordered by their ESG scores. We again use the return forecast in (40) to estimate $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. We choose η such that the unconstrained MVO portfolio has a realistic leverage.

Expected Return and Utility Decomposition. Figure B.4 demonstrates the decomposition of the expected return and utility of the portfolio into different components.

The upper panel of Figure B.4a shows that the expected utility of the optimal portfolio is positive between 2008 and 2018 and negative in other years. This utility is decomposed into three components in the lower panel using (27) in Proposition 5. The expected utility of the unconstrained MVO portfolio (blue) is positive over all years. Like the long/short portfolio in Figure B.2, the expected utility contribution of the three constraints (orange), treated as static, is negative. The expected utility contribution from information contained in the constraints (green), however, varies over time. The pattern is again consistent with the pattern of correlations between asset returns and ESG scores in Figure 6.



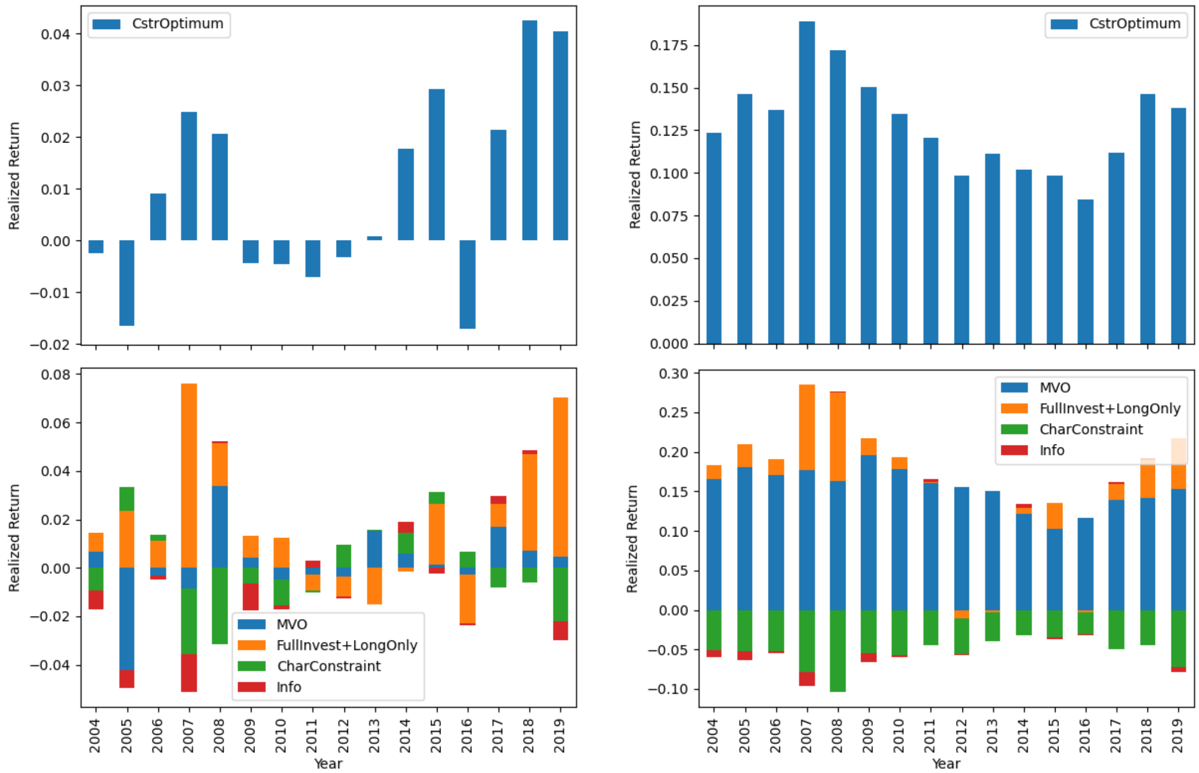
(a) Expected Utility

(b) Expected Return

Figure B.4: Expected return and utility and their decomposition, for the long-only portfolio defined in (B.7) with a constraint that excludes the bottom 500 assets ordered by their ESG scores. 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 (blue), all constraints treated as static (orange), and the information from the ESG constraint (green). 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 (blue), the full investment and long-only constraints combined together (orange), the ESG constraint treated as static (green), and the information from the ESG constraint (red).

Figure B.4b shows the expected utility of the optimal portfolio and its decomposition based on (26) in Proposition 5. The two constraints (orange and green) contribute negatively to expected returns. The expected return contribution from information is strongly negative in 2004, 2005, 2007, 2009, and 2019, and is strongly positive in 2011, 2014, and 2017. Together, the expected return of the constrained portfolio is lower than that of the unconstrained MVO portfolio in most years.

Realized Return Decomposition. Figure B.5 shows the realized returns of the optimal portfolio and *ex post* attribution of returns. Here we again compare a portfolio constructed without a return forecast ($p = 0$) in Figure B.5a with one based on a return forecast ($p = 0.1$) in Figure B.5b.



(a) No Return Predictability

(b) Return Predictability $p = 0.1$

Figure B.5: Realized return and their *ex post* decomposition, for the long-only portfolio defined in (B.7) with a constraint that excludes the bottom 500 assets ordered by their ESG scores. (a) corresponds to a return estimator in (40) with no ability to forecast future expected returns ($p = 0$), and (b) corresponds to a return estimator with some level of predictability ($p = 0.1$). 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 (blue), the full investment and long-only constraints combined together (orange), the ESG constraint treated as static (green), and the information from the ESG constraint (red).

The upper panel of Figure B.5a shows the realized residual returns in excess of the Fama-French five-factor model of the constrained portfolio without any return forecast. This is different from the realized residual returns of the unconstrained MVO portfolio in the lower panel (blue). To understand the difference between these returns, the lower panel further decomposes the realized return of the constrained portfolio based on Proposition 6.

In addition, Figure B.5b shows results parallel to those in Figure B.5a, but with a return forecast ($p = 0.1$) in portfolio construction. As expected, the realized residual returns for both the unconstrained and the constrained portfolio dramatically increase.

In both cases, the information component contributes negatively to realized returns before 2010, and positively in 2011, 2014, 2017, and 2018, consistent with the results of Proposition 6 and the pattern of correlations between asset returns and ESG scores in Figure 6.