Online Appendix for

Performance Attribution for Portfolio Constraints

Andrew W. Lo[1] and Ruixun Zhang[2]

A Proofs of Main Propositions

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:

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

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:

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

The first equation proves (2). Combining the two equations leads to the system of equations that the optimal Lagrange multipliers, $\boldsymbol{\lambda}^*$, should satisfy:

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

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:

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

[1] MIT Sloan School of Management; MIT Laboratory for Financial Engineering; MIT Computer Science and Artificial Intelligence Laboratory; and Santa Fe Institute. alo-admin@mit.edu (email).

[2] Peking University School of Mathematical Sciences; Peking University Center for Statistical Science; Peking University National Engineering Laboratory for Big Data Analysis and Applications; and Peking University Laboratory for Mathematical Economics and Quantitative Finance. zhangruixun@pku.edu.cn (email).

This completes the proof of (3).

To derive the expected return decomposition of (4), multiplying the expected return, $\boldsymbol{\mu}$, by the portfolio holdings of (3) leads directly to:

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

To derive the expected utility decomposition of (5), we have

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

The second term can be equivalently written as $-\frac{\gamma}{2}\left(-\frac{1}{\gamma}\boldsymbol{\Sigma}^{-1}\mathbf{A}'\boldsymbol{\lambda}^*\right)'\boldsymbol{\Sigma}\left(-\frac{1}{\gamma}\boldsymbol{\Sigma}^{-1}\mathbf{A}'\boldsymbol{\lambda}^*\right)$, in which $-\frac{1}{\gamma}\boldsymbol{\Sigma}^{-1}\mathbf{A}'\boldsymbol{\lambda}^*$ 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 (2) with static constraints $\mathbf{A}$ replaced by constraints that depend on characteristics $\mathbf{A}(\mathbf{X})$:

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

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

Therefore, the conditional expected return is given by:

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

which proves (6). To get explicit expressions for each term, we have:

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

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

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

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

Here step (1) follows from the fact that

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

This proves (7). To get explicit expressions for each term, we have:

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

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

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

A.3 Proof of Proposition 3

Normal case. We first prove the normal case under Assumption 2. The conditional distribution, $\mathbf{r}|\mathbf{X}$, is also 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

$$0 = \mathrm{Cov}\left(\mathbf{Z}, \mathbf{X}\right) = \mathrm{Cov}\left(\mathbf{r} - \mathbf{C}\mathbf{X}, \mathbf{X}\right) = \mathrm{Cov}(\mathbf{r}, \mathbf{X}) - \mathbf{C} \cdot \mathrm{Cov}(\mathbf{X}, \mathbf{X}), \tag{A.18}$$

which yields:

$$\mathbf{C} = \mathrm{Cov}(\mathbf{r}, \mathbf{X})\mathrm{Cov}(\mathbf{X}, \mathbf{X})^{-1}. \tag{A.19}$$

Therefore, the conditional expected return can be written as:

$$\begin{aligned}\boldsymbol{\mu}_{\mathbf{X}} = \mathbb{E}[\mathbf{r}|\mathbf{X}] &= \mathbb{E}[\mathbf{Z} + \mathbf{C}\mathbf{X}|\mathbf{X}] = \mathbb{E}[\mathbf{Z}|\mathbf{X}] + \mathbf{C}\mathbf{X} \\ &\overset{(1)}{=} \mathbb{E}[\mathbf{Z}] + \mathbf{C}\mathbf{X} = \mathbb{E}[\mathbf{r}] + \mathbf{C}(\mathbf{X} - \mathbb{E}[\mathbf{X}]) \\ &= \boldsymbol{\mu} + \mathrm{Cov}(\mathbf{r}, \mathbf{X})\mathrm{Cov}(\mathbf{X}, \mathbf{X})^{-1}(\mathbf{X} - \bar{\mathbf{X}}).\end{aligned} \tag{A.20}$$

Similarly, the conditional covariance matrix of $\mathbf{r}|\mathbf{X}$ can be written as:

$$\begin{aligned}\boldsymbol{\Sigma}_{\mathbf{X}} = \mathrm{Cov}(\mathbf{r}|\mathbf{X}) &= \mathrm{Cov}(\mathbf{Z} + \mathbf{C}\mathbf{X}|\mathbf{X}) = \mathrm{Cov}(\mathbf{Z}|\mathbf{X}) \\ &\overset{(1)}{=} \mathrm{Cov}(\mathbf{Z}) = \mathrm{Cov}(\mathbf{r} - \mathbf{C}\mathbf{X}) = \mathrm{Cov}(\mathbf{r}) - \mathbf{C}\mathrm{Cov}(\mathbf{X}, \mathbf{X})\mathbf{C}' \\ &= \boldsymbol{\Sigma} - \mathrm{Cov}(\mathbf{r}, \mathbf{X})\mathrm{Cov}(\mathbf{X}, \mathbf{X})^{-1}\mathrm{Cov}(\mathbf{r}, \mathbf{X})'.\end{aligned} \tag{A.21}$$

Here step (1) in (A.20) and (A.21) follows from the fact that $\mathbf{Z}$ and $\mathbf{X}$ are uncorrelated multivariate Gaussian random vectors and, therefore, are independent. In both cases, they depend on $\mathrm{Cov}(\mathbf{r}, \mathbf{X})$ and $\mathrm{Cov}(\mathbf{X}, \mathbf{X})$.

Under Assumption 3, we can further simplify (A.20):

$$\begin{aligned}\boldsymbol{\mu}_{\mathbf{X}} &= \boldsymbol{\mu} + \begin{pmatrix}\rho_1\sigma_{\mathbf{r}}\sigma_{\mathbf{x}_1}\mathbf{I} & \cdots & \rho_J\sigma_{\mathbf{r}}\sigma_{\mathbf{x}_J}\mathbf{I}\end{pmatrix}\begin{pmatrix}\frac{1}{\sigma_{\mathbf{x}_1}^2}\mathbf{I} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \frac{1}{\sigma_{\mathbf{x}_J}^2}\mathbf{I}\end{pmatrix}\begin{pmatrix}\mathbf{x}_1 - \bar{\mathbf{x}}_1 \\ \vdots \\ \mathbf{x}_J - \bar{\mathbf{x}}_J\end{pmatrix}, \\ &= \boldsymbol{\mu} + \sum_{j=1}^{J}\frac{\rho_j\sigma_{\mathbf{r}}(\mathbf{x}_j - \bar{\mathbf{x}}_j)}{\sigma_{\mathbf{x}_j}},\end{aligned} \tag{A.22}$$

which completes the proof of (12). Similarly, we can further simplify (A.21):

$$\begin{aligned}\boldsymbol{\Sigma}_{\mathbf{X}} &= \boldsymbol{\Sigma} - \begin{pmatrix}\rho_1\sigma_{\mathbf{r}}\sigma_{\mathbf{x}_1}\mathbf{I} & \cdots & \rho_J\sigma_{\mathbf{r}}\sigma_{\mathbf{x}_J}\mathbf{I}\end{pmatrix}\begin{pmatrix}\frac{1}{\sigma^2_{\mathbf{x}_1}}\mathbf{I} & 0 & 0\\ 0 & \ddots & 0\\ 0 & 0 & \frac{1}{\sigma^2_{\mathbf{x}_J}}\mathbf{I}\end{pmatrix}\begin{pmatrix}\rho_1\sigma_{\mathbf{r}}\sigma_{\mathbf{x}_1}\mathbf{I}\\ \vdots\\ \rho_J\sigma_{\mathbf{r}}\sigma_{\mathbf{x}_J}\mathbf{I}\end{pmatrix},\\ &= \boldsymbol{\Sigma} - \sum_{j=1}^{J}\frac{\rho_j^2\sigma_{\mathbf{r}}^2\sigma^2_{\mathbf{x}_j}\mathbf{I}}{\sigma^2_{\mathbf{x}_j}} = \boldsymbol{\Sigma} - \sum_{j=1}^{J}\rho_j^2\sigma_{\mathbf{r}}^2\mathbf{I},\end{aligned} \tag{A.23}$$

which completes the proof of (13).

MVT case. Next, we prove the MVT case under Assumption 2'. The properties of MVT distributions are outlined in Zellner (1971, p. 383–389) and Fang, Kotz, and Ng (1990, p. 42–47).

The conditional distribution of MVT is still MVT. In particular, Ding (2016, p. 294) shows that:

$$\mathbf{r}|\mathbf{X} \sim \text{MVT}\left(\boldsymbol{\mu} + \mathbf{V}_{\mathbf{r},\mathbf{X}}\mathbf{V}^{-1}_{\mathbf{X},\mathbf{X}}(\mathbf{X}-\bar{\mathbf{X}}),\ s(\mathbf{V}-\mathbf{V}_{\mathbf{r},\mathbf{X}}\mathbf{V}^{-1}_{\mathbf{X},\mathbf{X}}\mathbf{V}_{\mathbf{X},\mathbf{r}}),\ \nu + NJ\right), \tag{A.24}$$

where

$$s = \frac{\nu + d_1}{\nu + NJ} \quad \text{and} \quad d_1 = (\mathbf{X}-\bar{\mathbf{X}})'\mathbf{V}^{-1}_{\mathbf{X},\mathbf{X}}(\mathbf{X}-\bar{\mathbf{X}}). \tag{A.25}$$

Because $\text{Cov}(\mathbf{r},\mathbf{X}) = \frac{\nu}{\nu-2}\mathbf{V}_{\mathbf{r},\mathbf{X}}$ and $\text{Cov}(\mathbf{X},\mathbf{X}) = \frac{\nu}{\nu-2}\mathbf{V}_{\mathbf{X},\mathbf{X}}$, the conditional expected value is:

$$\begin{aligned}\boldsymbol{\mu}_{\mathbf{X}} = \mathbb{E}[\mathbf{r}|\mathbf{X}] &= \boldsymbol{\mu} + \mathbf{V}_{\mathbf{r},\mathbf{X}}\mathbf{V}^{-1}_{\mathbf{X},\mathbf{X}}(\mathbf{X}-\bar{\mathbf{X}})\\ &= \boldsymbol{\mu} + \frac{\nu-2}{\nu}\text{Cov}(\mathbf{r},\mathbf{X})\cdot\frac{\nu}{\nu-2}\text{Cov}(\mathbf{X},\mathbf{X})^{-1}(\mathbf{X}-\bar{\mathbf{X}})\\ &= \boldsymbol{\mu} + \text{Cov}(\mathbf{r},\mathbf{X})\text{Cov}(\mathbf{X},\mathbf{X})^{-1}(\mathbf{X}-\bar{\mathbf{X}}),\end{aligned} \tag{A.26}$$

which is the same form as the normal case in (A.20). The proof of (15) therefore follows from the same proof as the normal case in (A.22).

The conditional covariance matrix is:

$$\begin{aligned}\boldsymbol{\Sigma}_{\mathbf{X}} &= \text{Cov}(\mathbf{r}|\mathbf{X})\\ &= \frac{\nu+NJ}{\nu+NJ-2}\frac{\nu+d_1}{\nu+NJ}\left(\frac{\nu-2}{\nu}\boldsymbol{\Sigma} - \frac{\nu-2}{\nu}\text{Cov}(\mathbf{r},\mathbf{X})\frac{\nu}{\nu-2}\text{Cov}(\mathbf{X},\mathbf{X})^{-1}\frac{\nu-2}{\nu}\text{Cov}(\mathbf{X},\mathbf{r})\right)\\ &= \frac{(\nu+d_1)(\nu-2)}{(\nu+NJ-2)\nu}\left(\boldsymbol{\Sigma} - \text{Cov}(\mathbf{r},\mathbf{X})\text{Cov}(\mathbf{X},\mathbf{X})^{-1}\text{Cov}(\mathbf{r},\mathbf{X})'\right),\end{aligned} \tag{A.27}$$

which is the same form as the normal case in (A.21) except for the scaling factor $\frac{(\nu+d_1)(\nu-2)}{(\nu+NJ-2)\nu}$. Under

Assumption 3, $\mathbf{V}_{\mathbf{X},\mathbf{X}}$ is a diagonal matrix and

$$\begin{aligned} d_1 &= (\mathbf{X}-\bar{\mathbf{X}})'\mathbf{V}_{\mathbf{X},\mathbf{X}}^{-1}(\mathbf{X}-\bar{\mathbf{X}}) \\ &= N\left(\sigma_{\mathbf{x}_1}^{-2}\frac{1}{N}\sum_{i=1}^{N}(x_{1i}-\bar{\mathbf{x}}_1)^2+\cdots+\sigma_{\mathbf{x}_J}^{-2}\frac{1}{N}\sum_{i=1}^{N}(x_{Ji}-\bar{\mathbf{x}}_J)^2\right), \end{aligned} \tag{A.28}$$

which converges to NJ as N increases without bound. Therefore,

$$\frac{(\nu+d_1)(\nu-2)}{(\nu+NJ-2)\nu} \overset{p}{=} \frac{(\nu+NJ)(\nu-2)}{(\nu+NJ-2)\nu} = \frac{(1+\nu/NJ)}{1+(\nu-2)/NJ}(1-2/\nu) \overset{p}{=} (1-2/\nu) \tag{A.29}$$

where $\overset{p}{=}$ denotes equality in probability as N increases without bound. Combining (A.27) and (A.29) we have:

$$\boldsymbol{\Sigma}_{\mathbf{X}} = \text{Cov}(\mathbf{r}|\mathbf{X}) \overset{p}{=} (1-2/\nu)\left(\boldsymbol{\Sigma}-\text{Cov}(\mathbf{r},\mathbf{X})\text{Cov}(\mathbf{X},\mathbf{X})^{-1}\text{Cov}(\mathbf{r},\mathbf{X})'\right). \tag{A.30}$$

The proof of (16) therefore follows from the same proof as the normal case in (A.23).

A.4 Proof of Proposition 4

Substituting the excess return from information in Proposition 3 into the expected return decomposition of Proposition 2, we have:

$$\begin{aligned} \mathbb{E}\left[\boldsymbol{\omega}^{*\prime}\mathbf{r}|\mathbf{X}\right] = \boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}^* &= \boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}_{\text{MVO}} + \boldsymbol{\mu}'\boldsymbol{\omega}_{\text{CSTR}} + (\boldsymbol{\mu}_{\mathbf{X}}'-\boldsymbol{\mu}')\boldsymbol{\omega}_{\text{CSTR}} \\ &= \boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}_{\text{MVO}} + \boldsymbol{\mu}'\boldsymbol{\omega}_{\text{CSTR}} + \sum_{j=1}^{J}\rho_j\sigma_{\mathbf{r}}\frac{(\mathbf{x}_j-\bar{\mathbf{x}}_j)\boldsymbol{\omega}_{\text{CSTR}}}{\sigma_{\mathbf{x}_j}}, \end{aligned} \tag{A.31}$$

which completes the proof of (17).

Substituting the excess return and excess covariance from information in Proposition 3 into the expected utility decomposition of Proposition 2, we have:

$$\begin{aligned}
\boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}^* - \frac{\gamma}{2}\boldsymbol{\omega}^{*'}\boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{\omega}^* &= \boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}_{\text{MVO}} - \frac{\gamma}{2}\boldsymbol{\omega}_{\text{MVO}}'\boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{\omega}_{\text{MVO}} \\
&\quad - \frac{\gamma}{2}\boldsymbol{\omega}_{\text{CSTR}}'\boldsymbol{\Sigma}\boldsymbol{\omega}_{\text{CSTR}} \\
&\quad + (\boldsymbol{\mu}_{\mathbf{X}}' - \boldsymbol{\mu}')\boldsymbol{\omega}_{\text{CSTR}} - \gamma\boldsymbol{\omega}_{\text{SHR}}'(\boldsymbol{\Sigma}_{\mathbf{X}} - \boldsymbol{\Sigma})\boldsymbol{\omega}_{\text{CSTR}} \\
&= \boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}_{\text{MVO}} - \frac{\gamma}{2}\boldsymbol{\omega}_{\text{MVO}}'\boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{\omega}_{\text{MVO}} \\
&\quad - \frac{\gamma}{2}\boldsymbol{\omega}_{\text{CSTR}}'\boldsymbol{\Sigma}\boldsymbol{\omega}_{\text{CSTR}} \\
&\quad + \left(\sum_{j=1}^{J}\rho_j\sigma_{\mathbf{r}}\frac{(\mathbf{x}_j - \bar{\mathbf{x}}_j)\boldsymbol{\omega}_{\text{CSTR}}}{\sigma_{\mathbf{x}_j}} - \gamma\boldsymbol{\omega}_{\text{SHR}}'\left(-\sum_{j=1}^{J}\rho_j^2\sigma_{\mathbf{r}}^2\mathbf{I}\right)\boldsymbol{\omega}_{\text{CSTR}}\right) \qquad \text{(A.32)} \\
&= \boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}_{\text{MVO}} - \frac{\gamma}{2}\boldsymbol{\omega}_{\text{MVO}}'\boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{\omega}_{\text{MVO}} \\
&\quad - \frac{\gamma}{2}\boldsymbol{\omega}_{\text{CSTR}}'\boldsymbol{\Sigma}\boldsymbol{\omega}_{\text{CSTR}} \\
&\quad + \sum_{j=1}^{J}\left(\rho_j\sigma_{\mathbf{r}}\frac{(\mathbf{x}_j - \bar{\mathbf{x}}_j)\boldsymbol{\omega}_{\text{CSTR}}}{\sigma_{\mathbf{x}_j}} + \gamma\rho_j^2\sigma_{\mathbf{r}}^2\boldsymbol{\omega}_{\text{SHR}}'\boldsymbol{\omega}_{\text{CSTR}}\right),
\end{aligned}$$

which completes the proof of (18). When Assumption 2' is true, the expected utility decomposition holds when N and ν increase without bound.

A.5 Proof of Proposition 5

We first need the following two lemmas to prove Proposition 5.

Lemma A.1. *When $\boldsymbol{\mu}$ is unknown and $\boldsymbol{\Sigma}$ is known, consider the following prior on $\boldsymbol{\mu}$:*

$$\pi_0 : \boldsymbol{\mu} \sim N\left(\boldsymbol{\mu}_0, \frac{\boldsymbol{\Sigma}}{\tau}\right),$$

where $\boldsymbol{\mu}_0$ is the prior mean, and τ is a precision hyperparameter. The predictive distribution of the return is

$$\tilde{\mathbf{r}} \sim N\left(\tilde{\boldsymbol{\mu}}_0, \tilde{\boldsymbol{\Sigma}}_0\right),$$

where

$$\tilde{\boldsymbol{\mu}}_0 = \frac{T\hat{\boldsymbol{\mu}} + \tau\boldsymbol{\mu}_0}{T + \tau} \quad \textit{and} \quad \tilde{\boldsymbol{\Sigma}}_0 = \left(1 + \frac{1}{T + \tau}\right)\boldsymbol{\Sigma}.$$

Lemma A.2. *When $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are both unknown, consider the following prior on $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$:*

$$\pi_0 : \boldsymbol{\mu} \mid \boldsymbol{\Sigma} \sim N\left(\boldsymbol{\mu}_0, \frac{\boldsymbol{\Sigma}}{\tau}\right), \quad \boldsymbol{\Sigma} \sim IW(\boldsymbol{\Sigma}_0, v_0),$$

where $\boldsymbol{\mu}_0$ is the prior mean, $\boldsymbol{\Sigma}_0$ is the prior covariance matrices, τ and v_0 are hyperparameters, *and IW stands for the Inversed-Wishart distribution. The predictive distribution of the return is*

$$\tilde{\mathbf{r}} \sim MVT\left(\tilde{\boldsymbol{\mu}}_0,\ \frac{v_0+T-N-1}{v_0+T-N+1}\tilde{\boldsymbol{\Sigma}}_0,\ v_0+T-N+1\right),$$

where

$$\tilde{\boldsymbol{\mu}}_0 = \frac{T\hat{\boldsymbol{\mu}}+\tau\boldsymbol{\mu}_0}{T+\tau},$$

$$\tilde{\boldsymbol{\Sigma}}_0 = \left(1+\frac{1}{T+\tau}\right)\frac{1}{v_0+T-N-1}\left(\boldsymbol{\Sigma}_0 + T\hat{\boldsymbol{\Sigma}} + \frac{T\tau}{T+\tau}(\boldsymbol{\mu}_0-\hat{\boldsymbol{\mu}})(\boldsymbol{\mu}_0-\hat{\boldsymbol{\mu}})'\right).$$

Proof of Lemma A.1. Based on the prior π_0, we first calculate the predictive distribution of $\mathbf{r}$.

$$\begin{aligned}
\mathbb{P}(\mathbf{r}|\Phi_T) \propto& \exp\left\{-\frac{1}{2}\left\{tr\left[\boldsymbol{\Sigma}^{-1}(\mathbf{r}-\boldsymbol{\mu})(\mathbf{r}-\boldsymbol{\mu})'\right]\right\}\right\} \\
&\times \exp\left\{-\frac{1}{2}\left\{tr\left[\boldsymbol{\Sigma}^{-1}\tau(\boldsymbol{\mu}-\boldsymbol{\mu}_0)(\boldsymbol{\mu}-\boldsymbol{\mu}_0)'\right]\right\}\right\} \\
&\times \exp\left\{-\frac{1}{2}\left\{tr\left[\boldsymbol{\Sigma}^{-1}\sum_{t=1}^{T}(\mathbf{r}_t-\boldsymbol{\mu})(\mathbf{r}_t-\boldsymbol{\mu})'\right]\right\}\right\} \\
\propto& \exp\left\{-\frac{1}{2}\left\{tr\left[\boldsymbol{\Sigma}^{-1}\left(T(\boldsymbol{\mu}-\hat{\boldsymbol{\mu}})(\boldsymbol{\mu}-\hat{\boldsymbol{\mu}})' + \tau(\boldsymbol{\mu}-\boldsymbol{\mu}_0)(\boldsymbol{\mu}-\boldsymbol{\mu}_0)' + (\boldsymbol{\mu}-\mathbf{r})(\boldsymbol{\mu}-\mathbf{r})'\right)\right]\right\}\right\} \\
\propto& \exp\left\{-\frac{1}{2}\left\{tr\left[\boldsymbol{\Sigma}^{-1}\left((T+\tau+1)\left(\boldsymbol{\mu}-\frac{T\hat{\boldsymbol{\mu}}+\tau\boldsymbol{\mu}_0+\mathbf{r}}{T+\tau+1}\right)\left(\boldsymbol{\mu}-\frac{T\hat{\boldsymbol{\mu}}+\tau\boldsymbol{\mu}_0+\mathbf{r}}{T+\tau+1}\right)'\right)\right.\right.\right. \\
&\left.\left.\left. +\frac{T+\tau}{T+\tau+1}\left(\mathbf{r}-\frac{T\hat{\boldsymbol{\mu}}+\tau\boldsymbol{\mu}_0}{T+\tau}\right)\left(\mathbf{r}-\frac{T\hat{\boldsymbol{\mu}}+\tau\boldsymbol{\mu}_0}{T+\tau}\right)'\right]\right\}\right\},
\end{aligned}$$

where $\hat{\boldsymbol{\mu}} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{r}_t$. Integrating over the posterior distribution of $\boldsymbol{\mu}$, we derive:

$$\mathbb{P}(\mathbf{r}|\Phi_T) \propto \exp\left\{-\frac{1}{2}\left\{tr\left[\boldsymbol{\Sigma}^{-1}\frac{T+\tau}{T+\tau+1}\left(\mathbf{r}-\frac{T\hat{\boldsymbol{\mu}}+\tau\boldsymbol{\mu}_0}{T+\tau}\right)\left(\mathbf{r}-\frac{T\hat{\boldsymbol{\mu}}+\tau\boldsymbol{\mu}_0}{T+\tau}\right)'\right]\right\}\right\}.$$

This implies that the predictive distribution of $\mathbf{r}$ is also normally distributed with mean and covariance given by:

$$\tilde{\boldsymbol{\mu}}_0 = \frac{T\hat{\boldsymbol{\mu}}+\tau\boldsymbol{\mu}_0}{T+\tau}$$

$$\tilde{\boldsymbol{\Sigma}}_0 = \left(1+\frac{1}{T+\tau}\right)\boldsymbol{\Sigma}.$$

Proof of Lemma A.2. Based on the prior π_0, we first calculate the predictive distribution of $\mathbf{r}$.

$$\begin{aligned}\mathbb{P}(\mathbf{r}|\Phi_T) \propto\ & |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left\{tr\left[\boldsymbol{\Sigma}^{-1}(\mathbf{r}-\boldsymbol{\mu})(\mathbf{r}-\boldsymbol{\mu})'\right]\right\}\right\} \\ & \times |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left\{tr\left[\boldsymbol{\Sigma}^{-1}\tau(\boldsymbol{\mu}-\boldsymbol{\mu}_0)(\boldsymbol{\mu}-\boldsymbol{\mu}_0)'\right]\right\}\right\} \\ & \times |\boldsymbol{\Sigma}|^{-\frac{v_0+N+1}{2}} \exp\left\{-\frac{1}{2}\left\{tr\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_0\right]\right\}\right\} \\ & \times |\boldsymbol{\Sigma}|^{-\frac{T}{2}} \exp\left\{-\frac{1}{2}\left\{tr\left[\boldsymbol{\Sigma}^{-1}\sum_{t=1}^{T}(\mathbf{r}_t-\boldsymbol{\mu})(\mathbf{r}_t-\boldsymbol{\mu})'\right]\right\}\right\} \\ \propto\ & |\boldsymbol{\Sigma}|^{-\frac{v_0+N+T+3}{2}} \exp\left\{-\frac{1}{2}\left\{tr\left[\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\Sigma}_0+T\hat{\boldsymbol{\Sigma}}+\frac{T\tau}{T+\tau}(\boldsymbol{\mu}_0-\hat{\boldsymbol{\mu}})(\boldsymbol{\mu}_0-\hat{\boldsymbol{\mu}})'\right.\right.\right.\right. \\ & +(T+\tau+1)\left(\boldsymbol{\mu}-\frac{T\hat{\boldsymbol{\mu}}+\tau\boldsymbol{\mu}_0+\mathbf{r}}{T+\tau+1}\right)\left(\boldsymbol{\mu}-\frac{T\hat{\boldsymbol{\mu}}+\tau\boldsymbol{\mu}_0+\mathbf{r}}{T+\tau+1}\right)'\Bigg) \\ & \left.\left.\left.+\frac{T+\tau}{T+\tau+1}\left(\mathbf{r}-\frac{T\hat{\boldsymbol{\mu}}+\tau\boldsymbol{\mu}_0}{T+\tau}\right)\left(\mathbf{r}-\frac{T\hat{\boldsymbol{\mu}}+\tau\boldsymbol{\mu}_0}{T+\tau}\right)'\right]\right\}\right\}.\end{aligned}$$

where $\hat{\boldsymbol{\Sigma}} = \frac{1}{T}\sum_{t=1}^{T}(\mathbf{r}_t-\hat{\boldsymbol{\mu}})(\mathbf{r}_t-\hat{\boldsymbol{\mu}})'$. Integrating over the posterior distribution of $\boldsymbol{\mu}$, we derive:

$$\begin{aligned}\mathbb{P}(\mathbf{r}|\Phi_T) \propto\ & |\boldsymbol{\Sigma}|^{-\frac{v_0+N+T+2}{2}} \exp\left\{-\frac{1}{2}\left\{tr\left[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Sigma}_0+T\hat{\boldsymbol{\Sigma}}+\frac{T\tau}{T+\tau}(\boldsymbol{\mu}_0-\hat{\boldsymbol{\mu}})(\boldsymbol{\mu}_0-\hat{\boldsymbol{\mu}})'\right.\right.\right. \\ & \left.\left.\left.+\frac{T+\tau}{T+\tau+1}\left(\mathbf{r}-\frac{T\hat{\boldsymbol{\mu}}+\tau\boldsymbol{\mu}_0}{T+\tau}\right)\left(\mathbf{r}-\frac{T\hat{\boldsymbol{\mu}}+\tau\boldsymbol{\mu}_0}{T+\tau}\right)'\right]\right\}\right\}.\end{aligned}$$

The posterior distribution of $\boldsymbol{\Sigma}$ is Inversed-Wishart with a degree of freedom of v_0+T+1. Integrating over the posterior distribution of $\boldsymbol{\Sigma}$, we derive:

$$\begin{aligned}\mathbb{P}(\mathbf{r}|\Phi_T) \propto\ & \left|\boldsymbol{\Sigma}_0+T\hat{\boldsymbol{\Sigma}}+\frac{T\tau}{T+\tau}(\boldsymbol{\mu}_0-\hat{\boldsymbol{\mu}})(\boldsymbol{\mu}_0-\hat{\boldsymbol{\mu}})'\right. \\ & \left.+\frac{T+\tau}{T+\tau+1}\left(\mathbf{r}-\frac{T\hat{\boldsymbol{\mu}}+\tau\boldsymbol{\mu}_0}{T+\tau}\right)\left(\mathbf{r}-\frac{T\hat{\boldsymbol{\mu}}+\tau\boldsymbol{\mu}_0}{T+\tau}\right)'\right|^{-\frac{v_0+T+1}{2}}.\end{aligned}$$

This implies that the predictive distribution of $\mathbf{r}$ follows a MVT distribution with a degree of freedom of $v_0+T-N+1$, and its mean and covariance are given by:

$$\tilde{\boldsymbol{\mu}}_0 = \frac{T\hat{\boldsymbol{\mu}}+\tau\boldsymbol{\mu}_0}{T+\tau},$$
$$\tilde{\boldsymbol{\Sigma}}_0 = \left(1+\frac{1}{T+\tau}\right)\frac{1}{v_0+T-N-1}\left(\boldsymbol{\Sigma}_0+T\hat{\boldsymbol{\Sigma}}+\frac{T\tau}{T+\tau}(\boldsymbol{\mu}_0-\hat{\boldsymbol{\mu}})(\boldsymbol{\mu}_0-\hat{\boldsymbol{\mu}})'\right).$$

Proof of Proposition 5 with known $\boldsymbol{\Sigma}$. With Lemma A.1, we derive the predictive mean and covariance of $\tilde{\mathbf{r}}$ with the prior π_1:

$$\tilde{\boldsymbol{\mu}}_1 = \frac{T\hat{\boldsymbol{\mu}}+\tau\boldsymbol{\mu}_1}{T+\tau} \quad \text{and} \quad \tilde{\boldsymbol{\Sigma}}_1 = (1+\frac{1}{T+\tau})\boldsymbol{\Sigma}, \tag{A.33}$$

and the predictive mean and covariance of $\tilde{\mathbf{r}}$ with the prior π_2:

$$\tilde{\boldsymbol{\mu}}_2 = \frac{T\hat{\boldsymbol{\mu}} + \tau\boldsymbol{\mu}_2}{T+\tau} \quad \text{and} \quad \tilde{\boldsymbol{\Sigma}}_2 = (1 + \frac{1}{T+\tau})\boldsymbol{\Sigma}. \tag{A.34}$$

In order to have $\tilde{\boldsymbol{\omega}}_1^* = \tilde{\boldsymbol{\omega}}_2^*$, we combine (A.33)–(A.34) with (21)–(22) to derive

$$\tilde{\boldsymbol{\mu}}_1 - \mathbf{A}'\tilde{\boldsymbol{\lambda}} = \tilde{\boldsymbol{\mu}}_2.$$

Rearranging terms, we have

$$\begin{aligned}
\boldsymbol{\mu}_2 &= \boldsymbol{\mu}_1 - \mathbf{A}'\left(\mathbf{A}\boldsymbol{\Sigma}^{-1}\mathbf{A}'\right)^{-1}\left(\mathbf{A}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1 - \left(1+\frac{1}{\tau}\right)\gamma\mathbf{b}\right) - \frac{T}{\tau}\mathbf{A}'\left(\mathbf{A}\boldsymbol{\Sigma}^{-1}\mathbf{A}'\right)^{-1}\left(\mathbf{A}\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\mu}} - \gamma\mathbf{b}\right) \\
&= \boldsymbol{\mu}_1 - \mathbf{A}'\tilde{\boldsymbol{\lambda}}_{1,PR} - \frac{T}{\tau}\mathbf{A}'\tilde{\boldsymbol{\lambda}}_{1,DT},
\end{aligned}$$

which proves (23).

Proof of Proposition 5 with unknown $\boldsymbol{\Sigma}$. With Lemma A.2, we derive the predictive mean and covariance of $\tilde{\mathbf{r}}$ with the prior π_1:

$$\begin{aligned}
\tilde{\boldsymbol{\mu}}_1 &= \frac{T\hat{\boldsymbol{\mu}} + \tau\boldsymbol{\mu}_1}{T+\tau}, \\
\tilde{\boldsymbol{\Sigma}}_1 &= \left(1+\frac{1}{T+\tau}\right)\frac{1}{v_0+T-N-1}\left(\boldsymbol{\Sigma}_1 + T\hat{\boldsymbol{\Sigma}} + \frac{T\tau}{T+\tau}(\boldsymbol{\mu}_1-\hat{\boldsymbol{\mu}})(\boldsymbol{\mu}_1-\hat{\boldsymbol{\mu}})'\right),
\end{aligned} \tag{A.35}$$

and the predictive mean and covariance of $\tilde{\mathbf{r}}$ with the prior π_2:

$$\begin{aligned}
\tilde{\boldsymbol{\mu}}_2 &= \frac{T\hat{\boldsymbol{\mu}} + \tau\boldsymbol{\mu}_2}{T+\tau}, \\
\tilde{\boldsymbol{\Sigma}}_2 &= \left(1+\frac{1}{T+\tau}\right)\frac{1}{v_0+T-N-1}\left(\boldsymbol{\Sigma}_2 + T\hat{\boldsymbol{\Sigma}} + \frac{T\tau}{T+\tau}(\boldsymbol{\mu}_2-\hat{\boldsymbol{\mu}})(\boldsymbol{\mu}_2-\hat{\boldsymbol{\mu}})'\right).
\end{aligned} \tag{A.36}$$

We combine (A.35)–(A.36) with (21)–(22) to derive the following sufficient condition for $\tilde{\boldsymbol{\omega}}_1^* = \tilde{\boldsymbol{\omega}}_2^*$:

$$\tilde{\boldsymbol{\Sigma}}_1 = \tilde{\boldsymbol{\Sigma}}_2 \quad \text{and} \quad \tilde{\boldsymbol{\mu}}_1 - \mathbf{A}'\tilde{\boldsymbol{\lambda}} = \tilde{\boldsymbol{\mu}}_2.$$

In order for $\tilde{\boldsymbol{\Sigma}}_1 = \tilde{\boldsymbol{\Sigma}}_2$, we simply need

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

which proves (25). In order for $\tilde{\boldsymbol{\mu}}_1 - \mathbf{A}'\boldsymbol{\lambda}_1 = \tilde{\boldsymbol{\mu}}_2$, we plug in the predictive moments in (A.35)–(A.36) to derive

$$\begin{aligned}\boldsymbol{\mu}_2 &= \boldsymbol{\mu}_1 - \mathbf{A}'\left(\mathbf{A}\tilde{\boldsymbol{\Sigma}}_1^{-1}\mathbf{A}'\right)^{-1}\left(\mathbf{A}\tilde{\boldsymbol{\Sigma}}_1^{-1}\boldsymbol{\mu}_1 - \gamma\mathbf{b}\right)\\ &\quad - \frac{T}{\tau}\mathbf{A}'\left(\mathbf{A}\tilde{\boldsymbol{\Sigma}}_1^{-1}\mathbf{A}'\right)^{-1}\left(\mathbf{A}\tilde{\boldsymbol{\Sigma}}_1^{-1}\hat{\boldsymbol{\mu}} - \gamma\mathbf{b}\right)\\ &= \boldsymbol{\mu}_1 - \mathbf{A}'\tilde{\boldsymbol{\lambda}}_{2,PR} - \frac{T}{\tau}\mathbf{A}'\tilde{\boldsymbol{\lambda}}_{2,DT},\end{aligned}$$

which proves (24).

A.6 Proof of Proposition 6

When there is a single constraint $\mathbf{A}(\mathbf{x}) = \mathbf{x}'$, the Lagrange multiplier, λ^*, is a scalar:

$$\lambda^* = \frac{\mathbf{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \gamma b}{\mathbf{x}'\boldsymbol{\Sigma}^{-1}\mathbf{x}}. \tag{A.37}$$

Substituting this Lagrange multiplier into the decomposition of portfolio holdings in (3) yields:

$$\boldsymbol{\omega}^* = \frac{1}{\gamma}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{1}{\gamma}\lambda^*\boldsymbol{\Sigma}^{-1}\mathbf{x} = \frac{1}{\gamma}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{1}{\gamma}\frac{\mathbf{x}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \gamma b}{\mathbf{x}'\boldsymbol{\Sigma}^{-1}\mathbf{x}}\boldsymbol{\Sigma}^{-1}\mathbf{x}. \tag{A.38}$$

Substituting (A.38) into the last term (the information component) of the expected return decomposition in (17) and assuming the expected value $\bar{\mathbf{x}} = \mathbb{E}[\mathbf{x}] = 0$ yields:

$$\begin{aligned}\frac{\rho\sigma_{\mathbf{r}}}{\sigma_{\mathbf{x}}}\mathbf{x}'\boldsymbol{\omega}_{\text{CSTR}} &= \frac{\rho\sigma_{\mathbf{r}}}{\sigma_{\mathbf{x}}}\mathbf{x}'\left(-\frac{1}{\gamma}\lambda^*\boldsymbol{\Sigma}^{-1}\mathbf{x}\right)\\ &= \frac{\rho\sigma_{\mathbf{r}}}{\sigma_{\mathbf{x}}}\mathbf{x}'\left(-\frac{1}{\gamma}\frac{\mathbf{x}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \gamma b}{\mathbf{x}'\boldsymbol{\Sigma}^{-1}\mathbf{x}}\boldsymbol{\Sigma}^{-1}\mathbf{x}\right)\\ &= \frac{\rho\sigma_{\mathbf{r}}}{\sigma_{\mathbf{x}}}\left(-\frac{1}{\gamma}\frac{\mathbf{x}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \gamma b}{\mathbf{x}'\boldsymbol{\Sigma}^{-1}\mathbf{x}}\mathbf{x}'\boldsymbol{\Sigma}^{-1}\mathbf{x}\right)\\ &= \frac{\rho\sigma_{\mathbf{r}}}{\sigma_{\mathbf{x}}}\left(b - \mathbf{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}/\gamma\right).\end{aligned} \tag{A.39}$$

Therefore, the full decomposition of expected return in (17) reduces to:

$$\begin{aligned}\mathbb{E}\left[\boldsymbol{\omega}^{*\prime}\mathbf{r}|\mathbf{x}\right] &= \boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}_{\text{MVO}} + \boldsymbol{\mu}'\boldsymbol{\omega}_{\text{CSTR}} + \frac{\rho\sigma_{\mathbf{r}}}{\sigma_{\mathbf{x}}}\mathbf{x}'\boldsymbol{\omega}_{\text{CSTR}}\\ &= \frac{1}{\gamma}\boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{1}{\gamma}\lambda^*\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{x} + \frac{\rho\sigma_{\mathbf{r}}}{\sigma_{\mathbf{x}}}\left(b - \mathbf{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}/\gamma\right)\\ &= \frac{1}{\gamma}\boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \frac{1}{\gamma}\frac{\mathbf{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\mathbf{x}'\boldsymbol{\Sigma}^{-1}\mathbf{x}}\left(\gamma b - \mathbf{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right) + \frac{\rho\sigma_{\mathbf{r}}}{\sigma_{\mathbf{x}}}\left(b - \mathbf{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}/\gamma\right)\\ &= \frac{1}{\gamma}\boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \frac{\mathbf{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\mathbf{x}'\boldsymbol{\Sigma}^{-1}\mathbf{x}}\left(b - \mathbf{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}/\gamma\right) + \frac{\rho\sigma_{\mathbf{r}}}{\sigma_{\mathbf{x}}}\left(b - \mathbf{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}/\gamma\right).\end{aligned} \tag{A.40}$$

Note that both (A.39) and (A.40) assume that the Lagrange multiplier $\lambda^* \neq 0$. When the constraint is not binding for inequality constraints, (A.39) becomes zero and the last two terms of (A.40) vanish.

A.7 Proof of Proposition 7

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 (6) of Proposition 2. The i-th element of $\boldsymbol{\mu}_{\mathbf{x}} - \boldsymbol{\mu}$ is given by:

$$\begin{aligned} \mu_{i,\mathbf{x}} - \mu_i &= \mathbb{E}[r_i|\mathbf{x}] - \mathbb{E}[r_i] \\ &\overset{(1)}{=} \mathbb{E}[r_i|x_i] - \mathbb{E}[r_i] \\ &\overset{(2)}{=} x_i \left(\mathbb{E}[r_i|x_i = 1] - \mathbb{E}[r_i]\right) + (1 - x_i) \left(\mathbb{E}[r_i|x_i = 0] - \mathbb{E}[r_i]\right), \end{aligned} \tag{A.41}$$

for $i = 1, 2, \ldots, N$. Here step (1) follows from Assumption 3 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:

$$\begin{aligned} \rho \equiv \mathrm{Corr}(x_i, r_i) &= \frac{\mathrm{Cov}(x_i, r_i)}{\sqrt{\mathrm{Var}(x_i)\mathrm{Var}(r_i)}} \\ &= \frac{\mathbb{E}[x_i r_i] - \mathbb{E}[x_i]\mathbb{E}[r_i]}{\sqrt{\mathrm{Var}(x_i)\mathrm{Var}(r_i)}} \\ &\overset{(1)}{=} \frac{\mathbb{E}[r_i|x_i = 1]\mathbb{P}(x_i = 1) - \mathbb{P}(x_i = 1)\mathbb{E}[r_i]}{\sqrt{\mathbb{P}(x_i = 1)(1 - \mathbb{P}(x_i = 1))\mathrm{Var}(r_i)}} \\ &= \frac{\left(\mathbb{E}[r_i|x_i = 1] - \mathbb{E}[r_i]\right)\psi_{x_i=1}}{\sqrt{\psi_{x_i=1}(1 - \psi_{x_i=1})\sigma_{\mathbf{r}}^2}} \\ &= \frac{\mathbb{E}[r_i|x_i = 1] - \mathbb{E}[r_i]}{\sigma_{\mathbf{r}}}\sqrt{\frac{\psi_{x_i=1}}{\psi_{x_i=0}}}. \end{aligned} \tag{A.42}$$

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

Equation (A.42) implies that

$$\mathbb{E}[r_i|x_i = 1] - \mathbb{E}[r_i] = \rho\sigma_{\mathbf{r}}\sqrt{\frac{\psi_{x_i=0}}{\psi_{x_i=1}}}. \tag{A.43}$$

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

$$\mathbb{E}[r_i] = \mathbb{E}[r_i|x_i = 1]\psi_{x_i=1} + \mathbb{E}[r_i|x_i = 0]\psi_{x_i=0}, \tag{A.44}$$

which yields:

$$\mathbb{E}[r_i|x_i=0]-\mathbb{E}[r_i]=-\left(\mathbb{E}[r_i|x_i=1]-\mathbb{E}[r_i]\right)\frac{\psi_{x_i=1}}{\psi_{x_i=0}}=-\rho\sigma_{\mathbf{r}}\sqrt{\frac{\psi_{x_i=1}}{\psi_{x_i=0}}}, \tag{A.45}$$

Substituting (A.43) and (A.45) into (A.41) yields:

$$\mu_{i,\mathbf{x}}-\mu_i=\rho\sigma_{\mathbf{r}}\left(x_i\sqrt{\frac{\psi_{x_i=0}}{\psi_{x_i=1}}}-(1-x_i)\sqrt{\frac{\psi_{x_i=1}}{\psi_{x_i=0}}}\right) \tag{A.46}$$

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

$$\boldsymbol{\mu}_{\mathbf{x}}-\boldsymbol{\mu}=\rho\sigma_{\mathbf{r}}\left(\mathbf{x}\odot\mathbf{u}-(1-\mathbf{x})\odot\mathbf{v}\right), \tag{A.47}$$

where

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

Substituting this into (6) of Proposition 2 yields:

$$\begin{aligned}\mathbb{E}\left[\boldsymbol{\omega}^{*\prime}\mathbf{r}|\mathbf{x}\right]&=\boldsymbol{\mu}_{\mathbf{x}}'\boldsymbol{\omega}_{\mathrm{MVO}}+\boldsymbol{\mu}'\boldsymbol{\omega}_{\mathrm{CSTR}}+(\boldsymbol{\mu}_{\mathbf{x}}'-\boldsymbol{\mu}')\boldsymbol{\omega}_{\mathrm{CSTR}}\\&=\boldsymbol{\mu}_{\mathbf{x}}'\boldsymbol{\omega}_{\mathrm{MVO}}+\boldsymbol{\mu}'\boldsymbol{\omega}_{\mathrm{CSTR}}+\rho\sigma_{\mathbf{r}}\left(\mathbf{x}\odot\mathbf{u}-(1-\mathbf{x})\odot\mathbf{v}\right)'\boldsymbol{\omega}_{\mathrm{CSTR}},\end{aligned} \tag{A.48}$$

which completes the proof of (29).

B Additional Technical Results

In this Appendix, we provide additional technical results.

B.1 General Dependence between Returns and Characteristics

In this section, we relax Assumption 3 and the results in Section 2.3. In particular, we derive a version of Propositions 3–4 under Assumption 2 (normality). We allow for multiple characteristics of the same asset to be correlated with each other, and general dependence between returns and characteristics. In particular, we replace the covariance matrix of $[\,\mathbf{r}' \; \mathbf{x}_1' \; \cdots \; \mathbf{x}_J'\,]$ in (11) by the following:

Assumption B.1. *The covariance matrix of* $[\,\mathbf{r}' \; \mathbf{x}_1' \; \cdots \; \mathbf{x}_J'\,]$ *is given by:*

$$\begin{pmatrix} \boldsymbol{\Sigma} & \mathrm{Cov}\left(\mathbf{r},(\mathbf{x}_1',\ldots,\mathbf{x}_J')\right) \\ \mathrm{Cov}\left((\mathbf{x}_1',\ldots,\mathbf{x}_J'),\mathbf{r}\right) & \begin{pmatrix} \sigma_{\mathbf{x}_1}^2\mathbf{I} & \sigma_{\mathbf{x}_1}\sigma_{\mathbf{x}_2}\tau_{12}\mathbf{I} & \cdots & \sigma_{\mathbf{x}_1}\sigma_{\mathbf{x}_J}\tau_{1J}\mathbf{I} \\ \sigma_{\mathbf{x}_2}\sigma_{\mathbf{x}_1}\tau_{21}\mathbf{I} & \sigma_{\mathbf{x}_2}^2\mathbf{I} & \cdots & \sigma_{\mathbf{x}_2}\sigma_{\mathbf{x}_J}\tau_{2J}\mathbf{I} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{\mathbf{x}_J}\sigma_{\mathbf{x}_1}\tau_{J1}\mathbf{I} & \sigma_{\mathbf{x}_J}\sigma_{\mathbf{x}_2}\tau_{J,2}\mathbf{I} & \cdots & \sigma_{\mathbf{x}_J}^2\mathbf{I} \end{pmatrix} \end{pmatrix} \tag{B.1}$$

where $\tau_{ij} \equiv \mathrm{Corr}(r_{ik}, x_{jk})$ *is the correlation between the* i*-th and* j*-th characteristic values of the* k*-th asset, for* $i, j \in \{1, 2, \ldots, J\}$ *and* $k \in \{1, 2, \ldots, N\}$*.*

Recall that $\mathbf{X}$ represents the $(N \times J)$-dimensional vector $[\,\mathbf{x}_1' \; \cdots \; \mathbf{x}_J'\,]'$, and we use $\bar{\mathbf{X}} \equiv [\,\bar{\mathbf{x}}_1' \; \cdots \; \bar{\mathbf{x}}_J'\,]'$ to denote the expected value of $\mathbf{X}$. The following result generalizes Proposition 3 under the weaker Assumption B.1.

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

$$\boldsymbol{\mu}_{\mathbf{X}} = \mathbb{E}[\mathbf{r}|\mathbf{X}] = \boldsymbol{\mu} + \sum_{j=1}^{J} \frac{\mathrm{Cov}(\mathbf{r},\mathbf{x}_j)(\mathbf{x}_j - \bar{\mathbf{x}}_j)}{\sigma_{\mathbf{x}_j}^2(1-R_j^2)} - \sum_{\substack{i \neq j \\ 1 \le i,j \le J}} \frac{\beta_{ij}\mathrm{Cov}(\mathbf{r},\mathbf{x}_i)(\mathbf{x}_j - \bar{\mathbf{x}}_j)}{\sigma_{\mathbf{x}_i}^2(1-R_i^2)}, \tag{B.2}$$

and a covariance matrix given by:

$$\boldsymbol{\Sigma}_{\mathbf{X}} = \mathrm{Cov}(\mathbf{r}|\mathbf{X}) = \boldsymbol{\Sigma} - \sum_{j=1}^{J} \frac{\mathrm{Cov}(\mathbf{r},\mathbf{x}_j)\mathrm{Cov}(\mathbf{r},\mathbf{x}_j)'}{\sigma_{\mathbf{x}_j}^2(1-R_j^2)} + \sum_{\substack{i \neq j \\ 1 \le i,j \le J}} \frac{\beta_{ij}\mathrm{Cov}(\mathbf{r},\mathbf{x}_i)\mathrm{Cov}(\mathbf{r},\mathbf{x}_j)'}{\sigma_{\mathbf{x}_i}^2(1-R_i^2)}, \tag{B.3}$$

where $\mathrm{Cov}(\mathbf{r}, \mathbf{x}_j)$ *is the* $N \times N$ *covariance matrix between two* N*-dimensional vectors,* $\mathbf{r}$ *and* $\mathbf{x}_j$*, and* β_{ij} *and* R_i^2 *are defined as follows. Let*

$$\mathbf{x}_{[-i]} \equiv [\, x_1 \; \cdots \; x_{i-1} \; x_{i+1} \; \cdots \; x_J \,]' \tag{B.4}$$

denote the $(J-1)$-dimensional vector of an asset's $J-1$ characteristics not including the i-th one. We consider a hypothetical regression in which x_i is projected onto the $(J-1)$-dimensional space of $\mathbf{x}_{[-i]}$. Then,

$$\boldsymbol{\beta}_i = \left[\mathrm{Cov}\left(\mathbf{x}_{[-i]}, \mathbf{x}_{[-i]}\right)\right]^{-1} \mathrm{Cov}\left(\mathbf{x}_{[-i]}, x_i\right) \tag{B.5}$$

is the $(J-1)$-dimensional vector of regression coefficients and β_{ij} is the element of $\boldsymbol{\beta}_i$ that corresponds to x_j; and

$$R_i^2 = \frac{\mathrm{Cov}\left(\boldsymbol{\beta}_i'\mathbf{x}_{[-i]}, \boldsymbol{\beta}_i'\mathbf{x}_{[-i]}\right)}{\sigma^2_{\mathbf{x}_i}} = \frac{\mathrm{Cov}\left(\mathbf{x}_{[-i]}, x_i\right)' \left[\mathrm{Cov}\left(\mathbf{x}_{[-i]}, \mathbf{x}_{[-i]}\right)\right]^{-1} \mathrm{Cov}\left(\mathbf{x}_{[-i]}, x_i\right)}{\sigma^2_{\mathbf{x}_i}} \tag{B.6}$$

is the R-squared of the regression. Here

$$\mathrm{Cov}\left(\mathbf{x}_{[-i]}, x_i\right) = \begin{pmatrix} \sigma_{\mathbf{x}_1}\sigma_{\mathbf{x}_i}\tau_{1i} \\ \vdots \\ \sigma_{\mathbf{x}_{i-1}}\sigma_{\mathbf{x}_i}\tau_{i-1,i} \\ \sigma_{\mathbf{x}_{i+1}}\sigma_{\mathbf{x}_i}\tau_{i+1,i} \\ \vdots \\ \sigma_{\mathbf{x}_J}\sigma_{\mathbf{x}_i}\tau_{J,i} \end{pmatrix}$$

and

$$\mathrm{Cov}\left(\mathbf{x}_{[-i]}, \mathbf{x}_{[-i]}\right) = \begin{pmatrix} \sigma^2_{\mathbf{x}_1} & \cdots & \sigma_{\mathbf{x}_1}\sigma_{\mathbf{x}_{i-1}}\tau_{1,i-1} & \sigma_{\mathbf{x}_1}\sigma_{\mathbf{x}_{i+1}}\tau_{1,i+1} & \cdots & \sigma_{\mathbf{x}_1}\sigma_{\mathbf{x}_J}\tau_{1J} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{\mathbf{x}_{i-1}}\sigma_{\mathbf{x}_1}\tau_{i-1,1} & \cdots & \sigma^2_{\mathbf{x}_{i-1}} & \sigma_{\mathbf{x}_{i-1}}\sigma_{\mathbf{x}_{i+1}}\tau_{i-1,i+1} & \cdots & \sigma_{\mathbf{x}_{i-1}}\sigma_{\mathbf{x}_J}\tau_{i-1,J} \\ \sigma_{\mathbf{x}_{i+1}}\sigma_{\mathbf{x}_1}\tau_{i+1,1} & \cdots & \sigma_{\mathbf{x}_{i+1}}\sigma_{\mathbf{x}_{i-1}}\tau_{i+1,i-1} & \sigma^2_{\mathbf{x}_{i+1}} & \cdots & \sigma_{\mathbf{x}_{i+1}}\sigma_{\mathbf{x}_J}\tau_{i+1,J} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{\mathbf{x}_J}\sigma_{\mathbf{x}_1}\tau_{J1} & \cdots & \sigma_{\mathbf{x}_J}\sigma_{\mathbf{x}_{i-1}}\tau_{J,i-1} & \sigma_{\mathbf{x}_J}\sigma_{\mathbf{x}_{i+1}}\tau_{J,i+1} & \cdots & \sigma^2_{\mathbf{x}_J} \end{pmatrix}.$$

Proposition B.1 also allows for 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 4 under the weaker Assumption B.1.

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

$$\begin{aligned} \mathbb{E}\left[\boldsymbol{\omega}^{*\prime}\mathbf{r}|\mathbf{X}\right] &= \boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}^* = \boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}_{\mathrm{MVO}} + \boldsymbol{\mu}'\boldsymbol{\omega}_{\mathrm{CSTR}} \\ &\quad + \sum_{j=1}^{J} \boldsymbol{\omega}_{\mathrm{CSTR}}' \frac{\mathrm{Cov}(\mathbf{r}, \mathbf{x}_j)(\mathbf{x}_j - \bar{\mathbf{x}}_j)}{\sigma^2_{\mathbf{x}_j}(1-R_j^2)} - \sum_{\substack{i \neq j \\ 1 \le i,j \le J}} \boldsymbol{\omega}_{\mathrm{CSTR}}' \frac{\beta_{ij}\mathrm{Cov}(\mathbf{r}, \mathbf{x}_i)(\mathbf{x}_j - \bar{\mathbf{x}}_j)}{\sigma^2_{\mathbf{x}_i}(1-R_i^2)}. \end{aligned} \tag{B.7}$$

2. Expected utility decomposition.

$$
\begin{aligned}
\boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}^* - \frac{\gamma}{2}\boldsymbol{\omega}^{*\prime}\boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{\omega}^* &= \boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}_{\text{MVO}} - \frac{\gamma}{2}\boldsymbol{\omega}_{\text{MVO}}'\boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{\omega}_{\text{MVO}} - \frac{\gamma}{2}\boldsymbol{\omega}_{\text{CSTR}}'\boldsymbol{\Sigma}\boldsymbol{\omega}_{\text{CSTR}} \\
&+ \sum_{j=1}^{J}\left(\boldsymbol{\omega}_{\text{CSTR}}'\frac{\text{Cov}(\mathbf{r},\mathbf{x}_j)(\mathbf{x}_j-\bar{\mathbf{x}}_j)}{\sigma_{\mathbf{x}_j}^2(1-R_j^2)} + \gamma\boldsymbol{\omega}_{\text{SHR}}'\frac{\text{Cov}(\mathbf{r},\mathbf{x}_j)\text{Cov}(\mathbf{r},\mathbf{x}_j)'}{\sigma_{\mathbf{x}_j}^2(1-R_j^2)}\boldsymbol{\omega}_{\text{CSTR}}\right) \\
&- \sum_{\substack{i\neq j \\ 1\leq i,j\leq J}}\left(\boldsymbol{\omega}_{\text{CSTR}}'\frac{\beta_{ij}\text{Cov}(\mathbf{r},\mathbf{x}_i)(\mathbf{x}_j-\bar{\mathbf{x}}_j)}{\sigma_{\mathbf{x}_i}^2(1-R_i^2)} + \gamma\boldsymbol{\omega}_{\text{SHR}}'\frac{\beta_{ij}\text{Cov}(\mathbf{r},\mathbf{x}_i)\text{Cov}(\mathbf{r},\mathbf{x}_j)'}{\sigma_{\mathbf{x}_i}^2(1-R_i^2)}\boldsymbol{\omega}_{\text{CSTR}}\right).
\end{aligned}
\tag{B.8}
$$

Next, we state two corollaries of Proposition B.1. In particular, we provide additional intuition for the information contribution from portfolio constraints by separating the effect of general dependence between returns and characteristics from the effect of dependent characteristics of the same asset.

First, if we allow for general dependence between returns and characteristics, but multiple characteristics of the same asset are independent of each other, the covariance matrix of $[\,\mathbf{r}'\ \mathbf{x}_1'\ \dots\ \mathbf{x}_J'\,]$ can be written as:

$$
\begin{pmatrix}
\boldsymbol{\Sigma} & \text{Cov}\left(\mathbf{r},(\mathbf{x}_1',\dots,\mathbf{x}_J')\right) \\
\text{Cov}\left((\mathbf{x}_1',\dots,\mathbf{x}_J'),\mathbf{r}\right) & \begin{pmatrix} \sigma_{\mathbf{x}_1}^2\mathbf{I} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_{\mathbf{x}_J}^2\mathbf{I} \end{pmatrix}
\end{pmatrix}.
\tag{B.9}
$$

Corollary B.1. *Under the covariance structure in (B.9), $\mathbf{r}|\mathbf{X}$ is normally distributed with an expected value given by:*

$$
\boldsymbol{\mu}_{\mathbf{X}} = \mathbb{E}[\mathbf{r}|\mathbf{X}] = \boldsymbol{\mu} + \sum_{j=1}^{J}\frac{\text{Cov}(\mathbf{r},\mathbf{x}_j)}{\sigma_{\mathbf{x}_j}}\frac{(\mathbf{x}_j-\bar{\mathbf{x}}_j)}{\sigma_{\mathbf{x}_j}},
\tag{B.10}
$$

and a covariance matrix given by:

$$
\boldsymbol{\Sigma}_{\mathbf{X}} = \text{Cov}(\mathbf{r}|\mathbf{X}) = \boldsymbol{\Sigma} - \sum_{j=1}^{J}\frac{\text{Cov}(\mathbf{r},\mathbf{x}_j)\text{Cov}(\mathbf{r},\mathbf{x}_j)'}{\sigma_{\mathbf{x}_j}^2},
\tag{B.11}
$$

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

Second, if we allow for multiple characteristics of the same asset to be correlated with each other, but characteristics of one firm are independent of returns of another firm, the covariance matrix of $[\,\mathbf{r}'\ \mathbf{x}_1'\ \ldots\ \mathbf{x}_J'\,]$ can be written as:

$$\begin{pmatrix} \boldsymbol{\Sigma} & \rho_1\sigma_{\mathbf{r}}\sigma_{\mathbf{x}_1}\mathbf{I} \quad \rho_2\sigma_{\mathbf{r}}\sigma_{\mathbf{x}_2}\mathbf{I} \quad \cdots \quad \rho_J\sigma_{\mathbf{r}}\sigma_{\mathbf{x}_J}\mathbf{I} \\ \begin{matrix} \rho_1\sigma_{\mathbf{r}}\sigma_{\mathbf{x}_1}\mathbf{I} \\ \rho_2\sigma_{\mathbf{r}}\sigma_{\mathbf{x}_2}\mathbf{I} \\ \vdots \\ \rho_J\sigma_{\mathbf{r}}\sigma_{\mathbf{x}_J}\mathbf{I} \end{matrix} & \begin{pmatrix} \sigma^2_{\mathbf{x}_1}\mathbf{I} & \sigma_{\mathbf{x}_1}\sigma_{\mathbf{x}_2}\tau_{12}\mathbf{I} & \cdots & \sigma_{\mathbf{x}_1}\sigma_{\mathbf{x}_J}\tau_{1J}\mathbf{I} \\ \sigma_{\mathbf{x}_2}\sigma_{\mathbf{x}_1}\tau_{21}\mathbf{I} & \sigma^2_{\mathbf{x}_2}\mathbf{I} & \cdots & \sigma_{\mathbf{x}_2}\sigma_{\mathbf{x}_J}\tau_{2J}\mathbf{I} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{\mathbf{x}_J}\sigma_{\mathbf{x}_1}\tau_{J1}\mathbf{I} & \sigma_{\mathbf{x}_J}\sigma_{\mathbf{x}_2}\tau_{J,2}\mathbf{I} & \cdots & \sigma^2_{\mathbf{x}_J}\mathbf{I} \end{pmatrix} \end{pmatrix}. \tag{B.12}$$

Corollary B.2. *Under the covariance structure in (B.12), $\mathbf{r}|\mathbf{X}$ is normally distributed with an expected value given by:*

$$\boldsymbol{\mu}_{\mathbf{X}} = \mathbb{E}[\mathbf{r}|\mathbf{X}] = \boldsymbol{\mu} + \sum_{j=1}^{J} \frac{\rho_j\sigma_{\mathbf{r}}(\mathbf{x}_j - \bar{\mathbf{x}}_j)}{\sigma_{\mathbf{x}_j}(1-R_j^2)} - \sum_{\substack{i\neq j\\ 1\leq i,j\leq J}} \frac{\beta_{ij}\rho_i\sigma_{\mathbf{r}}(\mathbf{x}_j - \bar{\mathbf{x}}_j)}{\sigma_{\mathbf{x}_i}(1-R_i^2)}, \tag{B.13}$$

and a covariance matrix given by:

$$\boldsymbol{\Sigma}_{\mathbf{X}} = \mathrm{Cov}(\mathbf{r}|\mathbf{X}) = \boldsymbol{\Sigma} - \sum_{j=1}^{J} \frac{\rho_j^2\sigma_{\mathbf{r}}^2\mathbf{I}}{(1-R_j^2)} + \sum_{\substack{i\neq j\\ 1\leq i,j\leq J}} \frac{\beta_{ij}\rho_i\rho_j\sigma_{\mathbf{x}_j}\mathbf{I}}{\sigma_{\mathbf{x}_i}(1-R_i^2)}, \tag{B.14}$$

where β_{ij} and R_i^2 are defined in Proposition B.1.

Similarly, it is straightforward to derive corollaries of Proposition B.2 under the covariance structures in (B.9) and (B.12).

Proof of Proposition B.1. We recall from (A.20)–(A.21) in the proof of Proposition 3 in Section A.3 that both $\boldsymbol{\mu}_{\mathbf{X}}$ and $\boldsymbol{\Sigma}_{\mathbf{X}}$ depend critically on $\mathrm{Cov}(\mathbf{X},\mathbf{X})^{-1}$:

$$\begin{aligned} \boldsymbol{\mu}_{\mathbf{X}} &= \boldsymbol{\mu} + \mathrm{Cov}(\mathbf{r},\mathbf{X})\mathrm{Cov}(\mathbf{X},\mathbf{X})^{-1}(\mathbf{X}-\bar{\mathbf{X}}), \\ \boldsymbol{\Sigma}_{\mathbf{X}} &= \boldsymbol{\Sigma} - \mathrm{Cov}(\mathbf{r},\mathbf{X})\mathrm{Cov}(\mathbf{X},\mathbf{X})^{-1}\mathrm{Cov}(\mathbf{r},\mathbf{X})'. \end{aligned} \tag{B.15}$$

Therefore, the key to the proof of Proposition B.1 is to derive analytical expressions of $\mathrm{Cov}(\mathbf{X},\mathbf{X})^{-1}$ under covariance structure (B.1) in Assumption B.1.

For notational simplicity we define $Q \equiv \mathrm{Cov}(\mathbf{X},\mathbf{X})^{-1}$. Because $\mathrm{Cov}(\mathbf{X},\mathbf{X})$ is a block matrix under the covariance structure in (B.1), and linear operations of block matrices are equivalent to element-wise operations, we treat $\mathrm{Cov}(\mathbf{X},\mathbf{X})$ and Q as $J\times J$ matrices for notational simplicity.

We first calculate the explicit solution of the first row of Q. We express Q as:

$$Q = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix} \tag{B.16}$$

where the sizes of the sub-blocks Q_{11}, Q_{12}, Q_{21}, and Q_{22} are 1×1, $1 \times (J-1)$, $(J-1) \times 1$, and $(J-1) \times (J-1)$, respectively. Therefore, we have:

$$\begin{aligned}
\mathbf{I} &= Q \cdot \mathrm{Cov}(\mathbf{X}, \mathbf{X}) \\
&= \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix} \begin{pmatrix} \sigma^2_{\mathbf{x}_1} & \sigma_{\mathbf{x}_1}\sigma_{\mathbf{x}_2}\tau_{12} & \cdots & \sigma_{\mathbf{x}_1}\sigma_{\mathbf{x}_J}\tau_{1J} \\ \sigma_{\mathbf{x}_2}\sigma_{\mathbf{x}_1}\tau_{21} & \sigma^2_{\mathbf{x}_2} & \cdots & \sigma_{\mathbf{x}_2}\sigma_{\mathbf{x}_J}\tau_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{\mathbf{x}_J}\sigma_{\mathbf{x}_1}\tau_{J1} & \sigma_{\mathbf{x}_J}\sigma_{\mathbf{x}_2}\tau_{J,2} & \cdots & \sigma^2_{\mathbf{x}_J} \end{pmatrix} \\
&= \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix} \begin{pmatrix} \sigma^2_{\mathbf{x}_1} & d'_{1,J-1} \\ d_{1,J-1} & C_{J-1} \end{pmatrix},
\end{aligned} \tag{B.17}$$

where we write the $(J-1)$-dimensional vector $d_{1,J-1} \equiv \mathrm{Cov}\left(\mathbf{x}_{[-1]}, x_1\right)$ and the $(J-1) \times (J-1)$ matrix $C_{J-1} \equiv \mathrm{Cov}\left(\mathbf{x}_{[-1]}, \mathbf{x}_{[-1]}\right)$ for convenience.

The block-wise calculations of the first row in (B.17) leads to:

$$\begin{cases} Q_{11}\sigma^2_{\mathbf{x}_1} + Q_{12}d_{1,J-1} = 1, \\ Q_{11}d'_{1,J-1} + Q_{12}C_{J-1} = \mathbf{0}. \end{cases} \tag{B.18}$$

After rearranging terms, Q_{12} and Q_{11} can be expressed as:

$$\begin{cases} Q_{11} = (\sigma^2_{\mathbf{x}_1} - d'_{1,J-1}C^{-1}_{J-1}d_{1,J-1})^{-1}, \\ Q_{12} = -Q_{11}d'_{1,J-1}C^{-1}_{J-1}. \end{cases} \tag{B.19}$$

We observe that the terms in (B.19) correspond to an interpretation of a regression in which x_1 is projected onto the $(J-1)$-dimensional space of $\mathbf{x}_{[-1]}$. The $(J-1)$-dimensional vector of regression coefficients is given by (B.5):

$$\boldsymbol{\beta}_1 = \left[\mathrm{Cov}\left(\mathbf{x}_{[-1]}, \mathbf{x}_{[-1]}\right)\right]^{-1} \mathrm{Cov}\left(\mathbf{x}_{[-1]}, x_1\right) = C^{-1}_{J-1}d_{1,J-1}, \tag{B.20}$$

and the R-squared is given by (B.6):

$$R^2_1 = \frac{\mathrm{Cov}\left(\mathbf{x}_{[-1]}, x_1\right)' \left[\mathrm{Cov}\left(\mathbf{x}_{[-1]}, \mathbf{x}_{[-1]}\right)\right]^{-1} \mathrm{Cov}\left(\mathbf{x}_{[-1]}, x_1\right)}{\sigma^2_{\mathbf{x}_1}} = \frac{d'_{1,J-1}C^{-1}_{J-1}d_{1,J-1}}{\sigma^2_{\mathbf{x}_1}}. \tag{B.21}$$

Substituting (B.20)–(B.21) into (B.19) leads to:

$$\begin{cases} Q_{11} = \dfrac{1}{\sigma^2_{\mathbf{x}_1}(1 - R^2_1)}, \\ Q_{12} = \dfrac{-\boldsymbol{\beta}'_1}{\sigma^2_{\mathbf{x}_1}(1 - R^2_1)}. \end{cases} \tag{B.22}$$

Finally, we observe that the derivations in (B.16)–(B.22) apply to any row of the matrix Q. By symmetry, we have:

$$Q = \begin{pmatrix} \frac{1}{\sigma^2_{\mathbf{x}_1}(1-R_1^2)} & \frac{-\beta_{12}}{\sigma^2_{\mathbf{x}_1}(1-R_1^2)} & \cdots & \frac{-\beta_{1J}}{\sigma^2_{\mathbf{x}_1}(1-R_1^2)} \\ \frac{-\beta_{21}}{\sigma^2_{\mathbf{x}_2}(1-R_2^2)} & \frac{1}{\sigma^2_{\mathbf{x}_2}(1-R_2^2)} & \cdots & \frac{-\beta_{2J}}{\sigma^2_{\mathbf{x}_2}(1-R_2^2)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{-\beta_{J1}}{\sigma^2_{\mathbf{x}_J}(1-R_J^2)} & \frac{-\beta_{J2}}{\sigma^2_{\mathbf{x}_J}(1-R_J^2)} & \cdots & \frac{1}{\sigma^2_{\mathbf{x}_J}(1-R_J^2)}. \end{pmatrix} \tag{B.23}$$

Substituting (B.23) into (B.15) completes the proof.

Proof of Proposition B.2. Substituting the excess return and excess covariance from the information in Proposition B.1 into the decomposition of Proposition 2 yields the general decomposition in Proposition B.2 under Assumption B.1.

B.2 Additional Results for Attribution Using Bayesian Portfolio Analysis

In this section, we provide formal statements of our results in Section 3.

B.2.1 General Distributions of Predictive Returns

Proposition B.3 (Conditional Attribution with Information using Predictive Distribution)**.** *Under Assumption 1 and the Bayesian framework to solve for the optimal portfolio in (20), conditioned on information in* $\mathbf{X}$ *that is used to form constraints* $\mathbf{A}(\mathbf{X})$*, the following decompositions hold for the optimal portfolio,* $\tilde{\boldsymbol{\omega}}^*$*.*

1. *Decomposition of the expected predictive return.*

$$\mathbb{E}\left[\tilde{\boldsymbol{\omega}}^{*'}\tilde{\mathbf{r}}|\mathbf{X}\right] = \tilde{\boldsymbol{\mu}}'_{\mathbf{X}}\tilde{\boldsymbol{\omega}}^* = \tilde{\boldsymbol{\mu}}'_{\mathbf{X}}\tilde{\boldsymbol{\omega}}_{\text{MVO}} + \tilde{\boldsymbol{\mu}}'\tilde{\boldsymbol{\omega}}_{\text{CSTR}} + (\tilde{\boldsymbol{\mu}}'_{\mathbf{X}} - \tilde{\boldsymbol{\mu}}')\tilde{\boldsymbol{\omega}}_{\text{CSTR}}, \tag{B.24}$$

   where

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

   Here the Lagrange multipliers are given by:

$$\tilde{\boldsymbol{\lambda}}^* = \left(\mathbf{A}(\mathbf{X})\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{A}(\mathbf{X})'\right)^{-1}\left(\mathbf{A}(\mathbf{X})\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\boldsymbol{\mu}} - \gamma\mathbf{b}\right) \tag{B.25}$$

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

2. Decomposition of the expected utility with respect to the predictive distribution.

$$
\begin{aligned}
\tilde{\boldsymbol{\mu}}_{\mathbf{X}}'\tilde{\boldsymbol{\omega}}^* - \frac{\gamma}{2}\tilde{\boldsymbol{\omega}}^{*\prime}\tilde{\boldsymbol{\Sigma}}_{\mathbf{X}}\tilde{\boldsymbol{\omega}}^* = \tilde{\boldsymbol{\mu}}_{\mathbf{X}}'\tilde{\boldsymbol{\omega}}_{\text{MVO}} - \frac{\gamma}{2}\tilde{\boldsymbol{\omega}}_{\text{MVO}}'\tilde{\boldsymbol{\Sigma}}_{\mathbf{X}}\tilde{\boldsymbol{\omega}}_{\text{MVO}} \\
- \frac{\gamma}{2}\tilde{\boldsymbol{\omega}}_{\text{CSTR}}'\tilde{\boldsymbol{\Sigma}}\tilde{\boldsymbol{\omega}}_{\text{CSTR}} \\
+ (\tilde{\boldsymbol{\mu}}_{\mathbf{X}}' - \tilde{\boldsymbol{\mu}}')\tilde{\boldsymbol{\omega}}_{\text{CSTR}} - \gamma\tilde{\boldsymbol{\omega}}_{\text{SHR}}'(\tilde{\boldsymbol{\Sigma}}_{\mathbf{X}} - \tilde{\boldsymbol{\Sigma}})\tilde{\boldsymbol{\omega}}_{\text{CSTR}}.
\end{aligned}
\tag{B.26}
$$

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

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

$$
\tilde{\boldsymbol{\omega}}_{\text{SHR}} \equiv \tilde{\boldsymbol{\omega}}_{\text{MVO}} + \frac{1}{2}\tilde{\boldsymbol{\omega}}_{\text{CSTR}} = \tilde{\boldsymbol{\omega}}^* - \frac{1}{2}\tilde{\boldsymbol{\omega}}_{\text{CSTR}}. \tag{B.27}
$$

B.2.2 Specific Distributions of Predictive Returns

The following assumptions replace the population return vector $\mathbf{r}$ in Assumptions 2 and 2' with the predictive return vector $\tilde{\mathbf{r}}$.

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

When $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are both unknown and modeled as random parameters in the Bayesian framework, the predictive density of returns is typically a multivariate Student-t distribution,

$$
\tilde{\mathbf{r}} \sim \text{MVT}\left(\tilde{\boldsymbol{\mu}}, \tilde{\mathbf{V}}, \nu\right), \tag{B.28}
$$

where $\tilde{\boldsymbol{\mu}}$ and $\tilde{\mathbf{V}}$ are the location vector and scale matrix, respectively, and ν is the degree of freedom.

Assumption B.2'. *The predictive return and asset characteristics follow a multivariate Student-t distribution:*

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

The next assumption describes the dependence between characteristics and predictive returns.

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

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

2. *For the j-th constraint, the correlation between the predictive return and characteristic value of each asset is $\tilde{\rho}_j$, and there is no cross-correlation between the predictive return and characteristic value of different assets. In other words, the covariance between predictive returns $\tilde{\mathbf{r}}$ and characteristics $\mathbf{x}_j$ is given by*

$$\mathrm{Cov}(\tilde{\mathbf{r}}, \mathbf{x}_j) = \tilde{\rho}_j \sigma_{\tilde{\mathbf{r}}} \sigma_{\mathbf{x}_j} \mathbf{I}$$

, where $\sigma_{\tilde{\mathbf{r}}}$ is the cross-sectional standard deviation of predictive returns, $\sigma_{\mathbf{x}_j}$ is the cross-sectional standard deviation of the j-th characteristic, and $\mathbf{I}$ is the identity matrix.

This allows us to derive an explicit expression of the information component from predictive returns similar to that in Proposition 3.

Proposition B.4 (Information Decomposition Using Normal and MVT Predictive Distribution)**.** *Under Assumptions 1 and B.3,*

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

$$\tilde{\boldsymbol{\mu}}_{\mathbf{X}} = \mathbb{E}[\tilde{\mathbf{r}}|\mathbf{X}] = \tilde{\boldsymbol{\mu}} + \sum_{j=1}^{J} \tilde{\rho}_j \sigma_{\tilde{\mathbf{r}}} \frac{(\mathbf{x}_j - \bar{\mathbf{x}}_j)}{\sigma_{\mathbf{x}_j}}, \tag{B.29}$$

and a covariance matrix given by:

$$\tilde{\boldsymbol{\Sigma}}_{\mathbf{X}} = \mathrm{Cov}(\tilde{\mathbf{r}}|\mathbf{X}) = \tilde{\boldsymbol{\Sigma}} - \sum_{j=1}^{J} \tilde{\rho}_j^2 \sigma_{\tilde{\mathbf{r}}^2} \mathbf{I}. \tag{B.30}$$

- *if Assumption B.2' holds, $\tilde{\mathbf{r}}|\mathbf{X}$ follows a multivariate Student-t distribution:*

$$\mathrm{MVT}\left(\tilde{\boldsymbol{\mu}} + \tilde{\mathbf{V}}_{\mathbf{r},\mathbf{X}} \tilde{\mathbf{V}}_{\mathbf{X},\mathbf{X}}^{-1} (\mathbf{X} - \bar{\mathbf{X}}), s(\tilde{\mathbf{V}} - \tilde{\mathbf{V}}_{\mathbf{r},\mathbf{X}} \tilde{\mathbf{V}}_{\mathbf{X},\mathbf{X}}^{-1} \tilde{\mathbf{V}}_{\mathbf{X},\mathbf{r}}), \nu + NJ\right), \tag{B.31}$$

where

$$s = \frac{\nu + (\mathbf{X} - \bar{\mathbf{X}})' \tilde{\mathbf{V}}_{\mathbf{X},\mathbf{X}}^{-1} (\mathbf{X} - \bar{\mathbf{X}})}{\nu + NJ} \tag{B.32}$$

is a scaling parameter that approaches 1 as the number of assets, N, increases without bound. In particular, its expected value is given by:

$$\tilde{\boldsymbol{\mu}}_{\mathbf{X}} = \mathbb{E}[\tilde{\mathbf{r}}|\mathbf{X}] = \tilde{\boldsymbol{\mu}} + \sum_{j=1}^{J} \tilde{\rho}_j \sigma_{\tilde{\mathbf{r}}} \frac{(\mathbf{x}_j - \bar{\mathbf{x}}_j)}{\sigma_{\mathbf{x}_j}}, \tag{B.33}$$

for all N, and its covariance matrix is given by:

$$\tilde{\mathbf{\Sigma}}_{\mathbf{X}} = \text{Cov}(\tilde{\mathbf{r}}|\mathbf{X}) \stackrel{p}{=} (1 - 2/\nu) \left(\tilde{\mathbf{\Sigma}} - \sum_{j=1}^{J} \tilde{\rho}_j^2 \sigma_{\tilde{\mathbf{r}}^2} \mathbf{I} \right), \tag{B.34}$$

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

This result can be applied to several classical Bayesian portfolios with normal predictive distributions. For example, Klein and Bawa (1976) consider a model with a known covariance matrix and non-informative priors on the expected returns of a subset of assets. They derive a normally distributed predictive density. Jorion (1986) uses an informative conjugate prior and applies the James–Stein estimator to expected returns, which leads to a multivariate normal predictive density, conditioned on a known covariance matrix. The optimal portfolio is constructed using the predictive density with a plug-in covariance estimator following Zellner and Chetty (1965). Black and Litterman (1992) develop a quasi-Bayesian approach that allows investors to incorporate private views into the market view of expected returns.[3] Their framework is concerned with the uncertainty in the expected value but not the covariance matrix, in which case the predictive distribution is normal (Kolm and Ritter, 2017, p. 566).

Several other Bayesian and shrinkage portfolios have predictive distributions that follow multivariate Student-t distributions. Early studies such as Klein and Bawa (1976) and Brown (1978) use a non-informative diffuse prior, $\mathbb{P}(\boldsymbol{\mu}, \mathbf{\Sigma}) \propto |\mathbf{\Sigma}|^{-(N+1)/2}$, which leads to a predictive distribution with MVT $\left(\hat{\boldsymbol{\mu}}, \sqrt{1 + 1/T}\hat{\mathbf{\Sigma}}, T - N\right)$. The Bayesian optimal portfolio without constraints is shown to be $\tilde{\boldsymbol{\omega}}_{\text{MVO}} = \gamma^{-1}(T - N - 2)(T + 1)^{-1}\hat{\mathbf{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}$, where $\hat{\boldsymbol{\mu}}$ and $\hat{\mathbf{\Sigma}}$ are the sample mean and sample covariance matrix of the observed returns in the past T periods, respectively.

The conjugate prior is another popular class of priors on $(\boldsymbol{\mu}, \mathbf{\Sigma})$, which allows their posterior to take the same form of distribution as the prior. The most common specification considers a normal prior for $\boldsymbol{\mu}$ conditioned on $\mathbf{\Sigma}$ and an Inversed-Wishart prior for $\mathbf{\Sigma}$, which is adopted by, for example, Frost and Savarino (1986), Stambaugh (1997), Pástor (2000), Pástor and Stambaugh (2000), Zhou (2009), and Lai, Xing, and Chen (2011). In this case, the predictive distribution is still MVT, but the mean and covariance both have more shrinkage towards the fixed level specified in the prior.

Tu and Zhou (2010) analyze priors on portfolio weights instead of unknown parameters of the return distribution, a method termed "objective-based priors." They show that the prior on portfolio weights can be transformed into a prior on $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$. In particular, the objective-based prior, $\boldsymbol{\omega} \sim N(\boldsymbol{\omega}_0, \mathbf{\Sigma}_0\mathbf{\Sigma}^{-1}/\gamma)$, is equivalent to the prior on expected returns, $\boldsymbol{\mu} \sim N\left(\gamma\mathbf{\Sigma}\boldsymbol{\omega}_0, \sigma_\rho^2\mathbf{\Sigma}/s^2\right)$, where $\boldsymbol{\omega}_0$, $\mathbf{\Sigma}_0$, and σ_ρ are suitable prior constants and s^2 is the average diagonal elements of $\mathbf{\Sigma}$.

[3] Although usually considered a Bayesian approach, it is not entirely Bayesian because the data-generating process of returns is not fully specified explicitly. See discussions in Avramov and Zhou (2010, p. 30) and Jacquier and Polson (2011, p. 19). Nonetheless, we can interpret Black and Litterman's (1992) rule as an update of the prior of the equilibrium relationship with investors' private views (Zhou, 2009, p. 39).

In addition, DeMiguel et al. (2009) show that norm-constrained portfolios can be equivalently interpreted as Bayesian portfolios where investors have certain prior beliefs on portfolio weights $\boldsymbol{\omega}$. Their analysis applies very broadly to, for example, the no-short-sale portfolios (Jagannathan and Ma, 2003), the shrinkage covariance-based portfolios (Ledoit and Wolf, 2003, 2004), and the $1/N$ rule (DeMiguel, Garlappi, and Uppal, 2009).

The insights of DeMiguel et al. (2009) and Tu and Zhou (2010) provide an important link between the literature on robust portfolios with priors on portfolio weights[4] and the literature on Bayesian portfolio analysis. This implies that our attribution framework can, in principle, be applied to robust portfolio rules that directly impose priors or shrinkage on portfolio weights through their Bayesian equivalents. However, deriving the specific Bayesian formulation for each of them is not the focus and beyond the scope of this article.

B.3 Unconditional Attribution

The results in Proposition 2 are conditioned on $\mathbf{X}$. To obtain an unconditional decomposition over multiple time periods, we must compute the expectation of (6) and (7) 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 B.5 (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.*

1. *Expected return decomposition.*

$$\mathbb{E}\left[\boldsymbol{\omega}^{*\prime}\mathbf{r}\right] = \boldsymbol{\mu}'\boldsymbol{\omega}_{\text{MVO}} + \boldsymbol{\mu}'\mathbb{E}\left[\boldsymbol{\omega}_{\text{CSTR}}\right] + \mathbb{E}\left[(\boldsymbol{\mu}'_{\mathbf{X}} - \boldsymbol{\mu}')\boldsymbol{\omega}_{\text{CSTR}}\right]. \tag{B.35}$$

2. *Expected utility decomposition.*

$$\begin{aligned}\mathbb{E}\left[\boldsymbol{\omega}^{*\prime}\mathbf{r}\right] - \frac{\gamma}{2}\text{Var}\left(\boldsymbol{\omega}^{*\prime}\mathbf{r}\right) = {} & \boldsymbol{\mu}'\boldsymbol{\omega}_{\text{MVO}} - \frac{\gamma}{2}\boldsymbol{\omega}'_{\text{MVO}}\boldsymbol{\Sigma}\boldsymbol{\omega}_{\text{MVO}} \\ & - \frac{\gamma}{2}\mathbb{E}\left[\boldsymbol{\omega}'_{\text{CSTR}}\boldsymbol{\Sigma}\boldsymbol{\omega}_{\text{CSTR}}\right] \\ & + \mathbb{E}\left[(\boldsymbol{\mu}'_{\mathbf{X}} - \boldsymbol{\mu}')\boldsymbol{\omega}_{\text{CSTR}} - \boldsymbol{\omega}'_{\text{SHR}}(\boldsymbol{\Sigma}_{\mathbf{X}} - \boldsymbol{\Sigma})\boldsymbol{\omega}_{\text{CSTR}}\right] \\ & - \frac{\gamma}{2}\left(\text{Var}\left(\boldsymbol{\omega}'_{\text{CSTR}}\boldsymbol{\mu}_{\mathbf{X}}\right) + 2\text{Cov}\left(\boldsymbol{\omega}'_{\text{MVO}}\boldsymbol{\mu}_{\mathbf{X}}, \boldsymbol{\omega}'_{\text{CSTR}}\boldsymbol{\mu}_{\mathbf{X}}\right)\right).\end{aligned} \tag{B.36}$$

Proof of Proposition B.5. 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 $\mathbf{X}$.

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

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

$$\operatorname{Var}\left(\boldsymbol{\omega}^{*\prime}\mathbf{r}\right) = \mathbb{E}\left[\operatorname{Var}\left(\boldsymbol{\omega}^{*\prime}\mathbf{r}|\mathbf{X}\right)\right] + \operatorname{Var}\left(\mathbb{E}\left[\boldsymbol{\omega}^{*\prime}\mathbf{r}|\mathbf{X}\right]\right). \tag{B.37}$$

Therefore,

$$\begin{aligned}
\mathbb{E}\left[\boldsymbol{\omega}^{*\prime}\mathbf{r}\right] - \frac{\gamma}{2}\operatorname{Var}\left(\boldsymbol{\omega}^{*\prime}\mathbf{r}\right) &= \mathbb{E}\left[\mathbb{E}\left[\boldsymbol{\omega}^{*\prime}\mathbf{r}|\mathbf{X}\right]\right] - \frac{\gamma}{2}\mathbb{E}\left[\operatorname{Var}\left(\boldsymbol{\omega}^{*\prime}\mathbf{r}|\mathbf{X}\right)\right] - \frac{\gamma}{2}\operatorname{Var}\left(\mathbb{E}\left[\boldsymbol{\omega}^{*\prime}\mathbf{r}|\mathbf{X}\right]\right) \\
&= \mathbb{E}\left[\mathbb{E}\left[\boldsymbol{\omega}^{*\prime}\mathbf{r}|\mathbf{X}\right] - \frac{\gamma}{2}\operatorname{Var}\left(\boldsymbol{\omega}^{*\prime}\mathbf{r}|\mathbf{X}\right)\right] - \frac{\gamma}{2}\operatorname{Var}\left(\mathbb{E}\left[\boldsymbol{\omega}^{*\prime}\mathbf{r}|\mathbf{X}\right]\right) \\
&= \mathbb{E}\left[\boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}^* - \frac{\gamma}{2}\boldsymbol{\omega}^{*\prime}\boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{\omega}^*\right] - \frac{\gamma}{2}\operatorname{Var}\left(\boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}^*\right).
\end{aligned} \tag{B.38}$$

The first term follows from (7) in Proposition 2. The variance in the second term can be further decomposed into:

$$\begin{aligned}
\operatorname{Var}\left(\boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}^*\right) &= \operatorname{Var}\left(\boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}_{\mathrm{MVO}} + \boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}_{\mathrm{CSTR}}\right) \\
&= \boldsymbol{\omega}_{\mathrm{MVO}}'\operatorname{Var}\left(\boldsymbol{\mu}_{\mathbf{X}}'\right)\boldsymbol{\omega}_{\mathrm{MVO}} + \operatorname{Var}\left(\boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}_{\mathrm{CSTR}}\right) + 2\operatorname{Cov}\left(\boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}_{\mathrm{MVO}}, \boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}_{\mathrm{CSTR}}\right).
\end{aligned} \tag{B.39}$$

Substituting both terms back into (B.38) leads to:

$$\begin{aligned}
\mathbb{E}\left[\boldsymbol{\omega}^{*\prime}\mathbf{r}\right] - \frac{\gamma}{2}\operatorname{Var}\left(\boldsymbol{\omega}^{*\prime}\mathbf{r}\right) &= \mathbb{E}[\boldsymbol{\mu}_{\mathbf{X}}']\boldsymbol{\omega}_{\mathrm{MVO}} - \frac{\gamma}{2}\boldsymbol{\omega}_{\mathrm{MVO}}'\mathbb{E}[\boldsymbol{\Sigma}_{\mathbf{X}}]\boldsymbol{\omega}_{\mathrm{MVO}} - \frac{\gamma}{2}\boldsymbol{\omega}_{\mathrm{MVO}}'\operatorname{Var}\left(\boldsymbol{\mu}_{\mathbf{X}}'\right)\boldsymbol{\omega}_{\mathrm{MVO}} \\
&\quad - \frac{\gamma}{2}\mathbb{E}[\boldsymbol{\omega}_{\mathrm{CSTR}}'\boldsymbol{\Sigma}\boldsymbol{\omega}_{\mathrm{CSTR}}] \\
&\quad + \mathbb{E}\left[(\boldsymbol{\mu}_{\mathbf{X}}' - \boldsymbol{\mu}')\boldsymbol{\omega}_{\mathrm{CSTR}} - \boldsymbol{\omega}_{\mathrm{SHR}}'(\boldsymbol{\Sigma}_{\mathbf{X}} - \boldsymbol{\Sigma})\boldsymbol{\omega}_{\mathrm{CSTR}}\right] \\
&\quad - \frac{\gamma}{2}\operatorname{Var}\left(\boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}_{\mathrm{CSTR}}\right) - \frac{\gamma}{2}\operatorname{Cov}\left(\boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}_{\mathrm{MVO}}, \boldsymbol{\mu}_{\mathbf{X}}'\boldsymbol{\omega}_{\mathrm{CSTR}}\right) \\
&= \boldsymbol{\mu}'\boldsymbol{\omega}_{\mathrm{MVO}} - \frac{\gamma}{2}\boldsymbol{\omega}_{\mathrm{MVO}}'\boldsymbol{\Sigma}\boldsymbol{\omega}_{\mathrm{MVO}} \\
&\quad - \frac{\gamma}{2}\mathbb{E}\left[\boldsymbol{\omega}_{\mathrm{CSTR}}'\boldsymbol{\Sigma}\boldsymbol{\omega}_{\mathrm{CSTR}}\right] \\
&\quad + \mathbb{E}\left[(\boldsymbol{\mu}_{\mathbf{X}}' - \boldsymbol{\mu}')\boldsymbol{\omega}_{\mathrm{CSTR}} - \boldsymbol{\omega}_{\mathrm{SHR}}'(\boldsymbol{\Sigma}_{\mathbf{X}} - \boldsymbol{\Sigma})\boldsymbol{\omega}_{\mathrm{CSTR}}\right] \\
&\quad - \frac{\gamma}{2}\left(\operatorname{Var}\left(\boldsymbol{\omega}_{\mathrm{CSTR}}'\boldsymbol{\mu}_{\mathbf{X}}\right) + 2\operatorname{Cov}\left(\boldsymbol{\omega}_{\mathrm{MVO}}'\boldsymbol{\mu}_{\mathbf{X}}, \boldsymbol{\omega}_{\mathrm{CSTR}}'\boldsymbol{\mu}_{\mathbf{X}}\right)\right),
\end{aligned} \tag{B.40}$$

which completes the proof.

B.4 Ex-Post Return Attribution

Propositions 2–4 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, $\ddot{\mathbf{r}}$, to represent the realized returns of all assets. The goal for ex-post attribution is to decompose the realized portfolio return, $\ddot{\mathbf{r}}'\boldsymbol{\omega}^*$, into components attributable to the unconstrained MVO portfolio and each constraint.

If we treat constraints as static, (4) in Proposition 1 already provides such a decomposition as long as the *ex-ante* expected returns are replaced by *ex-post* realized returns:

$$\ddot{\mathbf{r}}'\boldsymbol{\omega}^* = \frac{1}{\gamma}\ddot{\mathbf{r}}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{1}{\gamma}\ddot{\mathbf{r}}'\boldsymbol{\Sigma}^{-1}\mathbf{A}'\boldsymbol{\lambda}^*. \tag{B.41}$$

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

However, this decomposition does not account for the information contained in each constraint. Equation (12) in Proposition 3 quantifies the excess return due to information, and we use the sample version of this decomposition to quantify realized returns attributable to information:

$$\ddot{\mathbf{r}}_{\text{Info}} = \sum_{j=1}^{J} \frac{\rho(\ddot{\mathbf{r}}, \ddot{\mathbf{x}}_j)\ddot{\sigma}_{\mathbf{r}}(\ddot{\mathbf{x}}_j - \ddot{\bar{\mathbf{x}}}_j)}{\ddot{\sigma}_{\mathbf{x}_j}}. \tag{B.42}$$

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

Proposition B.6 (Ex-Post Return Attribution)**.** *Under Assumptions 1, 2 (or 2'), and 3, realized portfolio returns can be decomposed into:*

$$\ddot{\mathbf{r}}'\boldsymbol{\omega}^* = \ddot{\mathbf{r}}'\boldsymbol{\omega}_{\text{MVO}} + \ddot{\mathbf{r}}'_{\text{Static}}\boldsymbol{\omega}_{\text{CSTR}} + \sum_{j=1}^{J}\rho(\ddot{\mathbf{r}}, \ddot{\mathbf{x}}_j)\ddot{\sigma}_{\mathbf{r}}\frac{(\ddot{\mathbf{x}}'_j - \ddot{\bar{\mathbf{x}}}'_j)\boldsymbol{\omega}_{\text{CSTR}}}{\ddot{\sigma}_{\mathbf{x}_j}}. \tag{B.43}$$

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

It is worth noting that in the last term of (B.43), 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.

Proof of Proposition B.6. Given the decomposition of portfolio holdings in (3), we have:

$$
\begin{aligned}
\ddot{\mathbf{r}}'\boldsymbol{\omega}^* &= \frac{1}{\gamma}\ddot{\mathbf{r}}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{1}{\gamma}\ddot{\mathbf{r}}'\boldsymbol{\Sigma}^{-1}\mathbf{A}'\boldsymbol{\lambda}^* \\
&= \frac{1}{\gamma}\ddot{\mathbf{r}}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{1}{\gamma}\ddot{\mathbf{r}}'_{\text{Static}}\boldsymbol{\Sigma}^{-1}\mathbf{A}'\boldsymbol{\lambda}^* - \frac{1}{\gamma}\ddot{\mathbf{r}}'_{\text{Info}}\boldsymbol{\Sigma}^{-1}\mathbf{A}'\boldsymbol{\lambda}^* \\
&\overset{(1)}{=} \frac{1}{\gamma}\ddot{\mathbf{r}}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{1}{\gamma}\ddot{\mathbf{r}}'_{\text{Static}}\boldsymbol{\Sigma}^{-1}\mathbf{A}'\boldsymbol{\lambda}^* - \frac{1}{\gamma}\left(\sum_{j=1}^{J}\frac{\rho(\ddot{\mathbf{r}},\ddot{\mathbf{x}}_j)\ddot{\sigma}_{\mathbf{r}}(\ddot{\mathbf{x}}'_j - \ddot{\bar{\mathbf{x}}}'_j)}{\ddot{\sigma}_{\mathbf{x}_j}}\right)\boldsymbol{\Sigma}^{-1}\mathbf{A}'\boldsymbol{\lambda}^* \\
&= \frac{1}{\gamma}\ddot{\mathbf{r}}'\boldsymbol{\omega}_{\text{MVO}} + \frac{1}{\gamma}\ddot{\mathbf{r}}'_{\text{Static}}\boldsymbol{\omega}_{\text{CSTR}} + \frac{1}{\gamma}\sum_{j=1}^{J}\rho(\ddot{\mathbf{r}},\ddot{\mathbf{x}}_j)\ddot{\sigma}_{\mathbf{r}}\frac{(\ddot{\mathbf{x}}'_j - \ddot{\bar{\mathbf{x}}}'_j)\boldsymbol{\omega}_{\text{CSTR}}}{\ddot{\sigma}_{\mathbf{x}_j}},
\end{aligned}
\tag{B.44}
$$

where step (1) follows from the definition of $\ddot{\mathbf{r}}_{\text{Info}}$ in (B.42).

B.5 Additional Simulation Results

Factor exposure. Figure B.1 demonstrates the attribution of expected utility for the example considered in Section 4.1.

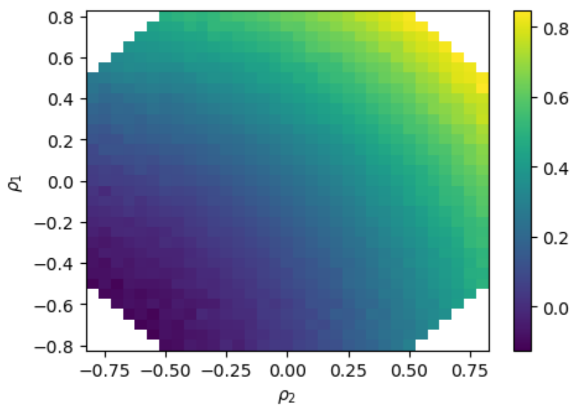
Figure B.1a shows the expected utility of the constrained portfolio, which varies between 0.8 and 2.2 as ρ_1 and ρ_2 vary between -0.8 and 0.8. Figure B.1b shows the expected utility of the unconstrained MVO problem, which is around 1.25 regardless of values of ρ_1 and ρ_2. When ρ_1 and ρ_2 are high, the expected utility of the constrained portfolio can actually be higher than that of the unconstrained portfolio.

The difference in expected utility between the unconstrained MVO and constrained portfolios is decomposed into the static and information components. Figure B.1c 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 B.1d 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.

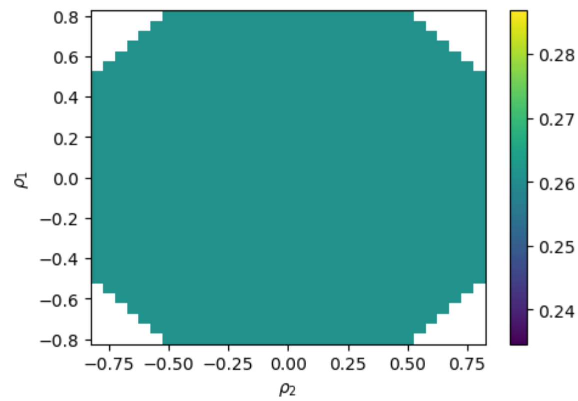
Exclusionary investing. Figure B.2 demonstrates the attribution of expected utility for the example considered in Section 4.2.

Figure B.2a shows the expected utility of the constrained portfolio, which varies between -0.5 and 2.0 as ρ_1 and ρ_2 vary between -0.8 and 0.8. Figure B.2b shows the expected utility of the unconstrained MVO problem, which is again around 1.25 regardless of values of ρ_1 and ρ_2.

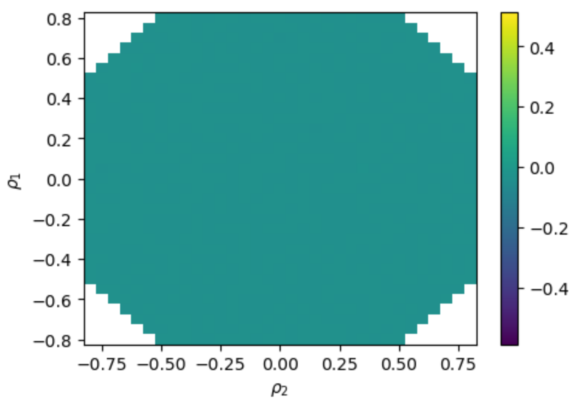
The difference in expected utility between the unconstrained MVO and constrained portfolios is decomposed into the static and information components. Figure B.2c shows the expected utility attributable to the exclusionary constraints as if they are static, which contributes to the expected utility negatively, ranging from -1.0 to -0.2. Figure B.2d shows the expected utility attributable to information, which increases as ρ increases, and has a similar pattern to Figure 3d.



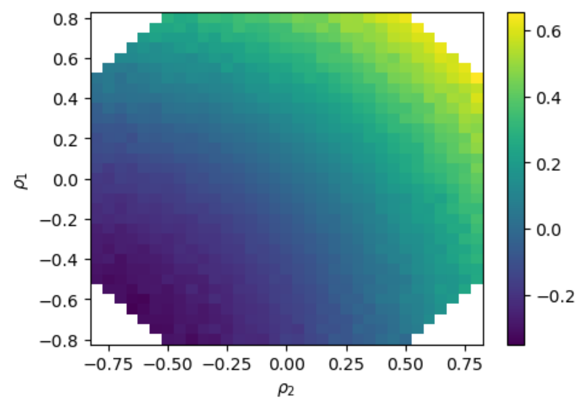
(a) Constrained Optimum



(b) Unconstrained MVO



(c) Decomposition: Constraint (Static)



(d) Decomposition: Constraint (Information)

Figure B.1: Decomposition of expected utility for the optimization problem in (27) with two constraints that depend on random characteristics, as correlations (ρ_1 and ρ_2) between the 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).

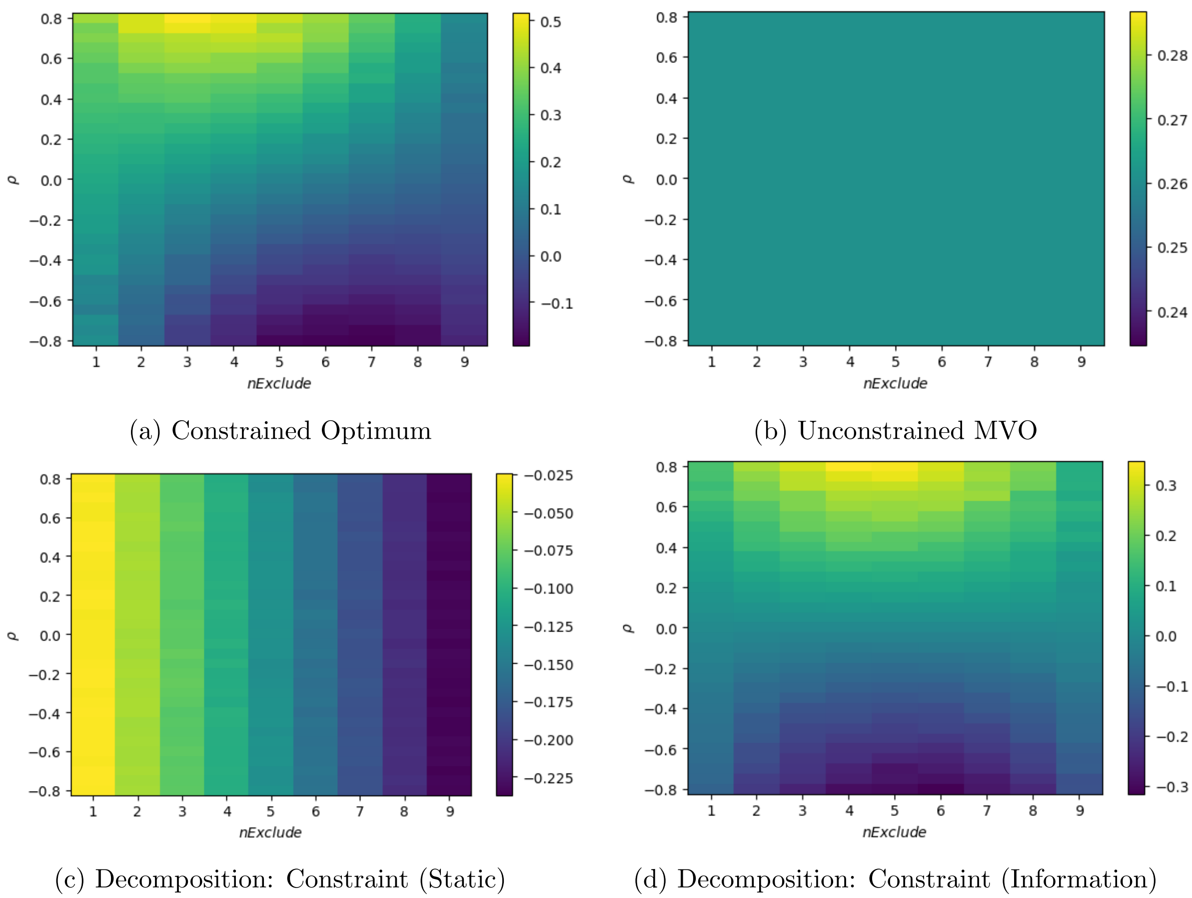


Figure B.2: Decomposition of expected utility for the problem in (30) with one exclusionary constraint that depends on random characteristics, as the number of excluded assets ($nExclude$) and the correlation (ρ) between the random characteristic and asset returns vary. The expected utility of the constrained portfolio (a) is decomposed into components corresponding to the unconstrained MVO portfolio (b), static constraints (c), and information in the constraints (d).

B.6 Additional Results for Empirical Analysis

B.6.1 Details of the Construction of ESG Scores

The raw ESG data in the MSCI KLD ESG dataset classifies environmental, social, and governance performance into 13 different categories, including seven qualitative categories (community, diversity, employee relations, environment, corporate governance, human rights, and product) and six controversial-business categories (alcohol, gambling, firearms, military, nuclear, and tobacco). 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.

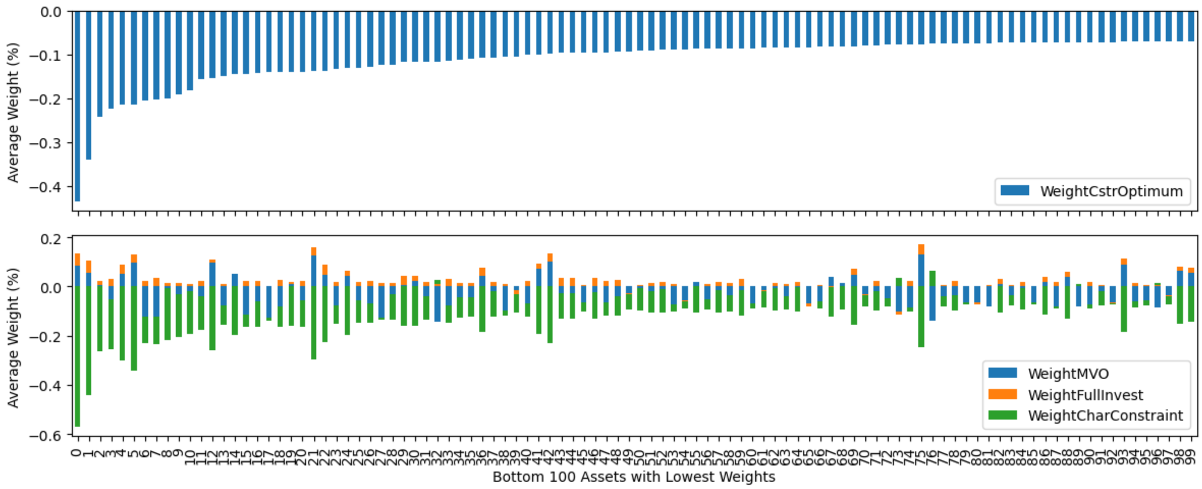
We follow Lins, Servaes, and Tamayo (2017) in aggregating the raw data into an ESG score. First, we mark all missing ratings as zero. As the maximum number of strengths and concerns for any given category will vary over time, we scale them for each category by dividing the number of strengths or concerns 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 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 $+7$.[5]

B.6.2 Portfolio Holdings Decomposition of the Main Empirical Example

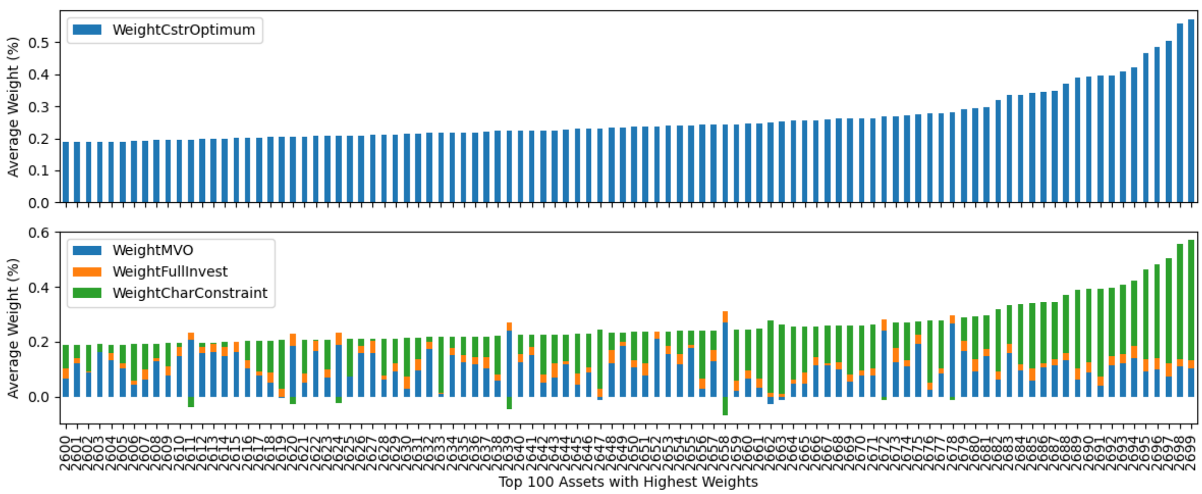
Taking Jorion's (1986) rule as an example, Figures B.3a and B.3b show the bottom and top 100 assets with the lowest and highest portfolio weights for the optimal portfolio solved from (32) and the estimators in (33), respectively, averaged over all years. We decompose the portfolio weights into components corresponding to the unconstrained MVO portfolio and two constraints, based on (3) in Proposition 1. In both the top and bottom assets, the full investment constraint (orange) typically gives positive weights.

For the bottom 100 assets in Figure B.3a, 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 B.3b, the unconstrained MVO portfolio (blue) generally gives positive weights and the ESG constraint (green) leads to additional positive weights.

[5] 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.



(a) Bottom assets with the lowest weights.



(b) Top assets with the highest weights.

Figure B.3: Average portfolio weights over all years and their decomposition, for the portfolio defined in (32) with a constraint on the average portfolio characteristic value ($\boldsymbol{\omega}'\mathbf{x}_{\text{ESG}} \geq 1.0$) and Jorion's (1986) estimates of predictive moments. (a) shows the 100 assets with the lowest weights and (b) shows the 100 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).

B.6.3 Long-Only ESG Portfolios

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

$$\begin{aligned} \max_{\boldsymbol{\omega}} \quad & \boldsymbol{\omega}'\boldsymbol{\mu} - \frac{\gamma}{2}\boldsymbol{\omega}'\boldsymbol{\Sigma}\boldsymbol{\omega} \\ \text{s.t.} \quad & \boldsymbol{\omega}'\mathbf{1} = 1 \\ & \boldsymbol{\omega}'\mathbf{x}_{\text{ESG}} \geq b \\ & \boldsymbol{\omega} \geq \mathbf{0}. \end{aligned} \tag{B.45}$$

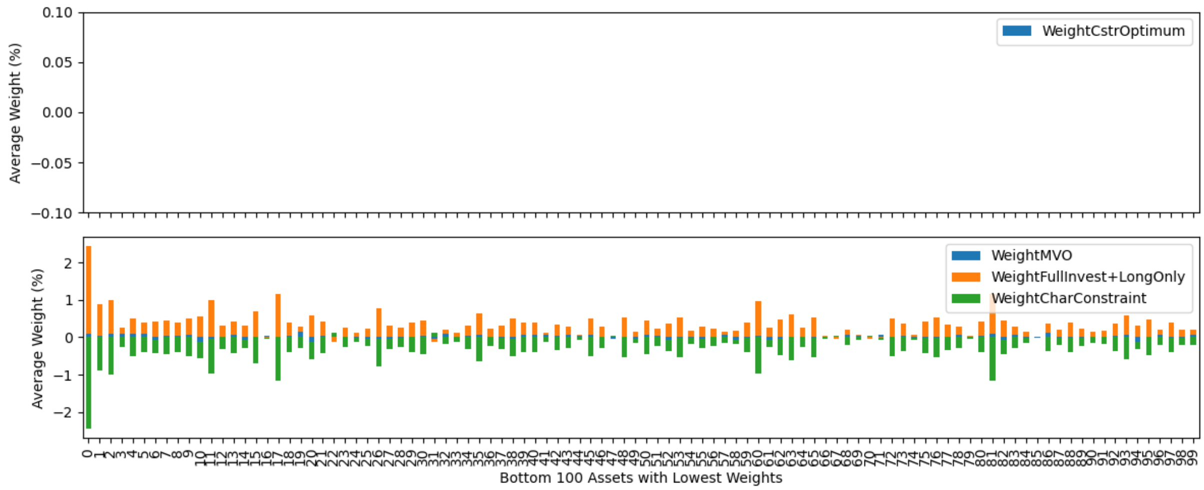
In contrast to (32), we have an additional constraint that all portfolio weights must be non-negative. We again set $b = 1$ as an example in our analysis, and use the return forecasts in (33)–(34) to estimate $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. We set $\gamma = 5$.

Portfolio holdings decomposition. Taking Jorion's (1986) portfolio as an example, Figure B.4 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 (3) in Proposition 1. For performance attributions of long-only portfolios, we always combine the full investment constraint and the long-only constraint for simplicity.

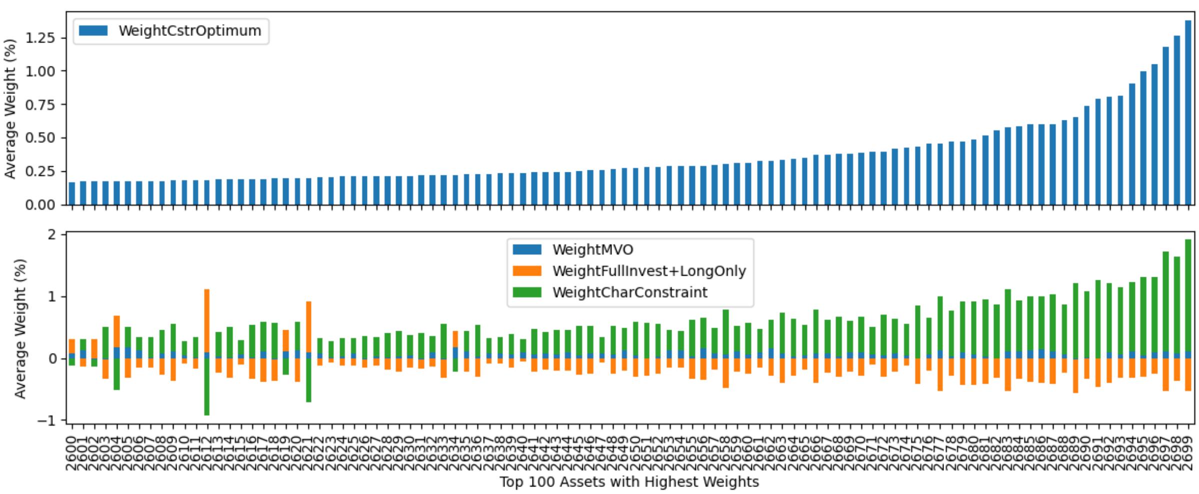
In Figure B.4a, the bottom assets have zero weights by design. This is a result of the negative contribution from the ESG constraint (green) combined with the positive contribution from the full investment and long-only constraints (orange). In other words, these assets tend to have low ESG scores, but the long-only constraint forces them to have zero weights instead of negative weights. However, for the top 100 assets in Figure B.4b, 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.5 demonstrates the decomposition of the expected utility and expected return of the long-only portfolio into different components, again for Jorion's (1986) portfolio as an example.

The upper panel of Figure B.5a 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 (18) in Proposition 4. The expected utility of the unconstrained MVO portfolio (blue) is positive over all years. Compared with the portfolio in Figure 5 that allows short positions, 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 portfolio with short positions, 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 4.



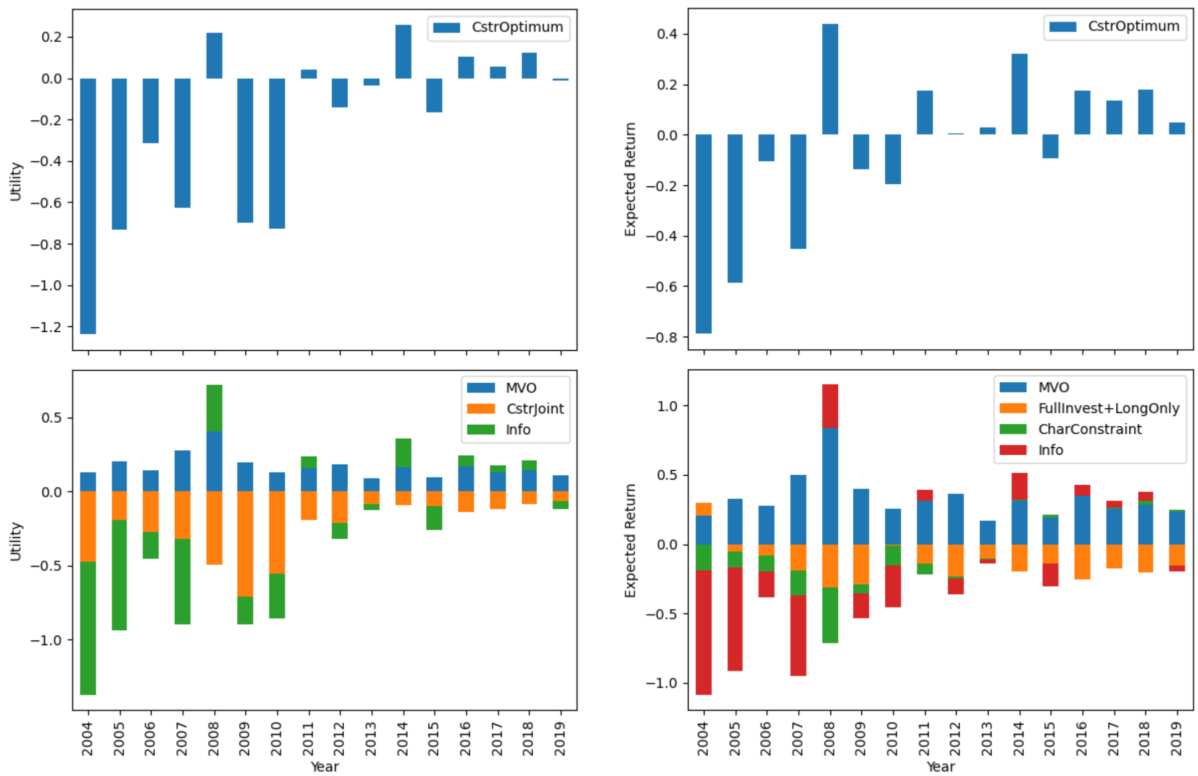
(a) Bottom assets with lowest weights.



(b) Top assets with highest weights.

Figure B.4: Average portfolio weights over all years and their decomposition, for the long-only portfolio defined in (B.45) with a constraint on the average portfolio characteristic value ($\boldsymbol{\omega}'\mathbf{x}_{\mathrm{ESG}} \geq 1.0$) and Jorion's (1986) estimates of predictive moments. (a) shows the 100 assets with the lowest weights and (b) shows the 100 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 ($\boldsymbol{\omega}'\mathbf{x}_{\mathrm{ESG}} \geq 1.0$, green).

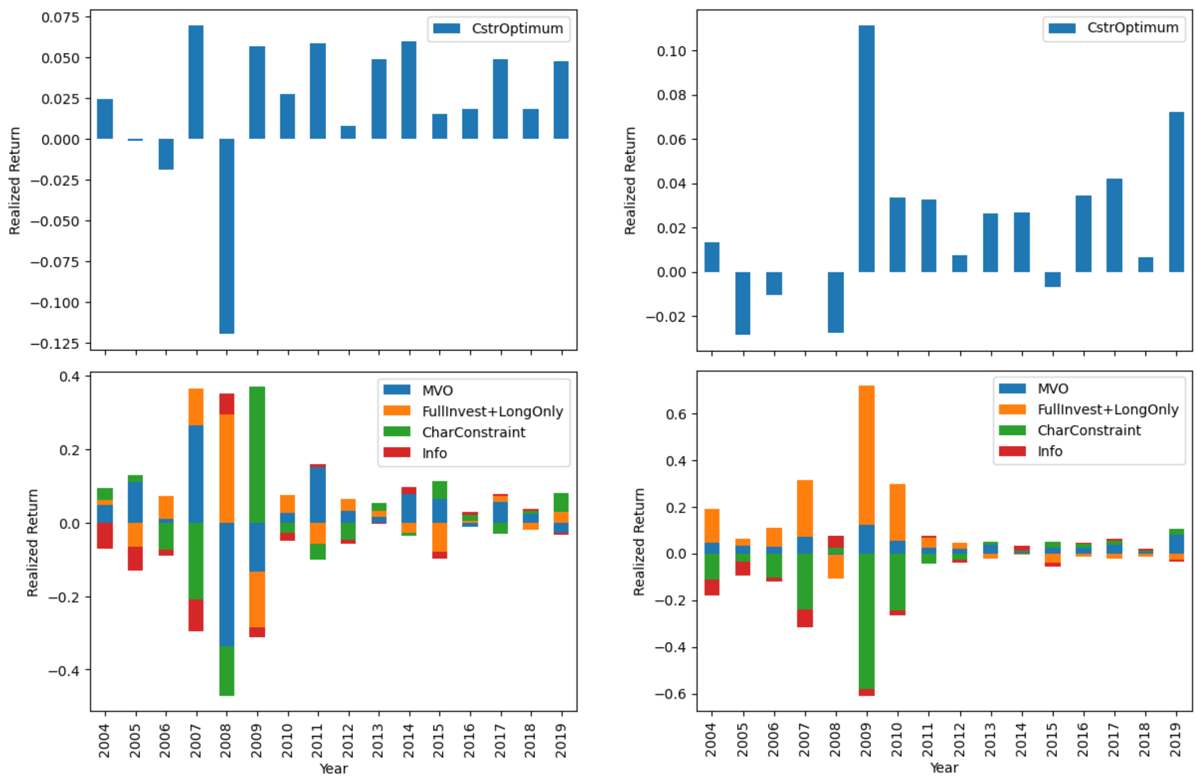
Figure B.5b shows the expected return of the optimal portfolio and its decomposition based on (17) in Proposition 4. The full investment and long-only constraints (orange) contribute negatively to expected returns. The ESG constraint (green) can contribute either positively or negatively to the expected returns, but generally on a small scale relative to other components. The expected return contribution from the information is very significant, which is negative before 2007 and positive in certain years after 2008. 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.



(a) Expected Utility (b) Expected Return

Figure B.5: Expected return and utility and their decomposition, for the long-only portfolio defined in (B.45) with a constraint on the average portfolio characteristic value ($\boldsymbol{\omega}'\mathbf{x}_{\text{ESG}} \geq 1.0$) and Jorion's (1986) estimates of predictive moments. 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 ($\boldsymbol{\omega}'\mathbf{x}_{\text{ESG}} \geq 1.0$) treated as static (green), and the information from the ESG constraint (red).

Realized return decomposition. Figure B.6 shows the realized returns of the optimal portfolio. Here we again compare Jorion's (1986) rule in Figure B.6a with the $1/N$ rule in Figure B.6b.



(a) Jorion (b) $1/N$

Figure B.6: Realized return and their decomposition, for the long-only portfolio defined in (B.45) with a constraint on the average portfolio characteristic value ($\boldsymbol{\omega}'\mathbf{x}_{\mathrm{ESG}} \geq 1.0$). (a) corresponds to Jorion's (1986) estimates of predictive moments in (33), and (b) corresponds to predictive moments consistent with the $1/N$ rule in (34). In each subfigure, the top panel shows the realized return in excess of the Fama–French five-factor model of the constrained portfolio, and the bottom panel shows its decomposition into components corresponding to the unconstrained MVO portfolio (blue), the full investment and long-only constraints combined together (orange), the ESG constraint ($\boldsymbol{\omega}'\mathbf{x}_{\mathrm{ESG}} \geq 1.0$) treated as static (green), and the information from the ESG constraint (red).

The upper panel of Figure B.6a shows that the realized residual returns of the constrained portfolio are positive in most years in our sample except for a large drawdown in 2008. The lower panel decomposes the realized return of the constrained portfolio based on Proposition B.6. The full investment and long-only constraints (orange) and the ESG constraint (green) can both contribute to the returns positively or negatively.

In addition, Figure B.6b shows results parallel to those in Figure B.6a, but with the $1/N$ rule as the unconstrained MVO portfolio. The realized residual returns, especially those for the unconstrained portfolio, are more stable over time.

In both cases, the information component contributes negatively to realized returns before 2007, and positively in certain years after 2008. Overall, these components explain the difference in residual returns between the unconstrained MVO and constrained portfolios.

B.6.4 Excluding Sin Stocks and Energy Stocks

Data. 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.

We follow Hong and Kacperczyk (2009) in identifying sin stocks as those that belong to the alcohol group (SIC codes 2100–2199) and the tobacco group (SIC codes 2080–2085). 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.

Our universe of stocks contains those with valid CRSP returns and industry labels. 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 B.1 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 3,000 stocks each year, of which 5.0% to 6.7% firms are excluded each year because they are labeled as either sin stocks or energy stocks as of the previous year.

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

$$x_i = \begin{cases} 0, & \text{if stock } i \text{ belongs to sin stocks or energy stocks} \\ 1, & \text{otherwise.} \end{cases} \tag{B.46}$$

Figure B.7 shows the year-over-year cross-sectional correlations between the residual returns and the inclusion variable (B.46), which tend to be negative before 2010 and positive after 2011. This implies that sin stocks and energy stocks tend to 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 increases, sin stocks and energy stocks tend to deliver lower excess returns.

Portfolio construction. Exclusionary investing usually does not consider short positions, because otherwise excluded assets can arguably be shorted. Therefore, we consider long-only portfolios

Table B.1: 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 the previous 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	2,326	5.5	12.9	57.7	−93.1	−15.5	3.7	26.1	957.5
2002	2,498	5.0	2.8	39.0	−92.3	−17.2	3.6	18.6	545.7
2003	2,519	5.4	9.4	36.6	−77.3	−10.4	3.7	19.4	469.3
2004	3,064	5.5	6.0	35.7	74.2	−11.3	3.0	16.3	811.1
2005	3,367	5.9	3.7	31.5	−93.2	−13.6	−0.1	15.9	288.1
2006	3,597	6.2	6.4	30.4	−77.0	−9.9	3.3	16.9	318.1
2007	3,910	6.5	7.0	44.2	−90.9	−16.5	0.0	21.2	511.0
2008	4,089	6.6	−12.6	44.9	−98.7	−39.9	18.1	7.0	647.1
2009	3,473	6.6	14.7	49.9	−95.9	−14.1	7.3	34.1	531.8
2010	3,856	6.7	4.6	29.6	−90.4	−11.0	2.4	16.4	332.2
2011	4,065	6.7	−0.1	30.1	−97.8	−16.4	1.3	15.7	309.6
2012	3,852	6.7	2.4	26.9	−97.3	−10.5	1.8	13.1	407.6
2013	3,831	6.7	0.4	31.3	−96.0	−15.5	−2.0	11.2	456.9
2014	3,900	6.1	0.3	26.8	−94.8	−12.5	1.0	13.4	295.5
2015	3,769	5.7	−1.1	28.4	−96.3	−15.4	0.9	13.4	194.9
2016	3,520	5.4	2.2	27.9	−97.1	−10.3	0.6	12.6	268.1
2017	3,451	5.8	4.4	27.9	−100.0	7.5	4.4	14.5	312.7
2018	3,346	6.1	−2.7	25.8	−100.0	−14.0	−3.0	9.1	199.5
2019	3,124	5.6	6.6	26.8	−95.2	−5.6	7.4	17.9	259.8
2020	3,078	5.3	4.3	47.6	−98.4	−15.4	1.6	16.4	1,432.2

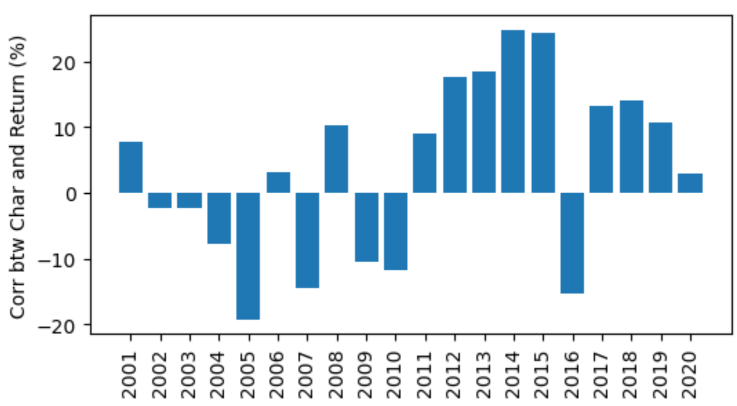


Figure B.7: Cross-sectional correlations between asset returns and the inclusion variable defined in (B.46) each year.

by solving the following problem each year.

$$
\begin{aligned}
\max_{\boldsymbol{\omega}} \quad & \boldsymbol{\omega}'\boldsymbol{\mu} - \frac{\gamma}{2}\boldsymbol{\omega}'\boldsymbol{\Sigma}\boldsymbol{\omega} \\
\text{s.t.} \quad & \boldsymbol{\omega}'\mathbf{1} = 1 \\
& \omega_i = 0 \quad \text{if } x_i = 0 \\
& \boldsymbol{\omega} \geq \mathbf{0}.
\end{aligned} \tag{B.47}
$$

We again use the return estimates in (33) and (34) to derive the predictive moments for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. We set $\gamma = 5$.

Expected return and utility decomposition. For Jorion's (1986) rule, Figure B.8 shows the decomposition of the expected utility and expected return of the portfolio into different components.

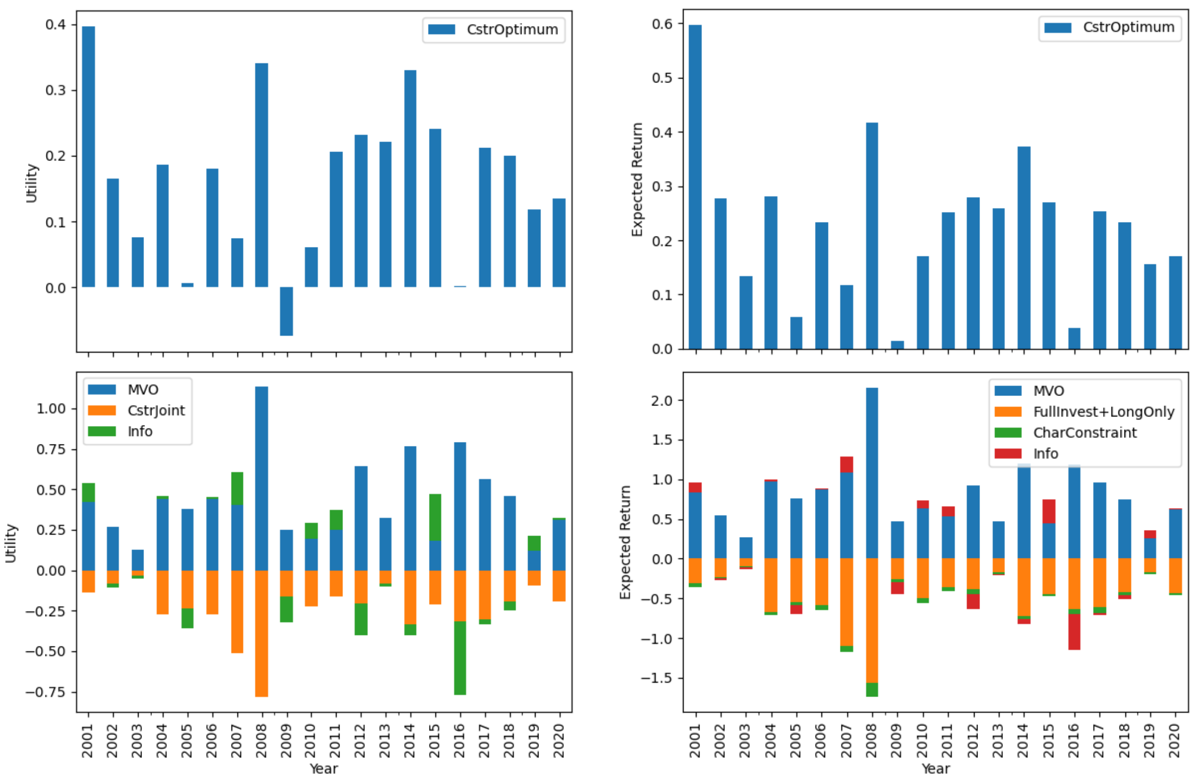
The upper panel of Figure B.8a shows that the expected utility of the optimal portfolio is positive through our 20-year sample except in 2009. This utility is decomposed into three components in the lower panel using (18) in Proposition 4 and its predictive return version in Appendix B.2.2. The expected utility of the unconstrained MVO portfolio (blue) is always positive, while the expected utility contribution of the three constraints (orange), treated as static, is always negative. The expected utility contribution from information contained in the constraints (green), however, varies over time. The magnitude of contribution from information is generally smaller in this case compared to the ESG constraints in Section 5.2 because the impact from exclusionary investing of sin stocks and stranded assets in our example is smaller due to the low percentage of excluded firms (see Table B.1).

Figure B.8b shows the expected return of the optimal portfolio and its decomposition based on (17) in Proposition 4 and its predictive return version in Appendix B.2.2. The two constraints (orange and green) contribute negatively to expected returns. The expected return contribution from information (red) is positive in certain years. 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 in Figure B.9, in which we compare two cases where the unconstrained MVO portfolio is Jorion's (1986) rule (Figure B.9a) and the $1/N$ rule (Figure B.9b).

The upper panel of Figure B.9a shows the realized residual returns for the constrained portfolio, which are decomposed into several components based on Proposition B.6 in Appendix B.4 in the lower panel. The contribution from the full investment and the long-only constraints (orange) is generally negative except for 2008, 2009, and 2019. The exclusionary investment constraint (green), treated as static, also contributes either positively or negatively over the 20-year period.

In addition, Figure B.9b shows results parallel to those in Figure B.9a, but with the $1/N$ rule as the unconstrained MVO portfolio. As expected, the impact from the full investment and the

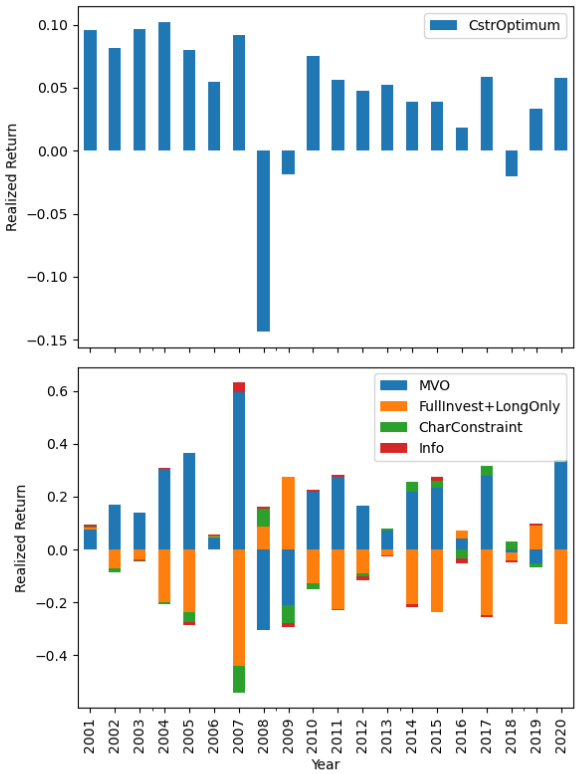


(a) Expected Utility

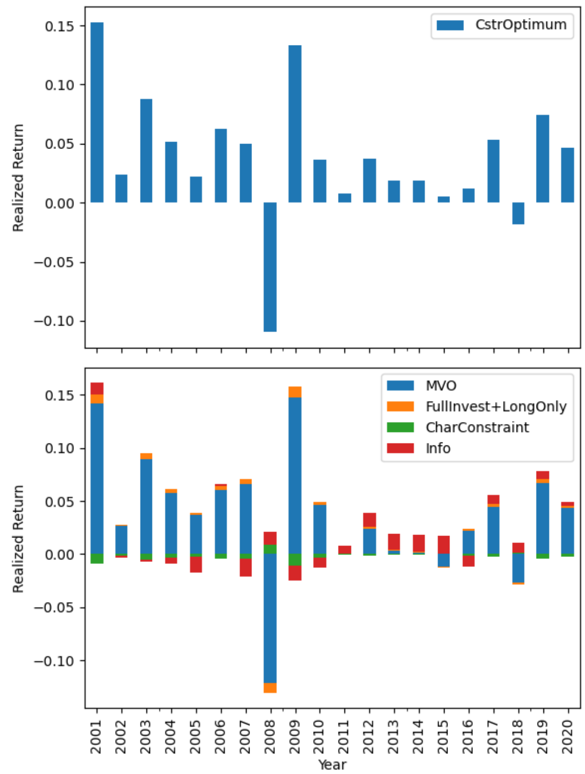
(b) Expected Return

Figure B.8: Expected return and utility and their decomposition, for the long-only portfolio defined in (B.47) with an exclusionary constraint based on the inclusion variable defined in (B.46) and Jorion's (1986) estimates of predictive moments. 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).

long-only constraints (orange) is much smaller. The contribution from information is positive in 2001, 2008, and in most years after 2011. However, the magnitude is insufficient to compensate for the large drawdown in 2008.



(a) Jorion's (1986) rule



(b) $1/N$ rule

Figure B.9: Realized return for the long-only portfolio defined in (B.47) with an exclusionary constraint based on the inclusion variable defined in (B.46). (a) corresponds to Jorion's (1986) estimates of predictive moments in (33), and (b) corresponds to predictive moments consistent with the $1/N$ rule in (34). In each subfigure, the top panel shows the realized return in excess of the Fama–French five-factor model of the constrained portfolio and the bottom panel shows its decomposition into components corresponding to the unconstrained MVO portfolio (blue), the full investment and long-only constraints combined together (orange), the exclusionary constraint treated as static (green), and information from the exclusionary constraint (red).

References

Avramov, D., and G. Zhou, 2010, Bayesian portfolio analysis, *Annual Review of Financial Economics* 2, 25–47.

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

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.

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

Brown, S. J., 1978, The portfolio choice problem: Comparison of certainty equivalence and optimal bayes portfolios, *Communications in Statistics-Simulation and Computation* 7, 321–334.

DeMiguel, V., L. Garlappi, F. J. Nogales, and R. Uppal, 2009, A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms, *Management Science* 55, 798–812.

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

Ding, P., 2016, On the conditional distribution of the multivariate t distribution, *The American Statistician* 70, 293–295.

Fan, J., Y. Li, and K. Yu, 2012, Vast volatility matrix estimation using high-frequency data for portfolio selection, *Journal of the American Statistical Association* 107, 412–428.

Fan, J., J. Zhang, and K. Yu, 2012, Vast portfolio selection with gross-exposure constraints, *Journal of the American Statistical Association* 107, 592–606.

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

Frost, P. A., and J. E. Savarino, 1986, An empirical bayes approach to efficient portfolio selection, *Journal of Financial and Quantitative Analysis* 21, 293–305.

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

Jacquier, E., and N. Polson, 2011, Bayesian methods in finance, in *The Oxford Handbook of Bayesian Econometrics*, chapter 9, 439–512 (Oxford University Press).

Jagannathan, R., and T. Ma, 2003, Risk reduction in large portfolios: Why imposing the wrong constraints helps, *The Journal of Finance* 58, 1651–1683.

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

Kan, R., and X. Wang, 2023, Optimal portfolio choice with unknown benchmark efficiency, *Management Science* .

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

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

Klein, R. W., and V. S. Bawa, 1976, The effect of estimation risk on optimal portfolio choice, *Journal of Financial Economics* 3, 215–231.

Kolm, P., and G. Ritter, 2017, On the bayesian interpretation of black–litterman, *European Journal of Operational Research* 258, 564–572.

Lai, T. L., H. Xing, and Z. Chen, 2011, Mean–variance portfolio optimization when means and covariances are unknown, *The Annals of Applied Statistics* 5, 798 – 823.

Ledoit, O., and M. Wolf, 2003, Improved estimation of the covariance matrix of stock returns with an application to portfolio selection, *Journal of Empirical Finance* 10, 603–621.

Ledoit, O., and M. Wolf, 2004, A well-conditioned estimator for large-dimensional covariance matrices, *Journal of Multivariate Analysis* 88, 365–411.

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.

Pástor, L., 2000, Portfolio selection and asset pricing models, *The Journal of Finance* 55, 179–223.

Pástor, L., and R. F. Stambaugh, 2000, Comparing asset pricing models: an investment perspective, *Journal of Financial Economics* 56, 335–381.

Stambaugh, R. F., 1997, Analyzing investments whose histories differ in length, *Journal of Financial Economics* 45, 285–331.

Tu, J., and G. Zhou, 2010, Incorporating economic objectives into bayesian priors: Portfolio choice under parameter uncertainty, *Journal of Financial and Quantitative Analysis* 45, 959–986.

Tu, J., and G. Zhou, 2011, Markowitz meets talmud: A combination of sophisticated and naive diversification strategies, *Journal of Financial Economics* 99, 204–215.

Zellner, A., 1971, *An Introduction to Bayesian Inference in Econometrics* (John Wiley & Sons, New York).

Zellner, A., and V. K. Chetty, 1965, Prediction and decision problems in regression models from the bayesian point of view, *Journal of the American Statistical Association* 60, 608–616.

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