Supplementary Materials

This electronic companion contains additional technical details and formal proofs of all results.

EC.1. Technical Appendix

EC.1.1. Combining Impact and Passive Portfolios

For impact investors who care about both passive and active returns, the *raw* return of the impact portfolio, instead of the residual return, needs to be taken into consideration. This appendix establishes a connection between the optimal weights derived from optimizing the residual return and those derived from optimizing the raw return. We accomplish this by briefly reviewing the portfolio theory and applying it specifically to the context of impact investing.

Under the linear multi-factor model in (1), one can define the expectation and covariance matrix of the *raw* returns of the N ranked assets as:

$$\tilde{\boldsymbol{\mu}} \equiv B\boldsymbol{\mu}_\Lambda + \boldsymbol{\mu}, \quad \tilde{\Sigma} \equiv B\Sigma_\Lambda B^\top + \Sigma, \tag{EC.1}$$

respectively, where $\boldsymbol{\mu}_\Lambda$ and Σ_Λ are the expectation and covariance matrix of $(\Lambda_1, \Lambda_2, \ldots, \Lambda_K)^\top$, B is an $N \times K$ matrix whose (i,k)-th entry is the beta of the i-th ranked asset on the k-th factor, and $\boldsymbol{\mu}$ and Σ are defined as in (7). Then the expectation and variance of the portfolio's raw return, r_p, are:

$$\mathbb{E}(r_p) = \boldsymbol{w}^\top \tilde{\boldsymbol{\mu}}, \quad \mathrm{Var}(r_p) = \boldsymbol{w}^\top \tilde{\Sigma} \boldsymbol{w}. \tag{EC.2}$$

As in Proposition 1, for impact investors who optimize the Sharpe ratio or the mean–variance utility of raw return, the following proposition characterizes the corresponding optimal portfolios.

Proposition EC.1. *Under the multi-factor model of (1), if investors construct portfolios based on N assets with frictionless borrowing and lending at the risk-free rate, and they maximize the Sharpe ratio of raw returns, $\mathrm{SR} = \mathbb{E}(r_p)/\sqrt{\mathrm{Var}(r_p)}$, or the mean–variance utility of raw returns, $\mathbb{E}(r_p) - 0.5\lambda\mathrm{Var}(r_p)$, with a constant risk-aversion parameter $\lambda > 0$, the optimal portfolio weights and the optimal Sharpe ratio are given by:*

$$\boldsymbol{w}^* \propto \tilde{\Sigma}^{-1}\tilde{\boldsymbol{\mu}}, \quad \textit{and} \quad \mathrm{SR}^* = \sqrt{\tilde{\boldsymbol{\mu}}^\top \tilde{\Sigma}^{-1} \tilde{\boldsymbol{\mu}}}. \tag{EC.3}$$

Proposition 1 gives the optimal weights for investors who only care about active returns, and Proposition EC.1 gives the optimal weights for investors who optimize for the overall risk-adjusted returns. Both sets of weights have the same mathematical form, albeit with different covariance matrices and expectation vectors. In fact, the two sets of weights are closely related, as demonstrated in the following results under a single-factor model setting.

PROPOSITION EC.2. *Consider the single-factor model:*

$$r_i = \beta_i r_M + \theta_i, \quad i = 1, 2, \ldots, N, \tag{EC.4}$$

where r_M is the return (beyond the risk-free rate) of a single factor (e.g., the market portfolio). The optimal weight of maximizing the Sharpe ratio of raw returns, $\mathrm{SR} = \mathbb{E}(r_p)/\sqrt{\mathrm{Var}(r_p)}$, or the mean–variance utility of raw returns, $\mathbb{E}(r_p) - 0.5\lambda\mathrm{Var}(r_p)$, with a constant risk-aversion parameter $\lambda > 0$, is:

$$\boldsymbol{w}^* \propto \tilde{\Sigma}^{-1}\tilde{\boldsymbol{\mu}} = \underbrace{\frac{\mu_M - \sigma_M^2 \boldsymbol{\beta}^\top \Sigma^{-1}\boldsymbol{\mu}}{1 + \sigma_M^2 \boldsymbol{\beta}^\top \Sigma^{-1}\boldsymbol{\beta}} \Sigma^{-1}\boldsymbol{\beta}}_{\textit{passive component}} + \underbrace{\Sigma^{-1}\boldsymbol{\mu}}_{\textit{active component}},$$

where $\mu_M \equiv \mathbb{E}(r_M)$, $\sigma_M^2 \equiv \mathrm{Var}(r_M)$, and $\boldsymbol{\beta}$ is the vector of β_i ranked by the impact factor $\boldsymbol{X}$. In addition, the squared optimal Sharpe ratio is:

$$\mathrm{SR}^{*2} = \tilde{\boldsymbol{\mu}}^\top \tilde{\Sigma}^{-1}\tilde{\boldsymbol{\mu}} = \underbrace{\frac{\mu_M^2 \boldsymbol{\beta}^\top \Sigma^{-1}\boldsymbol{\beta} + 2\mu_M \boldsymbol{\beta}^\top \Sigma^{-1}\boldsymbol{\mu} - \sigma_M^2 (\boldsymbol{\beta}^\top \Sigma^{-1}\boldsymbol{\mu})^2}{1 + \sigma_M^2 \boldsymbol{\beta}^\top \Sigma^{-1}\boldsymbol{\beta}}}_{\textit{passive component}} + \underbrace{\boldsymbol{\mu}^\top \Sigma^{-1}\boldsymbol{\mu}}_{\textit{active component}}.$$

Proposition EC.2 demonstrates that, under the single-factor framework, the optimal weights for maximizing the Sharpe ratio can be regarded as the optimal weights of maximizing the information ratio ($\Sigma^{-1}\boldsymbol{\mu}$) plus a passive component which is proportional to $\Sigma^{-1}\boldsymbol{\beta}$. The corresponding optimal Sharpe ratio can also be decomposed into the optimal information ratio and a passive component.

So far, we have characterized the optimal impact portfolios constructed based on the impact factor, $\boldsymbol{X}$. Moreover, one can combine the impact portfolio with any other portfolio. For example, we can add the impact portfolio to the suite of portfolios mimicking more traditional asset pricing factors such as value, size, and momentum. However, perhaps the most natural application is to combine the impact portfolio with a passive index fund such as the market portfolio, which we demonstrate under the single factor model, (EC.4). In particular, under the single factor model, we can define the expectation and covariance matrix of the $N+1$ assets (including N ranked assets and the market portfolio, r_M) as:

$$\hat{\boldsymbol{\mu}} \equiv \begin{pmatrix} \tilde{\boldsymbol{\mu}} \\ \mu_M \end{pmatrix} = \begin{pmatrix} \mu_M \boldsymbol{\beta} + \boldsymbol{\mu} \\ \mu_M \end{pmatrix}, \quad \hat{\Sigma} \equiv \begin{pmatrix} \tilde{\Sigma} & \sigma_M^2 \boldsymbol{\beta} \\ \sigma_M^2 \boldsymbol{\beta}^\top & \sigma_M^2 \end{pmatrix} = \begin{pmatrix} \sigma_M^2 \boldsymbol{\beta}\boldsymbol{\beta}^\top + \Sigma & \sigma_M^2 \boldsymbol{\beta} \\ \sigma_M^2 \boldsymbol{\beta}^\top & \sigma_M^2 \end{pmatrix}, \tag{EC.5}$$

where $\mu_M \equiv \mathbb{E}(r_M)$, $\sigma_M^2 \equiv \mathrm{Var}(r_M)$, $\boldsymbol{\beta}$ is the vector of β_i ranked by the impact factor $\boldsymbol{X}$, and $\boldsymbol{\mu}$ and Σ are defined as in (7). Denote by w_M the weight on the market portfolio, and $\hat{\boldsymbol{w}} \equiv \begin{pmatrix} \boldsymbol{w} \\ w_M \end{pmatrix}$ the weights on the $N+1$ assets. Then, under the single factor model, (EC.4), the expected value and variance of the return of the combined portfolio, $\hat{r}_p$, are:

$$\mathbb{E}(\hat{r}_p) = \hat{\boldsymbol{w}}^\top \hat{\boldsymbol{\mu}}, \quad \mathrm{Var}(\hat{r}_p) = \hat{\boldsymbol{w}}^\top \hat{\Sigma} \hat{\boldsymbol{w}}. \tag{EC.6}$$

The following proposition characterizes the optimal combined portfolio.

Proposition EC.3. *Under the single-factor model (EC.4) with frictionless borrowing and lending at the risk-free rate, if investors maximize the combined portfolio's Sharpe ratio of raw returns,* $\mathrm{SR} = \mathbb{E}(\hat{r}_p)/\sqrt{\mathrm{Var}(\hat{r}_p)}$*, or the mean–variance utility of raw returns,* $\mathbb{E}(\hat{r}_p) - 0.5\lambda \mathrm{Var}(\hat{r}_p)$*, with a constant risk-aversion parameter* $\lambda > 0$*, the optimal weights of the* N *assets and the market portfolio, and the squared optimal Sharpe ratio are given by:*

$$\hat{\boldsymbol{w}}^* = \begin{pmatrix} \boldsymbol{w}^* \\ w_M^* \end{pmatrix} \propto \hat{\Sigma}^{-1}\hat{\boldsymbol{\mu}} = \begin{pmatrix} \Sigma^{-1}\boldsymbol{\mu} \\ \frac{\mu_M}{\sigma_M^2} - \boldsymbol{\beta}^\top \Sigma^{-1}\boldsymbol{\mu} \end{pmatrix}, \quad \mathrm{SR}^{*2} = \underbrace{\frac{\mu_M^2}{\sigma_M^2}}_{\textit{passive component}} + \underbrace{\boldsymbol{\mu}^\top \Sigma^{-1}\boldsymbol{\mu}}_{\textit{active component}}.$$

When maximizing the Sharpe ratio of the combined portfolio, the optimal weights on the N assets, $\boldsymbol{w}^*$, are still proportional to $\Sigma^{-1}\boldsymbol{\mu}$. This is the same as the result of Proposition 1, which maximizes the information ratio of residual return. In other words, in the single-factor world with $N+1$ assets, maximizing the Sharpe ratio is equivalent to maximizing the information ratio. In addition, the optimal Sharpe ratio can be decomposed into the optimal information ratio (the active component) and the Sharpe ratio of the market portfolio (the passive component). Proposition EC.3 reduces to the special case of Treynor and Black (1973) if Σ is diagonal.

EC.1.2. Gaussian Optimal Impact Portfolios

In this appendix, we consider the case where the impact factors and residual returns are jointly normally distributed, i.e., the F in (3) is a bivariate normal distribution, and explicitly construct optimal impact portfolios and analyze their performance metrics. In particular, we assume that a special case of Assumption 1 holds:

Assumption EC.1 **(Bivariate Normality).** *The impact factors, X_i, and the residual returns, θ_i, $i=1,2,\dots,N$, are drawn from the following bivariate normal distribution:*

$$\begin{pmatrix} X_1 \\ \theta_1 \end{pmatrix}, \begin{pmatrix} X_2 \\ \theta_2 \end{pmatrix}, \dots, \begin{pmatrix} X_N \\ \theta_N \end{pmatrix} \overset{\text{IID}}{\sim} \mathcal{N}\left(\begin{pmatrix} \mu_X \\ \mu_\theta \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_\theta \\ \rho\sigma_X\sigma_\theta & \sigma_\theta^2 \end{pmatrix} \right),$$

where μ_X, μ_θ and $\sigma_X > 0, \sigma_\theta > 0$ are the expectations and standard deviations of X_i and θ_i, respectively, and $\rho \in (-1,1)$ is the correlation between X_i and θ_i.

Assumption EC.1 requires that $(X_i,\theta_i)^\top, i=1,2,\dots,N$ are IID jointly normally distributed. We relax the assumption of independence in Appendix EC.1.2.4.

EC.1.2.1. Moments of Impact Returns The following proposition characterizes the distribution of $\boldsymbol{\theta}_{[\boldsymbol{X}]}$ using the representation theorem, Theorem 1.

Proposition EC.4. *Under Assumption EC.1, we have:*

$$(\theta_{[1:N]},\dots,\theta_{[N:N]}) \overset{d}{=} \left(\sigma_\theta \left(\rho Y_{1:N} + \sqrt{1-\rho^2} Z_1 \right) + \mu_\theta, \dots, \sigma_\theta \left(\rho Y_{N:N} + \sqrt{1-\rho^2} Z_N \right) + \mu_\theta \right), \quad \text{(EC.7)}$$

where $Z_1, Z_2, \dots, Z_N \overset{\text{IID}}{\sim} \mathcal{N}(0,1)$, $Y_{1:N} \le Y_{2:N} \le \cdots \le Y_{N:N}$ are the order statistics of $Y_1, Y_2, \dots, Y_N \overset{\text{IID}}{\sim} \mathcal{N}(0,1)$, and random variables $\{Y_i\}_{i=1}^N$ and $\{Z_i\}_{i=1}^N$ are mutually independent. In particular, for $i,j=1,2,\dots,N$, and $i \neq j$, we have:

$$\mathbb{E}(\theta_{[i:N]}) = \sigma_\theta \cdot \rho \cdot \mathbb{E}(Y_{i:N}) + \mu_\theta, \quad \text{(EC.8)}$$

$$\mathrm{Var}(\theta_{[i:N]}) = \sigma_\theta^2 \cdot (1-\rho^2+\rho^2 \cdot \mathrm{Var}(Y_{i:N})), \quad \text{(EC.9)}$$

$$\mathrm{Cov}(\theta_{[i:N]}, \theta_{[j:N]}) = \sigma_\theta^2 \cdot \rho^2 \cdot \mathrm{Cov}(Y_{i:N}, Y_{j:N}). \quad \text{(EC.10)}$$

Equation (EC.7) shows that, under the joint normality assumption, the representation given by Theorem 1 reduces to a *linear* combination of order statistics of normal random variables, $Y_{i:N}$, and IID normal noise, Z_i. The weights of the two components are ρ and $\sqrt{1-\rho^2}$ respectively, which implies that a stronger dependence between the impact and residual returns (a higher $|\rho|$) leads to a larger weight on the order statistic component, $Y_{i:N}$, and a weaker dependence (a lower $|\rho|$) leads to a larger weight on the noise component, Z_i. This linear representation allows us to straightforwardly analyze the moments of induced order statistics.

Equation (EC.8) represents the expectation of $\theta_{[i:N]}$ in excess of the cross-sectional average residual return μ_θ as the product of three terms: the cross-sectional standard deviation of residual returns, σ_θ, the correlation between residual returns and impact factors, ρ, and a score representing the impact of the i-th asset relative to other assets, $\mathbb{E}(Y_{i:N})$. This is consistent with Grinold's (1994) insight that the alpha of active portfolio management equals "volatility times information coefficient (IC) times score," where the information coefficient represents the correlation between the active investment factor and the active return.

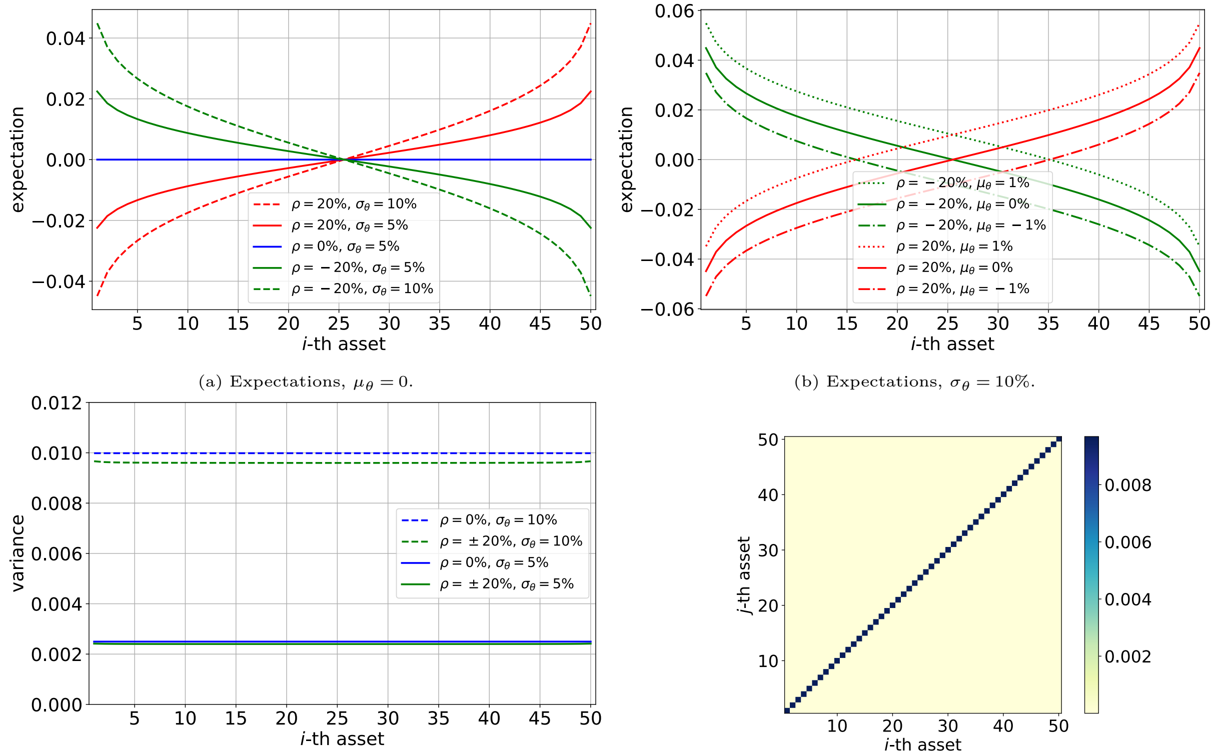
Equations (EC.9) and (EC.10) give the variances and covariances of induced order statistics. The variance of $\theta_{[i:N]}$ is proportional to $1-\rho^2+\rho^2\cdot\mathrm{Var}(Y_{i:N})$ and σ_θ^2. The correlation ρ determines the influence of the variance of order statistic, $\mathrm{Var}(Y_{i:N})$, on the variance of $\theta_{[i:N]}$. When $\rho=0$, the variance reduces to σ_θ^2 for any i. The covariance between $\theta_{[i:N]}$ and $\theta_{[j:N]}$ is proportional to $\mathrm{Cov}(Y_{i:N}, Y_{j:N})$ and $\sigma_\theta^2\cdot\rho^2$.

The moments of induced order statistics in Proposition EC.4 depend on $\mathbb{E}(Y_{i:N})$, $\mathrm{Var}(Y_{i:N})$, and $\mathrm{Cov}(Y_{i:N}, Y_{j:N})$, which do not have explicit expressions. Therefore we use Figure EC.1 to illustrate the results in Proposition EC.4 using a numerical example with 50 assets. Figure EC.1a and Figure EC.1b show the expected residual returns. When the correlation between the impact factor and residual return is positive ($\rho>0$), the expected value increases as the order i increases; when the correlation is negative ($\rho<0$), the expected value decreases as the order i increases. We prove a generalization of this result in Section EC.1.4.1. The dispersion in expected residual returns across assets is larger when σ_θ is larger. Different values of μ_θ lead to parallel shifts in the expected residual returns.

Figure EC.1c shows the variances of residual returns, which are primarily determined by the cross-sectional standard deviation of residual returns (σ_θ). In addition, stronger signals in the impact factor (as measured by higher values in $|\rho|$) lead to lower variances in residual returns. It is also worth noting that the variance as a function of order i is quite flat. Figure EC.1d shows the covariance matrix of the residual returns across all assets. The covariances between the residual returns of different assets, after ranking by the impact factor, are very close to zero.

There are two notable symmetries in the results of Proposition EC.4 and Figure EC.1. First, all quantities are symmetric for correlation coefficients with opposite signs (ρ and $-\rho$). Second, these quantities are also symmetric with respect to the median asset. We show that both of these symmetries hold under more general distributional conditions, in Appendices EC.1.4.4 and EC.1.5.2, respectively.

The following proposition provides the asymptotic distribution of $\boldsymbol{\theta}_{[\boldsymbol{X}]}$ under the joint normality assumption, Assumption EC.1. This is a corollary of the asymptotic version of the representation theorem, Theorem 3.



(a) Expectations, $\mu_\theta = 0$.

(b) Expectations, $\sigma_\theta = 10\%$.

(c) Variances, $\mu_\theta = 0$.

(d) Covariance matrix, $\rho = 20\%$, $\sigma_\theta = 10\%$.

Figure EC.1 Expectations, variances, and covariances of $\theta_{[1:N]}, \theta_{[2:N]}, \ldots, \theta_{[N:N]}$ under Assumption EC.1. We set $N = 50$ for illustrative purposes.

Proposition EC.5. *Under Assumption EC.1, we have:*

$$(\theta_{[i_1(N):N]}, \ldots, \theta_{[i_m(N):N]}) \xrightarrow{d} \left(\sigma_\theta\sqrt{1-\rho^2}Z_1 + \sigma_\theta \cdot \rho \cdot \Phi^{-1}(\xi_1) + \mu_\theta, \ldots, \sigma_\theta\sqrt{1-\rho^2}Z_m + \sigma_\theta \cdot \rho \cdot \Phi^{-1}(\xi_m) + \mu_\theta\right), \quad \text{(EC.11)}$$

where $Z_1, Z_2, \ldots, Z_N \overset{\text{IID}}{\sim} \mathcal{N}(0,1)$, Φ is the distribution function of $\mathcal{N}(0,1)$, and other notations are the same as in Theorem 3. In other words, $\theta_{[i_k(N):N]}$ converge in distribution to mutually independent normally distributed random variables with mean $\mu(\xi_k)$ and variance $\sigma^2(\xi_k)$, where

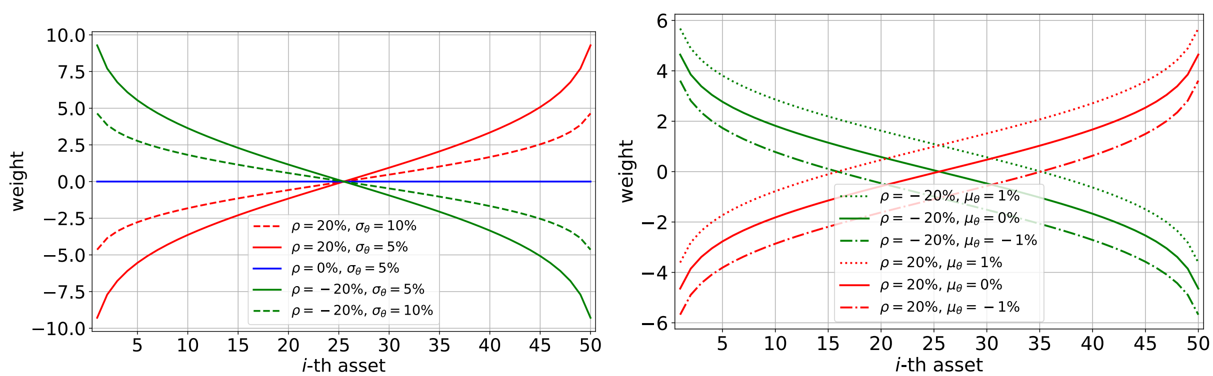
$$\mu(\xi_k) \equiv \sigma_\theta \cdot \rho \cdot \Phi^{-1}(\xi_k) + \mu_\theta, \quad \text{(EC.12)}$$

$$\sigma^2(\xi_k) \equiv \sigma_\theta^2 \cdot (1-\rho^2). \quad \text{(EC.13)}$$

The asymptotic result is consistent with the finite sample results. First, the asymptotic expectation of residual returns, $\mu(\xi_k)$, increases as the rank of asset, ξ_k, increases. Second, the asymptotic variance, $\sigma^2(\xi_k)$, is a constant that does not depend on ξ_k, which is consistent with our observation in Figure EC.1c. Third, the induced order statistics are asymptotically independent, which implies that the covariance matrix is approximately diagonal (Figure EC.1d).

EC.1.2.2. Characterization of Optimal Impact Portfolios Under Assumption EC.1, the explicit formulas for the moments of $\boldsymbol{\theta}_{[\boldsymbol{X}]}$ given by Proposition EC.4 allow us to investigate the weights and performance of optimal impact portfolios.

We begin by examining the optimal weights numerically. Figure EC.2 shows the optimal portfolio weights in terms of maximizing the information ratio. The weights are jointly determined by the expected residual returns and their variances, as in (9). When the correlation between the impact factor and residual return is positive ($\rho > 0$), assets with higher values of impact factor have higher weights; when the correlation is negative ($\rho < 0$), assets with higher values of impact factor have lower (negative) weights. The dispersion in optimal weights across assets decreases as σ_θ increases. In addition, different values of μ_θ lead to shifts in the optimal weights.



(a) Weights, $\mu_\theta = 0$.

(b) Weights, $\sigma_\theta = 10\%$.

Figure EC.2 Optimal impact portfolio weights under Assumption EC.1. We set $N = 50$ for illustrative purposes.

Now we study the optimal information ratio given by Proposition 1, which depends on the expected value and covariance matrix of residual returns after ranking based on the impact factor (see (9)). The explicit formulas for the moments of residual returns from Proposition EC.4 allow us to characterize the optimal information ratio.

Theorem EC.1 **(Optimal Performance).** *Under Assumption EC.1, if* $|\rho| \leq \sqrt{2}/2 \approx 70.71\%$, *the optimal information ratio of the impact portfolio,* IR^*, *as given in (9), satisfies the following bounds:*

$$\sqrt{\rho^2 q(N) + \frac{\mu_\theta^2}{\sigma_\theta^2} \cdot N} \leq \mathrm{IR}^* \leq \sqrt{\frac{\rho^2 q(N) + (\mu_\theta^2/\sigma_\theta^2) \cdot N}{1 - 2\rho^2}}, \tag{EC.14}$$

where

$$q(N) = \sum_{i=1}^{N} [\mathbb{E}(Y_{i:N})]^2, \tag{EC.15}$$

and $Y_{1:N} \leq Y_{2:N} \leq \cdots \leq Y_{N:N}$ *are the order statistics of* $Y_1, Y_2, \ldots, Y_N \overset{\mathrm{IID}}{\sim} \mathcal{N}(0,1)$.

Theorem EC.1 provides lower and upper bounds for IR^*. In practice, the correlation between the impact factor and residual returns, $|\rho|$, is usually small (Grinold 1989), which makes the bound in (EC.14) relatively tight. The technical condition $|\rho| \le \sqrt{2}/2$ not only ensures that the square root on the right-hand side of (EC.14) is well defined, but also aligns with Grinold's (1989) assumption that $|\rho|$ is small.

The following proposition shows a property of $q(N)$ defined in (EC.15):

Proposition EC.6. *For $q(N)$ defined in (EC.15) and $N \ge 1$, we have*

$$\lim_{N \to +\infty} \frac{q(N)}{N} = 1. \tag{EC.16}$$

Equation (EC.16) indicates that $q(N) \approx N$ when N is sufficiently large. Appendix EC.1.2.5 shows that this approximation is very good when N is greater than 100. Therefore, the bounds for IR^* given in (EC.14) can be approximated by:

$$\sqrt{\rho^2 + \mu_\theta^2/\sigma_\theta^2} \cdot \sqrt{N} \lesssim \text{IR}^* \lesssim \frac{\sqrt{\rho^2 + \mu_\theta^2/\sigma_\theta^2} \cdot \sqrt{N}}{\sqrt{1 - 2\rho^2}}. \tag{EC.17}$$

This approximation implies that the optimal information ratio of the impact portfolio is typically determined by three components: the correlation between the impact factor and residual returns, ρ, the inherent information ratio of the assets, μ_θ/σ_θ, and the number of assets included in the portfolio, N. In other words, impact investors can improve their portfolio performance by using impact factors with stronger signals (higher ρ), investing in more profitable assets (higher μ_θ/σ_θ), and including more assets in the portfolio (higher N).

When $\mu_\theta = 0$ and ρ is small, (EC.17) reduces to

$$\text{IR}^* \approx |\rho| \cdot \sqrt{N}. \tag{EC.18}$$

This is closely related to the characterization of optimal active investments, which Grinold (1989) terms the "Fundamental Law of Active Management (FLAM)."[23] The FLAM provides a simple approximation of the information ratio of an active portfolio using the product of the information coefficient (ρ in our notation) and the square root of the *breadth* of a strategy. In Grinold's (1989) framework, breadth is defined as the number of independent bets of a strategy in a given year. In our context, breadth is determined by the number of available assets in the universe.

Theorem EC.1 can be regarded as a generalization of the original FLAM in three aspects. First, the impact portfolio is constructed using only the *ranking* information of the impact factor, as opposed to the actual signal values in the FLAM. Second, the original FLAM is derived under the assumption that $|\rho|$ is small, while we provide bounds in (EC.14) for the optimal information ratio for the more general case of $|\rho| \leq \sqrt{2}/2$, which shrinks toward the original FLAM when $|\rho|$ is small. Third, we allow for nonzero expected residual returns, μ_θ, while the original FLAM assumes $\mu_\theta = 0$.

Theorem EC.1 characterizes the performance of an optimal impact portfolio that invests in all N assets in the universe. However, in reality, impact investors *may* face a tradeoff between the performance and the impact of their portfolios. If an investor desires a higher level of impact, or is restricted to investing in certain assets with low impact scores (also known as negative filtering), she may not be able to realize the optimal performance derived in Theorem EC.1. We discuss this performance–impact tradeoff in Appendix EC.1.2.6.

Furthermore, we generalize Theorem EC.1 in two ways. In Appendix EC.1.2.3, we extend our static framework to a dynamic one by analyzing the optimal impact portfolio performance with a time-varying ρ. In Appendix EC.1.2.4, we relax the IID assumption in Assumption EC.1 and allow for dependence between the N assets.

EC.1.2.3. Optimal Performance with Time-Varying Correlation Theorem EC.1 quantifies the performance of impact portfolios with respect to ρ, which it regards as a constant that measures the strength of the impact factor as a signal for residual returns. In practice, the strength of this signal can change over time, and we extend our static framework into a dynamic one by analyzing performance with a time-varying ρ in this appendix. We add the subscript t to the correlation, ρ_t, which is modeled as a random variable, and make the following assumption, which generalizes Assumption EC.1 to allow for time-varying correlations.

Assumption EC.2 **(Time-Varying Correlation)**. *The correlations, ρ_t, between impact factors, X_{it}, and residual returns, θ_{it}, over time $t = 1, 2, \ldots, T$, are IID random variables with a mean of $\overline{\rho}$ and a variance of σ_ρ^2. Given time t, X_{it} and θ_{it} are drawn from the following bivariate normal distribution across $i = 1, 2, \ldots, N$:*

$$\begin{pmatrix} X_{1t} \\ \theta_{1t} \end{pmatrix}, \begin{pmatrix} X_{2t} \\ \theta_{2t} \end{pmatrix}, \ldots, \begin{pmatrix} X_{Nt} \\ \theta_{Nt} \end{pmatrix} \Bigg| \; \rho_t \overset{\text{IID}}{\sim} \mathcal{N}\left(\begin{pmatrix} \mu_X \\ \mu_\theta \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \rho_t \cdot \sigma_X \cdot \sigma_\theta \\ \rho_t \cdot \sigma_X \cdot \sigma_\theta & \sigma_\theta^2 \end{pmatrix} \right),$$

where μ_X, μ_θ and $\sigma_X > 0, \sigma_\theta > 0$ are the expectations and standard deviations of X_{it} and θ_{it}, respectively.

The following result characterizes the optimal performance of the impact portfolio when correlations, ρ_t, vary over time. To derive this result, we need to first characterize the moments of the residual returns under Assumption EC.2. These details are given by Lemma EC.8 in Appendix EC.3.1.

Theorem EC.2 **(Optimal Performance with Time-Varying Correlation)**. *Under Assumption EC.2, if* $\sqrt{\overline{\rho}^2+\sigma_\rho^2} \le \sqrt{2}/2 \approx 70.71\%$*, the optimal information ratio of the impact portfolio,* $\text{IR}^*_{\text{TimeVary}}$*, as given in (9), will satisfy the following bounds:*

$$\sqrt{\frac{\overline{\rho}^2}{1/q(N)+\sigma_\rho^2}+\frac{\mu_\theta^2 N}{\sigma_\theta^2}} \le \text{IR}^*_{\text{TimeVary}} \le \sqrt{\frac{\overline{\rho}^2}{[1-2\cdot(\overline{\rho}^2+\sigma_\rho^2)]/q(N)+\sigma_\rho^2}+\frac{\mu_\theta^2 N}{\sigma_\theta^2\left[1-2\cdot(\overline{\rho}^2+\sigma_\rho^2)\right]}}, \tag{EC.19}$$

where $q(N)$ *is defined in (EC.15).*

Using Proposition EC.6, when N is sufficiently large, the bounds for $\text{IR}^*_{\text{TimeVary}}$ given in (EC.19) can be approximated by:

$$\sqrt{\frac{\overline{\rho}^2}{1/N+\sigma_\rho^2}+\frac{\mu_\theta^2 N}{\sigma_\theta^2}} \lesssim \text{IR}^*_{\text{TimeVary}} \lesssim \sqrt{\frac{\overline{\rho}^2}{[1-2\cdot(\overline{\rho}^2+\sigma_\rho^2)]/N+\sigma_\rho^2}+\frac{\mu_\theta^2 N}{\sigma_\theta^2\left[1-2\cdot(\overline{\rho}^2+\sigma_\rho^2)\right]}}. \tag{EC.20}$$

In particular, when $\overline{\rho}$ and σ_ρ are relatively small and $\mu_\theta = 0$, this approximation further reduces to:

$$\text{IR}^*_{\text{TimeVary}} \approx \frac{|\overline{\rho}|}{\sqrt{1/N+\sigma_\rho^2}}. \tag{EC.21}$$

Comparing this with the static case given by (EC.18), we have:

$$\text{IR}^*_{\text{TimeVary}} \approx \frac{|\overline{\rho}|}{\sqrt{1/N+\sigma_\rho^2}} < |\overline{\rho}|\cdot\sqrt{N} \approx \text{IR}^*,$$

as long as ρ_t is not a constant ($\sigma_\rho^2 > 0$). In other words, the variability in ρ_t leads to a lower information ratio than that from a constant ρ. When $\sigma_\rho^2 = 0$, Theorem EC.2 reduces to Theorem EC.1.

These results are also closely related to the literature on active portfolio management. The approximation given by (EC.21) is consistent with Ding and Martin (2017). Furthermore, when N increases without bound, (EC.21) converges to:

$$\text{IR}^*_{\text{TimeVary}} \approx \frac{|\overline{\rho}|}{\sigma_\rho},$$

which is the average correlation divided by its standard deviation. This is consistent with the approximation first proposed by Qian and Hua (2006). Our results show that these approximations hold under weaker conditions where investors only use the rank information—not the value—of the underlying signal (in our case, the impact factor).

EC.1.2.4. Optimal Performance with Cross-Sectional Dependence The optimal performance of impact portfolios given by Theorem EC.1 holds under Assumption EC.1, which requires that $(X_i, \theta_i)^\top, i = 1, 2, \ldots, N$ are drawn IID from a bivariate normal distribution. However, in practice, the impact factors, X_i, and the residual returns, θ_i, for different assets may be correlated, especially for firms in the same industry. In addition, the impact factor of asset i, X_i, may be correlated with not only the residual return of asset i, θ_i, but also that of other assets, θ_j, where $j \neq i$. Therefore, in this appendix, we consider the case when $(X_i, \theta_i)^\top$ is not IID, and make the following assumption, which is a generalization of Assumption EC.1 from a cross-sectional dependence perspective:

Assumption EC.3 **(Cross-Sectional Dependence)**. *The impact factors, X_i, and the residual returns, θ_i, $i = 1, 2, \ldots, N$, follow the following bivariate normal distribution:*

$$\begin{pmatrix} X_1 \\ \theta_1 \end{pmatrix}, \begin{pmatrix} X_2 \\ \theta_2 \end{pmatrix}, \ldots, \begin{pmatrix} X_N \\ \theta_N \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_X \\ \mu_\theta \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_\theta \\ \rho\sigma_X\sigma_\theta & \sigma_\theta^2 \end{pmatrix} \right),$$

where μ_X, μ_θ and $\sigma_X > 0, \sigma_\theta > 0$ are the expectations and standard deviations of X_i and θ_i, respectively. In addition, for $i \neq j$, the equicorrelated cross-sectional dependence is described by $\rho_X \equiv \mathrm{corr}(X_i, X_j)$, $\rho_\theta \equiv \mathrm{corr}(\theta_i, \theta_j)$, $\tilde{\rho} \equiv \mathrm{corr}(X_i, \theta_j)$, and $\rho \equiv \mathrm{corr}(X_i, \theta_i)$.

The following result characterizes the optimal performance of the impact portfolio with cross-sectional dependence in $(X_i, \theta_i)^\top$. The moments of the residual returns under Assumption EC.3 are given by Lemma EC.9 in Appendix EC.3.1.

Theorem EC.3 **(Optimal Performance with Cross-Sectional Dependence)**. *Under Assumption EC.3, if*

$$(1-\rho_\theta)(1-\rho_X) - 2(\rho-\tilde{\rho})^2 \geq 0, \tag{EC.22}$$

the optimal information ratio of the impact portfolio, $\mathrm{IR}^*_{\mathrm{Dependence}}$, as given in (9), will satisfy the following bounds:

$$\begin{aligned} &\sqrt{\frac{(\rho-\tilde{\rho})^2 q(N)}{(1-\rho_\theta)(1-\rho_X)} + \frac{\mu_\theta^2/\sigma_\theta^2}{(1-\rho_\theta)/N + \rho_\theta}} \leq \mathrm{IR}^*_{\mathrm{Dependence}} \\ &\leq \sqrt{\frac{(\rho-\tilde{\rho})^2 q(N)}{(1-\rho_\theta)(1-\rho_X) - 2(\rho-\tilde{\rho})^2} + \frac{\mu_\theta^2/\sigma_\theta^2}{[(1-\rho_\theta)(1-\rho_X) - 2(\rho-\tilde{\rho})^2]/[(1-\rho_X)N] + \rho_\theta}}, \end{aligned} \tag{EC.23}$$

where $q(N)$ is defined in (EC.15).

Using Proposition EC.6, when N is sufficiently large, the bounds for $\mathrm{IR}^*_{\mathrm{Dependence}}$ given in (EC.19) can be approximated by:

$$\begin{aligned} &\sqrt{\frac{(\rho-\tilde{\rho})^2 N}{(1-\rho_\theta)(1-\rho_X)} + \frac{\mu_\theta^2/\sigma_\theta^2}{(1-\rho_\theta)/N + \rho_\theta}} \lesssim \mathrm{IR}^*_{\mathrm{Dependence}} \\ &\lesssim \sqrt{\frac{(\rho-\tilde{\rho})^2 N}{(1-\rho_\theta)(1-\rho_X) - 2(\rho-\tilde{\rho})^2} + \frac{\mu_\theta^2/\sigma_\theta^2}{[(1-\rho_\theta)(1-\rho_X) - 2(\rho-\tilde{\rho})^2]/[(1-\rho_X)N] + \rho_\theta}}. \end{aligned}$$

In particular, when $\mu_\theta = 0$, this further reduces to

$$\frac{|\rho-\tilde{\rho}|}{\sqrt{(1-\rho_\theta)(1-\rho_X)}} \cdot \sqrt{N} \lesssim \mathrm{IR}^*_{\mathrm{Dependence}} \lesssim \frac{|\rho-\tilde{\rho}|}{\sqrt{(1-\rho_\theta)(1-\rho_X) - 2(\rho-\tilde{\rho})^2}} \cdot \sqrt{N}. \tag{EC.24}$$

With cross-sectional dependence specified in Assumption EC.3, in (EC.24), taking the lower bound as an example, the term $\frac{|\rho-\tilde{\rho}|}{\sqrt{(1-\rho_\theta)(1-\rho_X)}}$ replaces the simple correlation in (EC.18). We define this quantity as the "adjusted correlation":

$$\rho_{\text{adj}} \equiv \frac{\rho-\tilde{\rho}}{\sqrt{(1-\rho_\theta)(1-\rho_X)}}. \tag{EC.25}$$

In this sense, the optimal information ratio can still be approximated by the simple form:

$$\text{IR}^*_{\text{Dependence}} \approx \frac{|\rho-\tilde{\rho}|}{\sqrt{(1-\rho_\theta)(1-\rho_X)}} \cdot \sqrt{N} = |\rho_{\text{adj}}| \cdot \sqrt{N}.$$

When $\tilde{\rho} = \rho_\theta = \rho_X = 0$, we have $\rho_{\text{adj}} = \rho$, which reduces to (EC.18).

This result provides several intuitive relationships for how the optimal performance depends on the parameters in Assumption EC.3. First, a larger correlation between the impact and the residual return of the *same* asset, $\rho = \text{corr}(X_i, \theta_i)$, leads to a better performance, because it measures the strength of the i-th asset's signal on its residual returns. Second, a larger cross-stock correlation, $\tilde{\rho} = \text{corr}(X_i, \theta_j)$, leads to a lower information ratio. This can be rationalized if we view the cross-stock correlation as a measure of the degree of "signal leak" from the impact factor. Third, the cross-sectional correlations of both the impact factor, $\rho_X = \text{corr}(X_i, X_j)$, and the residual returns, $\rho_\theta = \text{corr}(\theta_i, \theta_j)$, contribute positively to portfolio performance, which can be explained by realizing that the cross-sectional dependence implies more signal sources and can provide more information.

EC.1.2.5. Optimal Portfolio Performance for a Small Investable Universe Theorems EC.1, EC.2, and EC.3 provide the approximate optimal portfolio performance when the number of assets, N, is large, using Proposition EC.6. In fact, this approximation, $q(N) \approx N$, is biased because

$$\frac{q(N)}{N} = \frac{\sum_{i=1}^N [\mathbb{E}(Y_{i:N})]^2}{N} < \frac{\sum_{i=1}^N \mathbb{E}[(Y_{i:N})^2]}{N} = \frac{\mathbb{E}[\sum_{i=1}^N (Y_{i:N})^2]}{N} = \frac{\mathbb{E}[\sum_{i=1}^N Y_i^2]}{N} = \frac{\sum_{i=1}^N \mathbb{E}Y_i^2}{N} = 1,$$

where the "$<$" holds due to Jensen's inequality. However, this approximation is very good when N is greater than 100. More precisely, Table EC.1 gives the smallest N such that the value of $q(N)/N$ is greater than a given threshold. For instance, to reach $q(N)/N \geq 0.95$, at least 52 assets should be included in the portfolio, and to reach $q(N)/N \geq 0.99$, at least 298 assets should be included.

Table EC.1 The smallest N such that $q(N)/N \geq c$.

c	0.5	0.6	0.7	0.8	0.9	0.95	0.99
Smallest N	4	5	7	11	24	52	298

EC.1.2.6. Tradeoff Between Performance and Impact Impact investors *may* face a tradeoff between the performance and the impact of their portfolios. The optimal impact portfolio we derived so far implies a particular level of impact, defined as the average impact score for all assets weighted by their portfolio weights. If an investor desires a higher level of impact, or is restricted to investing in certain assets with low impact scores (also known as negative filtering), she may not be able to realize the optimal performance derived in Theorem EC.1.

In this appendix, we provide the optimal information ratio for investors that are restricted to investing in a subset of all assets, and explicitly derive the tradeoff between a portfolio's investment performance and impact. See the following proposition.

PROPOSITION EC.7. *Under Assumption EC.1, assume that $\mu_\theta = 0$. Consider an investor who only invests in assets with impact scores X_i ranking within the $(\xi_1, \xi_2) \times 100$ percentile, where $0 \le \xi_1 < \xi_2 \le 1$. If N is sufficiently large and we approximate the moments of $\boldsymbol{\theta}_{[\boldsymbol{X}]}$ using Proposition EC.5, the maximum information ratio for her portfolio can be approximated by:*

$$\mathrm{IR} \approx \frac{|\rho| \cdot \sqrt{N}}{\sqrt{1-\rho^2}} \cdot \sqrt{\left(\xi_2 - \Phi^{-1}(\xi_2)\varphi\left(\Phi^{-1}(\xi_2)\right)\right) - \left(\xi_1 - \Phi^{-1}(\xi_1)\varphi\left(\Phi^{-1}(\xi_1)\right)\right)}, \tag{EC.26}$$

and its corresponding average impact, $\bar{X} = \mathbb{E}\left[\left(\sum_{i=1}^N w_i X_{i:N}\right) / \left(\sum_{i=1}^N |w_i|\right)\right]$, can be approximated by:

$$\bar{X} \approx \begin{cases} \operatorname{sign}(\rho) \cdot \dfrac{\mu_X\left[\varphi\left(\Phi^{-1}(\xi_1)\right) - \varphi\left(\Phi^{-1}(\xi_2)\right)\right] + \sigma_X\left[\left(\xi_2 - \Phi^{-1}(\xi_2)\varphi\left(\Phi^{-1}(\xi_2)\right)\right) - \left(\xi_1 - \Phi^{-1}(\xi_1)\varphi\left(\Phi^{-1}(\xi_1)\right)\right)\right]}{\left|\varphi\left(\Phi^{-1}(\xi_2)\right) - \varphi\left(\Phi^{-1}(\xi_1)\right)\right|} & \xi_2 \le 0.5 \text{ or } \xi_1 \ge 0.5, \\ \operatorname{sign}(\rho) \cdot \dfrac{\mu_X\left[\varphi\left(\Phi^{-1}(\xi_1)\right) - \varphi\left(\Phi^{-1}(\xi_2)\right)\right] + \sigma_X\left[\left(\xi_2 - \Phi^{-1}(\xi_2)\varphi\left(\Phi^{-1}(\xi_2)\right)\right) - \left(\xi_1 - \Phi^{-1}(\xi_1)\varphi\left(\Phi^{-1}(\xi_1)\right)\right)\right]}{2\varphi(0) - \varphi\left(\Phi^{-1}(\xi_2)\right) - \varphi\left(\Phi^{-1}(\xi_1)\right)} & \xi_1 < 0.5 < \xi_2, \end{cases} \tag{EC.27}$$

where Φ and φ are the distribution function and the density function of $\mathcal{N}(0,1)$, respectively.

As a comparison, we also derive the same set of results for equal-weighted portfolios.

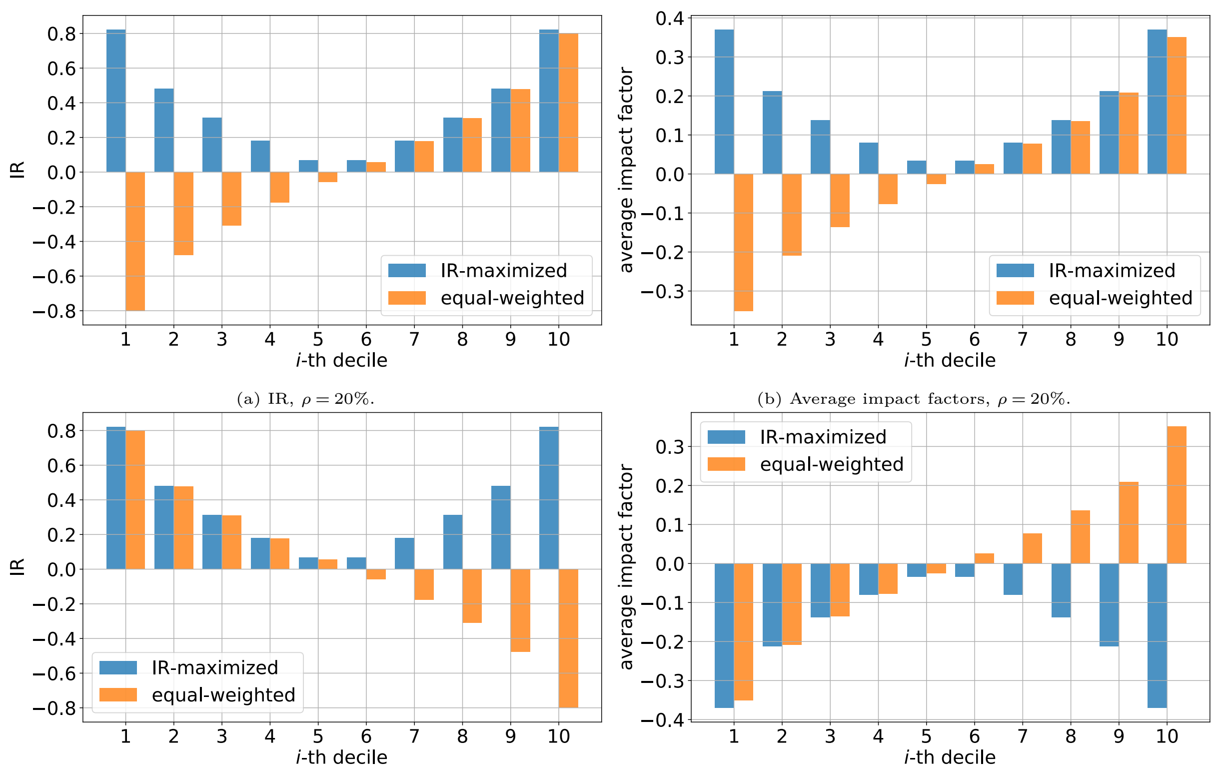
PROPOSITION EC.8. *Under Assumption EC.1, assume that $\mu_\theta = 0$. Consider an investor who only invests in assets with impact scores X_i ranking within the $(\xi_1, \xi_2) \times 100$ percentile, where $0 \le \xi_1 < \xi_2 \le 1$, and the investor is long only and puts equal weights on these assets. If N is sufficiently large and we approximate the moments of $\boldsymbol{\theta}_{[\boldsymbol{X}]}$ using Proposition EC.5, the information ratio of the portfolio can be approximated by:*

$$\mathrm{IR} \approx \frac{\rho \cdot \sqrt{N}}{\sqrt{1-\rho^2}} \cdot \frac{\varphi\left(\Phi^{-1}(\xi_1)\right) - \varphi\left(\Phi^{-1}(\xi_2)\right)}{\sqrt{\xi_2 - \xi_1}}, \tag{EC.28}$$

and its corresponding average impact, $\bar{X} = \mathbb{E}\left[\left(\sum_{i=1}^{N} w_i X_{i:N}\right) / \left(\sum_{i=1}^{N} |w_i|\right)\right]$*, can be approximated by:*

$$\bar{X} \approx \mu_X + \sigma_X \cdot \frac{\varphi\big(\Phi^{-1}(\xi_1)\big) - \varphi\big(\Phi^{-1}(\xi_2)\big)}{\xi_2 - \xi_1}, \qquad \text{(EC.29)}$$

where Φ *and* φ *are the distribution function and the density function of* $\mathcal{N}(0,1)$*, respectively.*



(a) IR, $\rho = 20\%$. (b) Average impact factors, $\rho = 20\%$.

(c) IR, $\rho = -20\%$. (d) Average impact factors, $\rho = -20\%$.

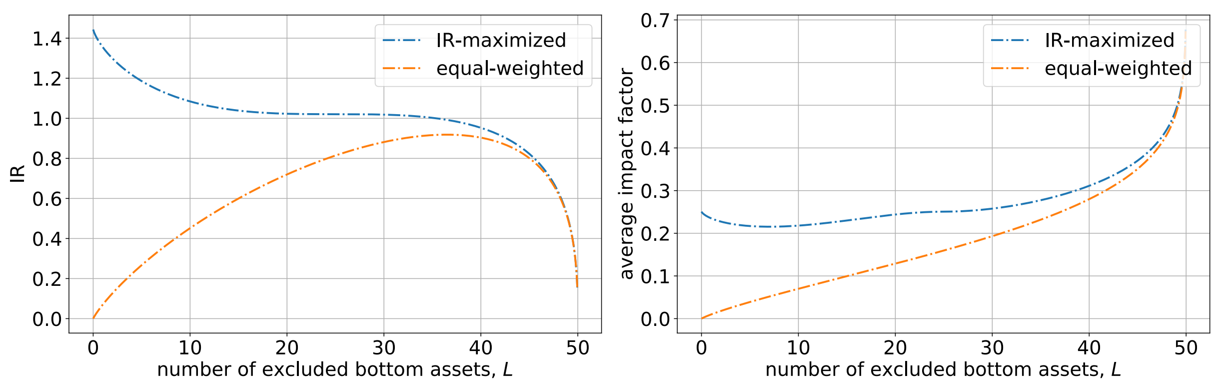
Figure EC.3 Information ratio (IR) and average impact scores, $\bar{X}$, for ten decile portfolios (with impact factors from the lowest to the highest). Blue bars are results for portfolios with optimal IR, and orange bars are results for portfolios with equal weights on assets within the decile. We set $\mu_\theta = 0\%$, $\mu_X = 0\%$, and $\sigma_X = 20\%$.

Proposition EC.7 and Proposition EC.8 provide the information ratio and the impact score for portfolios constructed from subsets of the universe of all assets. We apply these results to decile portfolios formed by dividing assets into ten deciles based on their impact scores. Figure EC.3 shows the information ratio (IR) and the average impact score of the IR-maximized portfolios and equal-weighted long-only portfolios when $\mu_\theta = 0\%$, $\mu_X = 0\%$, and $\sigma_X = 20\%$.

A performance–impact tradeoff is crucial for impact investors. Figure EC.3a and Figure EC.3b show the results when $\rho = 20\%$. These figures imply that impact investors can reach a win–win on both information ratio and impact scores when they invest in assets within the top deciles. However,

Figure EC.3c and Figure EC.3d show the results when $\rho = -20\%$. Unlike the case for a positive ρ, if an impact investor invests in top decile assets when $\rho = -20\%$ and maximizes the IR, she can earn a positive IR but a negative impact, while if she constructs an equal-weighted portfolio, she receives a positive impact but has to endure a negative IR. In other words, when the impact factor is negatively correlated with the residual return ($\rho < 0$), earning a better portfolio performance (as measured through IR) contradicts the goal of having a greater impact.

We further study the performance–impact tradeoff for a common impact investing strategy—negative screening. Negative screening investors exclude companies with low impact factors from the universe to form portfolios. Specifically, we assume that impact investors exclude assets with the lowest L values of impact factors. Figure EC.4 shows the IR and the average impact score of the IR-maximized portfolios and equal-weighted long-only portfolios as functions of L when $\rho = 20\%$, $\mu_\theta = 0\%$, $\mu_X = 0\%$, and $\sigma_X = 20\%$.



(a) IR. (b) Average impact factors.

Figure EC.4 Information ratio (IR) and average impact scores, $\bar{X}$, for portfolios constructed using assets excluding those with the bottom L impact scores. Blue lines are results for portfolios with optimal IR, and orange lines are results for portfolios with equal weights on assets included in the portfolio. We set $\rho = 20\%$, $\mu_\theta = 0\%$, $\mu_X = 0\%$, and $\sigma_X = 20\%$.

Figure EC.4 demonstrates that, for IR-maximized portfolios, as the number of bottom assets excluded from the universe, L, increases, the optimal IR decreases (Figure EC.4a), and the average impact score increases (Figure EC.4b). Therefore, impact investors face a tradeoff between performance and impact. For equal-weighted portfolios, the IR initially increases and then decreases as L increases. This pattern arises because buying the bottom assets decreases the IR, while buying the top assets increases the IR. In addition, the IR-maximized portfolios achieve both higher IR and higher average impact scores, implying a win–win situation for impact investors.

EC.1.3. Generalized Versions of the Representation Theorem

In our main article, we present the representation theorem under smoothness conditions (Theorem 1) as well as its generalized version (Theorem 2). In Appendix EC.1.3.1, we provide a more in-depth discussion of the technical details related to Theorem 2. Appendix EC.1.3.2 further extends the representation theorem to allow for cross-sectional heterogeneity.

EC.1.3.1. Additional Technical Details for the Representation Theorem Under General F This appendix provides technical details for Theorem 2 in our main paper.

When F_θ is not continuous, we need to clarify the definition of the inverse function, $F_\theta^{-1}(\cdot)$, in the mixture function given by (12). We adopt the following definition of the inverse function without loss of generality:

Definition EC.1 (Inverse Function). For any non-decreasing function $f(x)$, the inverse function of f is defined as $f^{-1}(y) = \inf\{x : f(x) \geq y\}$.

When C is non-differentiable, the function $g_u(\cdot)$ in (12), which is the inverse of $v \mapsto \frac{\partial C}{\partial u}(u,v)$, is not well defined. To address this issue, we introduce the concept of the modified partial Dini derivative proposed by Fang et al. (2020).

Definition EC.2 (Modified Partial Dini Derivative, (Fang et al. 2020)). The modified partial Dini derivative (MPDD) of a copula $C(u,v)$ with respect to u, denoted by $\mathscr{D}_1 C(u,v)$, is a bivariate function $[0,1]\times\mathbb{R}\to[0,1]$ defined as

$$\mathscr{D}_1 C(u,v) = \begin{cases} \inf_{y>v} D_1^+ C(u,y), & u\in[0,1), v\in[0,1), \\ \inf_{y>v} D_1^- C(1,y), & u=1, v\in[0,1), \\ 0, & u\in[0,1], v\in(-\infty,0), \\ 1, & u\in[0,1], v\in[1,+\infty), \end{cases} \tag{EC.30}$$

where $D_1^+ C(u,v)$ and $D_1^- C(u,v)$ are the partial right upper and left upper Dini derivatives of $C(u,v)$ with respect to u:

$$D_1^+ C(u,v) = \limsup_{h\to 0^+} \frac{C(u+h,v)-C(u,v)}{h}, \quad D_1^- C(u,v) = \limsup_{h\to 0^+} \frac{C(u,v)-C(u-h,v)}{h}. \tag{EC.31}$$

Remark EC.1. The MPDD is pointwisely defined for any copula. Fang et al. (2020, Theorem 2.1) demonstrate that $\mathscr{D}_1 C(u,v)$ defined by (EC.30) is a regular conditional distribution function of $C(u,v)$ given u. In other words, it satisfies:

(i) For any $u\in[0,1]$, $v\mapsto \mathscr{D}_1 C(u,v)$ is a distribution function;

(ii) For any $v\in[0,1]$, $u\mapsto \mathscr{D}_1 C(u,v)$ is Lebesgue measurable and $\int_0^a \mathscr{D}_1 C(u,v)\mathrm{d}u = C(a,v)$ for any $a\in[0,1]$.

Therefore, we can define $\mathbb{P}(V \le v | U = u) = \mathscr{D}_1 C(u,v)$. The subscript of $\mathscr{D}_1$ represents that the MPDD is calculated with respect to the first dimension, u. Fang et al. (2020, Theorem 2.1) also demonstrate that, when C has a density, we have $\mathscr{D}_1 C(u,v) = \frac{\partial C}{\partial u}(u,v)$, in which case the MPDD reduces to the simple partial derivative.

Finally, when F_X is not continuous, we demonstrate that the representation holds only when the copula of F is "linearly interpolating"; see Definition 1. The following result not only guarantees the existence and uniqueness of the linearly interpolating copula but also provides an explicit method for its construction.

Proposition EC.9. *For any bivariate distribution F, there exists a copula of F that satisfies (13) on $\overline{\mathcal{R}}_X^c \times \overline{\mathcal{R}}_\theta$. In addition, the copula is unique on $[0,1] \times \overline{\mathcal{R}}_\theta$. The construction of this copula is given in the proof.*

The linearly interpolating copula defined by Definition 1 plays a crucial role in the representation theorem. To build intuition, we provide examples to illustrate the linearly interpolating copula.

Example EC.1. Consider the case where both the marginal distributions of $\boldsymbol{X}$ and $\boldsymbol{\theta}$, F_X and F_θ, have discontinuity points. In particular, we assume that $\overline{\mathcal{R}}_X = [0,0.3] \cup [0.7,1]$ and $\overline{\mathcal{R}}_\theta = [0,0.2] \cup [0.5,1]$, which implies that $\overline{\mathcal{R}}_X^c = (0.3,0.7)$ and $\overline{\mathcal{R}}_\theta^c = (0.2,0.5)$. Figure EC.5 visualizes the linearly interpolating copula between $\boldsymbol{X}$ and $\boldsymbol{\theta}$. According to Sklar's theorem, the copula is uniquely determined in the purple regions: $[0,0.3] \times [0,0.2]$, $[0,0.3] \times [0.5,1]$, $[0.7,1] \times [0,0.2]$, and $[0.7,1] \times [0.5,1]$, and undetermined in the yellow regions: $(0.3,0.7) \times [0,0.2]$, $(0.3,0.7) \times [0.5,1]$, and $[0,1] \times (0.2,0.5)$. According to Definition 1, we say that a copula $C(u,v)$ is linearly interpolating on $\overline{\mathcal{R}}_X^c \times \overline{\mathcal{R}}_\theta$ with respect to u, indicating that it is linearly interpolating on $(0.3,0.7) \times [0,0.2]$ and $(0.3,0.7) \times [0.5,1]$ along $u \in \overline{\mathcal{R}}_X^c$. This is illustrated in Figure EC.5 by green solid straight lines. After linearly interpolating, the copula is determined on $[0,1] \times \overline{\mathcal{R}}_\theta$ but remains undetermined on $[0,1] \times \overline{\mathcal{R}}_\theta^c$.

Let us now consider a simpler example.

Example EC.2. Assume that $\boldsymbol{X}$ reduces to a constant, c, i.e., $X_i \equiv c$, and that $\theta_i \overset{\text{IID}}{\sim} \mathcal{N}(0,1)$. Then we have $\overline{\mathcal{R}}_X = \{0,1\}$ and $\overline{\mathcal{R}}_\theta = [0,1]$. By Sklar's theorem, the copula C is uniquely determined only on $\{0,1\} \times [0,1]$ by:

$$C(0,v) = 0, \quad C(1,v) = v, \quad v \in [0,1]. \tag{EC.32}$$

In fact, (EC.32) holds for any copula, so any copula can be considered as a copula between $\boldsymbol{X}$ and $\boldsymbol{\theta}$ satisfying (10). However, a copula that is linearly interpolating on $\overline{\mathcal{R}}_X^c \times \overline{\mathcal{R}}_\theta = (0,1) \times [0,1]$ is unique and given by:

$$C(u,v) = uv, \quad u \in [0,1], v \in [0,1].$$

This is the independence copula defined by Definition EC.5 in Appendix EC.1.4.3.

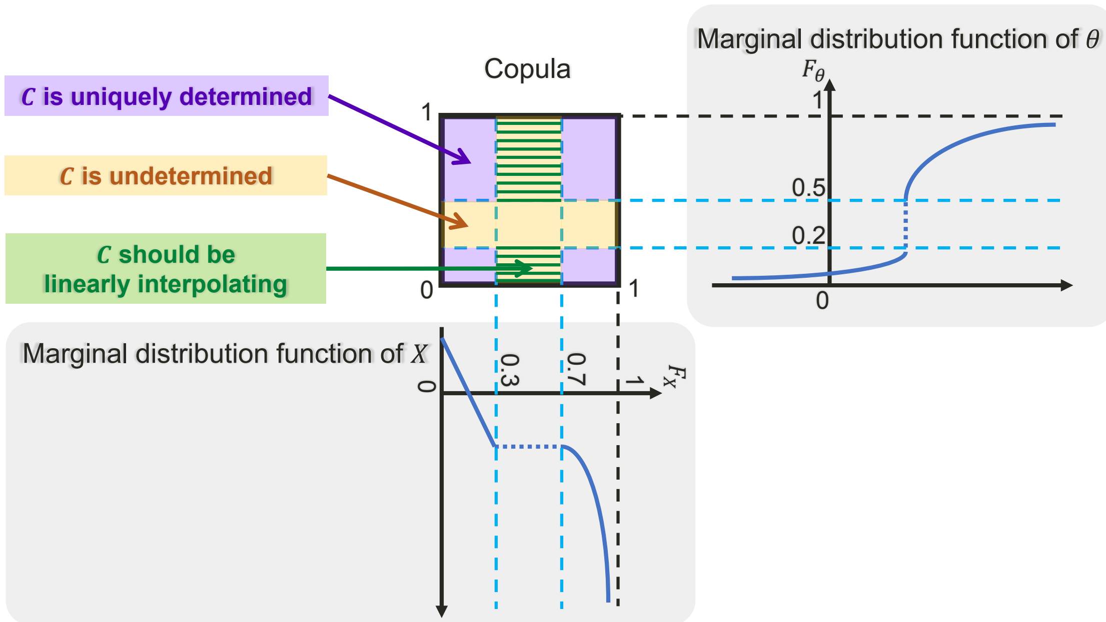


Figure EC.5 Illustration of a linear interpolating copula.

EC.1.3.2. Representation Theorem with Heterogeneous Distributions In practice, the joint distribution of the impact factor and residual returns may vary cross-sectionally. For example, brown stocks may have higher variances in returns than green assets due to regulatory uncertainty. In this part, we allow the impact factor, X_i, and residual returns, θ_i, to have heterogeneous cross-sectional distributions. We provide a version of the representation theorem with cross-sectional heterogeneity, discuss an application to markets with heterogeneous groups representing, for example, different industries, and illustrate the impact of heterogeneity on the optimal impact portfolio using numerical examples.

In particular, we assume that:

Assumption EC.4. *Bivariate vectors* $(X_1,\theta_1)^\top,(X_2,\theta_2)^\top,\dots,(X_N,\theta_N)^\top$ *are mutually independent and satisfy*

$$\begin{pmatrix} X_1 \\ \theta_1 \end{pmatrix} \sim F_1(\cdot,\cdot), \quad \begin{pmatrix} X_2 \\ \theta_2 \end{pmatrix} \sim F_2(\cdot,\cdot), \quad \dots, \quad \begin{pmatrix} X_N \\ \theta_N \end{pmatrix} \sim F_N(\cdot,\cdot), \tag{EC.33}$$

where $F_1(\cdot,\cdot),F_2(\cdot,\cdot),\dots,F_N(\cdot,\cdot)$ *are (potentially different) bivariate distribution functions with densities.*

We define the following notations:

- For $i=1,2,\dots,N$, the marginal distributions of X_i and θ_i are $F_{X,i}(\cdot)$ and $F_{\theta,i}(\cdot)$, respectively;
- For $i=1,2,\dots,N$, the marginal densities of X_i and θ_i are $f_{X,i}(\cdot)$ and $f_{\theta,i}(\cdot)$, respectively;
- For $i=1,2,\dots,N$, the copula of $F_i(\cdot,\cdot)$ is $C_i(\cdot,\cdot)$;
- For any permutation $(i_1,i_2,\dots,i_N)$ of $1,2,\dots,N$, let $p_{i_1,i_2,\dots,i_N}=\mathbb{P}(X_{i_1}\le X_{i_2}\le\cdots\le X_{i_N})$.

THEOREM EC.4 **(The Representation Theorem, Heterogeneous Distribution).** *Under Assumption EC.4, we have:*

$$(\theta_{[1:N]}, \theta_{[2:N]}, \ldots, \theta_{[N:N]}) \stackrel{d}{=} (h_1(\Pi, M_1, V_1), h_2(\Pi, M_2, V_2), \ldots, h_N(\Pi, M_N, V_N)), \tag{EC.34}$$

where "$\stackrel{d}{=}$" denotes equality in distribution. Here:

- *The random vector Π is a random permutation of $\{1, 2, \ldots, N\}$ satisfying*

$$\mathbb{P}(\Pi = (i_1, i_2, \ldots, i_N)) = p_{i_1, i_2, \ldots, i_N}$$

for any permutation $(i_1, i_2, \ldots, i_N)$ of $\{1, 2, \ldots, N\}$;

- *Given Π, the random vector $(M_1, M_2, \ldots, M_N)|\Pi = (i_1, i_2, \ldots, i_N)$ has a joint density of*

$$m_{i_1, i_2, \ldots, i_N}(x_1, x_2, \ldots, x_N) = \frac{1}{p_{i_1, i_2, \ldots, i_N}} \cdot \prod_{k=1}^{N} f_{X, i_k}(x_k) \cdot \mathbf{1}_{\{x_1 \le x_2 \le \cdots \le x_N\}}; \tag{EC.35}$$

- *The random variables $V_1, V_2 \ldots, V_N$ satisfy $V_1, V_2 \ldots, V_N \stackrel{\text{IID}}{\sim} \text{Uniform}(0, 1)$ and are independent of $\Pi, M_1, M_2, \ldots, M_N$;*
- *For $k = 1, 2, \ldots, N$, function h_k is defined as*

$$h_k(\pi, x, w) \equiv F_{\theta, i_k}^{-1} \circ h_{F_{X, i_k}(x)}^{i_k}(w) \tag{EC.36}$$

for any permutation $\pi = (i_1, i_2, \ldots, i_N)$ of $\{1, 2, \ldots, N\}$, and "$\circ$" represents function composition;

- *For $u \in [0, 1]$ and $i = 1, 2, \ldots, N$, function $h_u^i(w)$ is the inverse function of $v \mapsto \frac{\partial C_i}{\partial u}(u, v)$.*

Intuition and Implications. We make a few remarks about the representation theorem with heterogeneous distribution.

First, Theorem EC.4 reveals a similar representation to that in Theorem 1 because the distribution of induced order statistics, $\boldsymbol{\theta}_{[\boldsymbol{X}]}$, is a mixture of three components: a random permutation (Π), a set of random variables analogous to order statistics (M_i), and uniform noise (V_i). In comparison with Theorem 1, an additional random permutation component, Π, is required here. We refer to random variables M_i as "order statistics" because, by definition, they always satisfy $M_1 \le M_2 \le \cdots \le M_N$. In addition, if $X_1, X_2, \ldots, X_N$ have the same marginal densities, i.e., $f_{X,1} = f_{X,2} = \cdots = f_{X,N}$, one can easily verify that (EC.35) reduces to the density of order statistics of $X_1, X_2, \ldots, X_N$. Therefore, although M_i are not order statistics of a series of random variables themselves, they resemble the order statistics component in Theorem 1.

Second, because of the heterogeneity, the mixture functions defined by (EC.36), h_k, are different for $k = 1, 2, \ldots, N$. In addition, all three components of F_i—$F_{X,i}$, $F_{\theta,i}$, and C_i—contribute to the mixture function, h_k. This is different from Theorem 1 in which the mixture function g in (12) depends only on the marginal distribution of $\boldsymbol{\theta}$ and the copula. This is because Theorem EC.4 allows for heterogeneous distributions cross-sectionally and, therefore, the marginal distribution of X_i is also crucial for the mixture.

Third, the representation in Theorem EC.4 can be simplified when $F_1(\cdot,\cdot),\dots,F_N(\cdot,\cdot)$ are homogeneous cross-sectionally. In particular, one can easily verify that:

(i) If $F_\theta \equiv F_{\theta,1} = F_{\theta,2} = \cdots = F_{\theta,N}$, the mixture function (EC.36) reduces to

$$h_k(\pi,x,w) = F_\theta^{-1} \circ h^{i_k}_{F_{X,i_k}(x)}(w);$$

(ii) If $F_X \equiv F_{X,1} = F_{X,2} = \cdots = F_{X,N}$, the representation (EC.34) reduces to

$$(\theta_{[1:N]},\theta_{[2:N]},\dots,\theta_{[N:N]}) \stackrel{d}{=} (h_1(\Pi,U_{1:N},V_1), h_2(\Pi,U_{2:N},V_2),\dots,h_N(\Pi,U_{N:N},V_N)),$$

where $U_{1:N} \le U_{2:N} \le \cdots \le U_{N:N}$ and $V_1,V_2,\dots,V_N$ are defined as in Theorem 1, and the mixture function (EC.36) reduces to

$$h_k(\pi,x,w) = F^{-1}_{\theta,i_k} \circ h^{i_k}_x(w);$$

(iii) If $C \equiv C_1 = C_2 = \cdots = C_N$, the mixture function (EC.36) reduces to

$$h_k(\pi,x,w) = F^{-1}_{\theta,i_k} \circ h_{F_{X,i_k}(x)}(w),$$

where $h_u(w)$ is the inverse function of $v \mapsto \frac{\partial C}{\partial u}(u,v)$;

(iv) If $F_\theta \equiv F_{\theta,1} = F_{\theta,2} = \cdots = F_{\theta,N}$, $F_X \equiv F_{X,1} = F_{X,2} = \cdots = F_{X,N}$, and $C \equiv C_1 = C_2 = \cdots = C_N$, the representation (EC.34) reduces to (11).

Therefore, the representation theorem presented in the main paper can be regarded as a reduced form of Theorem EC.4 under the IID assumption.

Finally, like the IID version, the heterogeneous version allows us to efficiently compute the moments of $\boldsymbol{\theta}_{[\boldsymbol{X}]}$ through numerical integration. In particular, the first two moments of $\boldsymbol{\theta}_{[\boldsymbol{X}]}$ are given by the following proposition:

Proposition EC.10. *Under the assumptions of Theorem EC.4, we have:*

$$\mathbb{E}(\theta_{[i:N]}) = \sum_{k=1}^N \int_0^1 \int_0^1 \tilde{h}_k(u,w) H^i_k(u)\mathrm{d}u\mathrm{d}w, \tag{EC.37}$$

$$\mathbb{E}({\theta_{[i:N]}}^2) = \sum_{k=1}^N \int_0^1 \int_0^1 \left[\tilde{h}_k(u,w)\right]^2 H^i_k(u)\mathrm{d}u\mathrm{d}w, \tag{EC.38}$$

$$\mathbb{E}(\theta_{[i:N]}\theta_{[j:N]}) = \sum_{k=1}^N \sum_{l=1,l\neq k}^N \int_0^1 \int_0^1 \int_0^1 \int_0^1 \tilde{h}_k(u,w)\tilde{h}_l(z,s) J^{i,j}_{k,l}(u,z)\mathrm{d}z\mathrm{d}u\mathrm{d}w\mathrm{d}s, \tag{EC.39}$$

for $i,j=1,2,\dots,N$ and $i<j$, where functions $\tilde{h}_k(\cdot,\cdot)$, $H_k^i(\cdot)$, and $J_{k,l}^{i,j}(\cdot,\cdot)$ are defined by:

$$\tilde{h}_k(u,w) = F_{\theta,k}^{-1}\circ h_u^k(w), \tag{EC.40}$$

$$H_k^i(u) = \mathbb{P}\left(\xi_{i-1:N-1}^{[-k]} \le F_{X,k}^{-1}(u)\right) - \mathbb{P}\left(\xi_{i:N-1}^{[-k]} \le F_{X,k}^{-1}(u)\right), \tag{EC.41}$$

$$\begin{aligned} J_{k,l}^{i,j}(u,z) =& \mathbb{P}\left(\eta_{i-1:N-2}^{[-k,l]} \le F_{X,k}^{-1}(u), \eta_{j-2:N-2}^{[-k,l]} \le F_{X,l}^{-1}(z)\right) - \mathbb{P}\left(\eta_{i:N-2}^{[-k,l]} \le F_{X,k}^{-1}(u), \eta_{j-2:N-2}^{[-k,l]} \le F_{X,l}^{-1}(z)\right) \\ & - \mathbb{P}\left(\eta_{i-1:N-2}^{[-k,l]} \le F_{X,k}^{-1}(u), \eta_{j-1:N-2}^{[-k,l]} \le F_{X,l}^{-1}(z)\right) + \mathbb{P}\left(\eta_{i:N-2}^{[-k,l]} \le F_{X,k}^{-1}(u), \eta_{j-1:N-2}^{[-k,l]} \le F_{X,l}^{-1}(z)\right). \end{aligned} \tag{EC.42}$$

Here, $h_u^k(w)$ is defined in Theorem EC.4, $\xi_{1:N-1}^{[-k]} \le \xi_{2:N-1}^{[-k]} \le \cdots \le \xi_{N-1:N-1}^{[-k]}$ are order statistics of $\{X_s\}_{s=1,s\neq k}^N$, $\eta_{1:N-2}^{[-k,l]} \le \eta_{2:N-2}^{[-k,l]} \le \cdots \le \eta_{N-2:N-2}^{[-k,l]}$ are order statistics of $\{X_s\}_{s=1,s\neq k,s\neq l}^N$, $\xi_{0:N-1}^{[-k]} = \eta_{0:N-2}^{[-k,l]} = -\infty$, and $\xi_{N:N-1}^{[-k]} = \eta_{N-1:N-2}^{[-k,l]} = \eta_{N:N-2}^{[-k,l]} = +\infty$.

Functions $\tilde{h}_k$, H_k^i, and $J_{k,l}^{i,j}$ are entirely determined by the distributions of $\boldsymbol{X}$ and $\boldsymbol{\theta}$, and they can be computed using existing results of order statistics in, for example, Bapat and Beg (1989). Therefore, given the distributions of $\boldsymbol{X}$ and $\boldsymbol{\theta}$, we can use Proposition EC.10 to calculate the moments of $\boldsymbol{\theta}_{[\boldsymbol{X}]}$ through numerical integration.

Computational Cost. In comparison to the IID version (see Proposition 2), when considering $\boldsymbol{X}$ and $\boldsymbol{\theta}$ with completely different cross-sectional distributions, the computational cost increases because the summations in (EC.37)–(EC.39) imply that we need to calculate N double integrals for the expected value, N double integrals for the variance, and $N(N-1)$ quadruple integrals for covariances of $\theta_{[i:N]}$. This is not surprising because, in the most general case, each asset can have its own unique distribution without any cross-sectional structure.

An alternative to computing the moments of $\theta_{[i:N]}$ using numerical integrals in Proposition EC.10 is to use Monte Carlo simulation. One can start by simulating random samples of $\boldsymbol{X}$ and $\boldsymbol{\theta}$, sort $\boldsymbol{\theta}$ according to $\boldsymbol{X}$ to obtain the induced order statistics, $\boldsymbol{\theta}_{[\boldsymbol{X}]}$, and finally estimate their moments numerically using these samples. Computing moments using Proposition EC.10 may lead to smaller estimation errors and require less time especially when N is small.

We conduct numerical experiments to illustrate the computational performance of Proposition EC.10 when N is small. Under Assumption EC.4 and assuming that $F_1, F_2, \dots, F_N$ are all bivariate normal distributions:

$$\begin{pmatrix} X_i \\ \theta_i \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_{X,i} \\ \mu_{\theta,i} \end{pmatrix}, \begin{pmatrix} \sigma_{X,i}^2 & \rho_i\sigma_{X,i}\sigma_{\theta,i} \\ \rho_i\sigma_{X,i}\sigma_{\theta,i} & \sigma_{\theta,i}^2 \end{pmatrix} \right), \quad i=1,2,\dots,N,$$

where for all $i=1,2,\dots,N$, $\rho_i = 20\%$, $\sigma_{X,i} = 20\%$, $\mu_{\theta,i} = 0\%$, and parameters $\mu_{X,i}$ and $\sigma_{\theta,i}$ take values in Table EC.2, in which the N stocks all have different marginal distributions, and green stocks (stocks with higher expected values of $\boldsymbol{X}$) have lower variances in residual returns than brown stocks (stocks with higher expected values of $\boldsymbol{X}$).

Table EC.2 Simulation setups for computing numerical integrals using Proposition EC.10 and Monte Carlo simulations.

N	$\mu_{X,1}$	$\mu_{X,2}$	$\mu_{X,3}$	$\mu_{X,4}$	$\mu_{X,5}$	$\sigma_{\theta,1}$	$\sigma_{\theta,2}$	$\sigma_{\theta,3}$	$\sigma_{\theta,4}$	$\sigma_{\theta,5}$
2	-30.0%	30.0%	-	-	-	40.0%	10.0%	-	-	-
3	-30.0%	0.0%	30.0%	-	-	40.0%	25.0%	10.0%	-	-
4	-30.0%	-10.0%	10.0%	30.0%	-	40.0%	30.0%	20.0%	10.0%	-
5	-30.0%	-15.0%	0.0%	15.0%	30.0%	40.0%	32.5%	25.0%	17.5%	10.0%

We compute the expected value, variance, and optimal weights using two methods: (a) the representation theorem (calculating numerical integrals using Proposition EC.10) and (b) Monte Carlo simulations. We perform these computations using parameter values in Table EC.2, and record the errors and CPU times. The error is defined as $\sum_{i=1}^{N}(\hat{w}_i - w_i^*)^2$, where $(w_1^*, w_2^*, \ldots, w_N^*)^\top$ is the true optimal weights for the N assets given by (9),[24] and $(\hat{w}_1, \hat{w}_2, \ldots, \hat{w}_N)^\top$ is the optimal weights computed using either Method (a) or Method (b). The CPU time for Method (a) is the total time cost to calculate the numerical integrals for expectations and (co)variances, and obtain the optimal weights given by (9). The CPU time for Method (b) is the total time cost to simulate the random samples of induced order statistics, estimate the sample averages and (co)variances, and obtain the optimal weights.

The error and CPU time of both methods depend on the chosen degree of accuracy. In general, a higher degree of accuracy implies a lower error but requires a higher CPU time. For Method (a), the degree of accuracy is determined by the number of subintervals partitioned from the original interval for numerical integration, while for Method (b), it is determined by the number of random samples used in Monte Carlo simulations. We choose different degrees of accuracy for both methods and record their respective errors and CPU times. All experiments are conducted on a laptop with an Intel(R) Core(TM) i7-9750H CPU @ 2.60GHz.

Figure EC.6 shows the relationship between the error and the CPU time. The blue lines are results obtained using Method (a) (the representation theorem), and the orange lines are results obtained using Method (b) (Monte Carlo). Different dots on each line correspond to results obtained under different degrees of accuracy. We make the following observations from Figure EC.6.

First, there is a negative relationship between the error and the CPU time, which reflects the tradeoff between error and computational cost as the level of accuracy increases in both methods.

Second, blue lines generally lie on the lower left-hand side of orange lines, which indicates that Method (a), the representation theorem, achieves a given level of error with lower computational costs than Method (b), Monte Carlo simulations.

Third, the blue lines shift upwards as the number of assets, N, increases. This illustrates that the CPU time for the representation theorem increases as N grows.

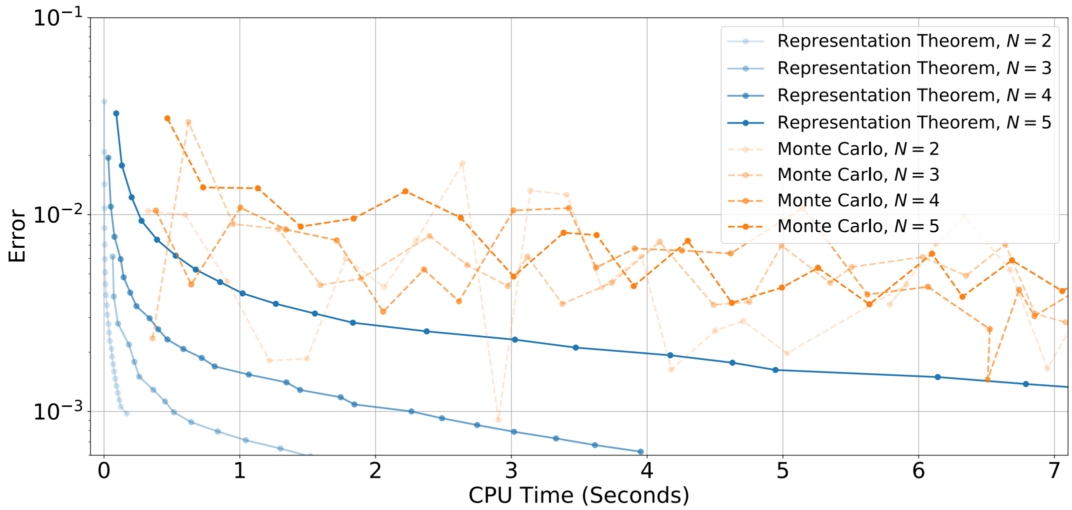


Figure EC.6 Errors and CPU times for computing optimal weights by numerical integrals using Proposition EC.10 and Monte Carlo simulations under parameter values in Table EC.2.

Finally, the errors in weights are much more stable when using the representation theorem, whereas the lines for Monte Carlo are noisier due to the random nature of simulations.

These observations demonstrate the computational effectiveness of the representation theorem, Proposition EC.10, especially when N is small. When N is large, we consider the following group structure which is more practically relevant.

Cross-Sectional Heterogeneity with Group Structure. In practice, one may model the market as the mixture of several groups with heterogeneity between groups but homogeneity within a group. The number of groups, denoted by G, is typically much smaller than the number of assets, N. For example, one may assume that all assets are divided into two groups: one "green" group with a higher expected value of impact factor and another "brown" group with a lower expected value of impact factor.

In general, we consider the following assumption, which is a special case of Assumption EC.4:

Assumption EC.5. *Bivariate vectors $(X_1, \theta_1)^\top, (X_2, \theta_2)^\top, \ldots, (X_N, \theta_N)^\top$ are mutually independent and satisfy*

$$\begin{aligned}
&\begin{pmatrix} X_1 \\ \theta_1 \end{pmatrix}, \begin{pmatrix} X_2 \\ \theta_2 \end{pmatrix}, \ldots, \begin{pmatrix} X_{N_1} \\ \theta_{N_1} \end{pmatrix} \overset{\text{IID}}{\sim} F_{(1)}(\cdot,\cdot), \\
&\begin{pmatrix} X_{N_1+1} \\ \theta_{N_1+1} \end{pmatrix}, \begin{pmatrix} X_{N_1+2} \\ \theta_{N_1+2} \end{pmatrix}, \ldots, \begin{pmatrix} X_{N_1+N_2} \\ \theta_{N_1+N_2} \end{pmatrix} \overset{\text{IID}}{\sim} F_{(2)}(\cdot,\cdot), \\
&\qquad\qquad\vdots \\
&\begin{pmatrix} X_{N_1+\cdots+N_{G-1}+1} \\ \theta_{N_1+\cdots+N_{G-1}+1} \end{pmatrix}, \begin{pmatrix} X_{N_1+\cdots+N_{G-1}+2} \\ \theta_{N_1+\cdots+N_{G-1}+2} \end{pmatrix}, \ldots, \begin{pmatrix} X_{N_1+\cdots+N_{G-1}+N_G} \\ \theta_{N_1+\cdots+N_{G-1}+N_G} \end{pmatrix} \overset{\text{IID}}{\sim} F_{(G)}(\cdot,\cdot),
\end{aligned} \tag{EC.43}$$

where $N_k \geq 1$ *for* $k = 1, 2, \ldots, G$, $N_1 + N_2 + \cdots + N_G = N$, *and* $F_{(1)}(\cdot,\cdot), F_{(2)}(\cdot,\cdot), \ldots, F_{(G)}(\cdot,\cdot)$ *are (different) bivariate distribution functions with densities.*

Assumption EC.5 separates all N assets into G groups, with assets in each group following the same distribution. Under this more practically relevant setup, one only needs to model the distributions for the G groups, $F_{(1)}, F_{(2)}, \ldots, F_{(G)}$, instead of the distributions of all N assets individually. This reduces the computational cost of calculating numerical integrals using Proposition EC.10, as shown in the following corollary.

Corollary EC.1. *Under Assumption EC.5, we have:*

$$\mathbb{E}(\theta_{[i:N]}) = \sum_{k=1}^{G} N_k \int_0^1 \int_0^1 \tilde{h}_{(k)}(u,w) H^i_{(k)}(u) \mathrm{d}u\mathrm{d}w, \tag{EC.44}$$

$$\mathbb{E}(\theta_{[i:N]}{}^2) = \sum_{k=1}^{G} N_k \int_0^1 \int_0^1 \left[\tilde{h}_{(k)}(u,w)\right]^2 H^i_{(k)}(u) \mathrm{d}u\mathrm{d}w, \tag{EC.45}$$

$$\begin{aligned}\mathbb{E}(\theta_{[i:N]}\theta_{[j:N]}) = &\sum_{k=1}^{G} \sum_{l=1, l\neq k}^{G} N_k N_l \int_0^1 \int_0^1 \int_0^1 \int_0^1 \tilde{h}_{(k)}(u,w) \tilde{h}_{(l)}(z,s) J^{i,j}_{(k),(l)}(u,z) \mathrm{d}z\mathrm{d}u\mathrm{d}w\mathrm{d}s \\ &+ \sum_{k=1}^{G} N_k(N_k - 1) \int_0^1 \int_0^1 \int_0^1 \int_0^1 \tilde{h}_{(k)}(u,w) \tilde{h}_{(k)}(z,s) J^{i,j}_{(k),(k)}(u,z) \mathrm{d}z\mathrm{d}u\mathrm{d}w\mathrm{d}s,\end{aligned} \tag{EC.46}$$

for $i, j = 1, 2, \ldots, N$ *and* $i < j$, *where functions* $\tilde{h}_{(k)}(\cdot,\cdot)$, $H^i_{(k)}(\cdot)$ *and* $J^{i,j}_{(k),(l)}(\cdot,\cdot)$ *are defined by:*

$$\tilde{h}_{(k)}(u,w) = F^{-1}_{\theta,(k)} \circ h^{(k)}_u(w), \tag{EC.47}$$

$$H^i_{(k)}(u) = \mathbb{P}\left(\xi^{[-(N_1+\cdots+N_k)]}_{i-1:N-1} \leq F^{-1}_{X,(k)}(u)\right) - \mathbb{P}\left(\xi^{[-(N_1+\cdots+N_k)]}_{i:N-1} \leq F^{-1}_{X,(k)}(u)\right), \tag{EC.48}$$

$$\begin{aligned}J^{i,j}_{(k),(l)}(u,z) = &\mathbb{P}\left(\eta^{[-(N_1+\cdots+N_k),(N_1+\cdots+N_l)]}_{i-1:N-2} \leq F^{-1}_{X,(k)}(u), \eta^{[-(N_1+\cdots+N_k),(N_1+\cdots+N_l)]}_{j-2:N-2} \leq F^{-1}_{X,(l)}(z)\right) \\ &- \mathbb{P}\left(\eta^{[-(N_1+\cdots+N_k),(N_1+\cdots+N_l)]}_{i:N-2} \leq F^{-1}_{X,(k)}(u), \eta^{[-(N_1+\cdots+N_k),(N_1+\cdots+N_l)]}_{j-2:N-2} \leq F^{-1}_{X,(l)}(z)\right) \\ &- \mathbb{P}\left(\eta^{[-(N_1+\cdots+N_k),(N_1+\cdots+N_l)]}_{i-1:N-2} \leq F^{-1}_{X,(k)}(u), \eta^{[-(N_1+\cdots+N_k),(N_1+\cdots+N_l)]}_{j-1:N-2} \leq F^{-1}_{X,(l)}(z)\right) \\ &+ \mathbb{P}\left(\eta^{[-(N_1+\cdots+N_k),(N_1+\cdots+N_l)]}_{i:N-2} \leq F^{-1}_{X,(k)}(u), \eta^{[-(N_1+\cdots+N_k),(N_1+\cdots+N_l)]}_{j-1:N-2} \leq F^{-1}_{X,(l)}(z)\right).\end{aligned} \tag{EC.49}$$

Here, $\xi^{[-(N_1+\cdots+N_k)]}_{k:N-1}$ *and* $\eta^{[-(N_1+\cdots+N_k),(N_1+\cdots+N_l)]}_{k:N-2}$ *are defined in Proposition EC.10,* $h^{(k)}_u(w)$ *is the inverse function of* $v \mapsto \frac{\partial C_{(k)}}{\partial u}(u,v)$, $C_{(k)}$ *is the copula of* $F_{(k)}$, *and* $F_{X,(k)}$ *and* $F_{\theta,(k)}$ *are marginal distributions of* $F_{(k)}$.

Corollary EC.1 allows us to efficiently compute the moments of $\boldsymbol{\theta}_{[\boldsymbol{X}]}$ and construct optimal portfolios using numerical integrals because the number of heterogeneous groups, G, is usually limited. In particular, only G double integrals are required for the expected value and variance, and G^2 quadruple integrals for the covariances of $\theta_{[i:N]}$. It is important that the computational cost increases linearly (quadratically) in the number of groups G—not the number of assets N—for the expected value and variance (covariances).

Numerical Examples. We use two examples, one with two groups and another with ten groups, to investigate how heterogeneity affects the moments of induced order statistics and the optimal weights of impact portfolios.

EXAMPLE EC.3. Consider $N=50$ assets in a universe divided into two groups: Group 1 with $N_1 = 25$ assets and Group 2 with $N_2=25$ assets. We further assume that

$$\text{Group 1: } \begin{pmatrix} X_1 \\ \theta_1 \end{pmatrix}, \begin{pmatrix} X_2 \\ \theta_2 \end{pmatrix}, \ldots, \begin{pmatrix} X_{25} \\ \theta_{25} \end{pmatrix} \overset{\text{IID}}{\sim} \mathcal{N}\left(\begin{pmatrix} \mu_X^- \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \rho\cdot\sigma_X\cdot\sigma_\theta^- \\ \rho\cdot\sigma_X\cdot\sigma_\theta^- & (\sigma_\theta^-)^2 \end{pmatrix} \right),$$
$$\text{Group 2: } \begin{pmatrix} X_{26} \\ \theta_{26} \end{pmatrix}, \begin{pmatrix} X_{27} \\ \theta_{27} \end{pmatrix}, \ldots, \begin{pmatrix} X_{50} \\ \theta_{50} \end{pmatrix} \overset{\text{IID}}{\sim} \mathcal{N}\left(\begin{pmatrix} \mu_X^+ \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \rho\cdot\sigma_X\cdot\sigma_\theta^+ \\ \rho\cdot\sigma_X\cdot\sigma_\theta^+ & (\sigma_\theta^+)^2 \end{pmatrix} \right),$$

where $\rho = 20\%$, $\sigma_X = 20\%$, and parameters μ_X^-, μ_X^+, σ_θ^-, and σ_θ^+ take the following five different setups:

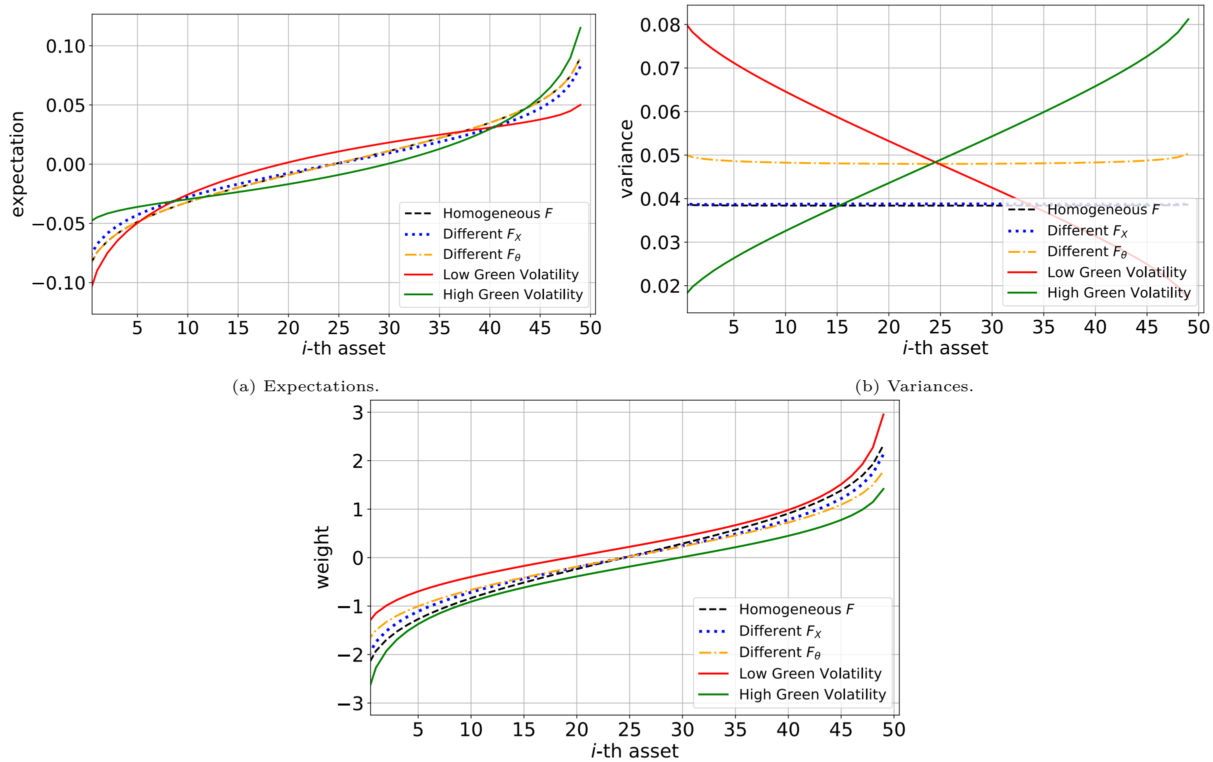
1. **Homogeneous** F: $\mu_X^- = \mu_X^+ = 0\%$, $\sigma_\theta^- = \sigma_\theta^+ = 20\%$;
2. **Different** F_X: $\mu_X^- = -10\%$, $\mu_X^+ = 10\%$, $\sigma_\theta^- = \sigma_\theta^+ = 20\%$;
3. **Different** F_θ: $\mu_X^- = \mu_X^+ = 0\%$, $\sigma_\theta^- = 30\%$, $\sigma_\theta^+ = 10\%$;
4. **Low Green Volatility**: $\mu_X^- = -10\%$, $\mu_X^+ = 10\%$, $\sigma_\theta^- = 30\%$, $\sigma_\theta^+ = 10\%$;
5. **High Green Volatility**: $\mu_X^- = -10\%$, $\mu_X^+ = 10\%$, $\sigma_\theta^- = 10\%$, $\sigma_\theta^+ = 30\%$.

Here Setup 1 is the homogeneous baseline which implies that $(X_i, \theta_i)^\top$ are IID cross-sectionally. Setup 2 assumes that Group 2 has a higher expected value (+10%) of impact factors ("green stocks"), while Group 1 has a negative expected value (−10%, "brown stocks"). The marginal distributions of θ_i are the same for both groups. Setup 3 considers two groups with the same marginal distributions of X_i but different marginal distributions of θ_i. Setup 4 assumes that green stocks (Group 1) have a lower variance in residual returns than brown stocks (Group 2), and Setup 5 considers the reverse case.

Figure EC.7 shows the expected value, variance, and optimal weights for the two groups of stocks under these setups using Corollary EC.1. First, because the impact factor is positively correlated with residual returns ($\rho > 0$), the optimal weight increases as i increases. Second, the expected value, variance, and optimal weights are symmetric with respect to the median asset for Setups 1, 2, and 3.

Third, the heterogeneity in cross-sectional distributions can lead to lower optimal weights compared to the homogeneous case. By comparing Setups 1 and 2, we observe that heterogeneity in the marginal distribution of $\boldsymbol{X}$ may reduce the magnitude of expected impact returns (Figure EC.7a), leading to lower magnitudes in optimal weights (Figure EC.7c). By comparing Setups 1 and 3, we find that heterogeneity in the marginal distribution of $\boldsymbol{\theta}$ may increase the variance of impact returns (Figure EC.7b), also resulting in lower magnitudes in optimal weights (Figure EC.7c).

Fourth, in contrast to Setup 1, the results of Setups 4 and 5 with heterogeneous return volatility show different patterns. Taking Setup 4 (Low Green Volatility) as an example, the expected impact



(a) Expectations.

(b) Variances.

(c) Weights.

Figure EC.7 Expectations, variances, and optimal weights for two groups of stocks.

returns for green stocks are lower than the negative expected impact returns for brown stocks (Figure EC.7a). In addition, the variances for green stocks are also much lower (Figure EC.7b), resulting in higher optimal weights (in absolute value) for green stocks compared to brown stocks (Figure EC.7c). The patterns are reserved for Setup 5, not surprisingly.

Example EC.4. Consider $N = 50$ assets in a universe divided into ten groups with each group containing $N_1 = N_2 = \cdots = N_{10} = 5$ assets (Groups 1 to 10, respectively). We further assume that

$$\text{Group 1: } \begin{pmatrix} X_1 \\ \theta_1 \end{pmatrix}, \begin{pmatrix} X_2 \\ \theta_2 \end{pmatrix}, \ldots, \begin{pmatrix} X_5 \\ \theta_5 \end{pmatrix} \overset{\text{IID}}{\sim} \mathcal{N}\left(\begin{pmatrix} \mu_X^{(1)} \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \rho \cdot \sigma_X \cdot \sigma_\theta^{(1)} \\ \rho \cdot \sigma_X \cdot \sigma_\theta^{(1)} & (\sigma_\theta^{(1)})^2 \end{pmatrix} \right),$$

$$\text{Group 2: } \begin{pmatrix} X_6 \\ \theta_6 \end{pmatrix}, \begin{pmatrix} X_7 \\ \theta_7 \end{pmatrix}, \ldots, \begin{pmatrix} X_{10} \\ \theta_{10} \end{pmatrix} \overset{\text{IID}}{\sim} \mathcal{N}\left(\begin{pmatrix} \mu_X^{(2)} \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \rho \cdot \sigma_X \cdot \sigma_\theta^{(2)} \\ \rho \cdot \sigma_X \cdot \sigma_\theta^{(2)} & (\sigma_\theta^{(2)})^2 \end{pmatrix} \right),$$

$$\vdots$$

$$\text{Group 10: } \begin{pmatrix} X_{46} \\ \theta_{46} \end{pmatrix}, \begin{pmatrix} X_{47} \\ \theta_{47} \end{pmatrix}, \ldots, \begin{pmatrix} X_{50} \\ \theta_{50} \end{pmatrix} \overset{\text{IID}}{\sim} \mathcal{N}\left(\begin{pmatrix} \mu_X^{(10)} \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \rho \cdot \sigma_X \cdot \sigma_\theta^{(10)} \\ \rho \cdot \sigma_X \cdot \sigma_\theta^{(10)} & (\sigma_\theta^{(10)})^2 \end{pmatrix} \right),$$

where $\rho = 20\%$, $\sigma_X = 20\%$, and parameters $\mu_X^{(j)}$ and $\sigma_\theta^{(j)}$ take the following five different setups. For $j = 1, 2, \ldots, 10$,

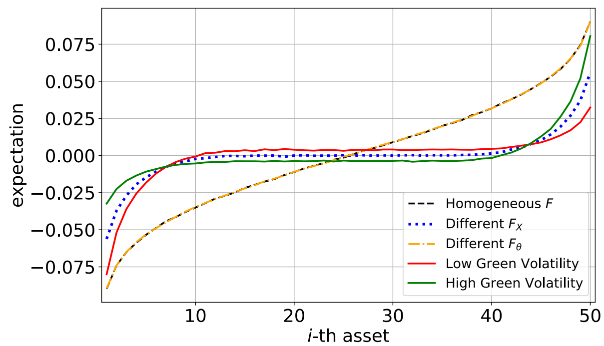
1. **Homogeneous** F: $\mu_X^{(j)} = 0\%$, $\sigma_\theta^{(j)} = 20\%$;

2. **Different** F_X: $\mu_X^{(j)} = -90\% + (j-1) \times 20\%$, $\sigma_\theta^{(j)} = 20\%$;
3. **Different** F_θ: $\mu_X^{(j)} = 0\%$, $\sigma_\theta^{(j)} = 29\% - (j-1) \times 2\%$;
4. **Low Green Volatility**: $\mu_X^{(j)} = -90\% + (j-1) \times 20\%$, $\sigma_\theta^{(j)} = 29\% - (j-1) \times 2\%$;
5. **High Green Volatility**: $\mu_X^{(j)} = -90\% + (j-1) \times 20\%$, $\sigma_\theta^{(j)} = 11\% + (j-1) \times 2\%$.

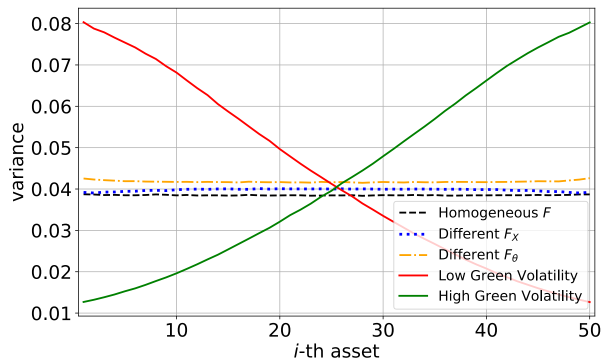
These setups are similar to those in Example EC.3, which represent various scenarios of cross-sectional heterogeneity. Setup 1 assumes a homogeneous cross-section. Setup 2 assumes that the distributions of X_i are different cross-sectionally, with expected values of $-90\%, -70\%, \ldots, 70\%, 90\%$ for the ten groups, respectively. Setup 3 assumes that the distributions of θ_i are different cross-sectionally, with idiosyncratic volatilities taking values of $11\%, 13\%, \ldots, 27\%, 29\%$, respectively. Setup 4 assumes that greener stocks have lower idiosyncratic volatilities than browner stocks, and Setup 5 considers the reverse case.

Figure EC.8 shows the expected value, variance, and optimal weights for the ten groups of stocks under these setups using Corollary EC.1. The results are similar to those for Example EC.3 in Figure EC.7. In particular, the cross-sectional heterogeneity leads to lower magnitudes of optimal weights. In addition, lower volatilities in greener stocks tend to yield higher optimal weights for them. By comparing Figure EC.8 with Figure EC.7, we can also find that the weights are flatter and closer to zero around the median asset as more groups are formed (Figure EC.8c).

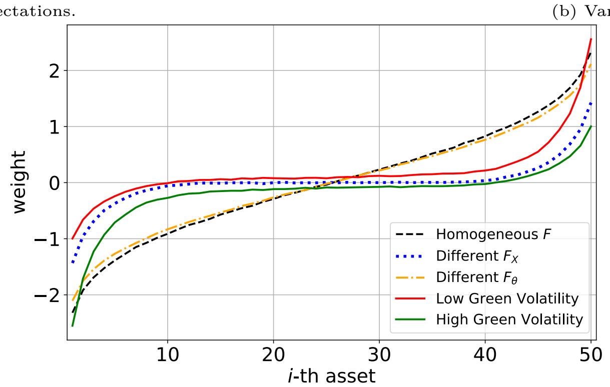
In summary, these results extend the representation theorem to accommodate cross-sectional heterogeneity, and the moments and optimal weights can be calculated efficiently, particularly when assets are divided into several groups.



(a) Expectations.



(b) Variances.



(c) Weights.

Figure EC.8 Expectations, variances, and optimal weights for ten groups of stocks.

EC.1.4. More Results for General Dependence via Copulas

This appendix is an extension of Section 4 and provides additional results regarding the influence of the copula on the impact returns.

EC.1.4.1. Moments of Impact Returns In this part, we study the moments of impact returns, $\boldsymbol{\theta}_{[\boldsymbol{X}]}$, under a general dependence structure, C. As observed in Figures EC.1a and EC.1b, under the joint normality assumption, the expectation of $\theta_{[i:N]}$ increases with the rank of the asset when the correlation between impact factor and asset returns, ρ, is positive, and decreases with the rank when ρ is negative. To demonstrate that this result holds for a general dependence structure, we need to generalize the concept of the positivity/negativity of ρ to a general copula.

Definition EC.3 (Stochastic Monotonicity (Nelsen 2007)). A copula $C(u,v)$ is *stochastically increasing* in u if $\mathscr{D}_1 C(u,v)$ is a non-increasing function of $u \in [0,1]$ for all $v \in [0,1]$, and is *stochastically decreasing* in u if $\mathscr{D}_1 C(u,v)$ is a non-decreasing function of $u \in [0,1]$ for all $v \in [0,1]$. Here, $\mathscr{D}_1 C(u,v)$ is defined by (EC.30).

Remark EC.2. Let (U,V) be a bivariate random vector with joint distribution function C. Because $\mathscr{D}_1 C(u,v)$ is a regular conditional distribution function of $C(u,v)$ given u (Remark EC.1), we can define $\mathbb{P}(V \le v | U = u) = \mathscr{D}_1 C(u,v)$. Therefore, a stochastically increasing C implies that, for a given v, $\mathbb{P}(V \le v | U = u)$ is a non-increasing function of u. In other words, $\mathbb{P}(V > v | U = u)$ is a non-decreasing function of u. This implies that a larger value of U corresponds to a higher probability that V also takes a larger value. Similarly, if C is stochastically decreasing, a larger value of U corresponds to a higher probability that V takes a lower value.

The concept of stochastic monotonicity in Definition EC.3 generalizes the notion of positivity/negativity of ρ for the bivariate normal distribution. A stochastically increasing copula implies a "positive dependence" in the distribution, indicating that large values of one variable tend to occur with large values of the other, while a stochastically decreasing copula implies a "negative dependence." For example, the copula for a bivariate normal distribution (the Gaussian copula, as we discuss in Section 4.1) is stochastically increasing if $\rho > 0$, and stochastically decreasing if $\rho < 0$. See Nelsen (2007, Section 5.2.3) for more discussions.

The following theorem characterizes the relationship between the stochastic monotonicity of the copula and the monotonicity of odd-order moments of $\theta_{[i:N]}$ with respect to the rank of the asset.

Theorem EC.5. *Under Assumption 1, assume that C is a copula of F. Then, for any $k = 0, 1, \ldots,$ if $C(u,v)$ is stochastically increasing in u, we have:*

$$\mathbb{E}\left(\theta_{[1:N]}^{2k+1}\right) \le \mathbb{E}\left(\theta_{[2:N]}^{2k+1}\right) \le \cdots \le \mathbb{E}\left(\theta_{[N:N]}^{2k+1}\right),$$

and if $C(u,v)$ is stochastically decreasing in u, we have:

$$\mathbb{E}\left(\theta_{[1:N]}^{2k+1}\right) \geq \mathbb{E}\left(\theta_{[2:N]}^{2k+1}\right) \geq \cdots \geq \mathbb{E}\left(\theta_{[N:N]}^{2k+1}\right).$$

Theorem EC.5 demonstrates that, the odd-order moments of $\theta_{[i:N]}$ increase with i if the copula is stochastically increasing, and decrease with i if the copula is stochastically decreasing. In particular, when $k=0$, this theorem gives the ranking of the expectations of induced order statistics. This implies that, if an impact investor forms portfolios only based on the expectation of impact returns, she may put higher weights on assets with higher values of the impact factor (larger i) when the impact factor has a positive dependence with residual returns. These findings explain our observations in Figures EC.1a and EC.1b under the joint normality assumption.

EC.1.4.2. Impact Portfolios Under Gumbel Copula In Section 4.2 of our main paper, we study the impact returns when the copula is a Clayton copula. In contrast to the Clayton copula, which characterizes lower tail dependence, the Gumbel copula characterizes upper tail dependence between $\boldsymbol{X}$ and $\boldsymbol{\theta}$.

Definition EC.4 (Gumbel Copula). The Gumbel copula with parameter $\gamma \in [1, +\infty)$ is defined as an Archimedean copula with generator function

$$\psi_\gamma^{\mathrm{Gu}}(u) \equiv (-\ln u)^\gamma, \quad u \in [0,1]. \tag{EC.50}$$

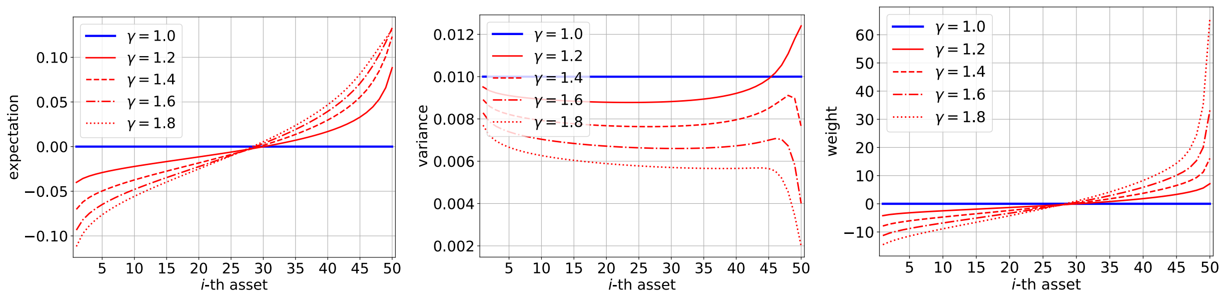
By replacing the generator function $\psi_\gamma^{\mathrm{Gu}}$ of the Gumbel copula into Proposition 5, we can derive the corresponding distribution of the induced order statistics, $\boldsymbol{\theta}_{[\boldsymbol{X}]}$ (for brevity, we write the i-th entry only):

$$\theta_{[i:N]} \stackrel{d}{=} F_\theta^{-1}\left[\exp\left(-\left[(\phi_\gamma^{\mathrm{Gu}})'^{-1}\left(-\frac{1}{\gamma}U_{i:N}V_i(-\ln U_{i:N})^{1-\gamma}\right) - (-\ln U_{i:N})^\gamma\right]^{1/\gamma}\right)\right], \tag{EC.51}$$

where $\phi_\gamma^{\mathrm{Gu}}(x) = (\psi_\gamma^{\mathrm{Gu}})^{-1}(x) = \exp(-x^{1/\gamma})$ for $x \geq 0$.[25]

By taking the limits of (EC.51) with respect to γ, we can show that the right-hand side of (EC.51) converges to $F_\theta^{-1}(U_{i:N})$ as γ increases without bound, and converges to $F_\theta^{-1}(V_i)$ as γ approaches 1. Therefore, like the Clayton copula, the parameter γ in the Gumbel copula determines the relative importance of the order statistics, $U_{i:N}$, and the uniform noise, V_i.

We then construct the optimal impact portfolios under the Gumbel copula using (EC.51) and Proposition 2. Figure EC.9 adopts the same parameter configurations as Figure 1, except that the copula is Gumbel rather than Clayton. Like the results for the Clayton copula, assets with higher impact factors have higher expectations (Figure EC.9a), which is consistent with Theorem EC.5, while the variances do not show monotonicity (Figure EC.9b). Assets with higher impact factors generally have higher optimal weights (Figure EC.9c). In addition, as the parameter γ increases, the dispersion of expected residual returns across assets increases (Figure EC.9a), and the level of variance decreases (Figure EC.9b). Together, this lead to more dispersed optimal weights for larger values of γ (Figure EC.9c).



(a) Expectation. (b) Variance. (c) Weight.

Figure EC.9 Expectations and variances of $\boldsymbol{\theta}_{[\boldsymbol{X}]}$, and optimal weights assuming a Gumbel copula. The marginal distribution of $\boldsymbol{\theta}$ is $\mathcal{N}(0,\sigma_\theta^2)$ with $\sigma_\theta = 10\%$. We set $N = 50$ for illustrative purposes.

Figure EC.9 also implies an asymmetric distribution of residual returns and optimal weights. However, compared to the results for the Clayton copula, the asymmetry is reversed—under the Gumbel copula, the magnitudes of weights for short positions are *lower* than those for long positions. This is consistent with the fact that the Gumbel copula characterizes upper tail dependence, meaning that the dependence between the impact factor and residual returns will be higher for assets with higher impact factors. As expected, this effect is opposite to the results for the Clayton copula.

EC.1.4.3. Impact Portfolios Under Fundamental Copula Yang et al. (2006) show that a bivariate copula can be decomposed into a convex combination of three fundamental copulas and an indecomposable part. These three fundamental copulas, namely the comonotonicity copula, the countermonotonicity copula, and the independence copula, are defined as follows.

Definition EC.5 (Fundamental Copulas). The comonotonicity copula, countermonotonicity copula, and independence copula are defined as:

$$C^+(u,v) \equiv \min\{u,v\}, \quad C^-(u,v) \equiv \max\{u+v-1,0\}, \quad C^\perp(u,v) \equiv uv,$$

for $u,v \in [0,1]$, respectively.

Remark EC.3. Assume that (U,V) is a bivariate random vector with both U and V uniformly distributed on $(0,1)$. One can verify that, if $U \equiv V$, (U,V) follows a comonotonicity copula; if $U \equiv -V$, (U,V) follows a countermonotonicity copula; and if U is independent of V, (U,V) follows an independence copula. Therefore, these three fundamental copulas correspond to three extreme cases: perfect positive correlation, perfect negative correlation, and independence.

The following proposition gives the distribution of $\boldsymbol{\theta}_{[\boldsymbol{X}]}$ under the three fundamental copulas using the representation theorem.

Proposition EC.11. *Under Assumption 1, if F_X is continuous, the distributions of $\boldsymbol{\theta}_{[\boldsymbol{X}]}$ are as follows for the three fundamental copulas:*

(a) For the comonotonicity copula, $(\theta_{[1:N]}, \theta_{[2:N]}, \ldots, \theta_{[N:N]}) \stackrel{d}{=} (\theta_{1:N}, \theta_{2:N}, \ldots, \theta_{N:N})$;

(b) For the countermonotonicity copula, $(\theta_{[1:N]}, \theta_{[2:N]}, \ldots, \theta_{[N:N]}) \stackrel{d}{=} (\theta_{N:N}, \theta_{N-1:N}, \ldots, \theta_{1:N})$;

(c) For the independence copula, $(\theta_{[1:N]}, \theta_{[2:N]}, \ldots, \theta_{[N:N]}) \stackrel{d}{=} (\theta_1, \theta_2, \ldots, \theta_N)$.

Proposition EC.11 is intuitive. When $\boldsymbol{X}$ and $\boldsymbol{\theta}$ are independent, ranking by $\boldsymbol{X}$ does not affect the order of $\boldsymbol{\theta}$. If $\boldsymbol{X}$ and $\boldsymbol{\theta}$ are comonotonic, ranking by $\boldsymbol{X}$ is the same as ranking by $\boldsymbol{\theta}$, and if $\boldsymbol{X}$ and $\boldsymbol{\theta}$ are countermonotonic, ranking by $\boldsymbol{X}$ is opposite to ranking by $\boldsymbol{\theta}$.

Although the fundamental copulas themselves are too extreme to be directly used in practice, they offer insights from two perspectives. First, they represent extreme cases of many widely used copulas. For example, the Gaussian copula converges to the comonotonicity (countermonotonicity) copula as ρ approaches $+1$ (-1), and reduces to an independence copula when $\rho = 0$. Second, because any bivariate copula can be decomposed into a convex combination of the three fundamental copulas and an indecomposable part (Yang et al. 2006), in practice, we can approximate any dependence structure between $\boldsymbol{X}$ and $\boldsymbol{\theta}$ by combining the three fundamental copulas and use the representation theorem to approximate the distribution of $\boldsymbol{\theta}_{[\boldsymbol{X}]}$.

EC.1.4.4. Impact Portfolios Under Elliptical Copula The elliptical copula is a generalization of the Gaussian copula.

Definition EC.6 (Elliptical Copula). The bivariate elliptical copula with generator $\boldsymbol{\Psi}_\rho$ and parameter $\rho \in (-1, 1)$ is defined as

$$C^{\text{El}}_{\boldsymbol{\Psi}_\rho}(u, v) \equiv \boldsymbol{\Psi}_\rho \left(\Psi^{-1}(u), \Psi^{-1}(v) \right),$$

where Ψ is the marginal distribution function of $\boldsymbol{\Psi}_\rho$, and $\boldsymbol{\Psi}_\rho$ is the joint distribution function of a bivariate elliptical distribution with density

$$f_{\boldsymbol{\Psi}_\rho}(x_1, x_2) \equiv c \cdot f\left(\frac{x_1^2 + x_2^2 - 2\rho x_1 x_2}{1 - \rho^2} \right), \quad x_1, x_2 \in \mathbb{R},$$

where $c > 0$ is the normalizing constant and function $f(x)$ satisfies $f(x) \geq 0$ for all x.

Remark EC.4. Many commonly used copulas are elliptical. For example, when $\boldsymbol{\Psi}_\rho$ is the joint distribution function of a bivariate normal distribution, $C^{\text{El}}_{\boldsymbol{\Psi}_\rho}$ is a Gaussian copula; and when $\boldsymbol{\Psi}_\rho$ is the joint distribution function of a bivariate t-distribution, $C^{\text{El}}_{\boldsymbol{\Psi}_\rho}$ is a t-copula. Parameter ρ characterizes the strength of dependence between $\boldsymbol{X}$ and $\boldsymbol{\theta}$.

The following proposition characterizes the distribution of $\boldsymbol{\theta}_{[\mathbf{X}]}$ when the copula is elliptical and the parameter ρ takes opposite signs.

PROPOSITION EC.12. *Consider two sets of IID bivariate random vectors,* $(X_i^{(1)}, \theta_i^{(1)})$ *and* $(X_i^{(2)}, \theta_i^{(2)})$, $i = 1, 2, \ldots, N$. *Assume that the marginal distributions of* $X_i^{(1)}$ *and* $X_i^{(2)}$ *are continuous. Denote their copulas and the marginal distributions of* $\theta_i^{(j)}$ *by* $C^{(j)}$ *and* $F_\theta^{(j)}$, *respectively, for* $j = 1, 2$. *Given* $\rho \in (-1, 1)$, *if* $F_\theta^{(1)} = F_\theta^{(2)}$, $C^{(1)} = C^{\mathrm{El}}_{\boldsymbol{\Psi}_\rho}$, *and* $C^{(2)} = C^{\mathrm{El}}_{\boldsymbol{\Psi}_{-\rho}}$, *we have:*

$$\left(\theta^{(1)}_{[1:N]}, \theta^{(1)}_{[2:N]}, \ldots, \theta^{(1)}_{[N:N]}\right) \stackrel{d}{=} \left(\theta^{(2)}_{[N:N]}, \theta^{(2)}_{[N-1:N]}, \ldots, \theta^{(2)}_{[1:N]}\right), \tag{EC.52}$$

where $(\theta^{(j)}_{[1:N]}, \theta^{(j)}_{[2:N]}, \ldots, \theta^{(j)}_{[N:N]})$ *are the induced order statistics of* $\theta_1^{(j)}, \theta_2^{(j)}, \ldots, \theta_N^{(j)}$ *ranked by* $X_1^{(j)}, X_2^{(j)}, \ldots, X_N^{(j)}$, *for* $j = 1, 2$. *Furthermore, the optimal weights of maximizing the information ratio under the two setups,* $(w_1^{*(j)}, w_2^{*(j)}, \ldots, w_N^{*(j)}), j = 1, 2$, *satisfy:*

$$(w_1^{*(1)}, w_2^{*(1)}, \ldots, w_N^{*(1)}) \propto (w_N^{*(2)}, w_{N-1}^{*(2)}, \ldots, w_1^{*(2)}). \tag{EC.53}$$

Proposition EC.12 implies that, when the dependence structure between $\boldsymbol{X}$ and $\boldsymbol{\theta}$ is elliptical, reversing the sign of the parameter ρ leads to a reversal in the distribution of residual returns and the optimal weights for the N ranked assets. This symmetry, as demonstrated in Figures EC.1 and EC.2, highlights that the distribution of residual returns and the optimal weights are symmetric with respect to $\pm\rho$. In particular, when ρ is negative, the optimal strategy for maximizing the information ratio is to go long on the assets with low impact factors, contradicting the goal of impact investing. Investors face a tradeoff between portfolio performance and impact when $\rho < 0$.

EC.1.5. More Results for General Return Distributions

This appendix is an extension of Section 5 and provides additional results regarding the influence of the marginal distribution of residual returns on the impact returns.

EC.1.5.1. Numerical Example for Skewed Returns To validate the theoretical results presented in Proposition 6 for skewed distributions, we consider one commonly used skewed distribution family, the skew-normal distribution (Azzalini and Capitanio 1999), which is defined as follows.

DEFINITION EC.7 (SCALED-SKEW-NORMAL DISTRIBUTION). A random variable θ follows a scaled-skew-normal distribution with parameters (α, σ_θ) if $\theta \overset{d}{=} \sigma_\theta \cdot [S - \mathbb{E}(S)]/\sqrt{\text{Var}(S)}$, where $\sigma_\theta > 0$, the random variable S follows the skew-normal distribution with density function:

$$f_S(x) = 2\varphi(x)\Phi(\alpha x), \quad x \in \mathbb{R},$$

and φ and Φ are the density function and distribution function of the standard normal distribution, respectively. We denote this by $\theta \sim \text{ScaleSkewNorm}(\alpha, \sigma_\theta)$.

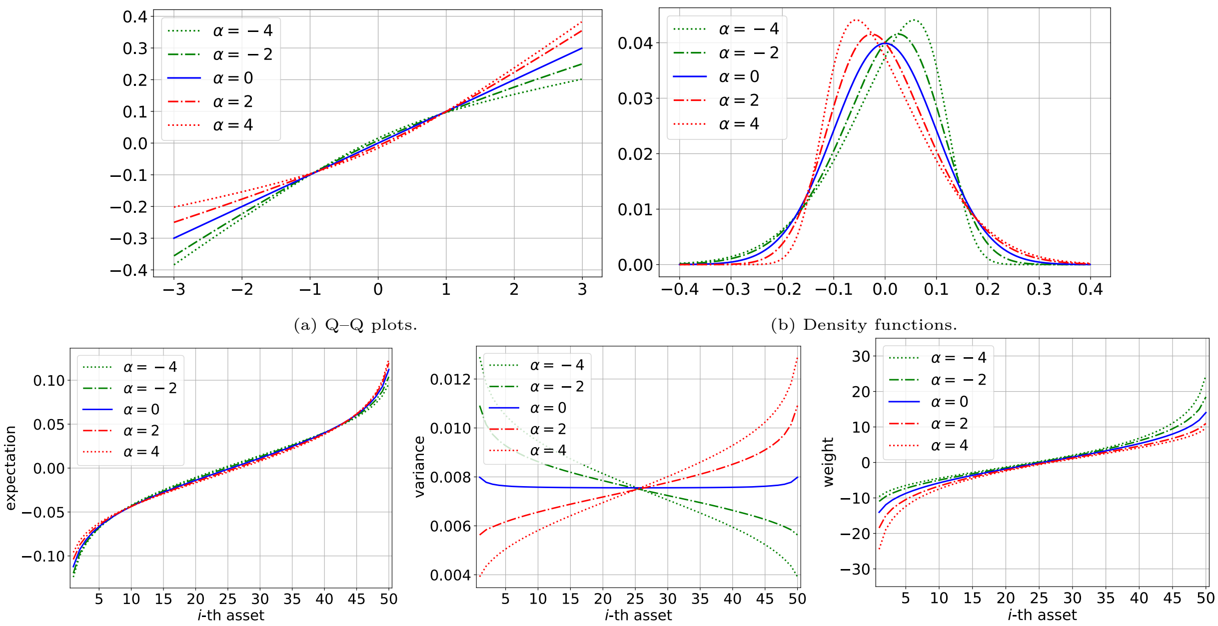
Figure EC.10a and Figure EC.10b are the Q–Q plot and the density function of the scaled-skew-normal distribution, respectively. The Q–Q plot is asymmetric on both tails for $\alpha \neq 0$. Figure EC.10b demonstrates that α parameterizes the skewness. The distribution is positively skewed when $\alpha > 0$, and negatively skewed when $\alpha < 0$. When $\alpha = 0$, by definition, it reduces to the normal distribution.

Figure EC.10c and Figure EC.10d show the expectations and variances of the induced order statistics, $\boldsymbol{\theta}_{[\boldsymbol{X}]}$, and Figure EC.10e shows the optimal weights, with the number of assets $N = 50$. The copula is set to be a Gaussian copula with parameter $\rho = 50\%$, and $\boldsymbol{\theta}$ is scaled-skew-normally distributed. In particular, Figure EC.10e demonstrates that, when $\alpha > 0$ (positively skewed), the optimal weights are smaller for top-ranking assets, and when $\alpha < 0$ (negatively skewed), the optimal weights are smaller for bottom-ranking assets. This finding is consistent with Proposition 6.

Furthermore, Figure EC.10c implies that the expectations of $\boldsymbol{\theta}_{[\boldsymbol{X}]}$ vary slightly for different α, but from Figure EC.10d, we find that different α leads to a strong dispersion in variances across assets, and the variances on the more skewed tail are larger. Therefore, the asymmetry of optimal weights is mainly driven by the dispersion in variance (risk), rather than the expectation.

EC.1.5.2. Impact Portfolios Under Symmetric Returns In practice, practitioners often use symmetric distributions, such as the normal distribution or the t-distribution, to model asset returns, which implies a skewness of zero. This section studies the distribution of $\boldsymbol{\theta}_{[\boldsymbol{X}]}$ and the optimal impact portfolio construction when the marginal distribution of $\boldsymbol{\theta}$ is symmetric.

The following proposition provides a sufficient condition under which the distribution of $\boldsymbol{\theta}_{[\boldsymbol{X}]}$ and the optimal weights are symmetric, given a symmetrically distributed $\boldsymbol{\theta}$.



(a) Q–Q plots. (b) Density functions.

(c) Expectations. (d) Variances. (e) Weights.

Figure EC.10 Figure EC.10a and Figure EC.10b are the Q–Q plot and density function of ScaleSkewNorm(α, σ_θ). Figures EC.10c–EC.10e are expectations, variances of $\boldsymbol{\theta}_{[\boldsymbol{X}]}$, and the optimal weights under Assumption 1. The copula is Gaussian with parameter ρ, and the marginal distribution of $\boldsymbol{\theta}$ is ScaleSkewNorm(α, σ_θ). We set $N = 50$, $\rho = 50\%$, and $\sigma_\theta = 10\%$ for illustrative purposes.

Proposition EC.13. *Under Assumption 1, assume that both F_X and F_θ are continuous functions, and C is a copula of F with a density. Further assume that F_θ is symmetric with respect to μ_θ:*

$$F_\theta(\mu_\theta + x) = 1 - F_\theta(\mu_\theta - x), \quad \forall x \in \mathbb{R}, \tag{EC.54}$$

and the copula C of F is radially symmetric:

$$C(u, v) = u + v - 1 + C(1-u, 1-v), \quad u, v \in [0,1]. \tag{EC.55}$$

Then the induced order statistics, $\boldsymbol{\theta}_{[\boldsymbol{X}]}$, satisfy:

$$(\theta_{[1:N]} - \mu_\theta, \theta_{[2:N]} - \mu_\theta, \ldots, \theta_{[N:N]} - \mu_\theta) \stackrel{d}{=} (\mu_\theta - \theta_{[N:N]}, \mu_\theta - \theta_{[N-1:N]}, \ldots, \mu_\theta - \theta_{[1:N]}).$$

Furthermore, the optimal weights to maximize the information ratio, $(w_1^*, w_2^*, \ldots, w_N^*)$, satisfy:

$$(w_1^*, w_2^*, \ldots, w_N^*)^\top + (w_N^*, w_{N-1}^*, \ldots, w_1^*)^\top \propto 2\mu_\theta \Sigma^{-1} \mathbf{1},$$

where Σ is the covariance matrix of $\boldsymbol{\theta}_{[\boldsymbol{X}]}$, and $\mathbf{1}$ is an all-one vector. In particular, if $\mu_\theta = 0$, we have:

$$(w_1^*, w_2^*, \ldots, w_N^*)^\top = -(w_N^*, w_{N-1}^*, \ldots, w_1^*)^\top.$$

Remark EC.5. If the joint distribution of a random vector (U,V) follows a radially symmetric copula, for any u and v in $[0,1]$, the probabilities of (U,V) being in the regions $[0,u]\times[0,v]$ and $[1-u,1]\times[1-v,1]$ are always equal. By definition, all elliptical copulas (e.g., the Gaussian copula, See Appendix EC.1.4.4) are radially symmetric. See Nelsen (2007).

Proposition EC.13 implies that the distribution of $\boldsymbol{\theta}_{[\boldsymbol{X}]}$ and the optimal weights are symmetric with respect to its long and short positions, when both the marginal distribution of $\boldsymbol{\theta}$ is symmetric and the copula is radially symmetric. The case of joint normality discussed in Section EC.1.2 is a special case of Proposition EC.13, and the symmetry of the distribution of $\boldsymbol{\theta}_{[\boldsymbol{X}]}$ and the optimal weights is illustrated in Figures EC.1 and EC.2. However, this symmetry does not hold when $\boldsymbol{\theta}$ is symmetrically distributed and the copula is not radially symmetric, as shown in Figures 1 and EC.9 for the Clayton and Gumbel copulas.

EC.2. Empirical Study

In this appendix, we use real data to demonstrate that asset returns and impact factors are both highly non-normal in practice, and that accounting for general marginal distributions and dependences using our framework can achieve superior impact portfolio performance.

EC.2.1. Data

Our study relies on two types of data: stock returns and Fama–French factor returns, and various measures of carbon emission for individual companies. The latter is an important subject of interest in the rapidly growing literature on climate finance (Bolton and Kacperczyk 2021, 2023) as well as a major component in ESG scores (Pástor et al. 2022).

The measures of carbon emission for individual companies come from the Trucost Environmental dataset.[26] It includes data for 3,969 US companies from 2005 to 2020, offering one of the most extensive historical datasets in this domain. Table EC.3 shows the number of covered companies each year, in which a sharp increase occurred in 2016 (from 1,066 to 2,894) due to the expansion in coverage from only large-cap companies before 2016 to more small- and mid-cap companies after 2016.

Table EC.3 The number of covered companies each year.

Year	2005	2006	2007	2008	2009	2010	2011	2012
Number	956	957	953	959	978	966	957	956
Year	2013	2014	2015	2016	2017	2018	2019	2020
Number	1,061	1,085	1,066	2,894	2,928	2,920	2,937	3,286

We use two categories of annual carbon emission measures included in the Trucost Environmental data: (1) the total level (in tons of CO_2) and (2) the intensity (the ratio of total emissions to revenue). Both (1) and (2) are further classified into three different sources of emissions according to the Greenhouse Gas Protocol—Scopes 1, 2, and 3.[27] Nearly 100% company–years in our sample have valid values for all of these measures.[28] To mitigate the impact of outliers, we follow Bolton and Kacperczyk (2021) and Lo et al. (2022) to winsorize (2) at the 2.5% level and take the natural logarithm of (1) to obtain the log level of carbon emission.

We obtain monthly dividend-adjusted return data for US companies from 2006 to 2021 from the CRSP dataset,[29] which covers monthly returns for US stocks listed on the NYSE, AMEX, and NASDAQ. We obtain monthly Fama–French five-factor (Fama and French 2015) data from Kenneth R. French's website,[30] which includes the time series of the market factor, the size factor (Small Minus Big, i.e., SMB), the value factor (High Minus Low, i.e., HML), the profitability factor (Robust

Minus Weak, i.e., RMW), and the investment factor (Conservative Minus Aggressive, i.e., CMA). The risk-free rate is also provided.

The datasets we use in this appendix are also used by recent studies in this literature (Bolton and Kacperczyk 2021, Lo et al. 2022, Pástor et al. 2022). In particular, Lo et al. (2022) form green portfolios using the bivariate normality framework of Lo and Zhang (2023). The framework in this study allows for much more general distributions of the impact factor and residual returns. The following sections illustrate how this increased flexibility enables us to better fit real data and achieve higher risk-adjusted impact returns.

EC.2.2. Distribution of Impact Factor and Residual Returns

Here we demonstrate that the distribution of $(X_i, \theta_i)^\top$ is highly non-normal, which motivates our framework to allow for general distributions and dependences. We use each measure of carbon emission as the impact factor, $\boldsymbol{X}$. The residual return of each stock, θ_i, is estimated by a rolling-window Fama–French five-factor regression using its monthly returns in the previous five years.

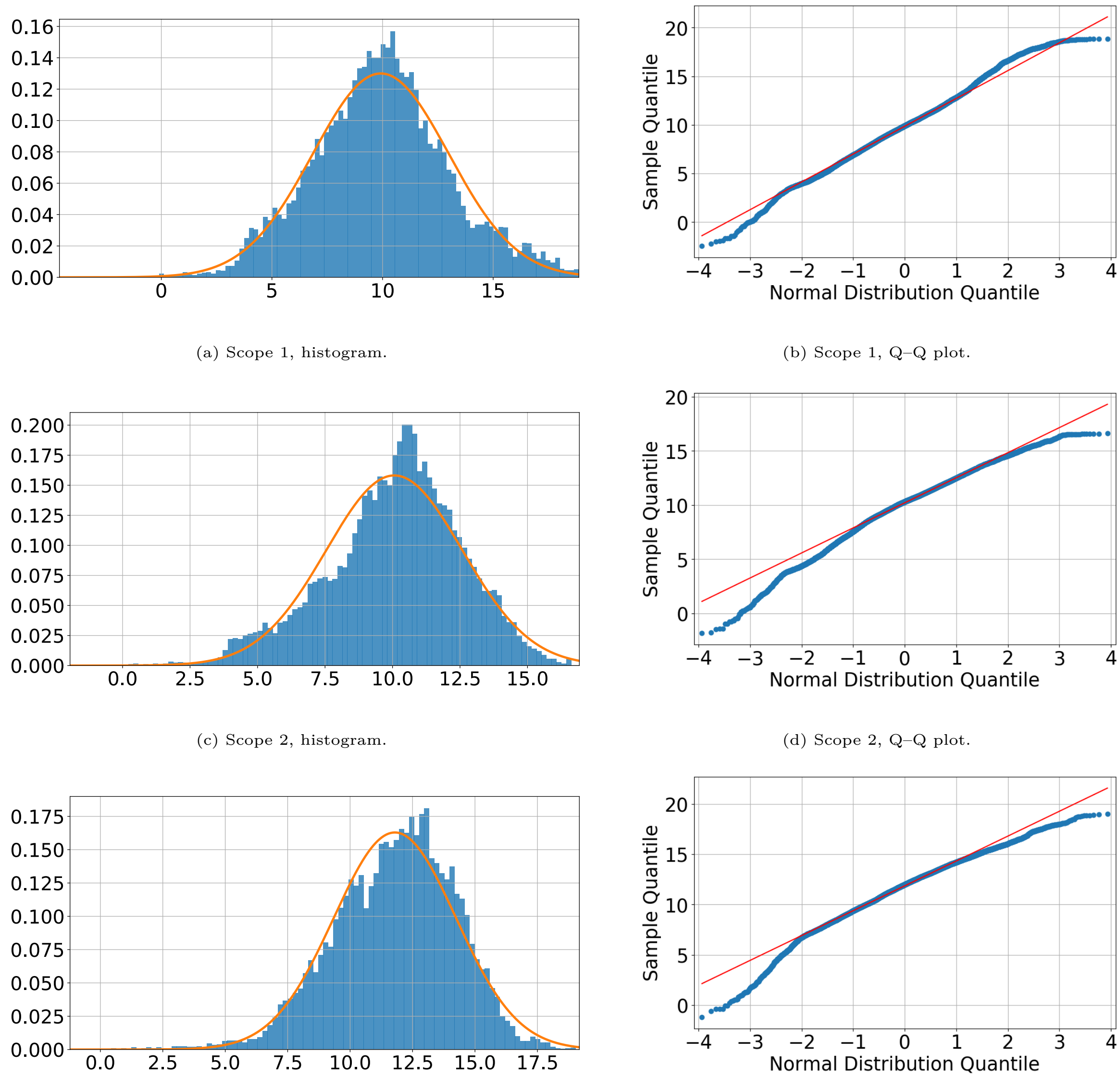
For each sample $(X_i, \theta_i)^\top$ in a particular year t, X_i is the value of the impact factor at the end of year $t-1$ and θ_i is the residual return in year t.[31] The one-year lag of the impact factor, $\boldsymbol{X}$, is to make sure that portfolios are constructed using information available in the previous year, because the impact factors, $\boldsymbol{X}$, in the Trucost Environmental data are updated annually.

Figure EC.11 and Figure EC.12 show histograms and Q–Q plots (with respect to the standard normal distribution) for the log levels and intensities of carbon emissions across the three scopes, respectively. In both figures, each sample corresponds to a company–year. The solid lines in all figures are the densities and Q–Q plots of normal distributions fitted using the samples.

Figures EC.11 and EC.12 reveal that the distributions of neither log carbon emissions nor carbon intensity follow a normal distribution. The distribution of log carbon emissions deviates from the normal distribution considerably at both tails. The distribution of carbon intensity departs even more significantly from the normal distribution.

Figure EC.13 shows the histogram and Q–Q plot of monthly residual returns, with each sample corresponding to a company–month. The solid lines represent the density and Q–Q plot of the normal distribution fitted using the data. The figure demonstrates that the residual returns have much heavier tails compared to the normal distribution. This is consistent with empirical findings in the literature; see, for example, Cont (2001).

Table EC.4 shows the summary statistics for the impact factors, $\boldsymbol{X}$, and residual returns, $\boldsymbol{\theta}$, including the average value, standard deviation, minimum value, 25th percentile, median (50th percentile), 75th percentile, maximum value, skewness, and kurtosis (in excess of the kurtosis of a normal distribution). We also conduct the Kolmogorov–Smirnov (KS) normality test for the marginal distributions



(a) Scope 1, histogram. (b) Scope 1, Q–Q plot.

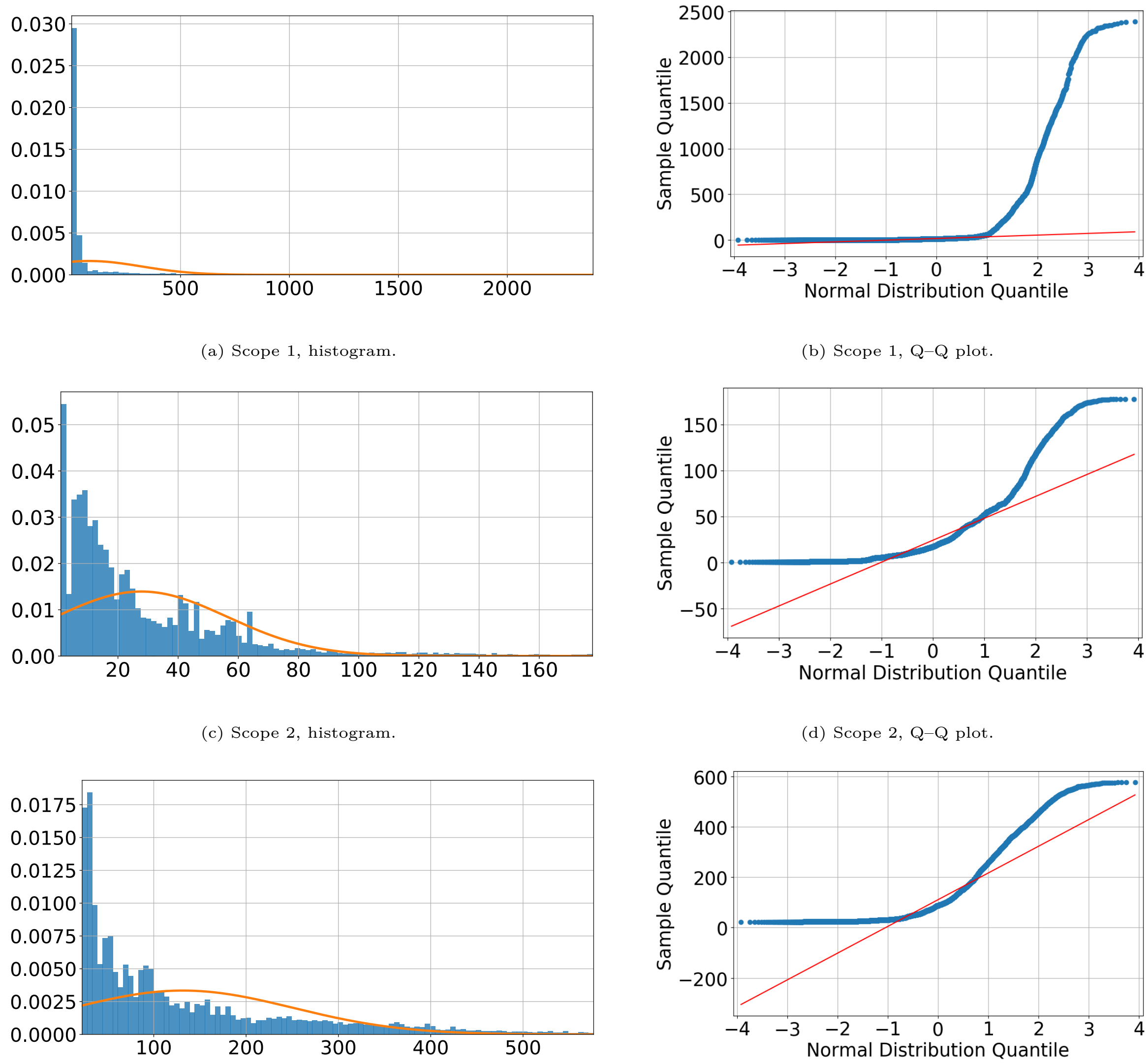
(c) Scope 2, histogram. (d) Scope 2, Q–Q plot.

(e) Scope 3, histogram. (f) Scope 3, Q–Q plot.

Figure EC.11 Histograms and Q–Q plots for log levels of carbon emissions. Solid lines are the densities and Q–Q plots of normal distributions fitted using the samples.

of the impact factor and residual returns. The test statistics and associated p-values are reported in the last two rows. The results indicate that the null hypothesis of normality is significantly rejected for all impact factors and residual returns in our sample.

To further confirm that the impact factor and residual returns, $(X_i, \theta_i)^\top$, are not jointly normally distributed throughout our entire sample period, we perform the Henze–Zirkler test (Henze and Zirkler 1990) for each month, where each sample corresponds to a company. Figure EC.14 shows the monthly time series of the test statistics and p-values.



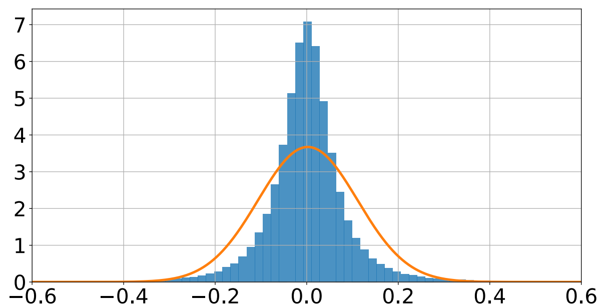
(a) Scope 1, histogram. (b) Scope 1, Q–Q plot.

(c) Scope 2, histogram. (d) Scope 2, Q–Q plot.

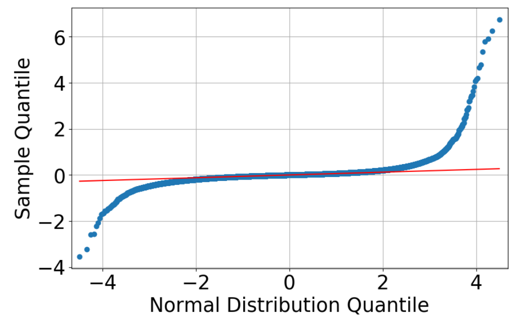
(e) Scope 3, histogram. (f) Scope 3, Q–Q plot.

Figure EC.12 Histograms and Q–Q plots for carbon emission intensity. Solid lines are the densities and Q–Q plots of normal distributions fitted using the samples.

We have three observations. First, at a 5% significance level, $(X_i, \theta_i)^\top$ is not normally distributed for all months in our sample and for all impact factors. Second, the test statistic tends to be higher when $\boldsymbol{X}$ represents carbon intensity measures compared to when $\boldsymbol{X}$ represents log levels of carbon emissions. This observation aligns with the findings from Figures EC.11–EC.12 and Table EC.4, which indicate that carbon intensity measures deviate further from the normal distribution than the log levels of carbon emissions. Third, the test statistic experiences a sharp increase from 2016 to 2017. This can be attributed to the inclusion of small- and mid-cap companies in the dataset after 2016 (see Table EC.3).



(a) Histogram.

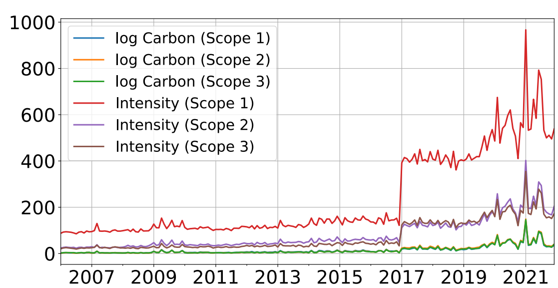


(b) Q–Q plot.

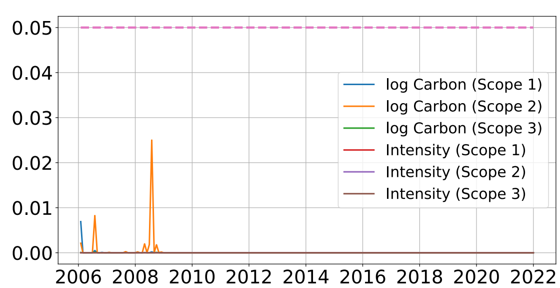
Figure EC.13 Histogram and Q–Q plot for residual returns. Solid lines are the density and Q–Q plot of the normal distribution fitted using the samples.

Table EC.4 Summary statistics and normality tests for impact factors, $\boldsymbol{X}$, and residual returns, $\boldsymbol{\theta}$. For $\boldsymbol{X}$, each sample corresponds to a company–year; for $\boldsymbol{\theta}$, each sample corresponds to a company–month.

	Log Carbon Emission			Carbon Emission Intensity			Residual Return
	Scope 1	Scope 2	Scope 3	Scope 1	Scope 2	Scope 3	
average	9.92	10.07	11.79	80.47	27.75	131.38	0.24%
std	3.07	2.52	2.45	240.70	28.61	119.38	10.86%
min	-2.43	-1.81	-1.19	0.47	1.02	22.43	-353.53%
25th percentile	7.92	8.64	10.19	4.26	8.26	38.94	-4.04%
median	9.91	10.31	11.99	13.87	17.46	87.50	0.04%
75th percentile	11.78	11.76	13.52	29.17	40.36	181.85	4.10%
max	18.86	16.63	19.03	2391.61	177.69	576.48	674.78%
skewness	0.08	-0.49	-0.49	5.38	2.06	1.41	6.25
ex. kurtosis	0.18	0.42	0.68	34.04	5.31	1.32	287.39
KS statistic	0.99	0.99	1.00	0.81	0.92	1.00	0.40
p-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00



(a) Henze–Zirkler test statistics.



(b) p-values.

Figure EC.14 Time series of the statistics and p-values of the Henze–Zirkler test.

EC.2.3. Performance of Impact Portfolios

We use both the log levels and intensities of carbon emissions for scopes 1, 2, and 3 to form impact portfolios, respectively. The *negative* values of each measure are used as the impact factor, $\boldsymbol{X}$, because lower levels of carbon emissions correspond to greener companies. We consider the following long/short portfolios.

Equal-Weighted Portfolios. In each year, we simply go long the top 50% of stocks with equal weights and short the bottom 50% of stocks with equal weights.

Bivariate Normality Portfolios. We assume that the joint normality assumption (Assumption EC.1) holds. If the estimated correlation $\rho > 0$, we use the optimal weights given by (9). Otherwise, we go long the top 50% of stocks with equal weights, and short the bottom 50% of stocks with equal weights, which reduces to the Equal-Weighted portfolio.

Gaussian Copula Portfolios. We assume that the general assumption (Assumption 1) holds with a Gaussian copula, and F_θ is a scaled-t-distribution with parameters $(1, \sigma_\theta)$. If the estimated parameter of the Gaussian copula $\rho > 0$, we use the optimal weights given by (9). Otherwise, we go back to the Equal-Weighted portfolio.

Clayton Copula Portfolios. We assume that the general assumption (Assumption 1) holds with a Clayton copula, and F_θ is a scaled-t-distribution with parameters $(1, \sigma_\theta)$. If the estimated parameter of the Clayton copula $\gamma > 0$, we use the optimal weights given by (9). Otherwise, we go back to the Equal-Weighted portfolio.

Gumbel Copula Portfolios. We assume that the general assumption (Assumption 1) holds with a Gumbel copula, and F_θ is a scaled-t-distribution with parameters $(1, \sigma_\theta)$. If the estimated parameter of the Gumbel copula $\gamma > 1$, we use the optimal weights given by (9). Otherwise, we go back to the Equal-Weighted portfolio.

Empirical Portfolios. We assume that the general assumption (Assumption 1) holds. Both the copula, C, and the marginal distribution, F_θ, are estimated nonparametrically using empirical distributions.[32]

Out of these six portfolios, the Equal-Weighted Portfolio and the Bivariate Normality Portfolio based on the joint normality assumption in Lo and Zhang (2023) serve as baselines. The Gaussian, Clayton, and Gumbel Copula Portfolios capture more general dependence structures and allow for heavy-tailed marginal distributions. The Empirical Portfolios are more flexible because they use the empirical marginal distributions and copulas and avoid choosing between various parametric copulas.

Next, we describe the procedure for estimating the parameters required for constructing the portfolios described above.[33] For the first five portfolios with parametric copulas, in each month, we calculate the maximum likelihood estimation of the parameters (for example, the correlation coefficient for the Bivariate Normality Portfolio, ρ) using the cross-sectional data, $(X_i, \theta_i)^\top$, $i = 1, 2, \ldots, N$. Then, for each year, we use the average estimated parameters over the previous five years to construct portfolios. For Empirical Portfolios, in each year, we use all cross-sectional data over the previous five years to estimate the empirical marginal distribution and the empirical copula. In addition, after determining the optimal weights, for each long/short portfolio, we standardize the optimal weights by requiring that $\sum_{i=1}^{N}|w_i|=1$ to ensure the same level of leverage across portfolios.

We test the profitability of all strategies from 2011 to 2021.[34] Tables EC.5 and EC.6 summarize the performance of portfolios constructed using the log levels and intensities of carbon emissions, respectively. In particular, we report their annualized raw return (return), standard deviation (std.), Sharpe ratio (SR), alpha from the Fama–French five-factor model (FF5 α), volatility of residual returns ($\sigma(\theta_p)$), information ratio (IR), and annual turnover.[35] Note that we report metrics related to both raw returns (return, std., and SR) and residual returns (FF5 α, $\sigma(\theta_p)$, and IR).[36] In addition to these performance metrics, Figure EC.15 visualizes the cumulative residual returns for these portfolios.

Table EC.5 Performance of impact portfolios constructed using the log levels of carbon emissions. All metrics in this exhibit are annualized.

	Equal-Weight	Bivariate Normality	Gaussian Copula	Clayton Copula	Gumbel Copula	Empirical
			Scope 1			
return	0.75%	1.03%	15.79%	3.27%	14.75%	15.97%
std.	3.17%	4.54%	18.96%	5.14%	19.16%	18.42%
SR	0.20	0.20	0.83	0.61	0.76	0.86
FF5 α	1.78%	2.38%	3.61%	2.39%	3.51%	2.85%
$\sigma(\theta_p)$	2.13%	3.06%	5.41%	3.03%	5.86%	4.47%
IR	0.83	0.78	0.67	0.79	0.60	0.64
turnover	43.99%	41.72%	36.00%	39.12%	96.11%	35.18%
			Scope 2			
return	0.56%	0.66%	16.01%	1.31%	15.32%	15.79%
std.	3.02%	4.36%	19.05%	4.05%	19.19%	18.49%
SR	0.15	0.12	0.83	0.29	0.79	0.85
FF5 α	1.81%	2.39%	4.09%	1.85%	4.57%	2.88%
$\sigma(\theta_p)$	1.82%	2.67%	5.89%	2.26%	6.49%	4.60%
IR	0.99	0.89	0.69	0.82	0.70	0.63
turnover	46.16%	50.57%	44.08%	50.27%	97.67%	37.16%
			Scope 3			
return	-0.26%	0.28%	15.28%	0.72%	15.71%	14.96%
std.	2.93%	4.52%	19.43%	4.21%	20.01%	18.78%
SR	-0.13	0.04	0.78	0.14	0.78	0.79
FF5 α	0.73%	1.46%	2.73%	1.15%	4.06%	1.81%
$\sigma(\theta_p)$	1.73%	2.93%	6.31%	2.60%	7.47%	4.65%
IR	0.42	0.50	0.43	0.44	0.54	0.39
turnover	44.58%	50.56%	43.14%	49.85%	95.03%	36.07%

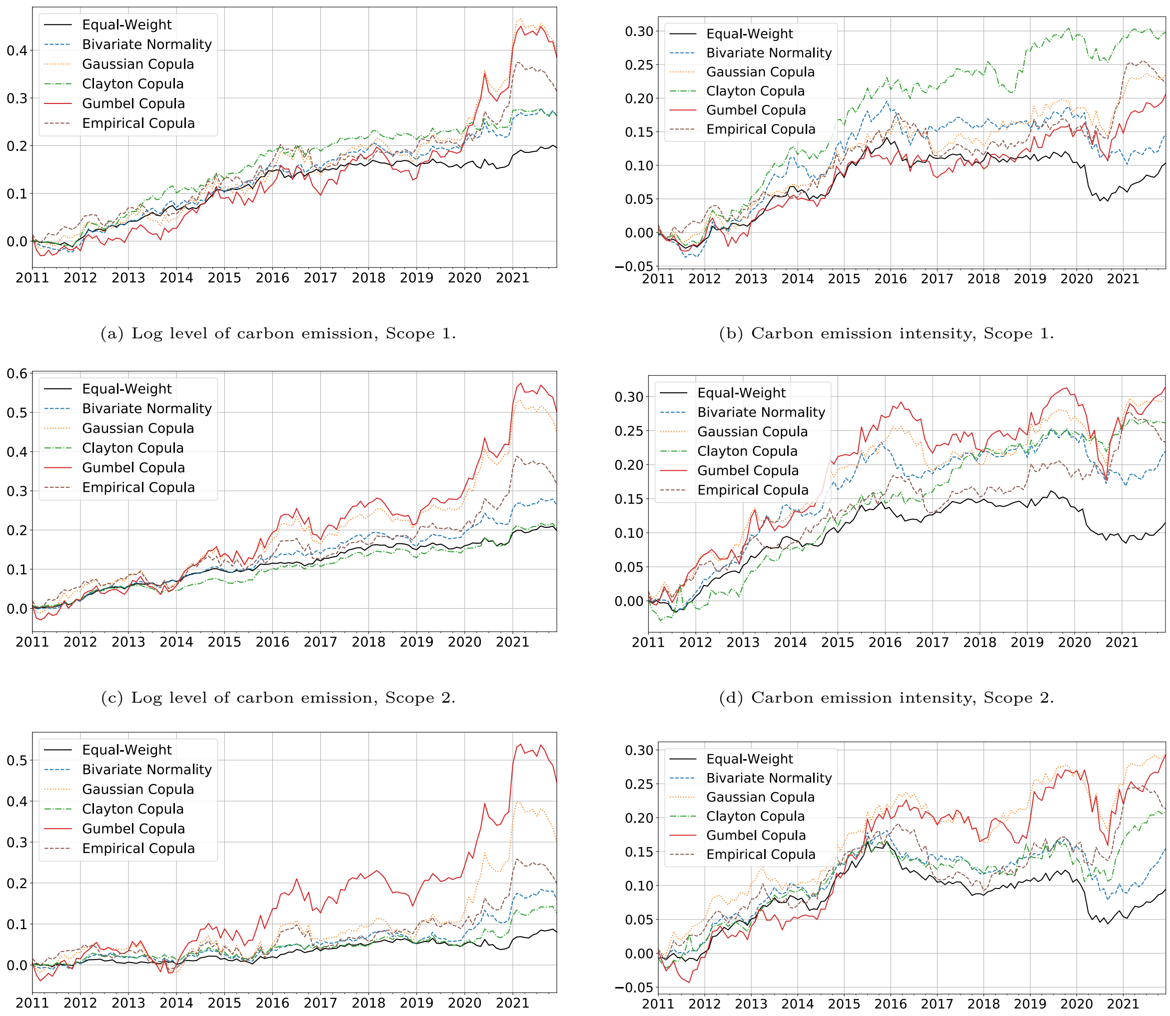
Table EC.5 shows that, for impact portfolios constructed using the log levels of carbon emissions, the Equal-Weighted Portfolio generally underperforms other portfolios in terms of annualized return, Sharpe ratio, and active α for all three scopes. This illustrates the superiority of our impact investing theory based on induced order statistics over the traditional Equal-Weighted Portfolio.

Table EC.6 Performance of impact portfolios constructed using carbon emission intensities. All metrics in this exhibit are annualized.

	Equal-Weight	Bivariate Normality	Gaussian Copula	Clayton Copula	Gumbel Copula	Empirical
			Scope 1			
return	0.51%	0.49%	14.94%	5.75%	14.21%	15.73%
std.	2.77%	4.00%	17.22%	5.42%	17.14%	17.50%
SR	0.14	0.09	0.86	1.04	0.82	0.89
FF5 α	0.94%	1.31%	2.12%	2.72%	1.97%	2.02%
$\sigma(\theta_p)$	2.55%	3.68%	3.40%	3.64%	3.80%	3.57%
IR	0.37	0.36	0.62	0.75	0.52	0.57
turnover	41.95%	45.81%	36.01%	41.23%	96.50%	30.91%
			Scope 2			
return	0.21%	0.49%	14.78%	5.09%	13.14%	15.44%
std.	2.51%	3.60%	16.57%	4.60%	16.22%	17.17%
SR	0.04	0.10	0.88	1.08	0.80	0.89
FF5 α	1.04%	1.96%	2.70%	2.40%	2.78%	2.12%
$\sigma(\theta_p)$	2.12%	3.14%	3.69%	3.14%	3.91%	3.32%
IR	0.49	0.62	0.73	0.76	0.71	0.64
turnover	43.01%	45.79%	34.69%	41.14%	102.47%	31.63%
			Scope 3			
return	0.28%	-0.09%	14.60%	3.09%	13.07%	15.50%
std.	2.90%	3.34%	17.10%	3.97%	18.08%	17.33%
SR	0.06	-0.06	0.85	0.75	0.72	0.89
FF5 α	0.86%	1.39%	2.70%	1.88%	2.71%	1.94%
$\sigma(\theta_p)$	2.63%	2.89%	3.97%	3.42%	5.16%	3.97%
IR	0.33	0.48	0.68	0.55	0.52	0.49
turnover	39.57%	48.91%	38.91%	46.08%	100.12%	27.76%

Table EC.5 also demonstrates that, when using the log levels of carbon emissions, the Gaussian Copula, Clayton Copula, Gumbel Copula, and Empirical Portfolios outperform the Bivariate Normality Portfolio in many cases in terms of annualized return, Sharpe ratio, and active α. This suggests that extending Lo and Zhang's (2023) joint normality assumption (Assumption EC.1) to a general distribution assumption (Assumption 1) may lead to higher risk-adjusted returns. However, we also acknowledge that in many cases, the Bivariate Normality Portfolio can achieve similar or higher information ratios compared to other portfolios. This can be attributed to the fact that the log levels of carbon emissions do not deviate from the normal distribution strongly enough (see Figure EC.11), and there, non-normal portfolios may not consistently improve performance.

On the other hand, when considering the performance of impact portfolios constructed using carbon intensity measures in Table EC.6, we find that the Gaussian Copula, Clayton Copula, Gumbel Copula, and Empirical Portfolios consistently and strongly outperform the Equal-Weighted Portfolio and the Bivariate Normality Portfolio in terms of annualized return, Sharpe Ratio, active α, and information ratio. For example, for Scope 1, the Equal-Weighted Portfolio and the Bivariate Normality Portfolio earn information ratios of 0.37 and 0.36, respectively, while the Gaussian Copula, Clayton Copula, Gumbel Copula, and Empirical Portfolios earn 0.62, 0.75, 0.52, and 0.57, respectively.



(a) Log level of carbon emission, Scope 1.

(b) Carbon emission intensity, Scope 1.

(c) Log level of carbon emission, Scope 2.

(d) Carbon emission intensity, Scope 2.

(e) Log level of carbon emission, Scope 3.

(f) Carbon emission intensity, Scope 3.

Figure EC.15 Cumulative residual returns for impact portfolios.

It is also worth noting that the Empirical Portfolio generally outperforms the Equal-Weighted Portfolio and the Bivariate Normality Portfolio in terms of the active α and information ratio, and has a competitive—but not necessarily better—performance compared to the Gaussian Copula, Clayton Copula, and Gumbel Copula Portfolios. Using empirical distributions does eliminate the need to choose a specific type of parametric copula, which is a desired property in practice. But we would also like to point out that it requires more data to estimate these nonparametric distributions well. In this sense, using specific parametric forms of copula can be regarded as a way to regularize nonparametric estimations especially when data is scarce. In practice, investors should weigh the pros and cons of nonparametric and parametric methods given the amount of data available in the specific application.

Overall, by allowing for a more flexible distributional assumption in our framework, impact investors can achieve much improved risk-adjusted returns. As Figure EC.12, Table EC.4, and Figure EC.14 demonstrate, the carbon intensity measures strongly deviate from a normal distribution. Although the marginal distribution of $\boldsymbol{X}$ does not affect the distribution of induced order statistics given the copula (see Theorem 1), it does affect the empirical estimation of the dependence between $\boldsymbol{X}$ and $\boldsymbol{\theta}$. In particular, when $\boldsymbol{X}$ significantly deviates from the normal distribution, the framework of Lo and Zhang (2023) is inadequate in describing this dependence using a single correlation ρ of the bivariate normal distribution. As a result, the portfolios constructed using the framework in this article are able to outperform significantly.

In summary, our empirical study demonstrates that, by extending the joint normality assumption to the general case (Assumption 1), impact investors can achieve higher profits, excess returns, Sharpe ratios, and information ratios, especially when the impact factor and residual returns deviate from the normal distribution significantly. This underscores the effectiveness of our general framework in practice.

Our empirical analysis is related to the growing literature on measuring the association between asset returns and sustainability. Bolton and Kacperczyk (2021, 2023) and Bolton et al. (2022) find that higher stock returns and lower P/E and market-to-book ratios are associated with higher levels and growth rates of carbon emissions, both in the US and internationally. On the other hand, several studies find the opposite results. Görgen et al. (2020) find an insignificantly negative carbon premium when they combine multiple carbon emission-related measures; Cheema-Fox et al. (2021) find that a portfolio going long in low-carbon intensity sectors and shorting high-carbon sectors delivered a positive alpha; Aswani et al. (2024) show that Bolton and Kacperczyk's (2021) carbon premiums disappear after accounting for disclosed versus estimated emissions; Zhang (2024) finds that Bolton and Kacperczyk's (2021) carbon premiums arise from the forward-looking sales information contained in emissions instead of the carbon risk premium. In the related literature on ESG investing, several studies find neutral or positive performance for green portfolios (Berg et al. 2021, Lindsey et al. 2021, Pástor et al. 2021). Pástor et al. (2022) and Ardia et al. (2023) further show that the high returns for green assets in recent years reflect unexpectedly strong increases in environmental concerns, not high expected returns.

Our empirical findings are generally consistent with this literature, but may appear at odds with the findings of Bolton and Kacperczyk (2021, 2023) in the US market. We provide several remarks on this difference. First and foremost, the focus of our empirical study is primarily to demonstrate the applicability of our portfolio construction framework using induced order statistics given a measure of impact, while studies such as Bolton and Kacperczyk (2021, 2023) and Pástor et al. (2021, 2022) focus on estimating the carbon risk premium and greenium in equilibrium using very different methodologies. For example, Bolton and Kacperczyk (2021, 2023) perform panel regressions of individual asset returns on carbon emission measures and a suite of control variables. We use a rolling window to estimate the dependence between returns and the impact factor, which can be regarded as a simple estimate of the dependence between the expected return and the impact factor next year. In principle, one can apply our framework to more sophisticated estimates of expected returns, such as the equilibrium model of Pástor et al. (2021), or those in the literature on robust portfolio with parameter uncertainty (Jorion 1986, Kan and Zhou 2007, DeMiguel et al. 2009, Kan et al. 2022). However, exploring different estimates of the expected return is beyond the scope of this paper and is left as future work.

Second, both Aswani et al. (2024) and Zhang (2024) find that the previously documented carbon premium can be explained, at least partially, by either estimation bias in emissions or forward-looking sales information contained in emissions.

Third, our results are consistent with Lo et al. (2022), who find a significant greenium in the US market for a much wider set of non-carbon environmental measures, including water consumption, waste disposal, land and water pollution, etc.

Finally, our results are in fact also consistent with Pástor et al.'s (2022) empirical findings in a similar time period. Lo et al. (2022) use the same dataset as ours and find that a significant portion of the realized greenium in the US market over the past decade can be explained by the unexpected increase in climate concerns. This is consistent with Pástor et al.'s (2022) findings and offers another way to reconcile our results with Bolton and Kacperczyk's (2021) carbon premium.

EC.3. Lemmas and Proofs

EC.3.1. Lemmas

LEMMA EC.1. *Let $C(u,v)$ be a copula. Then, the partial Dini derivatives $D_1^+C(u,v)$ and $D_1^-C(u,v)$ defined by (EC.31) are non-decreasing with respect to v for any $u \in (0,1)$.*

Proof of Lemma EC.1. Because C is a copula, for any $v_1 < v_2$ and $h > 0$, we have $C(u+h,v_1) - C(u,v_1) \le C(u+h,v_2) - C(u,v_2)$. Taking the limit superiors of both sides proves the result for $D_1^+C(u,v)$. Similar arguments apply to $D_1^-C(u,v)$. □

LEMMA EC.2. *Given a distribution function $f : \mathbb{R} \to [0,1]$, define $\underline{f}^{-1}(w) = \inf\{v : f(v) \ge w, v \in \overline{\mathcal{R}}_\theta\}$ for $w \in [0,1]$. Then, we have:*

(i) For any $w \in [0,1]$, $F_\theta^{-1} \circ \underline{f}^{-1}(w) = F_\theta^{-1} \circ f^{-1}(w)$;

(ii) For any $v \in \overline{\mathcal{R}}_\theta$, $\{w : \underline{f}^{-1}(w) \le v\} = \{w : w \le f(v)\}$.

Proof of Lemma EC.2. By definition, $\underline{f}^{-1}(w) \ge f^{-1}(w)$. If $\underline{f}^{-1}(w) > f^{-1}(w)$, we have $f^{-1}(w) \notin \overline{\mathcal{R}}_\theta$. Therefore, there must exist $d \in \Delta_\theta$ such that $f^{-1}(w) \in (F_\theta(d^-), F_\theta(d))$, where Δ_θ is the set of discontinuity points of F_θ. Then, by Definition EC.1, $F_\theta^{-1} \circ f^{-1}(w) = d$. In addition, we have $\underline{f}^{-1}(w) = F_\theta(d)$ and, therefore, $F_\theta^{-1} \circ \underline{f}^{-1}(w) = d$. This proves Part (i). Part (ii) can be verified by the definition of $\underline{f}^{-1}(w)$ and the right continuity of f. □

LEMMA EC.3. *Let $C(u,v)$ be a copula. For any fixed $0 \le a < b \le 1$ and fixed $v \in [0,1]$, define*

$$C^*(u,v) = \begin{cases} \frac{b-u}{b-a}C(a,v) + \frac{u-a}{b-a}C(b,v), & u \in (a,b), \\ C(u,v), & u \notin (a,b). \end{cases}$$

Then, if $C(u,v)$ is a concave function of u, $C^*(u,v)$ is also a concave function of u.

Proof of Lemma EC.3. We omit the proof because one can easily verify this by the definition of a concave function. □

LEMMA EC.4. *Let $W_{1:N} \le W_{2:N} \le \cdots \le W_{N:N}$ be the order statistics of $W_1, W_2, \ldots, W_N \overset{\text{IID}}{\sim} F_W$ with density function f_W. Then, the density of $W_{i:N}$ is:*

$$f_{W_{i:N}}(x) = \frac{N!}{(i-1)!(N-i)!} \cdot f_W(x)[F_W(x)]^{i-1}[1-F_W(x)]^{N-i},$$

and the joint density of $(W_{i_1:N}, W_{i_2:N}, \ldots, W_{i_m:N})$ is:

$$\begin{aligned} f_{W_{i_1:N},W_{i_2:N},\ldots,W_{i_m:N}}(x_1,x_2,\ldots,x_m) = & \frac{N!}{(i_1-1)!(i_2-i_1-1)!\cdots(N-i_m)!} \\ & \cdot [F_W(x_1)]^{i_1-1} f_W(x_1)[F_W(x_2)-F_W(x_1)]^{i_2-i_1-1} f_W(x_2) \\ & \cdots [1-F_W(x_m)]^{N-i_m} f_W(x_m) \cdot \mathbf{1}_{\{x_1<x_2<\cdots<x_m\}}, \end{aligned}$$

for $1 \le i_1 < i_2 < \cdots < i_m \le N$.

Proof of Lemma EC.4. See David and Nagaraja (2004, Section 2). □

LEMMA EC.5. *Let $Y_{1:N} \le Y_{2:N} \le \cdots \le Y_{N:N}$ be the order statistics of $Y_1, Y_2, \ldots, Y_N \overset{\text{IID}}{\sim} \mathcal{N}(0,1)$. Then, for any $i, j = 1, 2, \ldots, N$, we have:*

(i) $\mathbb{E}(Y_{i:N}) = -\mathbb{E}(Y_{N+1-i:N})$;

(ii) $\text{Cov}(Y_{i:N}, Y_{j:N}) = \text{Cov}(Y_{N+1-j:N}, Y_{N+1-i:N})$;

(iii) $\Phi^{-1}\left(\frac{i-1}{N}\right) \le \mathbb{E}(Y_{i:N}) \le \Phi^{-1}\left(\frac{i}{N}\right)$, *where Φ is the distribution function of $\mathcal{N}(0,1)$;*

(iv) $\text{Cov}(Y_{i:N}, Y_{j:N}) \ge 0$;

(v) $\text{Cov}\left(Y_{i:N}, \sum_{k\neq i} Y_{k:N}\right) = 1 - \text{Var}(Y_{i:N})$.

Proof of Lemma EC.5. Part (i) and Part (ii) are direct corollaries of Lemma EC.4. Part (iii) is shown in David and Nagaraja (2004, Section 4.5). Part (iv) is shown in Bickel (1967, Theorem 2.1). For Part (v), note that

$$\text{Cov}\left(Y_{i:N}, \sum_{k\neq i} Y_{k:N}\right) = \text{Cov}\left(Y_{i:N}, \sum_{k=1}^{N} Y_{k:N}\right) - \text{Var}(Y_{i:N})$$

$$= \text{Cov}\left(Y_{i:N}, \sum_{k=1}^{N} Y_{k}\right) - \text{Var}(Y_{i:N}) = \sum_{k=1}^{N} \text{Cov}(Y_{i:N}, Y_k) - \text{Var}(Y_{i:N}) = 1 - \text{Var}(Y_{i:N}),$$

where the last equation uses $\text{Cov}(Y_{i:N}, Y_k) = \frac{1}{N}$ (Wang et al. 1996, Theorem 1). □

LEMMA EC.6. *Let $Y_{1:N} \le Y_{2:N} \le \cdots \le Y_{N:N}$ be the order statistics of $Y_1, Y_2, \ldots, Y_N \overset{\text{IID}}{\sim} \mathcal{N}(0,1)$. Sequence $1 \le i_1(N) < i_2(N) \le N$ satisfies that, as $N \to +\infty$, $i_1(N)/N \to \xi_1$ and $i_2(N)/N \to \xi_2$ with constants ξ_1 and ξ_2 such that $0 \le \xi_1 \le \xi_2 \le 1$. Then:*

(i) $\lim_{N\to+\infty} \frac{\sum_{i=i_1(N)}^{i_2(N)} [\mathbb{E}(Y_{i:N})]^2}{N} = \left(\xi_2 - \Phi^{-1}(\xi_2)\varphi\big(\Phi^{-1}(\xi_2)\big)\right) - \left(\xi_1 - \Phi^{-1}(\xi_1)\varphi\big(\Phi^{-1}(\xi_1)\big)\right)$;

(ii) $\lim_{N\to+\infty} \frac{\sum_{i=i_1(N)}^{i_2(N)} |\mathbb{E}(Y_{i:N})|}{N} = \begin{cases} \left|\varphi\big(\Phi^{-1}(\xi_2)\big) - \varphi\big(\Phi^{-1}(\xi_1)\big)\right|, & \xi_2 \le 0.5 \text{ or } \xi_1 \ge 0.5, \\ 2\varphi(0) - \varphi\big(\Phi^{-1}(\xi_2)\big) - \varphi\big(\Phi^{-1}(\xi_1)\big), & \xi_1 < 0.5 < \xi_2; \end{cases}$

(iii) $\lim_{N\to+\infty} \frac{\sum_{i=i_1(N)}^{i_2(N)} \mathbb{E}(Y_{i:N})}{N} = \varphi\big(\Phi^{-1}(\xi_1)\big) - \varphi\big(\Phi^{-1}(\xi_2)\big)$,

where Φ and φ are the distribution function and density function of $\mathcal{N}(0,1)$, respectively.

Proof of Lemma EC.6. We first prove Part (i). By Part (iii) of Lemma EC.5, $\mathbb{E}(Y_{i:N})$ satisfies $\Phi^{-1}\left(\frac{i-1}{N}\right) \le \mathbb{E}(Y_{i:N}) \le \Phi^{-1}\left(\frac{i}{N}\right)$. Further note that $\Phi^{-1}\left(\frac{i-1}{N}\right) \ge 0$ when $i > N/2$, and $\Phi^{-1}\left(\frac{i}{N}\right) \le 0$ when $i \le N/2$. Hence, as $N \to +\infty$,

$$\begin{aligned}
\frac{\sum_{i=i_1(N)}^{i_2(N)} [\mathbb{E}(Y_{i:N})]^2}{N} &= \frac{\sum_{i_1(N)\le i\le i_2(N), i\le N/2} [\mathbb{E}(Y_{i:N})]^2}{N} + \frac{\sum_{i_1(N)\le i\le i_2(N), i> N/2} [\mathbb{E}(Y_{i:N})]^2}{N} \\
&\le \frac{\sum_{i_1(N)\le i\le i_2(N), i\le N/2} [\Phi^{-1}(\frac{i-1}{N})]^2}{N} + \frac{\sum_{i_1(N)\le i\le i_2(N), i> N/2} [\Phi^{-1}(\frac{i}{N})]^2}{N} \\
&\to \int_{\xi_1}^{\xi_2} [\Phi^{-1}(q)]^2 \mathrm{d}q \overset{x=\Phi^{-1}(q)}{=\!=\!=} \int_{\Phi^{-1}(\xi_1)}^{\Phi^{-1}(\xi_2)} x^2 \varphi(x) \mathrm{d}x \\
&= -\Phi^{-1}(\xi_2)\varphi\big(\Phi^{-1}(\xi_2)\big) + \Phi^{-1}(\xi_1)\varphi\big(\Phi^{-1}(\xi_1)\big) + \xi_2 - \xi_1,
\end{aligned}$$

where the limit holds because $\frac{\sum_{i_1(N)\le i\le i_2(N), i\le N/2}[\Phi^{-1}((i-1)/N)]^2}{N}$ and $\frac{\sum_{i_1(N)\le i\le i_2(N), i> N/2}[\Phi^{-1}(i/N)]^2}{N}$ are both Riemann sums. The same lower bound can also be obtained, and thus, Part (i) holds. Using similar approaches, we can show that

$$\frac{\sum_{i=i_1(N)}^{i_2(N)}|\mathbb{E}(Y_{i:N})|}{N}\to\int_{\xi_1}^{\xi_2}|\Phi^{-1}(q)|\mathrm{d}q \xlongequal{x=\Phi^{-1}(q)} \int_{\Phi^{-1}(\xi_1)}^{\Phi^{-1}(\xi_2)}|x|\varphi(x)\mathrm{d}x=\int_{\Phi^{-1}(\xi_1)}^{\Phi^{-1}(\xi_2)}\frac{|x|}{\sqrt{2\pi}}e^{-x^2/2}\mathrm{d}x$$

and

$$\frac{\sum_{i=i_1(N)}^{i_2(N)}\mathbb{E}(Y_{i:N})}{N}\to\int_{\xi_1}^{\xi_2}\Phi^{-1}(q)\mathrm{d}q \xlongequal{x=\Phi^{-1}(q)} \int_{\Phi^{-1}(\xi_1)}^{\Phi^{-1}(\xi_2)}x\varphi(x)\mathrm{d}x=\int_{\Phi^{-1}(\xi_1)}^{\Phi^{-1}(\xi_2)}\frac{x}{\sqrt{2\pi}}e^{-x^2/2}\mathrm{d}x$$

as $N\to+\infty$. Calculating the integrals completes the proofs of Parts (ii) and (iii). □

LEMMA EC.7. *For an $N\times N$ matrix $S=(S_{ij})_{i,j=1}^N$, the smallest and largest eigenvalues of S, $\lambda_{\min}(S)$ and $\lambda_{\max}(S)$, have the following bounds respectively:*

(i) $\lambda_{\min}(S)\ge\min_{i=1,2,\ldots,N}\left[S_{ii}-\sum_{j\ne i}|S_{ij}|\right]$;

(ii) $\lambda_{\max}(S)\le\max_{i=1,2,\ldots,N}\sum_{j=1}^N|S_{ij}|$.

Proof of Lemma EC.7. This is a corollary of the Gershgorin circle theorem, which states that for any eigenvalue λ of S, there exists i with $1\le i\le N$ such that $|\lambda-S_{ii}|\le\sum_{j\ne i}|S_{ij}|$. See, for example, Horn and Johnson (2012, Theorem 6.1.1). □

LEMMA EC.8. *Denote by $\theta_{[1:N],t},\theta_{[2:N],t},\ldots,\theta_{[N:N],t}$ the induced order statistics of $\theta_{1t},\theta_{2t},\ldots,\theta_{Nt}$ ranked by $X_{1t},X_{2t},\ldots,X_{Nt}$. Under Assumption EC.2, the expectation of $\theta_{[i:N],t}$ is:*

$$\mathbb{E}(\theta_{[i:N],t})=\mathbb{E}[\mathbb{E}(\theta_{[i:N],t}|\rho_t)]=\sigma_\theta\cdot\overline{\rho}\cdot\mathbb{E}(Y_{i:N})+\mu_\theta. \tag{EC.56}$$

The variance of $\theta_{[i:N],t}$ is:

$$\mathrm{Var}(\theta_{[i:N],t})=\mathbb{E}[\mathrm{Var}(\theta_{[i:N],t}|\rho_t)]+\mathrm{Var}[\mathbb{E}(\theta_{[i:N],t}|\rho_t)], \tag{EC.57}$$

where

$$\mathbb{E}[\mathrm{Var}(\theta_{[i:N],t}|\rho_t)]=\sigma_\theta^2\cdot[1-(\overline{\rho}^2+\sigma_\rho^2)+(\overline{\rho}^2+\sigma_\rho^2)\cdot\mathrm{Var}(Y_{i:N})], \tag{EC.58}$$

$$\mathrm{Var}[\mathbb{E}(\theta_{[i:N],t}|\rho_t)]=\sigma_\theta^2\cdot\sigma_\rho^2\cdot[\mathbb{E}(Y_{i:N})]^2. \tag{EC.59}$$

The covariance between $\theta_{[i:N],t}$ and $\theta_{[j:N],t}$ is:

$$\mathrm{Cov}(\theta_{[i:N],t},\theta_{[j:N],t})=\mathbb{E}[\mathrm{Cov}(\theta_{[i:N],t},\theta_{[j:N],t}|\rho_t)]+\mathrm{Cov}[\mathbb{E}(\theta_{[i:N],t}|\rho_t),\mathbb{E}(\theta_{[j:N],t}|\rho_t)], \tag{EC.60}$$

where

$$\mathbb{E}[\mathrm{Cov}(\theta_{[i:N],t},\theta_{[j:N],t}|\rho_t)]=\sigma_\theta^2\cdot(\overline{\rho}^2+\sigma_\rho^2)\cdot\mathrm{Cov}(Y_{i:N},Y_{j:N}), \tag{EC.61}$$

$$\mathrm{Cov}[\mathbb{E}(\theta_{[i:N],t}|\rho_t),\mathbb{E}(\theta_{[j:N],t}|\rho_t)]=\sigma_\theta^2\cdot\sigma_\rho^2\cdot\mathbb{E}(Y_{i:N})\cdot\mathbb{E}(Y_{j:N}). \tag{EC.62}$$

Here, $i,j=1,2,\ldots,N$ and $i\ne j$; $Y_{1:N}\le Y_{2:N}\le\cdots\le Y_{N:N}$ are the order statistics of $Y_1,Y_2,\ldots,Y_N\overset{\text{IID}}{\sim}\mathcal{N}(0,1)$.

Proof of Lemma EC.8 Given ρ_t, using Proposition EC.4, we have

$$\mathbb{E}(\theta_{[i:N],t}|\rho_t) = \sigma_\theta \cdot \rho_t \cdot \mathbb{E}(Y_{i:N}) + \mu_\theta,$$
$$\mathrm{Var}(\theta_{[i:N],t}|\rho_t) = \sigma_\theta^2 \cdot (1 - \rho_t^2 + \rho_t^2 \cdot \mathrm{Var}(Y_{i:N})),$$
$$\mathrm{Cov}(\theta_{[i:N],t}, \theta_{[j:N],t}|\rho_t) = \sigma_\theta^2 \cdot \rho_t^2 \cdot \mathrm{Cov}(Y_{i:N}, Y_{j:N}).$$

Then (EC.56), (EC.57), and (EC.60) hold by the law of iterated expectation, the law of total variance, and the law of total covariance, respectively. □

LEMMA EC.9. *Under Assumption EC.3, we have:*

$$\mathbb{E}(\theta_{[i:N]}) = \sigma_\theta \cdot \frac{\rho - \tilde{\rho}}{\sqrt{1-\rho_X}} \cdot \mathbb{E}(Y_{i:N}) + \mu_\theta, \tag{EC.63}$$

$$\mathrm{Var}(\theta_{[i:N]}) = \sigma_\theta^2 \cdot \left(1 - \frac{(\rho-\tilde{\rho})^2}{1-\rho_X} + \frac{(\rho-\tilde{\rho})^2}{1-\rho_X} \cdot \mathrm{Var}(Y_{i:N})\right), \tag{EC.64}$$

$$\mathrm{Cov}(\theta_{[i:N]}, \theta_{[j:N]}) = \sigma_\theta^2 \cdot \left(\rho_\theta + \frac{(\rho-\tilde{\rho})^2}{1-\rho_X} \cdot \mathrm{Cov}(Y_{i:N}, Y_{j:N})\right), \tag{EC.65}$$

for $i,j = 1,2,\ldots,N$ *and* $i \neq j$. *Here,* $Y_{1:N} \le Y_{2:N} \le \cdots \le Y_{N:N}$ *are the order statistics of* $Y_1, Y_2, \ldots, Y_N \overset{\text{IID}}{\sim} \mathcal{N}(0,1)$.

Proof of Lemma EC.9. See Lee and Viana (1999, Corollary 2.1). □

EC.3.2. Proofs for Section 2

Proof of Proposition 1. For investors to maximize the information ratio, let $\mathrm{IR}(\boldsymbol{w}) = \frac{\boldsymbol{w}^\top \boldsymbol{\mu}}{\sqrt{\boldsymbol{w}^\top \Sigma \boldsymbol{w}}}$. Then $\frac{\partial \mathrm{IR}}{\partial \boldsymbol{w}} = \frac{(\boldsymbol{w}^\top \Sigma \boldsymbol{w}) \cdot \boldsymbol{\mu} - (\boldsymbol{w}^\top \boldsymbol{\mu}) \cdot \Sigma \boldsymbol{w}}{(\boldsymbol{w}^\top \Sigma \boldsymbol{w})^{3/2}}$, and one can directly check that $\frac{\partial \mathrm{IR}}{\partial \boldsymbol{w}} = \boldsymbol{0}$ if and only if $\boldsymbol{w} \propto \Sigma^{-1}\boldsymbol{\mu}$. For investors to maximize the mean–variance utility, let $U(\boldsymbol{w}) = \boldsymbol{w}^\top \boldsymbol{\mu} - \frac{\lambda}{2}\boldsymbol{w}^\top \Sigma \boldsymbol{w}$. Then $\frac{\partial U}{\partial \boldsymbol{w}} = \boldsymbol{\mu} - \lambda \Sigma \boldsymbol{w}$, and thus $\frac{\partial U}{\partial \boldsymbol{w}} = \boldsymbol{0}$ if and only if $\boldsymbol{w} = \frac{1}{\lambda}\Sigma^{-1}\boldsymbol{\mu} \propto \Sigma^{-1}\boldsymbol{\mu}$. When $\boldsymbol{w} \propto \Sigma^{-1}\boldsymbol{\mu}$, we have $\mathrm{IR} = \frac{\boldsymbol{w}^\top \boldsymbol{\mu}}{\sqrt{\boldsymbol{w}^\top \Sigma \boldsymbol{w}}} = \frac{\boldsymbol{\mu}^\top \Sigma^{-1} \boldsymbol{\mu}}{\sqrt{\boldsymbol{\mu}^\top \Sigma^{-1} \Sigma \Sigma^{-1} \boldsymbol{\mu}}} = \sqrt{\boldsymbol{\mu}^\top \Sigma^{-1} \boldsymbol{\mu}}$. This completes the proof. □

EC.3.3. Proofs for Section 3

Proof of Theorem 1. This is a corollary of Theorem 2. □

Proof of Theorem 2. We first prove "if" followed by the proof for "only if." According to the footnote at the end of Definition 1, we have $\overline{\mathcal{R}}_X^c = \bigcup_{d \in \Delta_X} (F_X(d^-), F_X(d))$.

Proof of "if." The proof of "if" includes five main steps.

First, we calculate $\mathscr{D}_1 C(u,v)$ on $\overline{\mathcal{R}}_X^c \times \overline{\mathcal{R}}_\theta$. For any $d \in \Delta_X$ and $u \in (F_X(d^-), F_X(d))$, the linearly interpolating property implies that the partial right upper Dini derivative of $C(u,v)$ given by (EC.31) satisfies $D_1^+ C(u,v) = \frac{C(F_X(d),v) - C(F_X(d^-),v)}{F_X(d) - F_X(d^-)}$ for $v \in \overline{\mathcal{R}}_\theta$. This allows us to study $\mathscr{D}_1 C(u,v)$.

Let $A = \{v \in \overline{\mathcal{R}}_\theta : \exists \delta > 0 \text{ such that } \overline{\mathcal{R}}_\theta \cap (v, v+\delta) = \emptyset\}$. For any $v \in \overline{\mathcal{R}}_\theta \backslash A$, there should exist a sequence $\{v_n\}_{n=1}^{+\infty}$ such that $v_n \in \overline{\mathcal{R}}_\theta$, $v_n > v$, and $\lim_{n \to +\infty} v_n = v$. In addition, Lemma EC.1

demonstrates that $D_1^+C(u,v)$ is non-decreasing with respect to v. Hence, for any $u \in (F_X(d^-), F_X(d))$ and $v \in \overline{\mathcal{R}}_\theta \backslash A$, we have

$$\begin{aligned}\mathscr{D}_1 C(u,v) &= \inf_{y>v} D_1^+ C(u,y) = \lim_{n\to+\infty} D_1^+ C(u,v_n) \\ &= \lim_{n\to+\infty} \frac{C(F_X(d),v_n) - C(F_X(d^-),v_n)}{F_X(d)-F_X(d^-)} = \frac{C(F_X(d),v) - C(F_X(d^-),v)}{F_X(d)-F_X(d^-)},\end{aligned} \tag{EC.66}$$

where the last equality holds because of the continuity of C (Nelsen 2007, Theorem 2.2.4).

For any $v \in A$, Fang et al. (2020, Theorem 2.1) demonstrate that, there exists a set B_v with Lebesgue measure 0 such that $\mathscr{D}_1 C(u,v) = \frac{\partial C}{\partial u}(u,v)$ for $u \in [0,1] \backslash B_v$. In addition, by definition, there are at most countably infinite elements in A. Therefore, $E_1 \equiv \bigcup_{v\in A} B_v$ has a Lebesgue measure 0, and $\mathscr{D}_1 C(u,v) = \frac{\partial C}{\partial u}(u,v)$ for any $u \in [0,1]\backslash E_1$ and $v \in A$. Hence, (EC.66) holds for any $d \in \Delta_X$, $u \in (F_X(d^-), F_X(d)) \backslash E_1$ and $v \in \overline{\mathcal{R}}_\theta$.

Second, we define $\tilde{\mathscr{D}}_1 C(u,v) : [0,1]\times\mathbb{R} \to [0,1]$ as follows:

$$\tilde{\mathscr{D}}_1 C(u,v) = \begin{cases} \frac{C(F_X(d),v)-C(F_X(d^-),v)}{F_X(d)-F_X(d^-)}, & u = F_X(d), d \in \Delta_X, v \in \mathbb{R}, \\ \mathscr{D}_1 C(u,v), & \text{otherwise}, \end{cases} \tag{EC.67}$$

and build a relationship between functions $\tilde{\mathscr{D}}_1 C(u,v)$ and $\mathscr{D}_1 C(u,v)$. For any $d \in \Delta_X$, $u \in (F_X(d^-), F_X(d))\backslash E_1$, and $v \in \overline{\mathcal{R}}_\theta$, we have

$$\tilde{\mathscr{D}}_1 C\left(F_X(F_X^{-1}(u)), v\right) = \tilde{\mathscr{D}}_1 C\left(F_X(d), v\right) = \frac{C(F_X(d),v)-C(F_X(d^-),v)}{F_X(d)-F_X(d^-)} = \mathscr{D}_1 C(u,v), \tag{EC.68}$$

where the three equalities hold because of $F_X(F_X^{-1}(u)) = F_X(d)$ (by Definition EC.1), (EC.67), and (EC.66), respectively. Meanwhile, for any $u \notin \left(\bigcup_{d\in\Delta_X} [F_X(d^-), F_X(d)]\right) \cup E_1$ and $v \in \overline{\mathcal{R}}_\theta$, we have

$$\tilde{\mathscr{D}}_1 C\left(F_X(F_X^{-1}(u)), v\right) = \tilde{\mathscr{D}}_1 C\left(u, v\right) = \mathscr{D}_1 C(u,v), \tag{EC.69}$$

where the two equalities hold because of $F_X(F_X^{-1}(u)) = u$ (by Definition EC.1) and (EC.67), respectively. Therefore, by combining (EC.68) and (EC.69), we have

$$\tilde{\mathscr{D}}_1 C\left(F_X(F_X^{-1}(u)), v\right) = \mathscr{D}_1 C(u,v), \quad u \in [0,1]\backslash E_2, v \in \overline{\mathcal{R}}_\theta, \tag{EC.70}$$

where $E_2 \equiv E_1 \cup \{F_X(d) : d \in \Delta_X\} \cup \{F_X(d^-) : d \in \Delta_X\}$, which has a Lebesgue measure 0.

Third, we define

$$\tilde{g}(u,w) \equiv F_\theta^{-1} \circ \tilde{g}_u(w), \quad \tilde{g}_u(w) \equiv \inf\{v : \tilde{\mathscr{D}}_1 C(u,v) \geq w, v \in \overline{\mathcal{R}}_\theta\}, \tag{EC.71}$$

and build a relationship between $\tilde{g}(u,w)$ and $g(u,w)$. Let

$$\underline{g}(u,w) \equiv F_\theta^{-1} \circ \underline{g}_u(w), \quad \underline{g}_u(w) \equiv \inf\{v : \mathscr{D}_1 C(u,v) \geq w, v \in \overline{\mathcal{R}}_\theta\}. \tag{EC.72}$$

Thus, by (EC.70), we have $\tilde{g}_{F_X(F_X^{-1}(u))}(w) = \underline{g}_u(w)$ and, therefore, $\tilde{g}(F_X(F_X^{-1}(u)), w) = \underline{g}(u,w)$, when $u \in [0,1]\backslash E_2$ and $w \in [0,1]$. In addition, because $v \mapsto \mathscr{D}_1 C(u,v)$ is a distribution function (see Remark EC.1), Part (i) of Lemma EC.2 implies that $\underline{g}_u(w) = g_u(w)$ and, therefore, $\underline{g}(u,w) = g(u,w)$ for any $u \in [0,1]$ and $w \in [0,1]$. Hence,

$$\tilde{g}\left(F_X(F_X^{-1}(u)), w\right) = g(u,w), \quad u \in [0,1]\backslash E_2, w \in [0,1]. \tag{EC.73}$$

Fourth, we prove that $(X_i, \theta_i) \stackrel{d}{=} \left(F_X^{-1}(U_i), \tilde{g}(F_X(F_X^{-1}(U_i)), V_i)\right)$. For any $x, y \in [0,1]$, the independence between U_i and V_i implies that

$$\begin{aligned}&\mathbb{P}\left(F_X^{-1}(U_i) \le x, \tilde{g}(F_X(F_X^{-1}(U_i)), V_i) \le y\right) = \mathbb{P}\left(U_i \le F_X(x), \tilde{g}(F_X(F_X^{-1}(U_i)), V_i) \le y\right)\\ =&\int_0^{F_X(x)} \mathbb{P}\left(\tilde{g}(F_X(F_X^{-1}(u)), V_i) \le y\right) \mathrm{d}u. \end{aligned}\tag{EC.74}$$

Furthermore, by the definition of $\tilde{g}$ in (EC.71) and Part (ii) of Lemma EC.2, we have

$$\begin{aligned}&\int_0^{F_X(x)} \mathbb{P}\left(\tilde{g}(F_X(F_X^{-1}(u)), V_i) \le y\right) \mathrm{d}u = \int_0^{F_X(x)} \mathbb{P}\left(\tilde{g}_{F_X(F_X^{-1}(u))}(V_i) \le F_\theta(y)\right) \mathrm{d}u\\ =&\int_0^{F_X(x)} \mathbb{P}\left(V_i \le \tilde{\mathscr{D}}_1 C(F_X(F_X^{-1}(u)), F_\theta(y))\right) \mathrm{d}u. \end{aligned}\tag{EC.75}$$

By combining (EC.70), (EC.74), (EC.75), and Remark EC.1, we have

$$\begin{aligned}&\mathbb{P}(F_X^{-1}(U_i) \le x, \tilde{g}(F_X(F_X^{-1}(U_i)), V_i) \le y) = \int_0^{F_X(x)} \mathbb{P}\left(V_i \le \tilde{\mathscr{D}}_1 C(F_X(F_X^{-1}(u)), F_\theta(y))\right) \mathrm{d}u\\ =&\int_0^{F_X(x)} \mathscr{D}_1 C(u, F_\theta(y)) \mathrm{d}u = C(F_X(x), F_\theta(y)) = \mathbb{P}(X_i \le x, \theta_i \le y).\end{aligned}$$

Therefore, $(X_i, \theta_i) \stackrel{d}{=} \left(F_X^{-1}(U_i), \tilde{g}(F_X(F_X^{-1}(U_i)), V_i)\right)$.

Fifth, we prove that

$$\left(\theta_{[1:N]}, \ldots, \theta_{[N:N]}\right) \stackrel{d}{=} \left(\tilde{g}(F_X(F_X^{-1}(U_{1:N})), V_1), \ldots, \tilde{g}(F_X(F_X^{-1}(U_{N:N})), V_N)\right). \tag{EC.76}$$

Let $H_i \equiv F_X^{-1}(U_i)$ and $W_i \equiv \tilde{g}(F_X(F_X^{-1}(U_i)), V_i) = \tilde{g}(F_X(H_i), V_i)$, then by the fourth step, we have

$$\left(\theta_{[1:N]}, \ldots, \theta_{[N:N]}\right) \stackrel{d}{=} \left(W_{[1:N]}, \ldots, W_{[N:N]}\right), \tag{EC.77}$$

where $\{W_{[i:N]}\}_{i=1}^N$ are induced order statistics of $\{W_i\}_{i=1}^N$ ranked by $\{H_i\}_{i=1}^N$. We claim that

$$\left(W_{[1:N]}, \ldots, W_{[N:N]}\right) \stackrel{d}{=} \left(\tilde{g}(F_X(H_{1:N}), V_1), \ldots, \tilde{g}(F_X(H_{N:N}), V_N)\right). \tag{EC.78}$$

To prove this claim, taking the case of $H_1 \le \cdots \le H_N$ as an example, we have

$$\begin{aligned}&\mathbb{P}\left(W_{[1:N]} \le x_1, \ldots, W_{[N:N]} \le x_N, H_1 \le \cdots \le H_N\right)\\ =&\mathbb{P}\left(W_1 \le x_1, \ldots, W_N \le x_N, H_1 \le \cdots \le H_N\right)\\ =&\mathbb{P}\left(\tilde{g}(F_X(F_X^{-1}(U_1)), V_1) \le x_1, \ldots, \tilde{g}(F_X(F_X^{-1}(U_N)), V_N) \le x_N, F_X^{-1}(U_1) \le \cdots \le F_X^{-1}(U_N)\right)\\ =&\int_0^1 \cdots \int_0^1 \mathbf{1}_{\{\tilde{g}(F_X(F_X^{-1}(u_1)), w_1) \le x_1, \ldots, \tilde{g}(F_X(F_X^{-1}(u_N)), w_N) \le x_N\}} \cdot\\ &\qquad \mathbf{1}_{\{F_X^{-1}(u_1) \le \cdots \le F_X^{-1}(u_N)\}} \mathrm{d}u_1 \cdots \mathrm{d}u_N \mathrm{d}w_1 \cdots \mathrm{d}w_N.\end{aligned}\tag{EC.79}$$

In addition,

$$\begin{aligned}
&\mathbb{P}\left(\tilde{g}(F_X(H_{1:N}),V_1)\le x_1,\ldots,\tilde{g}(F_X(H_{N:N}),V_N)\le x_N, H_1\le\cdots\le H_N\right)\\
=&\mathbb{P}\left(\tilde{g}(F_X(H_1),V_1)\le x_1,\ldots,\tilde{g}(F_X(H_N),V_N)\le x_N, H_1\le\cdots\le H_N\right)\\
=&\mathbb{P}\left(\tilde{g}(F_X(F_X^{-1}(U_1)),V_1)\le x_1,\ldots,\tilde{g}(F_X(F_X^{-1}(U_N)),V_N)\le x_N, F_X^{-1}(U_1)\le\cdots\le F_X^{-1}(U_N)\right)\\
=&\int_0^1\cdots\int_0^1 \mathbf{1}_{\{\tilde{g}(F_X(F_X^{-1}(u_1)),w_1)\le x_1,\ldots,\tilde{g}(F_X(F_X^{-1}(u_N)),w_N)\le x_N\}}\cdot\\
&\qquad\qquad \mathbf{1}_{\{F_X^{-1}(u_1)\le\cdots\le F_X^{-1}(u_N)\}}\mathrm{d}u_1\cdots\mathrm{d}u_N\mathrm{d}w_1\cdots\mathrm{d}w_N,
\end{aligned}\tag{EC.80}$$

which is equal to (EC.79). The arguments above hold for any ordering of $H_1,\ldots,H_N$. Taking the summation over all orderings, we obtain

$$\mathbb{P}\left(W_{[1:N]}\le x_1,\ldots,W_{[N:N]}\le x_N\right)=\mathbb{P}\left(\tilde{g}(F_X(H_{1:N}),V_1)\le x_1,\ldots,\tilde{g}(F_X(H_{N:N}),V_N)\le x_N\right),$$

which proves the claim. In addition, because $H_{i:N}=F_X^{-1}(U_{i:N})$, by combining (EC.77) and (EC.78), we have (EC.76) holds.

Finally, combining (EC.73) and (EC.76) proves (14).

Proof of "only if." Let $\hat{C}$ be a copula of F linearly interpolating on $\overline{\mathcal{R}}_X^c\times\overline{\mathcal{R}}_\theta$. We aim to prove that $C\equiv\hat{C}$ on $[0,1]\times\overline{\mathcal{R}}_\theta$. Let

$$\hat{g}(u,w)\equiv F_\theta^{-1}\circ\hat{g}_u(w),\tag{EC.81}$$

where $\hat{g}_u(w)$ is the inverse function of $v\mapsto\mathscr{D}_1\hat{C}(u,v)$. Based on the result of Part "if," we have $\theta_{[N:N]}\overset{d}{=}\hat{g}(U_{N:N},V_N)$. Furthermore, because (14) holds, we have $\theta_{[N:N]}\overset{d}{=}g(U_{N:N},V_N)$ and, therefore, $\hat{g}(U_{N:N},V_N)\overset{d}{=}g(U_{N:N},V_N)$. Note that for any $y\in\mathbb{R}$, by Lemma EC.4,

$$\begin{aligned}
&\mathbb{P}\left(\hat{g}(U_{N:N},V_N)\le y\right)=\int_0^1\mathbb{P}\left(\hat{g}(u,V_N)\le y\right)Nu^{N-1}\mathrm{d}u=\int_0^1\mathbb{P}\left(\hat{g}_u(V_N)\le F_\theta(y)\right)Nu^{N-1}\mathrm{d}u\\
&=\int_0^1\mathbb{P}\left(V_N\le\mathscr{D}_1\hat{C}(u,F_\theta(y))\right)Nu^{N-1}\mathrm{d}u=\int_0^1\mathscr{D}_1\hat{C}(u,F_\theta(y))Nu^{N-1}\mathrm{d}u,
\end{aligned}$$

and similarly, $\mathbb{P}\left(g(U_{N:N},V_N)\le y\right)=\int_0^1\mathscr{D}_1C(u,F_\theta(y))Nu^{N-1}\mathrm{d}u$. Therefore,

$$\begin{aligned}
0&=\mathbb{P}\left(\hat{g}(U_{N:N},V_N)\le y\right)-\mathbb{P}(g(U_{N:N},V_N)\le y)\\
&=\int_0^1\left[\mathscr{D}_1\hat{C}(u,F_\theta(y))-\mathscr{D}_1C(u,F_\theta(y))\right]Nu^{N-1}\mathrm{d}u
\end{aligned}$$

holds for any $N\ge1$. Hence,

$$\int_0^1\left[\mathscr{D}_1\hat{C}(u,F_\theta(y))-\mathscr{D}_1C(u,F_\theta(y))\right]u^N\mathrm{d}u=0,\quad\forall N\ge0.\tag{EC.82}$$

Because both $\mathscr{D}_1\hat{C}$ and $\mathscr{D}_1C$ are regular conditional distribution functions (Remark EC.1), for any $y\in\mathbb{R}$, $h_y(u)\equiv\mathscr{D}_1\hat{C}(u,F_\theta(y))-\mathscr{D}_1C(u,F_\theta(y))$ is Lebesgue measurable and bounded. Hence, $h_y(u)\in$

$L^2[0,1]$. In addition, (EC.82) implies that $h_y(u)$ is orthogonal to all polynomials, which are dense in $L^2[0,1]$. Therefore, by the property of the L^2 space, $h_y(u)=0$ almost everywhere for $u\in[0,1]$. In other words, for any $y\in\mathbb{R}$,

$$\mathscr{D}_1\hat{C}(u,F_\theta(y))=\mathscr{D}_1 C(u,F_\theta(y)),\quad \text{almost everywhere,}\quad u\in[0,1].$$

For any $a\in[0,1]$, by integrating the equation above with respect to u from 0 to a, according to Remark EC.1, we have $\hat{C}(a,F_\theta(y))=C(a,F_\theta(y))$. Hence, $C\equiv\hat{C}$ holds on $[0,1]\times\mathcal{R}_\theta$, where $\mathcal{R}_\theta$ is the range of F_θ. The continuity of copula (Nelsen 2007, Theorem 2.2.4) further implies that $C\equiv\hat{C}$ holds on $[0,1]\times\overline{\mathcal{R}}_\theta$. This completes the proof. $\square$

Proof of Proposition 2. This is a corollary of Theorem 2 given Lemma EC.4. $\square$

Proof of Theorem 3. This is a corollary of Theorem 4. $\square$

Proof of Theorem 4. We first prove "if" followed by the proof for "only if." For notational simplicity, we abbreviate $i_k(N)$ as i_k.

Proof of "if." According to Theorem 2 and Lemma EC.4, for any fixed $x_1,\dots,x_m$, we have

$$\begin{aligned}
&\mathbb{P}\left(\theta_{[i_1:N]}\le x_1,\dots,\theta_{[i_m:N]}\le x_m\right)=\mathbb{P}\left(g(U_{i_1:N},V_{i_1})\le x_1,\dots,g(U_{i_m:N},V_{i_m})\le x_m\right)\\
&=\int_0^1\cdots\int_0^{u_3}\int_0^{u_2}\mathbb{P}\left(g_{u_1}(V_{i_1})\le F_\theta(x_1),\dots,g_{u_m}(V_{i_m})\le F_\theta(x_m)\right)\\
&\qquad\cdot\frac{u_1^{i_1-1}(u_2-u_1)^{i_2-i_1-1}\cdots(1-u_m)^{N-i_m}N!}{(i_1-1)!(i_2-i_1-1)!\cdots(N-i_m)!}\mathrm{d}u_1\mathrm{d}u_2\cdots\mathrm{d}u_m\\
&=\int_0^1\cdots\int_0^{u_3}\int_0^{u_2}\prod_{k=1}^m\mathscr{D}_1C(u_k,F_\theta(x_k))\\
&\qquad\cdot\frac{u_1^{i_1-1}(u_2-u_1)^{i_2-i_1-1}\cdots(1-u_m)^{N-i_m}N!}{(i_1-1)!(i_2-i_1-1)!\cdots(N-i_m)!}\mathrm{d}u_1\mathrm{d}u_2\cdots\mathrm{d}u_m\\
&=\int_0^1\cdots\int_0^{u_3}\left[\int_0^{u_2}\mathscr{D}_1C(u_1,F_\theta(x_1))u_1^{i_1-1}(u_2-u_1)^{i_2-i_1-1}\frac{(i_2-1)!}{(i_1-1)!(i_2-i_1-1)!}\mathrm{d}u_1\right]\\
&\qquad\cdot\prod_{k=2}^m\mathscr{D}_1C(u_k,F_\theta(x_k))\cdot\frac{(u_3-u_2)^{i_3-i_2-1}\cdots(1-u_m)^{N-i_m}N!}{(i_2-1)!(i_3-i_2-1)!\cdots(N-i_m)!}\mathrm{d}u_2\cdots\mathrm{d}u_m.
\end{aligned}\tag{EC.83}$$

Now we study the integral in the square brackets in the equation above. We have

$$\begin{aligned}
&\int_0^{u_2}\mathscr{D}_1C(u_1,F_\theta(x_1))u_1^{i_1-1}(u_2-u_1)^{i_2-i_1-1}\frac{(i_2-1)!}{(i_1-1)!(i_2-i_1-1)!}\mathrm{d}u_1\\
&\overset{u_1=u_2t}{=\!=}\;u_2^{i_2-1}\int_0^1\mathscr{D}_1C(u_2t,F_\theta(x_1))t^{i_1-1}(1-t)^{i_2-i_1-1}\frac{(i_2-1)!}{(i_1-1)!(i_2-i_1-1)!}\mathrm{d}t.
\end{aligned}\tag{EC.84}$$

Let $b_N(t)=t^{i_1-1}(1-t)^{i_2-i_1-1}\frac{(i_2-1)!}{(i_1-1)!(i_2-i_1-1)!}$, which is the density of a beta distribution. Then, we have $\int_0^1 b_N(t)\mathrm{d}t=1$, and $b_N(t)$ increases with t when $t\in[0,\frac{i_1-1}{i_2-2}]$ and decreases when $t\in[\frac{i_1-1}{i_2-2},1]$. Let $t^*=\frac{\xi_1}{\xi_2}$, which satisfies that $\lim_{N\to+\infty}\frac{i_1-1}{i_2-2}=t^*$. Because $\mathscr{D}_1C(u,v)$ is continuous with respect to u on $[0,1]$, it is also uniformly continuous. Thus, for any $\varepsilon>0$, there exists $\delta>0$ (which does not depend on $u_2,\dots,u_m$) such that $|\mathscr{D}_1 C(u_2 t, F_\theta(x_1)) - \mathscr{D}_1 C(u_2 t^*, F_\theta(x_1))| < \varepsilon$ for any $u_2 \in [0,1]$ and $|t-t^*|<\delta$. Therefore,

$$\begin{aligned}
&\left|\int_0^1 \mathscr{D}_1 C(u_2 t, F_\theta(x_1)) t^{i_1-1}(1-t)^{i_2-i_1-1}\frac{(i_2-1)!}{(i_1-1)!(i_2-i_1-1)!}\mathrm{d}t - \mathscr{D}_1 C(u_2 t^*, F_\theta(x_1))\right| \\
\le& \int_0^1 |\mathscr{D}_1 C(u_2 t, F_\theta(x_1)) - \mathscr{D}_1 C(u_2 t^*, F_\theta(x_1))|\, b_N(t)\mathrm{d}t \\
=& \int_0^{t^*-\delta}\cdots\mathrm{d}t + \int_{t^*-\delta}^{t^*+\delta}\cdots\mathrm{d}t + \int_{t^*+\delta}^{1}\cdots\mathrm{d}t = \text{(I)}+\text{(II)}+\text{(III)}.
\end{aligned} \tag{EC.85}$$

Term (I) satisfies that (I) $\le 2\int_0^{t^*-\delta} b_N(t)\mathrm{d}t \le 2(t^*-\delta)b_N(t^*-\delta) \le 2b_N(t^*-\delta)$ when N is sufficiently large because of $0 \le \mathscr{D}_1 C \le 1$ (Remark EC.1) and the monotonicity of $b_N(t)$. We claim that $\lim_{N\to+\infty} b_N(t^*-\delta)=0$. To prove this claim, let us consider a random variable η_N following the beta distribution with density $b_N(t)$. For any $\tau>0$, by Chebyshev's inequality, we have $\mathbb{P}(|\eta_N - t^*| > \tau) \le \mathbb{P}(|\eta_N - \mathbb{E}(\eta_N)| > \tau/2) + \mathbb{P}(|\mathbb{E}(\eta_N) - t^*| > \tau/2) \le \mathrm{Var}(\eta_N)/(\tau/2)^2 + \mathbb{P}(|\mathbb{E}(\eta_N)-t^*|>\tau/2)$. It is straightforward to show that $\mathrm{Var}(\eta_N)\to 0$ and $\mathbb{E}(\eta_N)\to t^*$ as $N\to+\infty$ and, therefore, $\mathbb{P}(|\eta_N - t^*| > \tau)\to 0$, implying that η_N converges to t^* in probability. This demonstrates the claim, which implies that (I) $\to 0$ as $N\to+\infty$. Similarly, we have (III) $\to 0$. Term (II) satisfies that (II) $\le \varepsilon\int_{t^*-\delta}^{t^*+\delta} b_N(t)\mathrm{d}t \le \varepsilon\int_0^1 b_N(t)\mathrm{d}t = \varepsilon$. Therefore, (EC.85) is less than 2ε when N is sufficiently large, and we can estimate (EC.84) by

$$\begin{aligned}
&(\mathscr{D}_1 C(u_2 t^*, F_\theta(x_1)) - 2\varepsilon)u_2^{i_2-1} \le \\
&\qquad \int_0^{u_2} \mathscr{D}_1 C(u_1, F_\theta(x_1)) u_1^{i_1-1}(u_2-u_1)^{i_2-i_1-1}\frac{(i_2-1)!}{(i_1-1)!(i_2-i_1-1)!}\mathrm{d}u_1 \\
&\qquad\qquad\qquad\qquad\qquad\qquad \le (\mathscr{D}_1 C(u_2 t^*, F_\theta(x_1)) + 2\varepsilon)u_2^{i_2-1}.
\end{aligned}$$

Using similar approaches above, we can deductively derive similar bounds for the iterated integrals in (EC.83) and conclude that (EC.83) converges to $\prod_{k=1}^m \mathscr{D}_1 C(\xi_k, F_\theta(x_k))$ as N increases without bound thanks to the arbitrariness of ε. In addition, we also have

$$\begin{aligned}
&\mathbb{P}\left(g(\xi_1,V_1)\le x_1,\dots,g(\xi_m,V_m)\le x_m\right) = \mathbb{P}\left(g_{\xi_1}(V_1)\le F_\theta(x_1),\dots,g_{\xi_m}(V_m)\le F_\theta(x_m)\right) \\
=&\mathbb{P}\left(V_1 \le \mathscr{D}_1 C(\xi_1, F_\theta(x_1)),\dots,V_m \le \mathscr{D}_1 C(\xi_m, F_\theta(x_m))\right) = \prod_{k=1}^m \mathscr{D}_1 C(\xi_k, F_\theta(x_k)).
\end{aligned}$$

Therefore, we prove (17).

Proof of "only if." Let $\hat{C}$ be a copula of F linearly interpolating on $\overline{\mathcal{R}}_X^c \times \overline{\mathcal{R}}_\theta$. We aim to prove that $C \equiv \hat{C}$ on $[0,1]\times\overline{\mathcal{R}}_\theta$. Let $\hat{g}(u,v)$ be defined by (EC.81). Based on the result of Part "if," we have $\theta_{[i_k(N):N]} \xrightarrow{d} \hat{g}(\xi_k, V_k)$ for any $k=1,2,\dots,m$. Furthermore, because (17) holds, we have $\theta_{[i_k(N):N]} \xrightarrow{d} g(\xi_k, V_k)$. Therefore, $\hat{g}(\xi_k,V_k) \stackrel{d}{=} g(\xi_k,V_k)$. Note that for any $y\in\mathbb{R}$,

$$\mathbb{P}\left(\hat{g}(\xi_k,V_k)\le y\right) = \mathbb{P}\left(\hat{g}_{\xi_k}(V_k)\le F_\theta(y)\right) = \mathbb{P}\left(V_k \le \mathscr{D}_1\hat{C}(\xi_k, F_\theta(y))\right) = \mathscr{D}_1\hat{C}(\xi_k, F_\theta(y)),$$

and similarly, we have $\mathbb{P}(g(\xi_k, V_k) \le y) = \mathscr{D}_1 C(\xi_k, F_\theta(y))$. Therefore, because of the arbitrariness of $\xi_k \in (0,1)$, for any $y \in \mathbb{R}$, we have

$$\mathscr{D}_1 \hat{C}(\xi_k, F_\theta(y)) = \mathscr{D}_1 C(\xi_k, F_\theta(y)), \quad \forall \xi_k \in (0,1).$$

For any $a \in [0,1]$, by integrating the equation above with respect to ξ_k from 0 to a, according to Remark EC.1, we have $\hat{C}(a, F_\theta(y)) = C(a, F_\theta(y))$. Hence, $C \equiv \hat{C}$ holds on $[0,1] \times \mathcal{R}_\theta$, where $\mathcal{R}_\theta$ is the range of F_θ. The continuity of copula (Nelsen 2007, Theorem 2.2.4) further implies that $C \equiv \hat{C}$ holds on $[0,1] \times \overline{\mathcal{R}}_\theta$. This completes the proof. $\square$

EC.3.4. Proofs for Section 4

Proof of Proposition 3. This is a corollary of Theorem 2 because of the mixture function $g(u,w) = F_\theta^{-1} \circ \Phi\left(\rho\Phi^{-1}(u) + \sqrt{1-\rho^2}\Phi^{-1}(w)\right)$ and the fact that $\Phi^{-1}(U)$ follows $\mathcal{N}(0,1)$ if U follows Uniform$(0,1)$. $\square$

Proof of Proposition 4. This is a corollary of Theorem 4, and the proof is similar to the proof of Proposition 3. $\square$

Proof of Proposition 5. This is a corollary of Theorem 2 because of the mixture function $g(u,w) = F_\theta^{-1} \circ \phi\Big[\phi'^{-1}\left[\phi' \circ \phi^{-1}(u) \cdot w\right] - \phi^{-1}(u)\Big]$. $\square$

EC.3.5. Proofs for Section 5

Proof of Proposition 6. By (20), we have $w^{\mathrm{Ga}}(\xi) = \frac{\mathbb{E}\left[Q\left(\rho\Phi^{-1}(\xi)+\sqrt{1-\rho^2}Z\right)\right]}{\mathrm{Var}\left[Q\left(\rho\Phi^{-1}(\xi)+\sqrt{1-\rho^2}Z\right)\right]}$, where $Z \sim \mathcal{N}(0,1)$ and $Q(x) \equiv Q_{a,b}^{\mathrm{Skew}}(x)$ is given by (23). Thus, we need to study the first two moments of $Q\left(\rho\Phi^{-1}(\xi) + \sqrt{1-\rho^2}Z\right)$.

For notational simplicity, let $s \equiv \Phi^{-1}(\xi)$. The expectation is

$$\begin{aligned}
&\mathbb{E}\left[Q\left(\rho s + \sqrt{1-\rho^2}Z\right)\right] \\
&= a\int_{-\frac{\rho s}{\sqrt{1-\rho^2}}}^{+\infty} \left(\rho s + \sqrt{1-\rho^2}x\right)\frac{e^{-x^2/2}}{\sqrt{2\pi}}\mathrm{d}x + b\int_{-\infty}^{-\frac{\rho s}{\sqrt{1-\rho^2}}} \left(\rho s + \sqrt{1-\rho^2}x\right)\frac{e^{-x^2/2}}{\sqrt{2\pi}}\mathrm{d}x \\
&= a\rho s + (b-a)\left[\rho s\Phi\left(\frac{-\rho s}{\sqrt{1-\rho^2}}\right) - \sqrt{1-\rho^2}\varphi\left(\frac{-\rho s}{\sqrt{1-\rho^2}}\right)\right],
\end{aligned}$$

where φ is the density of $\mathcal{N}(0,1)$. Therefore, we have

$$\lim_{s\to+\infty} \frac{\mathbb{E}\left[Q\left(\rho s + \sqrt{1-\rho^2}Z\right)\right]}{a\rho s} = 1, \quad \lim_{s\to-\infty} \frac{\mathbb{E}\left[Q\left(\rho s + \sqrt{1-\rho^2}Z\right)\right]}{b\rho s} = 1. \tag{EC.86}$$

Similarly, direct calculation shows that the second-order moment is

$$\begin{aligned}
&\mathbb{E}\left[Q\left(\rho s + \sqrt{1-\rho^2}Z\right)\right]^2 \\
&= a^2\rho^2 s^2 + a^2(1-\rho^2) \\
&\quad + (b^2 - a^2)\left[(1-\rho^2)\zeta\left(\frac{-\rho s}{\sqrt{1-\rho^2}}\right) - 2\rho\sqrt{1-\rho^2}s\varphi\left(\frac{-\rho s}{\sqrt{1-\rho^2}}\right) + \rho^2 s^2\Phi\left(\frac{-\rho s}{\sqrt{1-\rho^2}}\right)\right],
\end{aligned}$$

where $\zeta(t)=\int_{-\infty}^{t} x^2 \frac{e^{-x^2/2}}{\sqrt{2\pi}}\mathrm{d}x$. Thus,

$$\lim_{s\to+\infty}\frac{\mathrm{Var}\left[Q\left(\rho s+\sqrt{1-\rho^2}Z\right)\right]}{a^2(1-\rho^2)}=1,\qquad \lim_{s\to-\infty}\frac{\mathrm{Var}\left[Q\left(\rho s+\sqrt{1-\rho^2}Z\right)\right]}{b^2(1-\rho^2)}=1. \tag{EC.87}$$

Combining (EC.86) and (EC.87) and replacing s with $\Phi^{-1}(\xi)$ completes the proof. $\square$

Proof of Proposition 7. Without loss of generality, we only prove the case of $\rho\in(0,1)$. By (20), we have $w^{\mathrm{Ga}}(\xi)=\frac{\mathbb{E}\left[Q\left(\rho\Phi^{-1}(\xi)+\sqrt{1-\rho^2}Z\right)\right]}{\mathrm{Var}\left[Q\left(\rho\Phi^{-1}(\xi)+\sqrt{1-\rho^2}Z\right)\right]}$, where $Z\sim\mathcal{N}(0,1)$ and $Q(x)\equiv Q^{\mathrm{HeavyTail}}_{\sigma,\tau,\beta}(x)$ is given by (24). Thus, we need to study the first two moments of $Q\left(\rho\Phi^{-1}(\xi)+\sqrt{1-\rho^2}Z\right)$. Since Q is an odd function, when $\xi=0.5$, we have

$$\mathbb{E}\left[Q\left(\rho\Phi^{-1}(\xi)+\sqrt{1-\rho^2}Z\right)\right]=\mathbb{E}\left[Q\left(\sqrt{1-\rho^2}Z\right)\right]=0,$$

which further implies that $w^{\mathrm{Ga}}(0.5)=0$.

Next, without loss of generality, we prove the results for $\xi>0.5$. Let us prove that, for $\xi\in(0.5,1)$, we have $w^{\mathrm{Ga}}(\xi)>0$. For notational simplicity, let $h_\beta(x)\equiv\begin{cases}|x|^\beta, & x\ge 0,\\ -|x|^\beta, & x<0,\end{cases}$ and $t\equiv\frac{1}{\Phi^{-1}(\xi)}>0$. Then,

$$\begin{aligned}\mathbb{E}\left[Q\left(\frac{\rho}{t}+\sqrt{1-\rho^2}Z\right)\right]&=\frac{\rho\sigma}{t}+\tau\int_{-\infty}^{+\infty}h_\beta\left(\frac{\rho}{t}+\sqrt{1-\rho^2}x\right)\frac{e^{-x^2/2}}{\sqrt{2\pi}}\mathrm{d}x\\&=\frac{\rho\sigma}{t}+\tau\left(\frac{\rho}{t}\right)^\beta\int_{-\infty}^{+\infty}h_\beta\left(1+\frac{\sqrt{1-\rho^2}}{\rho}tx\right)\frac{e^{-x^2/2}}{\sqrt{2\pi}}\mathrm{d}x.\end{aligned} \tag{EC.88}$$

We claim that this expectation is greater than zero. This is true since

$$\begin{aligned}&\int_{-\infty}^{+\infty}h_\beta\left(1+\frac{\sqrt{1-\rho^2}}{\rho}tx\right)\frac{e^{-x^2/2}}{\sqrt{2\pi}}\mathrm{d}x\\&=\int_{-\frac{\rho}{t\sqrt{1-\rho^2}}}^{+\infty}\left|1+\frac{\sqrt{1-\rho^2}}{\rho}tx\right|^\beta\frac{e^{-x^2/2}}{\sqrt{2\pi}}\mathrm{d}x-\int_{-\infty}^{-\frac{\rho}{t\sqrt{1-\rho^2}}}\left|1+\frac{\sqrt{1-\rho^2}}{\rho}tx\right|^\beta\frac{e^{-x^2/2}}{\sqrt{2\pi}}\mathrm{d}x\\&\overset{s=1+\sqrt{1-\rho^2}tx/\rho}{=\!=\!=\!=}\int_0^{+\infty}|s|^\beta\frac{e^{-(s-1)^2c^2/2}}{\sqrt{2\pi}}c\mathrm{d}s-\int_{-\infty}^{0}|s|^\beta\frac{e^{-(s-1)^2c^2/2}}{\sqrt{2\pi}}c\mathrm{d}s\\&=\int_0^{+\infty}|s|^\beta\frac{1}{\sqrt{2\pi}}c\left[e^{-(s-1)^2c^2/2}-e^{-(-s-1)^2c^2/2}\right]\mathrm{d}s=\int_0^{+\infty}|s|^\beta\frac{1}{\sqrt{2\pi}}ce^{-(s^2+1)c^2/2}\left(e^{sc^2}-e^{-sc^2}\right)\mathrm{d}s>0,\end{aligned}$$

where $c=\frac{\rho}{t\sqrt{1-\rho^2}}$. Hence, for $\xi\in(0.5,1)$, we have $w^{\mathrm{Ga}}(\xi)>0$.

We finally argue that when $\xi\to1^-$, i.e., $t=1/\Phi^{-1}(\xi)\to0^+$, we have $w^{\mathrm{Ga}}(\xi)\to0$. Let $f_\beta(t,x)=h_\beta\left(1+\frac{\sqrt{1-\rho^2}}{\rho}tx\right)\frac{e^{-x^2/2}}{\sqrt{2\pi}}$ and $u_\beta(t)=\int_{-\infty}^{+\infty}f_\beta(t,x)\mathrm{d}x$. We can easily check that, when $\beta>2$, for any $\delta>0$ and $t\in(-\delta,\delta)$, the following three properties hold:

(i) $\int_{-\infty}^{+\infty}|f_\beta(t,x)|\mathrm{d}x<+\infty$;

(ii) For a fixed x, $\frac{\partial f_\beta}{\partial t}(t,x)$ is a continuous function of t;

(iii) $\int_{-\infty}^{+\infty}\sup_{s\in[-\delta,\delta]}\left|\frac{\partial f_\beta}{\partial t}(s,x)\right|\mathrm{d}x<+\infty$.

Therefore, by Durrett (2019, Theorem A.5.3), $u'_\beta(0)=0$. With similar arguments, we can further show that $u''_\beta(0)=\frac{1-\rho^2}{\rho^2}\beta(\beta-1)$. Hence, when $\beta>2$, the second-order Taylor expansion of $u_\beta(t)$ is given by:

$$u_\beta(t)=1+\frac{1-\rho^2}{2\rho^2}\beta(\beta-1)t^2+o(t^2),\quad t\to 0^+. \tag{EC.89}$$

By combining (EC.88) and (EC.89), we have

$$\mathbb{E}\left[Q\left(\frac{\rho}{t}+\sqrt{1-\rho^2}Z\right)\right]=\frac{\rho\sigma}{t}+\tau\left(\frac{\rho}{t}\right)^\beta[1+o(1)],\quad t\to 0^+. \tag{EC.90}$$

Before examining the variance, let us make some more preparations. Let $\tilde{u}_\beta(t)=\int_{-\infty}^{+\infty}|f_\beta(t,x)|\mathrm{d}x$. Using similar approaches, when $\beta>2$, we can show that the second-order Taylor expansion of $\tilde{u}_\beta(t)$ is given by:

$$\tilde{u}_\beta(t)=1+\frac{1-\rho^2}{2\rho^2}\beta(\beta-1)t^2+o(t^2),\quad t\to 0^+. \tag{EC.91}$$

In addition, by direct calculation, we have

$$u_1(t)\equiv\int_{-\infty}^{+\infty}f_1(t,x)\mathrm{d}x=\int_{-\infty}^{+\infty}\left(1+\frac{\sqrt{1-\rho^2}}{\rho}tx\right)\frac{e^{-x^2/2}}{\sqrt{2\pi}}\mathrm{d}x=1. \tag{EC.92}$$

Hence,

$$\begin{aligned}
&\mathrm{Var}\left[Q\left(\frac{\rho}{t}+\sqrt{1-\rho^2}Z\right)\right]=\mathrm{Var}\left[\sigma\left(\frac{\rho}{t}+\sqrt{1-\rho^2}Z\right)+\tau h_\beta\left(\frac{\rho}{t}+\sqrt{1-\rho^2}Z\right)\right]\\
&=\sigma^2(1-\rho^2)+\tau^2\mathrm{Var}\left[h_\beta\left(\frac{\rho}{t}+\sqrt{1-\rho^2}Z\right)\right]+2\sigma\tau\mathrm{Cov}\left[\left(\frac{\rho}{t}+\sqrt{1-\rho^2}Z\right),h_\beta\left(\frac{\rho}{t}+\sqrt{1-\rho^2}Z\right)\right]\\
&=\sigma^2(1-\rho^2)+\tau^2\left(\frac{\rho}{t}\right)^{2\beta}\left[\tilde{u}_{2\beta}(t)-u_\beta(t)^2\right]+2\sigma\tau\left(\frac{\rho}{t}\right)^{\beta+1}\left[\tilde{u}_{\beta+1}(t)-u_1(t)u_\beta(t)\right]\\
&=\sigma^2(1-\rho^2)+\tau^2\frac{\rho^{2\beta}}{t^{2\beta-2}}\left[\frac{1-\rho^2}{\rho^2}\beta^2+o(1)\right]+2\sigma\tau\frac{\rho^{\beta+1}}{t^{\beta-1}}\left[\frac{1-\rho^2}{\rho^2}\beta+o(1)\right],\quad t\to 0^+,
\end{aligned} \tag{EC.93}$$

where the last equality holds due to (EC.89), (EC.91), and (EC.92). Therefore, by combining (EC.90) and (EC.93), as $t\to 0^+$, we have

$$\begin{aligned}
w^{\mathrm{Ga}}(\xi)&=\frac{\frac{\rho\sigma}{t}+\tau\left(\frac{\rho}{t}\right)^\beta[1+o(1)]}{\sigma^2(1-\rho^2)+\tau^2\frac{\rho^{2\beta}}{t^{2\beta-2}}\left[\frac{1-\rho^2}{\rho^2}\beta^2+o(1)\right]+2\sigma\tau\frac{\rho^{\beta+1}}{t^{\beta-1}}\left[\frac{1-\rho^2}{\rho^2}\beta+o(1)\right]}\\
&=\frac{\frac{1}{t^\beta}\left[\rho\sigma t^{\beta-1}+\tau\rho^\beta[1+o(1)]\right]}{\frac{1}{t^{2\beta-2}}\left[\sigma^2(1-\rho^2)t^{2\beta-2}+\tau^2\rho^{2\beta}\left[\frac{1-\rho^2}{\rho^2}\beta^2+o(1)\right]+2\sigma\tau\rho^{\beta+1}t^{\beta-1}\left[\frac{1-\rho^2}{\rho^2}\beta+o(1)\right]\right]}\to 0.
\end{aligned}$$

Hence, when $\xi\to 1^-$, we have $t=1/\Phi^{-1}(\xi)\to 0^+$ and $w^{\mathrm{Ga}}(\xi)\to 0$. □

EC.3.6. Proofs for Appendix EC.1.1

Proof of Proposition EC.1. The same as the proof of Proposition 1. □

Proof of Proposition EC.2. Under the single-factor model, (EC.4), the expectation and covariance matrix of raw returns, (EC.1), reduce to $\tilde{\boldsymbol{\mu}} = \mu_M \boldsymbol{\beta} + \boldsymbol{\mu}$ and $\tilde{\Sigma} = \Sigma + \sigma_M^2 \boldsymbol{\beta}\boldsymbol{\beta}^\top$. Hence, by the Sherman–Morrison formula, we have

$$\begin{aligned}\tilde{\Sigma}^{-1}\tilde{\boldsymbol{\mu}} = (\Sigma + \sigma_M^2\boldsymbol{\beta}\boldsymbol{\beta}^\top)^{-1}(\mu_M\boldsymbol{\beta}+\boldsymbol{\mu}) &= \left(\Sigma^{-1} - \frac{\sigma_M^2\Sigma^{-1}\boldsymbol{\beta}\boldsymbol{\beta}^\top\Sigma^{-1}}{1+\sigma_M^2\boldsymbol{\beta}^\top\Sigma^{-1}\boldsymbol{\beta}}\right)(\mu_M\boldsymbol{\beta}+\boldsymbol{\mu}) \\ &= \frac{\mu_M - \sigma_M^2\boldsymbol{\beta}^\top\Sigma^{-1}\boldsymbol{\mu}}{1+\sigma_M^2\boldsymbol{\beta}^\top\Sigma^{-1}\boldsymbol{\beta}}\Sigma^{-1}\boldsymbol{\beta} + \Sigma^{-1}\boldsymbol{\mu}.\end{aligned}$$

Furthermore, direct calculation shows that

$$\begin{aligned}\tilde{\boldsymbol{\mu}}^\top\tilde{\Sigma}^{-1}\tilde{\boldsymbol{\mu}} &= (\mu_M\boldsymbol{\beta}+\boldsymbol{\mu})^\top\left(\frac{\mu_M - \sigma_M^2\boldsymbol{\beta}^\top\Sigma^{-1}\boldsymbol{\mu}}{1+\sigma_M^2\boldsymbol{\beta}^\top\Sigma^{-1}\boldsymbol{\beta}}\Sigma^{-1}\boldsymbol{\beta} + \Sigma^{-1}\boldsymbol{\mu}\right) \\ &= \frac{\mu_M^2\boldsymbol{\beta}^\top\Sigma^{-1}\boldsymbol{\beta} + 2\mu_M\boldsymbol{\beta}^\top\Sigma^{-1}\boldsymbol{\mu} - \sigma_M^2(\boldsymbol{\beta}^\top\Sigma^{-1}\boldsymbol{\mu})^2}{1+\sigma_M^2\boldsymbol{\beta}^\top\Sigma^{-1}\boldsymbol{\beta}} + \boldsymbol{\mu}^\top\Sigma^{-1}\boldsymbol{\mu}.\end{aligned}$$

This completes the proof. □

Proof of Proposition EC.3. Similar to the proof of Proposition 1, the optimal weight vector for the $N+1$ assets is $\hat{\boldsymbol{w}} = \begin{pmatrix}\boldsymbol{w}\\ w_M\end{pmatrix} \propto \hat{\Sigma}^{-1}\hat{\boldsymbol{\mu}}$, and the optimal Sharpe ratio is $\sqrt{\hat{\boldsymbol{\mu}}^\top\hat{\Sigma}^{-1}\hat{\boldsymbol{\mu}}}$, where $\hat{\boldsymbol{\mu}}$ and $\hat{\Sigma}$ are given by (EC.5). Note that the inverse of $\hat{\Sigma}$ is

$$\hat{\Sigma}^{-1} = \begin{pmatrix}\sigma_M^2\boldsymbol{\beta}\boldsymbol{\beta}^\top + \Sigma & \sigma_M^2\boldsymbol{\beta}\\ \sigma_M^2\boldsymbol{\beta}^\top & \sigma_M^2\end{pmatrix}^{-1} = \begin{pmatrix}\Sigma^{-1} & -\Sigma^{-1}\boldsymbol{\beta}\\ -\boldsymbol{\beta}^\top\Sigma^{-1} & \frac{1+\sigma_M^2\boldsymbol{\beta}^\top\Sigma^{-1}\boldsymbol{\beta}}{\sigma_M^2}\end{pmatrix},$$

and thus,

$$\hat{\Sigma}^{-1}\hat{\boldsymbol{\mu}} = \begin{pmatrix}\Sigma^{-1} & -\Sigma^{-1}\boldsymbol{\beta}\\ -\boldsymbol{\beta}^\top\Sigma^{-1} & \frac{1+\sigma_M^2\boldsymbol{\beta}^\top\Sigma^{-1}\boldsymbol{\beta}}{\sigma_M^2}\end{pmatrix}\begin{pmatrix}\mu_M\boldsymbol{\beta}+\boldsymbol{\mu}\\ \mu_M\end{pmatrix} = \begin{pmatrix}\Sigma^{-1}\boldsymbol{\mu}\\ \frac{\mu_M}{\sigma_M^2} - \boldsymbol{\beta}^\top\Sigma^{-1}\boldsymbol{\mu}\end{pmatrix}.$$

Furthermore, the squared optimal Sharpe ratio is

$$\hat{\boldsymbol{\mu}}^\top\hat{\Sigma}^{-1}\hat{\boldsymbol{\mu}} = \begin{pmatrix}\mu_M\boldsymbol{\beta}^\top + \boldsymbol{\mu}^\top & \mu_M\end{pmatrix}\begin{pmatrix}\Sigma^{-1}\boldsymbol{\mu}\\ \frac{\mu_M}{\sigma_M^2} - \boldsymbol{\beta}^\top\Sigma^{-1}\boldsymbol{\mu}\end{pmatrix} = \frac{\mu_M^2}{\sigma_M^2} + \boldsymbol{\mu}^\top\Sigma^{-1}\boldsymbol{\mu}.$$

This completes the proof. □

EC.3.7. Proofs for Appendix EC.1.2

Proof of Proposition EC.4. This is a corollary of Theorem 1 because of the mixture function $g(u,w) = \sigma_\theta\left(\rho\Phi^{-1}(u) + \sqrt{1-\rho^2}\Phi^{-1}(w)\right) + \mu_\theta$ and the fact that $\Phi^{-1}(U)$ follows $\mathcal{N}(0,1)$ if U follows Uniform$(0,1)$. It is also a corollary of Proposition 3 when $F_\theta(y) = \Phi((y-\mu_\theta)/\sigma_\theta)$. See also David and Nagaraja (2004, Section 6.8). □

Proof of Proposition EC.5. Like the proof of Proposition EC.4, this is a corollary of Theorem 3 and also a corollary of Proposition 4. See also Lo and MacKinlay (1990). □

Proof of Theorem EC.1. This is a special case of Theorem EC.2 when $\sigma_\rho = 0$, and also a special case of Theorem EC.3 when $\tilde{\rho} = \rho_X = \rho_\theta = 0$. ☐

Proof of Proposition EC.6. This is a corollary of Part (i) of Lemma EC.6. ☐

Proof of Theorem EC.2. The optimal information ratio given by (9) is $\sqrt{\boldsymbol{\mu}^\top \Sigma^{-1} \boldsymbol{\mu}}$. Therefore, we first need to specify the expressions of $\boldsymbol{\mu}$ and Σ under Assumption EC.2. By Lemma EC.8, the i-th entry of $\boldsymbol{\mu}$ is given by (EC.56). In addition, the covariance matrix Σ can be decomposed into $\Sigma = A + B$, where, for $i, j = 1, 2, \ldots, N$ and $i \neq j$, the (i,i)-entry of A is given by (EC.58), and the (i,j)-entry of A is given by (EC.61); the (i,i)-entry of B is given by (EC.59), and the (i,j)-entry of B is given by (EC.62).

Let $\boldsymbol{\eta} = (\mathbb{E}(Y_{1:N}), \mathbb{E}(Y_{2:N}), \ldots, \mathbb{E}(Y_{N:N}))^\top$, then the matrix B can be written as $B = \sigma_\rho^2 \cdot \sigma_\theta^2 \cdot \boldsymbol{\eta}\boldsymbol{\eta}^\top$. Thus, by the Sherman–Morrison formula,

$$\Sigma^{-1} = (A+B)^{-1} = A^{-1} - \frac{\sigma_\rho^2 \cdot \sigma_\theta^2 \cdot A^{-1}\boldsymbol{\eta}\boldsymbol{\eta}^\top A^{-1}}{1 + \sigma_\rho^2 \cdot \sigma_\theta^2 \cdot \boldsymbol{\eta}^\top A^{-1}\boldsymbol{\eta}}.$$

In addition, by (EC.56), we have $\boldsymbol{\mu} = \overline{\rho} \cdot \sigma_\theta \cdot \boldsymbol{\eta} + \mu_\theta \mathbf{1}$, where $\mathbf{1} \in \mathbb{R}^N$ is an all-one vector, and thus,

$$\begin{aligned}(\mathrm{IR}^*_{\mathrm{TimeVary}})^2 = \boldsymbol{\mu}^\top \Sigma^{-1} \boldsymbol{\mu} = \overline{\rho}^2 \cdot \sigma_\theta^2 \cdot \boldsymbol{\eta}^\top A^{-1}\boldsymbol{\eta} + 2\overline{\rho} \cdot \sigma_\theta \cdot \mu_\theta \cdot \boldsymbol{\eta}^\top A^{-1}\mathbf{1} + \mu_\theta^2 \mathbf{1}^\top A^{-1}\mathbf{1} \\ - \frac{\sigma_\rho^2 \cdot \sigma_\theta^2 \cdot (\overline{\rho} \cdot \sigma_\theta \cdot \boldsymbol{\eta}^\top A^{-1}\boldsymbol{\eta} + \mu_\theta \mathbf{1}^\top A^{-1}\boldsymbol{\eta})^2}{1 + \sigma_\rho^2 \cdot \sigma_\theta^2 \cdot \boldsymbol{\eta}^\top A^{-1}\boldsymbol{\eta}}.\end{aligned} \tag{EC.94}$$

Now we claim that

$$\mathbf{1}^\top A^{-1}\boldsymbol{\eta} = 0. \tag{EC.95}$$

To prove this claim, note that A is a symmetric matrix, and hence A^{-1} is also symmetric. Meanwhile, Part (ii) of Lemma EC.5 implies that A is also a persymmetric matrix, i.e., $A_{ij} = A_{N+1-j,N+1-i}$, for any $i, j = 1, 2, \ldots, N$. By the property of persymmetric matrices, A^{-1} is also persymmetric (Horn and Johnson 2012, Page 36). Thus, with the help of Part (i) of Lemma EC.5,

$$\begin{aligned}\mathbf{1}^\top A^{-1}\boldsymbol{\eta} &= \sum_{j=1}^N \sum_{i=1}^N (A^{-1})_{ij}\mathbb{E}(Y_{j:N}) = \sum_{j=1}^N \sum_{i=1}^N (A^{-1})_{N+1-j,N+1-i}[-\mathbb{E}(Y_{N+1-j:N})] \\ &= -\sum_{j=1}^N \sum_{i=1}^N (A^{-1})_{N+1-i,N+1-j}\mathbb{E}(Y_{N+1-j:N}) = -\sum_{k=1}^N \sum_{s=1}^N (A^{-1})_{s,k}\mathbb{E}(Y_{k:N}) = -\mathbf{1}^\top A^{-1}\boldsymbol{\eta},\end{aligned}$$

which proves the claim. Therefore, (EC.94) reduces to

$$\begin{aligned}(\mathrm{IR}^*_{\mathrm{TimeVary}})^2 &= \overline{\rho}^2 \cdot \sigma_\theta^2 \cdot \boldsymbol{\eta}^\top A^{-1}\boldsymbol{\eta} + \mu_\theta^2 \mathbf{1}^\top A^{-1}\mathbf{1} - \frac{\sigma_\rho^2 \cdot \sigma_\theta^4 \cdot \overline{\rho}^2 (\boldsymbol{\eta}^\top A^{-1}\boldsymbol{\eta})^2}{1 + \sigma_\rho^2 \cdot \sigma_\theta^2 \cdot \boldsymbol{\eta}^\top A^{-1}\boldsymbol{\eta}} \\ &= \frac{\overline{\rho}^2}{1/(\sigma_\theta^2 \cdot \boldsymbol{\eta}^\top A^{-1}\boldsymbol{\eta}) + \sigma_\rho^2} + \mu_\theta^2 \mathbf{1}^\top A^{-1}\mathbf{1}.\end{aligned} \tag{EC.96}$$

Let $A = Q^\top \Lambda Q$ be the eigendecomposition of A, where Q is orthogonal and Λ is a diagonal matrix whose diagonal entries are eigenvalues of A.

We first prove the upper bound of $\mathrm{IR}^*_{\mathrm{TimeVary}}$. Denote by $\lambda_{\min}(A)$ the smallest eigenvalue of A, and Lemma EC.7 implies that $\lambda_{\min}(A) \geq \min_{i=1,2,\ldots,N}\left[A_{ii} - \sum_{j\neq i}|A_{ij}|\right]$, where A_{ij} is the (i,j)-entry of A, $i,j=1,2,\ldots,N$. Furthermore, by Parts (iv) and (v) of Lemma EC.5 and the assumption of $\sqrt{\overline{\rho}^2+\sigma_\rho^2} \leq \frac{\sqrt{2}}{2}$, for any $i=1,2,\ldots,N$, we have

$$
\begin{aligned}
A_{ii} - \sum_{j\neq i}|A_{ij}| &= \sigma_\theta^2 \cdot \left[1-(\overline{\rho}^2+\sigma_\rho^2)+(\overline{\rho}^2+\sigma_\rho^2)\cdot \mathrm{Var}(Y_{i:N})\right] - \sum_{j\neq i}\sigma_\theta^2\cdot(\overline{\rho}^2+\sigma_\rho^2)\cdot|\mathrm{Cov}(Y_{i:N},Y_{j:N})| \\
&= \sigma_\theta^2\cdot\left[1-(\overline{\rho}^2+\sigma_\rho^2)+(\overline{\rho}^2+\sigma_\rho^2)\cdot\mathrm{Var}(Y_{i:N}) - (\overline{\rho}^2+\sigma_\rho^2)\cdot\mathrm{Cov}\left(Y_{i:N},\sum_{j\neq i}Y_{j:N}\right)\right] \\
&= \sigma_\theta^2\left[1-2\cdot(\overline{\rho}^2+\sigma_\rho^2)\cdot(1-\mathrm{Var}(Y_{i:N}))\right] \geq \sigma_\theta^2\left[1-2\cdot(\overline{\rho}^2+\sigma_\rho^2)\right]\geq 0. \qquad \text{(EC.97)}
\end{aligned}
$$

Therefore, we have $\lambda_{\min}(A)\geq\sigma_\theta^2\left[1-2\cdot(\overline{\rho}^2+\sigma_\rho^2)\right]\geq 0$. Combining this with (EC.96) leads to the upper bound for IR:

$$
\begin{aligned}
(\mathrm{IR}^*_{\mathrm{TimeVary}})^2 &\leq \frac{\overline{\rho}^2}{\lambda_{\min}(A)/(\sigma_\theta^2\cdot\boldsymbol{\eta}^\top QQ^\top\boldsymbol{\eta})+\sigma_\rho^2} + \frac{\mu_\theta^2\mathbf{1}^\top QQ^\top\mathbf{1}}{\lambda_{\min}(A)} \\
&= \frac{\overline{\rho}^2}{\lambda_{\min}(A)/(\sigma_\theta^2\cdot q(N))+\sigma_\rho^2} + \frac{\mu_\theta^2 N}{\lambda_{\min}(A)} \\
&\leq \frac{\overline{\rho}^2}{[1-2\cdot(\overline{\rho}^2+\sigma_\rho^2)]/q(N)+\sigma_\rho^2} + \frac{\mu_\theta^2 N}{\sigma_\theta^2\left[1-2\cdot(\overline{\rho}^2+\sigma_\rho^2)\right]}. \qquad \text{(EC.98)}
\end{aligned}
$$

We then prove the lower bound of $\mathrm{IR}^*_{\mathrm{TimeVary}}$. Denote by $\lambda_{\max}(A)$ the largest eigenvalue of A, and Lemma EC.7 implies that $\lambda_{\max}(A)\leq\max_{i=1,2,\ldots,N}\sum_{j=1}^N|A_{ij}|$. Using similar arguments to (EC.97), we can prove that $\sum_{j=1}^N|A_{ij}|=\sigma_\theta^2$. Hence, $\lambda_{\max}(A)\leq\sigma_\theta^2$. Then, through proofs similar to (EC.98), we can obtain the lower bound:

$$
(\mathrm{IR}^*_{\mathrm{TimeVary}})^2 \geq \frac{\overline{\rho}^2}{1/q(N)+\sigma_\rho^2} + \frac{\mu_\theta^2 N}{\sigma_\theta^2}.
$$

This completes the proof. ☐

Proof of Theorem EC.3. The optimal information ratio given by (9) is $\sqrt{\boldsymbol{\mu}^\top\Sigma^{-1}\boldsymbol{\mu}}$. Therefore, we first need to specify the expressions of $\boldsymbol{\mu}$ and Σ under Assumption EC.3. By Lemma EC.9, $\boldsymbol{\mu}=\frac{\rho-\tilde{\rho}}{\sqrt{1-\rho_X}}\cdot\sigma_\theta\cdot\boldsymbol{\eta}+\mu_\theta\mathbf{1}$, where $\boldsymbol{\eta}\equiv(\mathbb{E}[Y_{1:N}],\mathbb{E}[Y_{2:N}],\ldots,\mathbb{E}[Y_{N:N}])^\top$. In addition, the covariance matrix Σ can be decomposed into $\Sigma=P+J$, where $J=\sigma_\theta^2\cdot\rho_\theta\cdot\mathbf{1}\mathbf{1}^\top$, $\mathbf{1}\in\mathbb{R}^N$ is an all-one vector, and $P=(P_{ij})_{i,j=1}^N$ satisfies

$$
P_{ii}=\sigma_\theta^2\cdot\left(1-\rho_\theta-\frac{(\rho-\tilde{\rho})^2}{1-\rho_X}+\frac{(\rho-\tilde{\rho})^2}{1-\rho_X}\cdot\mathrm{Var}(Y_{i:N})\right), \quad P_{ij}=\sigma_\theta^2\cdot\frac{(\rho-\tilde{\rho})^2}{1-\rho_X}\mathrm{Cov}(Y_{i:N},Y_{j:N}), \qquad \text{(EC.99)}
$$

for $i,j=1,2,\ldots,N$ and $i\neq j$. By the Sherman–Morrison formula,

$$
\Sigma^{-1}=(P+J)^{-1}=P^{-1}-\frac{\sigma_\theta^2\cdot\rho_\theta\cdot P^{-1}\mathbf{1}\mathbf{1}^\top P^{-1}}{1+\sigma_\theta^2\cdot\rho_\theta\cdot\mathbf{1}^\top P^{-1}\mathbf{1}}.
$$

Hence, the optimal information ratio satisfies

$$(\mathrm{IR}^*_{\mathrm{Dependence}})^2 = \boldsymbol{\mu}^\top \Sigma^{-1} \boldsymbol{\mu} = \frac{(\rho-\tilde{\rho})^2}{1-\rho_X} \cdot \sigma_\theta^2 \cdot \boldsymbol{\eta}^\top P^{-1} \boldsymbol{\eta} + 2\frac{\rho-\tilde{\rho}}{\sqrt{1-\rho_X}} \cdot \sigma_\theta \cdot \mu_\theta \cdot \mathbf{1}^\top P^{-1} \boldsymbol{\eta} + \mu_\theta^2 \mathbf{1}^\top P^{-1} \mathbf{1} - \frac{\sigma_\theta^2 \cdot \rho_\theta \cdot \left(\frac{\rho-\tilde{\rho}}{\sqrt{1-\rho_X}} \cdot \sigma_\theta \cdot \mathbf{1}^\top P^{-1} \boldsymbol{\eta} + \mu_\theta \mathbf{1}^\top P^{-1} \mathbf{1} \right)^2}{1+\sigma_\theta^2 \cdot \rho_\theta \cdot \mathbf{1}^\top P^{-1} \mathbf{1}}. \quad \text{(EC.100)}$$

Similar to the proof of (EC.95), we can prove that $\mathbf{1}^\top P^{-1} \boldsymbol{\eta} = 0$. Therefore, (EC.100) reduces to

$$\begin{aligned}(\mathrm{IR}^*_{\mathrm{Dependence}})^2 &= \frac{(\rho-\tilde{\rho})^2}{1-\rho_X} \cdot \sigma_\theta^2 \cdot \boldsymbol{\eta}^\top P^{-1} \boldsymbol{\eta} + \mu_\theta^2 \mathbf{1}^\top P^{-1} \mathbf{1} - \frac{\sigma_\theta^2 \cdot \rho_\theta \cdot \mu_\theta^2 \cdot (\mathbf{1}^\top P^{-1} \mathbf{1})^2}{1+\sigma_\theta^2 \cdot \rho_\theta \cdot \mathbf{1}^\top P^{-1} \mathbf{1}} \\ &= \frac{(\rho-\tilde{\rho})^2}{1-\rho_X} \cdot \sigma_\theta^2 \cdot \boldsymbol{\eta}^\top P^{-1} \boldsymbol{\eta} + \frac{\mu_\theta^2}{1/(\mathbf{1}^\top P^{-1} \mathbf{1}) + \sigma_\theta^2 \cdot \rho_\theta}. \end{aligned} \quad \text{(EC.101)}$$

Let $P = Q^\top \Lambda Q$ be the eigendecomposition of P, where Q is orthogonal and Λ is a diagonal matrix whose diagonal entries are eigenvalues of P.

We first prove the upper bound of $\mathrm{IR}^*_{\mathrm{Dependence}}$. Denote by $\lambda_{\min}(P)$ the smallest eigenvalue of P, and Lemma EC.7 implies that $\lambda_{\min}(P) \ge \min_{i=1,2,\dots,N} \left[P_{ii} - \sum_{j\ne i} |P_{ij}| \right]$, where P_{ii} and P_{ij} are given by (EC.99). In addition, by Parts (iv) and (v) of Lemma EC.5 and (EC.22), for any $i = 1,2,\dots,N$, we have

$$\begin{aligned} P_{ii} - \sum_{j\ne i} |P_{ij}| &= \sigma_\theta^2 \cdot \left[1 - \rho_\theta - \frac{(\rho-\tilde{\rho})^2}{1-\rho_X} + \frac{(\rho-\tilde{\rho})^2}{1-\rho_X} \cdot \mathrm{Var}(Y_{i:N}) \right] - \sum_{j \ne i} \sigma_\theta^2 \cdot \frac{(\rho-\tilde{\rho})^2}{1-\rho_X} \left| \mathrm{Cov}(Y_{i:N}, Y_{j:N}) \right| \\ &= \sigma_\theta^2 \cdot \left[1 - \rho_\theta - \frac{(\rho-\tilde{\rho})^2}{1-\rho_X} + \frac{(\rho-\tilde{\rho})^2}{1-\rho_X} \cdot \mathrm{Var}(Y_{i:N}) - \frac{(\rho-\tilde{\rho})^2}{1-\rho_X} \cdot \mathrm{Cov}\left(Y_{i:N}, \sum_{j\ne i} Y_{j:N} \right) \right] \\ &= \sigma_\theta^2 \left[1 - \rho_\theta - 2 \cdot \frac{(\rho-\tilde{\rho})^2}{1-\rho_X} \cdot (1 - \mathrm{Var}(Y_{i:N})) \right] \ge \sigma_\theta^2 \left[1 - \rho_\theta - 2 \cdot \frac{(\rho-\tilde{\rho})^2}{1-\rho_X} \right]. \end{aligned} \quad \text{(EC.102)}$$

Hence,

$$\lambda_{\min}(P) \ge \sigma_\theta^2 \left[1 - \rho_\theta - 2 \cdot \frac{(\rho-\tilde{\rho})^2}{1-\rho_X} \right] = \sigma_\theta^2 \cdot \frac{(1-\rho_\theta)(1-\rho_X) - 2(\rho-\tilde{\rho})^2}{1-\rho_X} \ge 0. \quad \text{(EC.103)}$$

Combining this with (EC.101), we have

$$\begin{aligned} (\mathrm{IR}^*_{\mathrm{Dependence}})^2 &= \frac{(\rho-\tilde{\rho})^2}{1-\rho_X} \cdot \sigma_\theta^2 \cdot \boldsymbol{\eta}^\top P^{-1} \boldsymbol{\eta} + \frac{\mu_\theta^2}{1/(\mathbf{1}^\top P^{-1} \mathbf{1}) + \sigma_\theta^2 \cdot \rho_\theta} \\ &\le \frac{(\rho-\tilde{\rho})^2}{1-\rho_X} \cdot \sigma_\theta^2 \cdot \frac{1}{\lambda_{\min}(P)} \boldsymbol{\eta}^\top Q Q^\top \boldsymbol{\eta} + \frac{\mu_\theta^2}{\lambda_{\min}(P)/(\mathbf{1}^\top Q Q^\top \mathbf{1}) + \sigma_\theta^2 \cdot \rho_\theta} \\ &= \frac{(\rho-\tilde{\rho})^2}{1-\rho_X} \cdot \sigma_\theta^2 \cdot \frac{1}{\lambda_{\min}(P)} q(N) + \frac{\mu_\theta^2}{\lambda_{\min}(P)/N + \sigma_\theta^2 \cdot \rho_\theta} \\ &\le \frac{(\rho-\tilde{\rho})^2 q(N)}{(1-\rho_\theta)(1-\rho_X) - 2(\rho-\tilde{\rho})^2} + \frac{\mu_\theta^2/\sigma_\theta^2}{[(1-\rho_\theta)(1-\rho_X) - 2(\rho-\tilde{\rho})^2]/[(1-\rho_X)N] + \rho_\theta}, \end{aligned} \quad \text{(EC.104)}$$

which proves the upper bound.

Now we prove the lower bound of $\mathrm{IR}^*_{\text{Dependence}}$. Denote by $\lambda_{\max}(P)$ the largest eigenvalue of P, and Lemma EC.7 implies that $\lambda_{\max}(P) \le \max_{i=1,2,\ldots,N} \sum_{j=1}^N |P_{ij}|$. Using arguments similar to (EC.102), we can derive that $\sum_{j=1}^N |P_{ij}| = \sigma_\theta^2 \cdot (1-\rho_\theta)$. Hence, $\lambda_{\max}(P) \le \sigma_\theta^2 \cdot (1-\rho_\theta)$, and similar to the proof of (EC.104), we can obtain that

$$(\mathrm{IR}^*_{\text{Dependence}})^2 \ge \frac{(\rho-\tilde{\rho})^2 q(N)}{(1-\rho_\theta)(1-\rho_X)} + \frac{\mu_\theta^2/\sigma_\theta^2}{(1-\rho_\theta)/N+\rho_\theta},$$

which proves the lower bound. This completes the proof. $\square$

Proof of Proposition EC.7. By (9), the optimal weights to maximize the information ratio have the form of $\Sigma^{-1}\boldsymbol{\mu}$. Let $Y_{1:N} \le Y_{2:N} \le \cdots \le Y_{N:N}$ be the order statistics of $Y_1, Y_2, \ldots, Y_N \overset{\text{IID}}{\sim} \mathcal{N}(0,1)$. Using Proposition EC.5, we approximate the variances and covariances of induced order statistics by $\mathrm{Var}(\theta_{[i:N]}) \approx \sigma_\theta^2 \cdot (1-\rho^2)$ and $\mathrm{Cov}(\theta_{[i:N]}, \theta_{[j:N]}) \approx 0$. In addition, when $\mu_\theta = 0$, (EC.8) states that $\mathbb{E}(\theta_{[i:N]}) = \sigma_\theta \cdot \rho \cdot \mathbb{E}(Y_{i:N})$. Thus, the optimal weight to maximize the IR, $\bar{\boldsymbol{w}}^* \equiv (0,\ldots,0,w^*_{\lfloor N\xi_1\rfloor}, w^*_{\lfloor N\xi_1+1\rfloor}, \ldots, w^*_{\lfloor N\xi_2\rfloor}, 0, \ldots, 0)$, can be approximated by

$$w_i^* \approx \frac{\mathbb{E}(Y_{i:N})}{\mathrm{Var}(\theta_{[i:N]})} \approx \frac{\sigma_\theta \cdot \rho \cdot \mathbb{E}(Y_{i:N})}{\sigma_\theta^2 \cdot (1-\rho^2)} = \frac{\rho \cdot \mathbb{E}(Y_{i:N})}{\sigma_\theta \cdot (1-\rho^2)}, \quad \lfloor N\xi_1 \rfloor \le i \le \lfloor N\xi_2 \rfloor,$$

where $\lfloor x \rfloor$ is the largest integer not greater than x. Hence, the optimal information ratio satisfies

$$\begin{aligned}
\mathrm{IR} = \frac{\bar{\boldsymbol{w}}^{*\top}\boldsymbol{\mu}}{\sqrt{\bar{\boldsymbol{w}}^{*\top}\Sigma\bar{\boldsymbol{w}}^*}} &\approx \frac{\sum_{i=\lfloor N\xi_1\rfloor}^{\lfloor N\xi_2\rfloor}\left[\frac{\rho\cdot\mathbb{E}(Y_{i:N})}{\sigma_\theta\cdot(1-\rho^2)}\right]\cdot[\sigma_\theta\cdot\rho\cdot\mathbb{E}(Y_{i:N})]}{\sqrt{\sum_{i=\lfloor N\xi_1\rfloor}^{\lfloor N\xi_2\rfloor}\left[\frac{\rho\cdot\mathbb{E}(Y_{i:N})}{\sigma_\theta\cdot(1-\rho^2)}\right]^2\cdot[\sigma_\theta^2\cdot(1-\rho^2)]}} \\
&= \sqrt{\sum_{i=\lfloor N\xi_1\rfloor}^{\lfloor N\xi_2\rfloor}\frac{\rho^2\cdot[\mathbb{E}(Y_{i:N})]^2}{1-\rho^2}} = \frac{|\rho|\cdot\sqrt{N}}{\sqrt{1-\rho^2}}\cdot\sqrt{\frac{\sum_{i=\lfloor N\xi_1\rfloor}^{\lfloor N\xi_2\rfloor}[\mathbb{E}(Y_{i:N})]^2}{N}} \\
&\to \frac{|\rho|\cdot\sqrt{N}}{\sqrt{1-\rho^2}}\cdot\sqrt{\left(\xi_2-\Phi^{-1}(\xi_2)\varphi\big(\Phi^{-1}(\xi_2)\big)\right)-\left(\xi_1-\Phi^{-1}(\xi_1)\varphi\big(\Phi^{-1}(\xi_1)\big)\right)},
\end{aligned}$$

as $N\to+\infty$, where the limit holds due to Part (i) of Lemma EC.6. Thus, (EC.26) holds. Meanwhile, since $X_i \sim \mathcal{N}(\mu_X, \sigma_X^2)$, the average impact factor satisfies

$$\begin{aligned}
\bar{X} &= \mathbb{E}\left[\frac{\sum_{i=\lfloor N\xi_1\rfloor}^{\lfloor N\xi_2\rfloor} w_i^* X_{i:N}}{\sum_{i=\lfloor N\xi_1\rfloor}^{\lfloor N\xi_2\rfloor}|w_i^*|}\right] = \mathbb{E}\left[\frac{\sum_{i=\lfloor N\xi_1\rfloor}^{\lfloor N\xi_2\rfloor} w_i^* X_{i:N}}{\sum_{i=\lfloor N\xi_1\rfloor}^{\lfloor N\xi_2\rfloor}|w_i^*|}\right] = \frac{\mu_X\sum_{i=\lfloor N\xi_1\rfloor}^{\lfloor N\xi_2\rfloor} w_i^* + \sigma_X\sum_{i=\lfloor N\xi_1\rfloor}^{\lfloor N\xi_2\rfloor} w_i^*\mathbb{E}(Y_{i:N})}{\sum_{i=\lfloor N\xi_1\rfloor}^{\lfloor N\xi_2\rfloor}|w_i^*|} \\
&\approx \mathrm{sign}(\rho)\cdot\frac{\mu_X\sum_{i=\lfloor N\xi_1\rfloor}^{\lfloor N\xi_2\rfloor}\mathbb{E}(Y_{i:N}) + \sigma_X\sum_{i=\lfloor N\xi_1\rfloor}^{\lfloor N\xi_2\rfloor}\mathbb{E}(Y_{i:N})\mathbb{E}(Y_{i:N})}{\sum_{i=\lfloor N\xi_1\rfloor}^{\lfloor N\xi_2\rfloor}|\mathbb{E}(Y_{i:N})|},
\end{aligned}$$

and thus, (EC.27) holds thanks to Parts (i)–(iii) of Lemma EC.6. $\square$

Proof of Proposition EC.8. As in the proof of Proposition EC.7, we have $\mathrm{Var}(\theta_{[i:N]}) \approx \sigma_\theta^2 \cdot (1-\rho^2)$ and $\mathrm{Cov}(\theta_{[i:N]}, \theta_{[j:N]}) \approx 0$. In addition, when $\mu_\theta = 0$, (EC.8) implies that $\mathbb{E}(\theta_{[i:N]}) = \sigma_\theta \cdot \rho \cdot \mathbb{E}(Y_{i:N})$. Hence, the information ratio of the equal-weighted portfolio can be approximated by

$$\mathrm{IR} \approx \frac{\sum_{i=\lfloor N\xi_1\rfloor}^{\lfloor N\xi_2\rfloor}\frac{1}{\lfloor N\xi_2\rfloor-\lfloor N\xi_1\rfloor+1}\cdot[\sigma_\theta\cdot\rho\cdot\mathbb{E}(Y_{i:N})]}{\sqrt{\sum_{i=\lfloor N\xi_1\rfloor}^{\lfloor N\xi_2\rfloor}\left[\frac{1}{\lfloor N\xi_2\rfloor-\lfloor N\xi_1\rfloor+1}\right]^2\cdot[\sigma_\theta^2\cdot(1-\rho^2)]}}$$

$$= \frac{\rho}{\sqrt{1-\rho^2}} \cdot \frac{1}{\sqrt{\lfloor N\xi_2 \rfloor - \lfloor N\xi_1 \rfloor + 1}} \cdot \sum_{i=\lfloor N\xi_1 \rfloor}^{\lfloor N\xi_2 \rfloor} \mathbb{E}(Y_{i:N}) \approx \frac{\rho}{\sqrt{1-\rho^2}} \cdot \frac{\sqrt{N}}{\sqrt{\xi_2 - \xi_1}} \cdot \sum_{i=\lfloor N\xi_1 \rfloor}^{\lfloor N\xi_2 \rfloor} \frac{\mathbb{E}(Y_{i:N})}{N}$$
$$\to \frac{\rho \cdot \sqrt{N}}{\sqrt{1-\rho^2}} \cdot \frac{\varphi\big(\Phi^{-1}(\xi_1)\big) - \varphi\big(\Phi^{-1}(\xi_2)\big)}{\sqrt{\xi_2 - \xi_1}},$$

as $N \to +\infty$, where the limit holds due to Part (iii) of Lemma EC.6. Thus, (EC.28) holds. Meanwhile, since $X_i \sim \mathcal{N}(\mu_X, \sigma_X^2)$, the average impact factor satisfies

$$\bar{X} = \mathbb{E}\left[\frac{\sum_{i=\lfloor N\xi_1 \rfloor}^{\lfloor N\xi_2 \rfloor} \frac{1}{\lfloor N\xi_2 \rfloor - \lfloor N\xi_1 \rfloor + 1} X_{i:N}}{\sum_{i=\lfloor N\xi_1 \rfloor}^{\lfloor N\xi_2 \rfloor} \frac{1}{\lfloor N\xi_2 \rfloor - \lfloor N\xi_1 \rfloor + 1}}\right] = \frac{\sum_{i=\lfloor N\xi_1 \rfloor}^{\lfloor N\xi_2 \rfloor} \mathbb{E}(X_{i:N})}{\lfloor N\xi_2 \rfloor - \lfloor N\xi_1 \rfloor + 1} = \mu_X + \sigma_X \cdot \frac{\sum_{i=\lfloor N\xi_1 \rfloor}^{\lfloor N\xi_2 \rfloor} \mathbb{E}(Y_{i:N})}{\lfloor N\xi_2 \rfloor - \lfloor N\xi_1 \rfloor + 1}$$
$$\approx \mu_X + \sigma_X \cdot \frac{\left[\sum_{i=\lfloor N\xi_1 \rfloor}^{\lfloor N\xi_2 \rfloor} \mathbb{E}(Y_{i:N})\right]/N}{\xi_2 - \xi_1} \to \mu_X + \sigma_X \cdot \frac{\varphi\big(\Phi^{-1}(\xi_1)\big) - \varphi\big(\Phi^{-1}(\xi_2)\big)}{\xi_2 - \xi_1},$$

as $N \to +\infty$, where the limit holds due to Part (iii) of Lemma EC.6. Thus, (EC.29) holds. □

EC.3.8. Proofs for Appendix EC.1.3

Proof of Proposition EC.9. Let C be a copula of F. Define $\tilde{C}(u,v) =$

$$\begin{cases} \frac{F_X(d)-u}{F_X(d)-F_X(d^-)} C(F_X(d^-), v) + \frac{u-F_X(d^-)}{F_X(d)-F_X(d^-)} C(F_X(d), v), & (u,v) \in (F_X(d^-), F_X(d)) \times [0,1], d \in \Delta_X, \\ C(u,v), & \text{otherwise.} \end{cases}$$

We can show by definition that $\tilde{C}(u,v)$ is also a copula of F, and it satisfies (13). This proves its existence. In addition, by Sklar's (1959) theorem, C is uniquely determined on $\overline{\mathcal{R}}_X \times \overline{\mathcal{R}}_\theta$. Hence, the copula satisfying (13) is uniquely determined on $[0,1] \times \overline{\mathcal{R}}_\theta$. □

Proof of Theorem EC.4. Let $\mathcal{P}$ be the set of all permutations of $1, 2, \ldots, N$. For any $y_1, y_2, \ldots, y_N$,

$$\begin{aligned} &\mathbb{P}\left(\theta_{[1:N]} \le y_1, \ldots, \theta_{[N:N]} \le y_N\right) \\ &= \sum_{(i_1,\ldots,i_N)\in\mathcal{P}} \mathbb{P}\left(\theta_{i_1} \le y_1, \ldots, \theta_{i_N} \le y_N, X_{i_1} \le \cdots \le X_{i_N}\right) \\ &= \sum_{(i_1,\ldots,i_N)\in\mathcal{P}} \int \cdots \int \mathbb{P}\left(\theta_{i_1} \le y_1, \ldots, \theta_{i_N} \le y_N | X_{i_1} = x_1, \ldots, X_{i_N} = x_N\right) \\ &\qquad \cdot \prod_{k=1}^N f_{X,i_k}(x_k) \cdot \mathbf{1}_{\{x_1 \le \cdots \le x_N\}} \mathrm{d}x_1 \cdots \mathrm{d}x_N \\ &= \sum_{(i_1,\ldots,i_N)\in\mathcal{P}} \int \cdots \int \prod_{k=1}^N \mathbb{P}\left(\theta_{i_k} \le y_k | X_{i_k} = x_k\right) \cdot \prod_{k=1}^N f_{X,i_k}(x_k) \cdot \mathbf{1}_{\{x_1 \le \cdots \le x_N\}} \mathrm{d}x_1 \cdots \mathrm{d}x_N, \end{aligned} \tag{EC.105}$$

where the last equation uses the independence between $(X_1, \theta_1)^\top, (X_2, \theta_2)^\top, \ldots, (X_N, \theta_N)^\top$.

In addition, because of $V_1, V_2 \ldots, V_N \overset{\text{IID}}{\sim} \text{Uniform}(0,1)$ and the independence between $V_1, V_2 \ldots, V_N$ and $\Pi, M_1, M_2, \ldots, M_N$,

$$\mathbb{P}\left(h_1(\Pi, M_1, V_1) \le y_1, \ldots, h_N(\Pi, M_N, V_N) \le y_N\right)$$

$$
\begin{aligned}
&= \sum_{\pi=(i_1,\ldots,i_N)\in\mathcal{P}} p_{i_1,\ldots,i_N} \mathbb{P}\left(h_1(\pi, M_1, V_1) \le y_1, \ldots, h_N(\pi, M_N, V_N) \le y_N | \Pi = \pi\right) \\
&= \sum_{\pi=(i_1,\ldots,i_N)\in\mathcal{P}} p_{i_1,\ldots,i_N} \cdot \\
&\qquad \int \cdots \int \mathbb{P}\left(h_1(\pi, x_1, V_1) \le y_1, \ldots, h_N(\pi, x_N, V_N) \le y_N | \Pi = \pi, X_1 = x_1, \ldots, X_N = x_N\right) \\
&\qquad\qquad \cdot m_{i_1,i_2,\ldots,i_N}(x_1, x_2, \ldots, x_N) \mathrm{d}x_1 \cdots \mathrm{d}x_N \\
&= \sum_{\pi=(i_1,\ldots,i_N)\in\mathcal{P}} \int \cdots \int \mathbb{P}\left(h_1(\pi, x_1, V_1) \le y_1, \ldots, h_N(\pi, x_N, V_N) \le y_N\right) \\
&\qquad\qquad \cdot \prod_{k=1}^N f_{X,i_k}(x_k) \cdot \mathbf{1}_{\{x_1 \le \cdots \le x_N\}} \mathrm{d}x_1 \cdots \mathrm{d}x_N \\
&= \sum_{\pi=(i_1,\ldots,i_N)\in\mathcal{P}} \int \cdots \int \prod_{k=1}^N \mathbb{P}\left(h_k(\pi, x_k, V_k) \le y_k\right) \cdot \prod_{k=1}^N f_{X,i_k}(x_k) \cdot \mathbf{1}_{\{x_1 \le \cdots \le x_N\}} \mathrm{d}x_1 \cdots \mathrm{d}x_N. \qquad \text{(EC.106)}
\end{aligned}
$$

By the definition of h_k,

$$
\begin{aligned}
\mathbb{P}\left(h_k(\pi, x_k, V_k) \le y_k\right) &= \mathbb{P}\left(F_{\theta,i_k}^{-1} \circ h^{i_k}_{F_{X,i_k}(x_k)}(V_k) \le y_k\right) = \mathbb{P}\left(h^{i_k}_{F_{X,i_k}(x_k)}(V_k) \le F_{\theta,i_k}(y_k)\right) \\
&= \mathbb{P}\left(V_k \le \frac{\partial C_{i_k}}{\partial u}\left(F_{X,i_k}(x_k), F_{\theta,i_k}(y_k)\right)\right) = \frac{\partial C_{i_k}}{\partial u}\left(F_{X,i_k}(x_k), F_{\theta,i_k}(y_k)\right),
\end{aligned}
$$

where $\frac{\partial C_{i_k}}{\partial u}\left(F_{X,i_k}(x_k), F_{\theta,i_k}(y_k)\right) = \frac{\partial C_{i_k}}{\partial u}(u, v)\big|_{u=F_{X,i_k}(x_k), v=F_{\theta,i_k}(y_k)}$. Furthermore, by Sklar's theorem, $\left(F_{X,i_k}(X_{i_k}), F_{\theta,i_k}(\theta_{i_k})\right) \sim C_{i_k}(\cdot,\cdot)$. Hence,

$$
\begin{aligned}
\mathbb{P}\left(\theta_{i_k} \le y_k | X_{i_k} = x_k\right) &= \mathbb{P}\left(F_{\theta,i_k}(\theta_{i_k}) \le F_{\theta,i_k}(y_k) \middle| F_{X,i_k}(X_{i_k}) = F_{X,i_k}(x_k)\right) \\
&= \frac{\partial C_{i_k}}{\partial u}\left(F_{X,i_k}(x_k), F_{\theta,i_k}(y_k)\right) = \mathbb{P}\left(h_k(\pi, x_k, V_k) \le y_k\right).
\end{aligned}
$$

This implies that (EC.105) and (EC.106) take the same value, which completes the proof. $\square$

Proof of Proposition EC.10. Using Theorem EC.4, we have

$$
\begin{aligned}
&\mathbb{E}(\theta_{[i:N]}) = \mathbb{E}(h_i(\Pi, M_i, V_i)) \\
&= \sum_{\pi=(j_1,j_2,\ldots,j_N)} p_{j_1,j_2,\ldots,j_N} \int_0^1 \int_{-\infty}^{+\infty} \Bigg[F_{\theta,j_i}^{-1} \circ h^{j_i}_{F_{X,j_i}(x)}(w) \cdot \frac{1}{p_{j_1,j_2,\ldots,j_N}} f_{X,j_i}(x) \\
&\qquad \cdot \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \prod_{s=1,s\ne i}^N f_{X,j_s}(x_s) \mathbf{1}_{\{x_1 \le \cdots \le x_{i-1} \le x \le x_{i+1} \le \cdots \le x_N\}} \mathrm{d}x_1 \cdots \mathrm{d}x_{i-1} \mathrm{d}x_{i+1} \mathrm{d}x_N \Bigg] \mathrm{d}x \mathrm{d}w \\
&= \sum_{k=1}^N \int_0^1 \int_{-\infty}^{+\infty} \Bigg[F_{\theta,j_i}^{-1} \circ h^{j_i}_{F_{X,j_i}(x)}(w) f_{X,j_i}(x) \cdot \sum_{\pi=(j_1,j_2,\ldots,j_N), j_i = k} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \\
&\qquad \prod_{s=1,s\ne i}^N f_{X,j_s}(x_s) \mathbf{1}_{\{x_1 \le \cdots \le x_{i-1} \le x \le x_{i+1} \le \cdots \le x_N\}} \mathrm{d}x_1 \cdots \mathrm{d}x_{i-1} \mathrm{d}x_{i+1} \mathrm{d}x_N \Bigg] \mathrm{d}x \mathrm{d}w \\
&= \sum_{k=1}^N \int_0^1 \int_{-\infty}^{+\infty} \left[F_{\theta,j_i}^{-1} \circ h^{j_i}_{F_{X,j_i}(x)}(w) f_{X,j_i}(x) \cdot H_k^i(u) \right] \mathrm{d}x \mathrm{d}w = \sum_{k=1}^N \int_0^1 \int_0^1 \tilde{h}_k(u, w) H_k^i(u) \mathrm{d}u \mathrm{d}w,
\end{aligned}
$$

which proves (EC.37). The proofs of (EC.38) and (EC.39) are similar, which we omit here. $\square$

Proof of Corollary EC.1. This is a direct corollary of Proposition EC.10. □

EC.3.9. Proofs for Appendix EC.1.4

Proof of Theorem EC.5. We only prove the case that the copula of F is stochastically increasing. Let C be a linearly interpolating copula of F constructed using Proposition EC.9. Nelsen (2007, Corollary 5.2.11) demonstrates that a copula $\tilde{C}(u,v)$ is stochastically increasing if and only if, for any $v\in[0,1]$, $\tilde{C}(u,v)$ is a concave function of u. Therefore, by Lemma EC.3, the linearly interpolating copula C is also stochastically increasing.

By Theorem 2, for $i=1,2,\ldots,N$, $\theta_{[i:N]}\overset{d}{=}g(U_{i:N},V_i)$, where g is defined as (15). According to the definition of stochastically increasing, $\mathscr{D}_1C(u,v)$ is non-increasing with u. We claim that this implies that the function $g(u,w)=F_\theta^{-1}\circ g_u(w)$ is non-decreasing with u. To prove this claim, consider $0\le u_1<u_2\le 1$, then for any fixed v, we have $\mathscr{D}_1C(u_1,v)\ge\mathscr{D}_1C(u_2,v)$. Because $g_{u_1}(w)=\inf\{v:\mathscr{D}_1C(u_1,v)\ge w\}$ and $g_{u_2}(w)=\inf\{v:\mathscr{D}_1C(u_2,v)\ge w\}$, we immediately have $g_{u_1}(w)\le g_{u_2}(w)$ for any fixed w. Further because F_θ is non-decreasing, $g(u_1,w)\le g(u_2,w)$ holds. This proves the claim.

Through the representation $\theta_{[i:N]}\overset{d}{=}g(U_{i:N},V_i)$ and Lemma EC.4, we have

$$\mathbb{E}\left(\theta_{[i:N]}^{2k+1}\right)=\int_0^1\int_0^1[g(u,w)]^{2k+1}\frac{N!u^{i-1}(1-u)^{N-i}}{(i-1)!(N-i)!}\mathrm{d}u\mathrm{d}w.$$

Let $b_i(u)=\frac{N!u^{i-1}(1-u)^{N-i}}{(i-1)!(N-i)!}$, which is the density of a beta distribution. Thus,

$$\int_0^1 b_i(u)\mathrm{d}u=1,\quad i=1,2,\ldots,N. \tag{EC.107}$$

In addition, it is easy to verify that, for $u\in(0,1)$, $b_{i+1}(u)-b_i(u)>0$ if and only if $u>\frac{i}{N}$, and $b_{i+1}(u)-b_i(u)<0$ if and only if $u<\frac{i}{N}$. Therefore,

$$\begin{aligned}
&\mathbb{E}\left(\theta_{[i+1:N]}^{2k+1}\right)-\mathbb{E}\left(\theta_{[i:N]}^{2k+1}\right)=\int_0^1\int_0^1[g(u,w)]^{2k+1}[b_{i+1}(u)-b_i(u)]\mathrm{d}u\mathrm{d}w\\
&=\int_0^1\left[\int_{\frac{i}{N}}^1[g(u,w)]^{2k+1}[b_{i+1}(u)-b_i(u)]\mathrm{d}u-\int_0^{\frac{i}{N}}[g(u,w)]^{2k+1}[b_i(u)-b_{i+1}(u)]\mathrm{d}u\right]\mathrm{d}w\\
&\ge\int_0^1\left[\int_{\frac{i}{N}}^1\left[g\left(\frac{i}{N},w\right)\right]^{2k+1}[b_{i+1}(u)-b_i(u)]\mathrm{d}u-\int_0^{\frac{i}{N}}\left[g\left(\frac{i}{N},w\right)\right]^{2k+1}[b_i(u)-b_{i+1}(u)]\mathrm{d}u\right]\mathrm{d}w\\
&=\int_0^1\left[\left[g\left(\frac{i}{N},w\right)\right]^{2k+1}\int_0^1[b_{i+1}(u)-b_i(u)]\mathrm{d}u\right]\mathrm{d}w=0,
\end{aligned}$$

where the inequality uses the fact that $g(u,w)$ is non-decreasing with u, and the last equality holds because of (EC.107). This completes the proof. □

Proof of Proposition EC.11. We use Theorem 2 to prove the result, and all notations are the same as in Theorem 2. If C is a comonotonicity copula, we have $\mathscr{D}_1 C(u,v) = \begin{cases} 1, & v \geq u, \\ 0, & v < u, \end{cases}$ and therefore $g_u(w) = u$ for $w \in (0,1]$. This implies that

$$(g(U_{1:N}, V_1), \ldots, g(U_{N:N}, V_N)) = \left(F_\theta^{-1}(U_{1:N}), \ldots, F_\theta^{-1}(U_{N:N})\right) \overset{d}{=} (\theta_{1:N}, \ldots, \theta_{N:N}).$$

If C is a countermonotonicity copula, we have $\mathscr{D}_1 C(u,v) = \begin{cases} 1, & v \geq 1-u, \\ 0, & v < 1-u, \end{cases}$ and therefore $g_u(w) = 1-u$ for $w \in (0,1]$. This implies that

$$(g(U_{1:N}, V_1), \ldots, g(U_{N:N}, V_N)) = \left(F_\theta^{-1}(1-U_{1:N}), \ldots, F_\theta^{-1}(1-U_{N:N})\right) \overset{d}{=} (\theta_{N:N}, \ldots, \theta_{1:N}).$$

If C is an independence copula, we have $\mathscr{D}_1 C(u,v) = v$ and therefore $g_u(w) = w$ for $w \in [0,1]$. This implies that

$$(g(U_{1:N}, V_1), \ldots, g(U_{N:N}, V_N)) = \left(F_\theta^{-1}(V_1), \ldots, F_\theta^{-1}(V_N)\right) \overset{d}{=} (\theta_1, \ldots, \theta_N).$$

This completes the proof. □

Proof of Proposition EC.12. We use Theorem 1 to prove the result, and all notations are the same as in Theorem 1. Let $F_\theta \equiv F_\theta^{(1)} = F_\theta^{(2)}$. By definition, one can verify that $\frac{\partial}{\partial u} C^{\mathrm{El}}_{\mathbf{\Psi}_\rho}(u,v) = \frac{\partial}{\partial u} C^{\mathrm{El}}_{\mathbf{\Psi}_{-\rho}}(1-u,v)$. Let $g_u^\rho(w)$ and $g_u^{-\rho}(w)$ be the inverse functions of $v \mapsto \frac{\partial}{\partial u} C^{\mathrm{El}}_{\mathbf{\Psi}_\rho}(u,v)$ and $v \mapsto \frac{\partial}{\partial u} C^{\mathrm{El}}_{\mathbf{\Psi}_{-\rho}}(u,v)$, respectively. They therefore satisfy $g_u^\rho(w) = g_{1-u}^{-\rho}(w)$ for any $u, w \in [0,1]$. Then, let $g^\rho(u,w) = F_\theta \circ g_u^\rho(w)$ and $g^{-\rho}(u,w) = F_\theta \circ g_u^{-\rho}(w)$, we have $g^\rho(u,w) = g^\rho(1-u,w)$. This implies that

$$\begin{aligned}(g^\rho(U_{1:N}, V_1), \ldots, g^\rho(U_{N:N}, V_N)) &= (g^{-\rho}(1-U_{1:N}, V_1), \ldots, g^{-\rho}(1-U_{N:N}, V_N)) \\ &\overset{d}{=} (g^{-\rho}(U_{N:N}, V_N), \ldots, g^{-\rho}(U_{1:N}, V_1)),\end{aligned}$$

which proves (EC.52). Furthermore, by (9), the optimal weights satisfy (EC.53). This completes the proof. □

EC.3.10. Proofs for Appendix EC.1.5

Proof of Proposition EC.13. We use Theorem 1 to prove the result, and all notations are the same as in Theorem 1. By Theorem 1, we have

$$\begin{aligned}&\mathbb{P}\left(\theta_{[1:N]} - \mu_\theta \leq x_1, \ldots, \theta_{[N:N]} - \mu_\theta \leq x_N\right) \\ =&\mathbb{P}\left(g_{U_{1:N}}(V_1) \leq F_\theta(\mu_\theta + x_1), \ldots, g_{U_{N:N}}(V_N) \leq F_\theta(\mu_\theta + x_N)\right) \\ =&\int_0^1 \cdots \int_0^1 \mathbb{P}\left(g_{u_1}(V_1) \leq F_\theta(\mu_\theta + x_1), \ldots, g_{u_N}(V_N) \leq F_\theta(\mu_\theta + x_N)\right) N! \mathbf{1}_{\{u_1 < \cdots < u_N\}} \mathrm{d}u_1 \cdots \mathrm{d}u_N \\ =&\int_0^1 \cdots \int_0^1 \prod_{i=1}^N \left[\frac{\partial C}{\partial u}(u_i, F_\theta(\mu_\theta + x_i))\right] N! \mathbf{1}_{\{u_1 < \cdots < u_N\}} \mathrm{d}u_1 \cdots \mathrm{d}u_N,\end{aligned} \tag{EC.108}$$

and similarly, with the help of the continuity of F_θ and $v \mapsto \frac{\partial C}{\partial u}(u,v)$,

$$\begin{aligned}
&\mathbb{P}\left(\mu_\theta - \theta_{[N:N]} \le x_1, \ldots, \mu_\theta - \theta_{[1:N]} \le x_N\right) \\
&= \int_0^1 \cdots \int_0^1 \prod_{i=1}^N \left[1 - \frac{\partial C}{\partial u}(u_i, F_\theta(\mu_\theta - x_i))\right] N! \mathbf{1}_{\{u_1 > \cdots > u_N\}} \mathrm{d}u_1 \cdots \mathrm{d}u_N \\
&= \int_0^1 \cdots \int_0^1 \prod_{i=1}^N \left[1 - \frac{\partial C}{\partial u}(1 - u_i, F_\theta(\mu_\theta - x_i))\right] N! \mathbf{1}_{\{u_1 < \cdots < u_N\}} \mathrm{d}u_1 \cdots \mathrm{d}u_N, \qquad \text{(EC.109)}
\end{aligned}$$

where the last equality replaces u_i with $1 - u_i$.

Furthermore, Nelsen (2007, Theorem 2.7.3) demonstrates that, if C is the joint distribution of (U,V), we have $(U,V) \overset{d}{=} (1-U, 1-V)$ and, therefore, $\mathbb{P}(V \le v | U = u) = \mathbb{P}(1 - V \le v | 1 - U = u) = \mathbb{P}(V \ge 1 - v | U = 1 - u)$. This implies that $\frac{\partial C}{\partial u}(u,v) = 1 - \frac{\partial C}{\partial u}(1-u, 1-v)$. Combining this with (EC.108), (EC.109), and (EC.54) completes the proof. □

Endnotes

[23] See also Grinold and Kahn (1999, Chapter 6) and Grinold and Kahn (2019, Chapters 4 and 5) for recent developments.

[24] The true optimal weights are computed by using a sufficiently large number of subintervals partitioned from the original interval for the numerical integration. In our numerical experiments, we set the number of subintervals to 1,000 to compute the true optimal weights.

[25] The derivative of $\phi_\gamma^{\mathrm{Gu}}$ is $(\phi_\gamma^{\mathrm{Gu}})'(x) = -\exp(-x^{1/\gamma}) x^{1/\gamma - 1}/\gamma$, whose inverse function $(\phi_\gamma^{\mathrm{Gu}})'^{-1}$ cannot be written explicitly. However, one can still calculate it numerically in practice.

[26] See `https://www.marketplace.spglobal.com/en/datasets/trucost-environmental-(46)`.

[27] Scope 1 emissions cover greenhouse gas emissions from operations that are owned or controlled by the company. Scope 2 emissions cover emissions from the consumption of purchased electricity, heat, or steam by the company. Scope 3 emissions cover other indirect emissions not covered in Scope 2, such as from the extraction and production of purchased materials and fuels, transport-related activities in vehicles not owned or controlled by the reporting entity, electricity-related activities, outsourced activities, waste disposal, etc. See `https://ghgprotocol.org/corporate-standard`.

[28] We treat both null and zero values in the data as invalid values.

[29] We obtain the CRSP data from the Wharton Research Data Service.

[30]See https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

[31]We match the Trucost Environmental data with the CRSP data by ISIN for all stocks issued in the US. The ISINs of stocks issued in the US begin with US. We only use stocks with valid current-month residual returns and last year's impact factor.

[32]In our empirical study, we use the Deheuvels' empirical copula. See, for example, Cherubini et al. (2004, Section 5.5.1).

[33]For simplicity, we assume that $\sigma_\theta = 0$.

[34]The carbon emission data start in 2005, which we use to correlate with the residual returns starting in 2006. By the end of 2010, we have five years of data to estimate ρ.

[35]We define the annual turnover as

$$\text{turnover} = \frac{1}{T}\sum_{t=1}^{T}\left(\sum_{i=1}^{N}\left|w_{i,t+1} - \frac{w_{i,t}(1+r_{i,t+1})}{1+\sum_{j=1}^{N} w_{j,t} r_{j,t+1}}\right|\right),$$

where $w_{i,t}$ and $r_{i,t}$ are the weight and return of stock i in the portfolio in year t, respectively. The portfolio alpha, α, is the intercept term from the Fama–French five-factor regression, and the volatility of residual returns, $\sigma(\theta_p)$, is the standard deviation of the regression's residual returns. The information ratio is defined as the ratio of α to $\sigma(\theta_p)$.

[36]Although our methodology is based on residual returns in excess of asset pricing factors, the portfolios have nonzero exposures to these factors and therefore gain factor risk premiums. Therefore we report metrics related to both raw and residual returns.

References

Ardia D, Bluteau K, Boudt K, Inghelbrecht K (2023) Climate change concerns and the performance of green vs. brown stocks. *Management Science* 69(12):7607–7632.

Aswani J, Raghunandan A, Rajgopal S (2024) Are carbon emissions associated with stock returns? *Review of Finance* 28(1):75–106.

Azzalini A, Capitanio A (1999) Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 61(3):579–602.

Bapat R, Beg M (1989) Order statistics for nonidentically distributed variables and permanents. *Sankhyā: The Indian Journal of Statistics, Series A* 79–93.

Berg F, Kölbel JF, Pavlova A, Rigobon R (2021) ESG confusion and stock returns: Tackling the problem of noise, available at SSRN 3941514.

Bickel PJ (1967) Some contributions to the theory of order statistics. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, 575–591 (University of California Press).

Bolton P, Halem Z, Kacperczyk M (2022) The financial cost of carbon. *Journal of Applied Corporate Finance* 34(2):17–29.

Bolton P, Kacperczyk M (2021) Do investors care about carbon risk? *Journal of Financial Economics* 142(2):517–549.

Bolton P, Kacperczyk M (2023) Global pricing of carbon-transition risk. *The Journal of Finance* 78(6):3677–3754.

Cheema-Fox A, LaPerla BR, Serafeim G, Turkington D, Wang HS (2021) Decarbonization factors. *The Journal of Impact and ESG Investing* 2(1):47–73.

Cherubini U, Luciano E, Vecchiato W (2004) *Copula Methods in Finance* (John Wiley & Sons).

Cont R (2001) Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance* 1(2):223–236.

David HA, Nagaraja HN (2004) *Order Statistics* (John Wiley & Sons).

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

Ding Z, Martin RD (2017) The fundamental law of active management: Redux. *Journal of Empirical Finance* 43:91–114.

Durrett R (2019) *Probability: Theory and Examples* (Cambridge University Press).

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

Fang J, Jiang F, Liu Y, Yang J (2020) Copula-based Markov process. *Insurance: Mathematics and Economics* 91:166–187.

Görgen M, Jacob A, Nerlinger M, Riordan R, Rohleder M, Wilkens M (2020) Carbon risk, available at SSRN 2930897.

Grinold RC (1989) The fundamental law of active management. *The Journal of Portfolio Management* 15(3):30–37.

Grinold RC (1994) Alpha is volatility times IC times score. *The Journal of Portfolio Management* 20(4):9–16.

Grinold RC, Kahn RN (1999) *Active Portfolio Management: A Quantitative Approach for Producing Superior Returns and Controlling Risk* (New York: McGraw-Hill Education), 2nd edition.

Grinold RC, Kahn RN (2019) *Advances in Active Portfolio Management: New Developments in Quantitative Investing* (McGraw Hill Professional).

Henze N, Zirkler B (1990) A class of invariant consistent tests for multivariate normality. *Communications in Statistics—Theory and Methods* 19(10):3595–3617.

Horn RA, Johnson CR (2012) *Matrix Analysis* (Cambridge University Press).

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

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

Kan R, Zhou G (2007) Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis* 42(3):621–656.

Lee HM, Viana M (1999) The joint covariance structure of ordered symmetrically dependent observations and their concomitants of order statistics. *Statistics & Probability Letters* 43(4):411–414.

Lindsey LA, Pruitt S, Schiller C (2021) The cost of ESG investing, available at SSRN 3975077.

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

Lo AW, Zhang R (2023) Quantifying the impact of impact investing. *Management Science* forthcoming.

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

Nelsen RB (2007) *An Introduction to Copulas* (Springer Science & Business Media).

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

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

Qian E, Hua R (2006) Active risk and information ratio. *The World of Risk Management*, 151–167 (World Scientific).

Sklar M (1959) Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut Statistique de l'Université de Paris* 8:229–231.

Treynor JL, Black F (1973) How to use security analysis to improve portfolio selection. *The Journal of Business* 46(1):66–86.

Wang W, Sarkar SK, Bai Z (1996) Some new results on covariances involving order statistics from dependent random variables. *Journal of Multivariate Analysis* 59(2):308–316.

Yang J, Cheng S, Zhang L (2006) Bivariate copula decomposition in terms of comonotonicity, countermonotonicity and independence. *Insurance: Mathematics and Economics* 39(2):267–284.

Zhang S (2024) Carbon returns across the globe. *The Journal of Finance* Forthcoming.