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 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 $\times$ matrix whose (,)-th entry is the beta of the -th ranked asset on the -th factor, and $\boldsymbol{\mu}$ and Σ are defined as in (7). Then the expectation and variance of the portfolio's raw return, $_p$, are:

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

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 assets with frictionless borrowing and lending at the risk-free rate, and they maximize the Sharpe ratio of raw returns,* $\mathrm{SR} = \mathbb{E}(\ _p)/\sqrt{\mathrm{Var}(\ _p)}$*, or the mean–variance utility of raw returns,* $\mathbb{E}(\ _p) - 0.5\lambda\mathrm{Var}(\ _p)$*, 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:*

$$ {}_i = \beta_i \ {}_M + \theta_i, \qquad = 1, 2, \ldots, \quad , \tag{EC.4}$$

where ${}_M$ *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,* $\mathrm{SR} = \mathbb{E}(\ {}_p)/\sqrt{\mathrm{Var}(\ {}_p)}$*, or the mean–variance utility of raw returns,* $\mathbb{E}(\ {}_p) - 0.5\lambda \mathrm{Var}(\ {}_p)$*, 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 $\mu_M \equiv \mathbb{E}(\ {}_M)$*,* $\sigma_M^2 \equiv \mathrm{Var}(\ {}_M)$*, 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 $\ + 1$ assets (including $\ $ ranked assets and the market portfolio, ${}_M$) 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 $\mu_M \equiv \mathbb{E}(\ {}_M)$, $\sigma_M^2 \equiv \mathrm{Var}(\ {}_M)$, $\boldsymbol{\beta}$ 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 ${}_M$ the weight on the market portfolio, and $\hat{\boldsymbol{w}} \equiv \begin{pmatrix} \boldsymbol{w} \\ {}_M \end{pmatrix}$ the weights on the $\ + 1$ assets. Then, under the single factor model, (EC.4), the expected value and variance of the return of the combined portfolio, $\hat{\ }_p$, are:

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

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,* $\mathrm{SR} = \mathbb{E}(\hat{\ }_p)/\sqrt{\mathrm{Var}(\hat{\ }_p)}$*, or the mean–variance utility of raw returns,* $\mathbb{E}(\hat{\ }_p) - 0.5\lambda \mathrm{Var}(\hat{\ }_p)$*, with a constant risk-aversion parameter* $\lambda > 0$*, the optimal weights of the assets and the market portfolio, and the squared optimal Sharpe ratio are given by:*

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

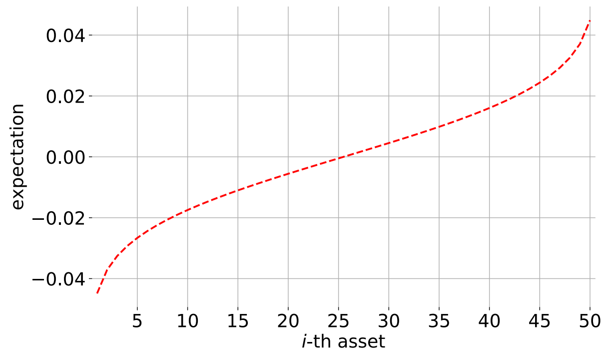
When maximizing the Sharpe ratio of the combined portfolio, the optimal weights on the assets, $\boldsymbol{w}^*$, are still proportional to $\Sigma^{-1}\boldsymbol{\mu}$. This is the same as the result of Proposition 1, which maximizes the information ratio of residual return. In other words, in the single-factor world with $+1$ 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

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 -th asset relative to other assets, $\mathbb{E}(\ {}_{i:N})$. 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 $1 - \rho^2 + \rho^2 \cdot \mathrm{Var}($


0.04
0.02
0.00
−0.02
−0.04
expectation
5
10
15
20
25
30
35
40
45
50
i-th asset

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 $|\rho| \leq \sqrt{2}/2$ 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 ( ) defined in (EC.15):

Proposition EC.6. *For ( ) defined in (EC.15) and ≥ 1, we have*

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

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

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

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, . In other words, impact investors can improve their portfolio per In th o a general case of $|\rho| \le \sqrt{2}/2$, 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 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 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 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, $_{it}$, and residual returns, θ_{it}, over time $= 1, 2, \ldots,$, are IID random variables with a mean of $\overline{\rho}$ and a variance of σ_ρ^2. Given time , $_{it}$ and θ_{it} are drawn from the following bivariate normal distribution across $= 1, 2, \ldots,$:*

$$\begin{pmatrix} {}_{1t} \\ \theta_{1t} \end{pmatrix}, \begin{pmatrix} {}_{2t} \\ \theta_{2t} \end{pmatrix}, \ldots, \begin{pmatrix} {}_{Nt} \\ \theta_{Nt} \end{pmatrix} \Bigg| \; \rho_t \overset{\text{IID}}{\sim} \mathcal{N}\Bigg(\Bigg($$

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,* $\mathrm{IR}^*_{\mathrm{T}\quad\mathrm{Va}}$, *as given in (9), will satisfy the following bounds:*

$$\sqrt{\frac{\overline{\rho}^2}{1/\;(\quad)+\sigma_\rho^2}+\frac{\mu_\theta^2}{\sigma_\theta^2}} \le \mathrm{IR}^*_{\mathrm{T}\quad\mathrm{Va}} \le \sqrt{\frac{\overline{\rho}^2}{[1-2\cdot(\overline{\rho}^2+\sigma_\rho^2)]/\;(\quad)+\sigma_\rho^2}+\frac{\mu_\theta^2}{\sigma_\theta^2\left[1-2\cdot(\overline{\rho}^2+\sigma_\rho^2)\right]}},$$

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 $i \neq j$. 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 $(X_i, \theta_i)^\top$

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":

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

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

$$\mathrm{IR}^*_{\mathrm{D}\quad \mathrm{d}\ \ \mathrm{c}} \approx \frac{|\rho-\tilde{\rho}|}{\sqrt{(1-\rho_\theta)(1-\rho_X)}} \cdot \sqrt{\quad} = |\rho_{\mathsf{adj}}| \cdot \sqrt{\quad}.$$

When $\tilde{\rho} = \rho_\theta = \rho_X = 0$, we have $\rho_{\mathsf{adj}} = \rho$, 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, $\rho = \mathrm{corr}(\ \ _i, \theta_i)$, leads to a better performance, because it measures the strength of the -th asset's signal on its residual returns. Second, a larger cross-stock correlation, $\tilde{\rho} = \mathrm{corr}(\ \ _i, \theta_j)$, 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, $\rho_X = \mathrm{corr}(\ \ _i, \ \ _j)$, and the residual returns, $\rho_\theta = \mathrm{corr}(\theta_i, \theta_j)$, 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, , is large, using Proposition EC.6. In fact, this approximation, $(\ \ ) \approx$, is biased because

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

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

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

	0.5	0.6	0.7	0.8	0.9	0.95	0.99
Smallest	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 $_i$ ranking within the $(\xi_1, \xi_2) \times 100$ percentile, where $0 \le \xi_1 < \xi_2 \le 1$. If 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:*

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

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

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

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 $_i$ 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 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:*

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

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

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

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

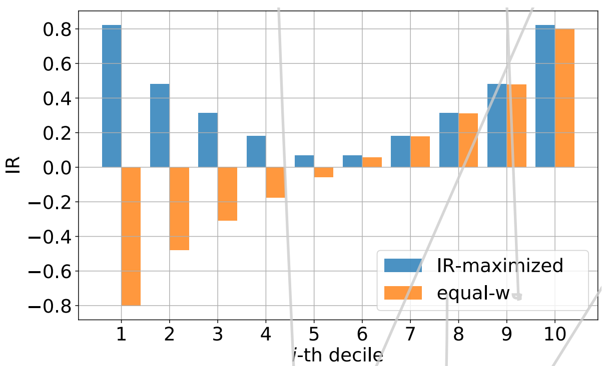


Figure EC.3c and Figure EC.3d show the results when $\rho = -20\%$. Unlike the case for a positive ρ, if an impact investor invests in top decile assets when $\rho = -20\%$ and maximizes the IR, she can earn a positive IR but a negative impact, while if she constructs an equal-weighted portfolio, she 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 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 when $\rho = 20\%$, $\mu_\theta = 0\%$, $\mu_X = 0\%$, and

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 ( ), the inverse function of is defined as $^{-1}(\ ) = \inf\{\ :\ (\ ) \geq\ \}$.

When C is non-differentiable, the function $_u(\cdot)$ in (12), which is the inverse of $\mapsto \frac{\partial C}{\partial u}(\ ,\ )$, 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 $C(\ ,\ )$ with respect to , denoted by $\mathscr{D}_1 C(\ ,\ )$, is a bivariate function $[0,1] \times \mathbb{R} \to [0,1]$ defined as

$$\mathscr{D}_1 C($$

Therefore, we can define $\mathbb{P}(\ \leq \ | \ = \ ) = \mathscr{D}_1 C(\ , \ )$. The subscript of $\mathscr{D}_1$ represents that the MPDD is calculated with respect to the first dimension, . Fang et al. (2020, Theorem 2.1) also demonstrate that, when C has a density, we have $\mathscr{D}_1 C(\ , \ ) = \frac{\partial C}{\partial u}(\ , \ )$, 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 $C(\ , \ )$ is linearly interpolating on $\overline{\mathcal{R}}_X^c \times \overline{\mathcal{R}}_\theta$ with respect to , 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 $\ \in \overline{\mathcal{R}}_X^c$. 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, , i.e., $\ _i \equiv \ $, 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:

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

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 $\overline{\mathcal{R}}_X^c \times \overline{\mathcal{R}}$

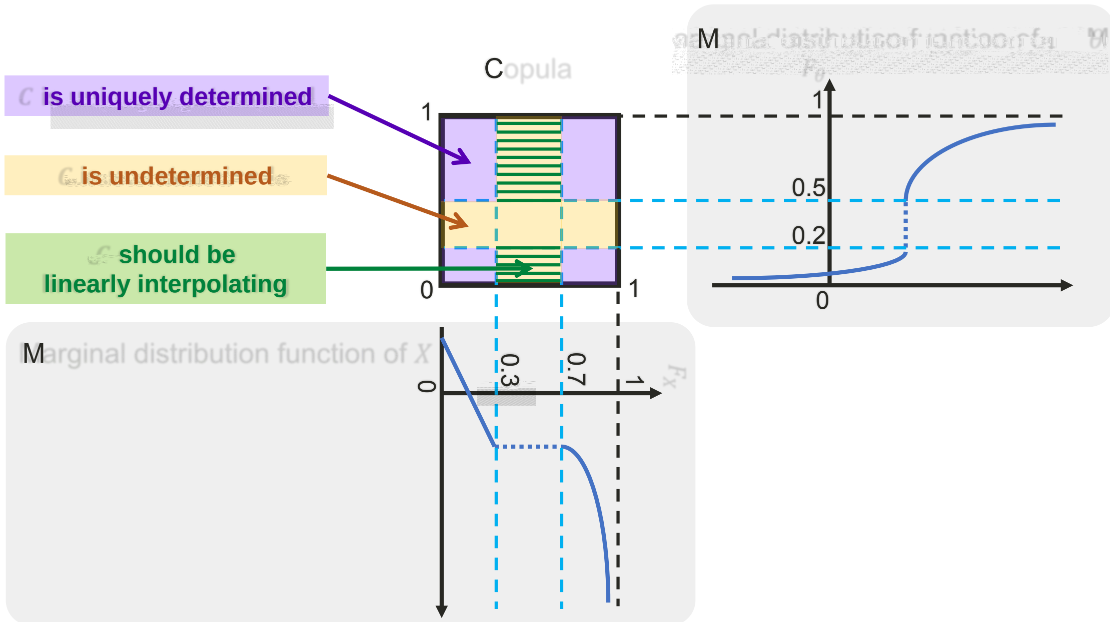


Figure EC.5 Illustration of a linear interpolating copula.

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, $_i$, 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* $(\ _1, \theta_1)^\top, (\ _2, \theta_2)^\top, \dots, (\ _N, \theta_N)^\top$ *are mutually independent and satisfy*

$$\begin{pmatrix} \ _1 \\ \theta_1 \end{pmatrix} \sim F_1(\cdot,\cdot), \quad \begin{pmatrix} \ _2 \\ \theta_2 \end{pmatrix} \sim F_2(\cdot,\cdot), \quad \dots, \quad \begin{pmatrix} \ _N \\ \theta_N \end{pmatrix} \sim F_N(\cdot,\cdot), \tag{EC.33}$$

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 $\ = 1, 2, \dots,\ $, the marginal distributions of $\ _i$ and θ_i are $F_{X,i}(\cdot)$ and $F_{\theta,i}(\cdot)$, respectively;
- For $\ = 1, 2, \dots,\ $, the marginal densities of $\ _i$ and θ_i are $\ _{X,i}(\cdot)$ and $\ _{\theta,i}(\cdot)$, respectively;
- For $\ = 1, 2, \dots,\ $, the copula of $F_i(\cdot,\cdot)$ is $C_i(\cdot,\cdot)$;
- For any permutation $(\ _1,\ _2, \dots,\ _N)$ of $1, 2, \dots,\ $, let $\ _{i_1,i_2,\dots,i_N} = \mathbb{P}(\ _{i_1} \le \ _{i_2} \le \cdots \le \ _{i_N})$.

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

$$(\theta_{[1:N]}, \theta_{[2:N]}, \ldots, \theta_{[N:N]}) \stackrel{d}{=} (\ _1(\Pi, \ \ _1, \ _1), \ _2(\Pi, \ \ _2, \ _2), \ldots, \ _N(\Pi, \ \ _N, \ _N)), \tag{EC.34}$$

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

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

$$\mathbb{P}(\Pi = (\ _1, \ _2, \ldots, \ _N)) = \ _{i_1, i_2, \ldots, i_N}$$

for any permutation $(\ _1, \ _2, \ldots, \ _N)$ *of* $\{1, 2, \ldots, \ \}$;

- *Given* Π, *the random vector* $(\ \ _1, \ \ _2, \ldots, \ \ _N)|\Pi = (\ _1, \ _2, \ldots, \ _N)$ *has a joint density of*

$$\ _{i_1, i_2, \ldots, i_N}(\ \ _1, \ \ _2, \ldots, \ \ _N) = \frac{1}{\ _{i_1, i_2, \ldots, i_N}} \cdot \prod_{k=1}^{N} \ _{X, i_k}(\ _k) \cdot \mathbf{1}_{\{\ _1 \leq \ _2 \leq \cdots \leq \ _N\}}; \tag{EC.35}$$

- *The random variables* $\ _1, \ _2 \ldots, \ _N$ *satisfy* $\ _1, \ _2 \ldots, \ _N \stackrel{\text{IID}}{\sim} \text{Uniform}(0, 1)$ *and are independent of* $\Pi, \ \ _1, \ \ _2, \ldots, \ \ _N$;
- *For* $\ = 1, 2, \ldots, \ $, *function* $\ _k$ *is defined as*

$$\ _k(\pi, \ , \ ) \equiv F^{-1}_{\theta, i_k} \circ \ ^{i_k}_{F_{X, i_k}(\ )}(\ ) \tag{EC.36}$$

for any permutation $\pi = (\ _1, \ _2, \ldots, \ _N)$ *of* $\{1, 2, \ldots, \ \}$, *and "$\circ$" represents function composition;*

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

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 ($\ _i$), and uniform noise ($\ _i$). In comparison with Theorem 1, an additional random permutation component, Π, is required here. We refer to random variables $\ _i$ as "order statistics" because, by definition, they always satisfy $\ _1 \leq \ _2 \leq \cdots \leq \ _N$. In addition, if $\ _1, \ _2, \ldots, \ _N$ have the same marginal densities, i.e., $\ _{X,1} = \ _{X,2} = \cdots = \ _{X,N}$, one can easily verify that (EC.35) reduces to the density of order statistics of $\ _1, \ _2, \ldots, \ _N$. Therefore, although $\ _i$ 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), $\ _k$, are different for $\ = 1, 2, \ldots, \ $. In addition, all three components of F_i—$F_{X,i}$, $F_{\theta,i}$, and C_i—contribute to the mixture function, $\ _k$. This is different from Theorem 1 in which the mixture function 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 $_i$ is also crucial for the mixture.

Third, the representation in Theorem EC.4 can be simplified when $F_1(\cdot,\cdot),\ldots,F_N(\cdot,\cdot)$ 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

$$_k(\pi,\ ,\ ) = F_\theta^{-1} \circ \ ^{i_k}_{F_{X,i_k}(\ )}(\ );$$

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

$$(\theta_{[1:N]},\theta_{[2:N]},\ldots,\theta_{[N:N]}) \overset{d}{=} (\ _1(\Pi,\ _{1:N},\ _1),\ _2(\Pi,\ _{2:N},\ _2),\ldots,\ _N(\Pi,\ _{N:N},\ _N)),$$

where $_{1:N} \le\ _{2:N} \le \cdots \le\ _{N:N}$ and $_1,\ _2,\ldots,\ _N$ are defined as in Theorem 1, and the mixture function (EC.36) reduces to

$$_k(\pi,\ ,\ ) = F_{\theta,i_k}^{-1} \circ \ ^{i_k}(\ );$$

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

$$_k(\pi,\ ,\ ) = F_{\theta,i_k}^{-1} \circ \ _{F_{X,i_k}(\ )}(\ ),$$

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

(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:*

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

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

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

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

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

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

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

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

Functions $\tilde{\ }_k$, H_k^i, and $J_{k,l}^{i,j}$ are entirely determined by the distributions of $\boldsymbol{X}$ and

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

	$\mu_{X,1}$	$\mu_{X,2}$	$\mu_{X,3}$	$\mu_{X,4}$	$\mu_{X,5}$	$\sigma_{\theta,1}$	$\sigma_{\theta,2}$	$\sigma_{\theta,3}$	$\sigma_{\theta,4}$	$\sigma_{\theta,5}$
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 $\sum_{i=1}^{N}(\hat{\ }_i - {}_i^*)^2$, where $({}_1^*, {}_2^*, \ldots, {}_N^*)^\top$ is the true optimal weights for the assets given by (9),[24] and $(\hat{\ }_1, \hat{\ }_2, \ldots, \hat{\ }_N)^\top$ 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.

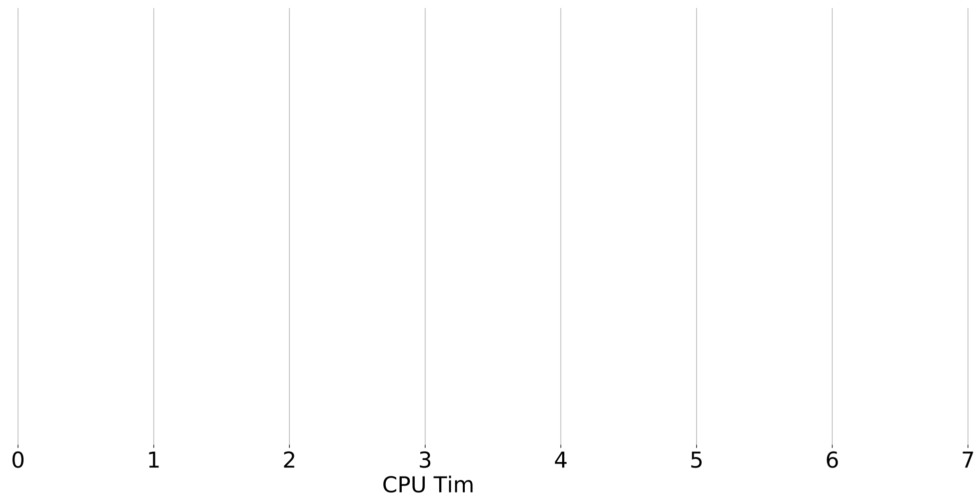
The error and CPU time of both methods depend on the chosen degree of accuracy. In general, a higher degree of accuracy implies a lower error but requires a higher CPU time. For Method (a), the degree of accuracy is determined by the number of subintervals partitioned from the original interval for numerical 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, , increases. This illustrates that the CPU time for the representation theorem increases as grows.


0
1
2
3
4
5
6
7
CPU Tim

where $_k \geq 1$ *for* $= 1, 2, \ldots, G$, $_1 + \ _2 + \cdots + \ _G =$, *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 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 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:*

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

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

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

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

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

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

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

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,* $^{(k)}_{u}(\ )$ *is the inverse function of* $\mapsto \frac{\partial C_{(k)}}{\partial u}(\ , \ )$, $C_{(k)}$ *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

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 $\;=50$ assets in a universe divided into two groups: Group 1 with $\;_1=$ 25 assets and Group 2 with $\;_2=25$ assets. We further assume that

$$\text{Group 1: } \begin{pmatrix} \;_1 \\ \theta_1 \end{pmatrix}, \begin{pmatrix} \;_2 \\ \theta_2 \end{pmatrix}, \ldots, \begin{pmatrix} \;_{25} \\ \theta_{25} \end{pmatrix} \overset{\text{IID}}{\sim} \mathcal{N}\left(\begin{pmatrix} \mu_X^- \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \rho \cdot \sigma_X \cdot \sigma_\theta^- \\ \rho \cdot \sigma_X \cdot \sigma_\theta^- & (\sigma_\theta^-)^2 \end{pmatrix} \right),$$
$$\text{Group 2: } \begin{pmatrix} \;_{26} \\ \theta_{26} \end{pmatrix}, \begin{pmatrix} \;_{27} \\ \theta_{27} \end{pmatrix}, \ldots, \begin{pmatrix} \;_{50} \\ \theta_{50} \end{pmatrix} \overset{\text{IID}}{\sim} \mathcal{N}\left(\begin{pmatrix} \mu_X^+ \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \rho \cdot \sigma_X \cdot \sigma_\theta^+ \\ \rho \cdot \sigma_X \cdot \sigma_\theta^+ & (\sigma_\theta^+)^2 \end{pmatrix} \right),$$

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

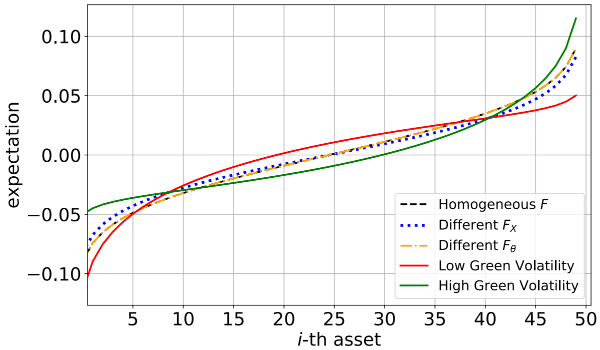
1. **Homogeneous** F: $\mu_X^-=\mu_X^+=0\%$, $\sigma_\theta^-=\sigma_\theta^+=20\%$;
2. **Different** F_X: $\mu_X^-=-10\%$, $\mu_X^+=10\%$, $\sigma_\theta^-=\sigma_\theta^+=20\%$;
3. **Different** F_θ: $\mu_X^-=\mu_X^+=0\%$, $\sigma_\theta^-=30\%$, $\sigma_\theta^+=10\%$;
4. **Low Green Volatility**: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

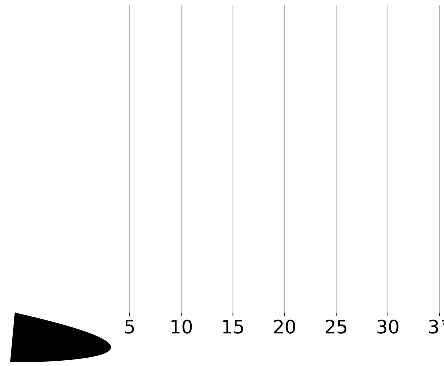
Figure EC.7 shows the expected value, variance, and optimal weights for the two groups of stocks under these setups using Corollary EC.1. First, because the impact factor is positively correlated with residual returns ($\rho>0$), the optimal weight increases as $\;$ increases. Second, the expected value,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

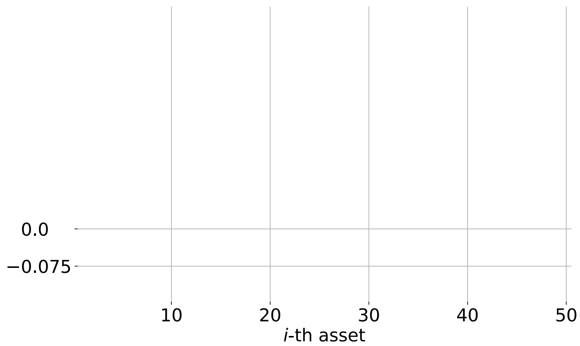


2. **Different** F_X: $\mu_X^{(j)} = -90\% + (\quad - 1) \times 20\%$, $\sigma_\theta^{(j)} = 20\%$;
3. **Different** F_θ: $\mu_X^{(j)} = 0\%$, $\sigma_\theta^{(j)} = 29\% - (\quad - 1) \times 2\%$;
4. **Low Green Volatility**: $\mu_X^{(j)} = -90\% + (\quad - 1) \times 20\%$, $\sigma_\theta^{(j)} = 29\%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

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 $\quad_i$ 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.


0.0
−0.075
10
20
30
40
50
i-th asset

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 $C(\ ,\ )$ is *stochastically increasing* in if $\mathscr{D}_1 C(\ ,\ )$ is a non-increasing function of $\in [0,1]$ for all $\in [0,1]$, and is *stochastically decreasing* in if $\mathscr{D}_1 C(\ ,\ )$ is a non-decreasing function of $\in [0,1]$ for all $\in [0,1]$. Here, $\mathscr{D}_1 C(\ ,\ )$ is defined by (EC.30).

Remark EC.2. Let $(\ ,\ )$ be a bivariate random vector with joint distribution function C. Because $\mathscr{D}_1 C(\ ,\ )$ is a regular conditional distribution function of $C(\ ,\ )$ given (Remark EC.1), we can define $\mathbb{P}(\ \le\ \mid\ =\ ) = \mathscr{D}_1 C(\ ,\ )$. Therefore, a stochastically increasing C implies that, for a given , $\mathbb{P}(\ \le\ \mid\ =\ )$ is a non-increasing function of . In other words, $\mathbb{P}(\ >\ \mid\ =\ )$ is a non-decreasing function of . This implies that a larger value of corresponds to a higher probability that also takes a larger value. Similarly, if C is stochastically decreasing, a larger value of corresponds to a higher probability that 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 $= 0, 1, \ldots,$ if $C(\ ,\ )$ is stochastically increasing in , we have:*

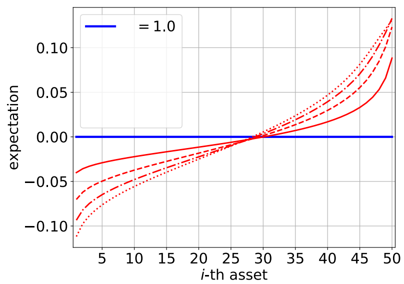
$$\mathbb{E}\left(\theta_{[1:N]}^{2k+1}\right) \le \mathbb{E}\left(\theta_{[2:N]}^{2k+1}\right) \le \cdots \le \mathbb{E}\left(\theta_{[N:N]}^{2k+1}\right),$$

and if $C(\ ,\ )$ is stochastically decreasing in , 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

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).



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

$$C^{\mathrm{E}}_{\boldsymbol{\Psi}_\rho}(\ ,\ ) \equiv \boldsymbol{\Psi}_\rho\left(\Psi^{-1}(\ ), \Psi^{-1}(\ )\right)$$

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

Proposition EC.12. *Consider two sets of IID bivariate random vectors,* $(\ {}^{(1)}_{i}, \theta_i^{(1)})$ *and* $(\ {}^{(2)}_{i}, \theta_i^{(2)})$, $\ = 1, 2, \ldots, \ $ *. Assume that the marginal distributions of* ${}^{(1)}_{i}$ *and* ${}^{(2)}_{i}$ *are continuous. Denote their copulas and the marginal distributions of* $\theta_i^{(j)}$ *by* $C^{(j)}$ *and* $F_\theta^{(j)}$*, respectively, for* $\ = 1, 2$*. Given* $\rho \in (-1, 1)$*, if* $F_\theta^{(1)} = F_\theta^{(2)}$*,* $C^{(1)} = C^{\mathrm{E}}_{\boldsymbol{\Psi}_\rho}$*, and* $C^{(2)} = C^{\mathrm{E}}_{\boldsymbol{\Psi}_{-\rho}}$*, 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* ${}^{(j)}_{1}, \ {}^{(j)}_{2}, \ldots, \ {}^{(j)}_{N}$*, for* $\ = 1, 2$*. Furthermore, the optimal weights of maximizing the information ratio under the two setups,* $(\ {}^{*(j)}_{1}, \ {}^{*(j)}_{2}, \ldots, \ {}^{*(j)}_{N})$, $\ = 1, 2$*, satisfy:*

$$(\ {}^{*(1)}_{1}, \ {}^{*(1)}_{2}, \ldots, \ {}^{*(1)}_{N}) \propto (\ {}^{*(2)}_{N}, \ {}^{*(2)}_{N-1}, \ldots, \ {}^{*(2)}_{1}). \tag{EC.53}$$

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 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 $\theta \overset{d}{=} \sigma_\theta \cdot [\ \ - \mathbb{E}(\ \ )]/\sqrt{\mathrm{Var}(\ \ )}$, where $\sigma_\theta > 0$, the random variable follows the skew-normal distribution with density function:

$$ _S(\ \ ) = 2\varphi(\ \ )\Phi(\alpha\ \ ), \qquad \in \mathbb{R}, $$

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

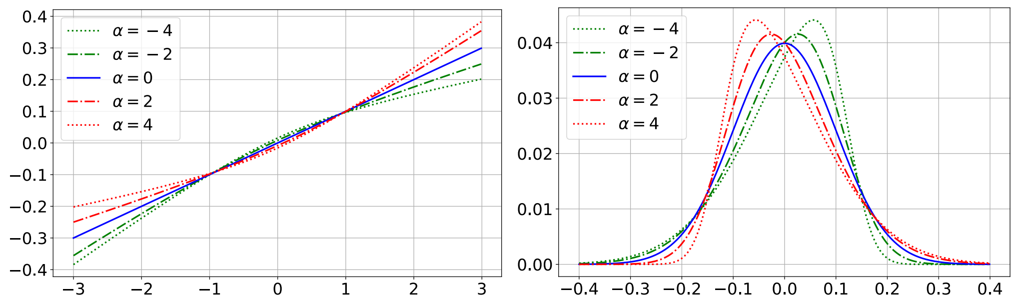
Figure EC.10a and Figure EC.10b are the Q–Q plot and the density function of the scaled-skew-normal distribution, respectively. The Q–Q plot is asymmetric on both tails for $\alpha \neq 0$. Figure EC.10b demonstrates that α parameterizes the skewness. The distribution is positively skewed when $\alpha > 0$, 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 $\ \ = 50$. 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.

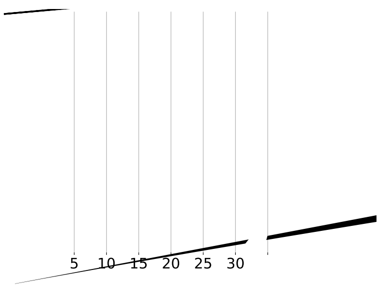
Furthermore, Figure EC.10c implies that the expectations of $\boldsymbol{\theta}_{[\boldsymbol{X}]}$ vary slightly for different α, but from Figure EC.10d, we find that different α leads to a strong dispersion in variances across assets, and the variances on the more skewed tail are larger. Therefore, 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 -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.



Remark EC.5. If the joint distribution of a random vector (,) follows a radially symmetric copula, for any and in $[0,1]$, the probabilities of (,) being in the regions $[0, \ ] \times [0, \ ]$ and $[1- \ ,1] \times [1- \ ,1]$ 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

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

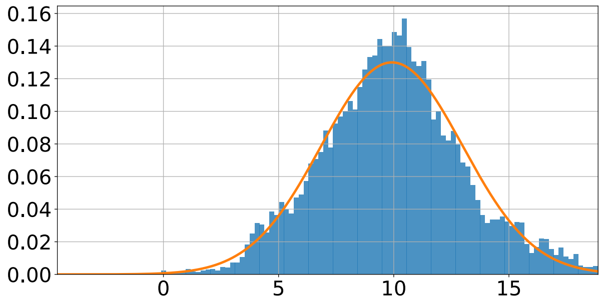
Here we demonstrate that the distribution of $(\ \ _i, \theta_i)^\top$ is highly non-normal, which motivates our framework to allow for general distributions and dependences. We use each measure of carbon emission as the impact factor, $\boldsymbol{X}$. The residual return of each stock, θ_i, is estimated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

For each sample $(\ \ _i, \theta_i)^\top$ in a particular year , $\ \ _i$ is the value of the impact factor at the end of year $\ -1$ and θ_i is the residual return in year .[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.

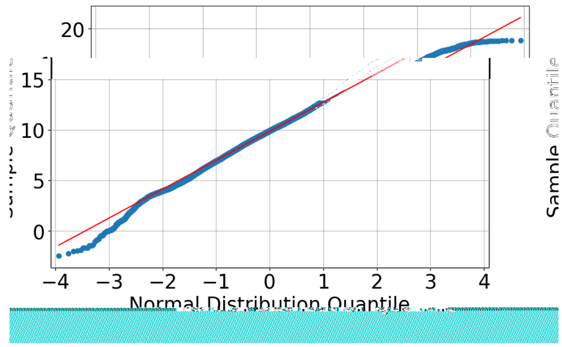
Figure EC.11 and Figure EC.12 show histograms and Q–Q plots (with respect to the standard normal distribution) for the log levels and intensities of carbon emissions across the three scopes, respectively. In both figures, each sample corresponds to a company–year. The solid lines in all figures are the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

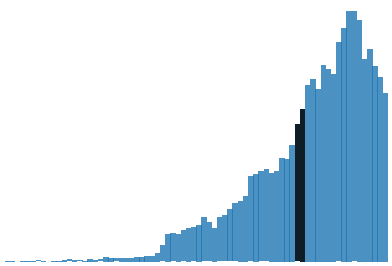
Figure EC.13 shows the histogram and Q–Q plot of monthly residual returns, with each sample

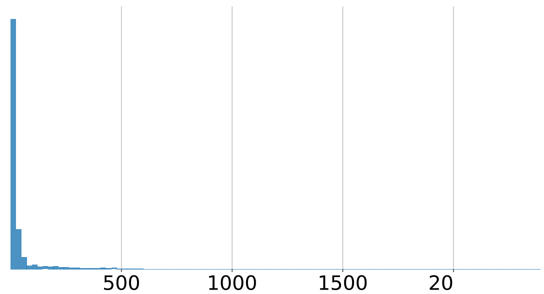


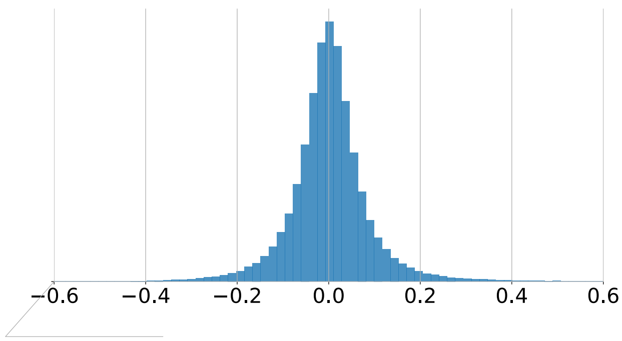
(a) Scope 1, histogram.



(b) Scope 1, Q–Q plot.




500
1000
1500
20


−0.6
−0.4
−0.2
0.0
0.2
0.4
0.6

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- -distribution with parameters $(1, \sigma_\theta)$

determining the optimal weights, for each long/short portfolio, we standardize the optimal weights by requiring that $\sum_{i=1}^{N} | \quad _i| = 1$ to ensure the same level of leverage across portfolios.

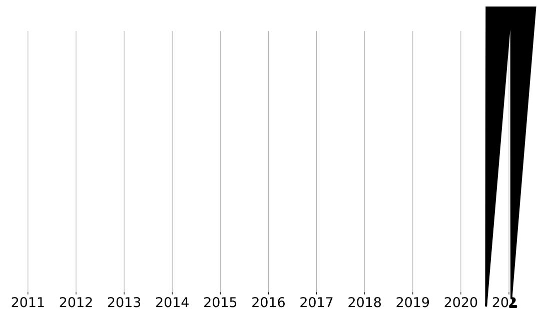
We test the profitability of all strategies from 2011 to 2021.[34] Tables EC.5 and EC.6 summarize the performance of portfolios constructed using the log levels and intensities of carbon emissions, respectively. In particular, we report their annualized raw return (return), standard deviation (std.), 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.


2011
2012
2013
2014
2015
2016
2017
2018
2019
2020

ev B r

Boltom Kacperczyk

ure 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.

) peufoum e

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 sF p hige-carbon s ctora dElivc-ud

) s W

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

Proof of Lemma EC.1. Because C is a copula, for any $_1 < \ _2$ and > 0, we have $C(\ +\ ,\ _1) - C(\ ,\ _1) \leq C(\ +\ ,\ _2) - C(\ ,\ _2)$. Taking the limit superiors of both sides proves the result for $D_1^+C(\ ,\ )$. Similar arguments apply to $D_1^-C(\ ,\ )$. □

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

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

(ii) For any $\in \overline{\mathcal{R}}_\theta$, $\{\ :\ \_^{-1}(\ ) \leq\ \} = \{\ :\ \leq\ (\ )\}$.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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, $\mathbb{E}({}_{i:N})$ satisfies $\Phi^{-1}\left(\frac{i-1}{N}\right) \le \mathbb{E}({}_{i:N}) \le \Phi^{-1}\left(\frac{i}{N}\right)$. Further note that $\Phi^{-1}\left(\frac{i-1}{N}\right) \ge 0$ when $\ >\ /2$, and $\Phi^{-1}\left(\frac{i}{N}\right) \le 0$ when $\ \le\ /2$. Hence, as $\ \to +\infty$,

$$\frac{\sum_{i=i_1(N)}^{i_2(N)} [\mathbb{E}({}_{i:N})]^2}{}$$

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

$$\frac{\sum_{i=i_1(N)}^{i_2(N)} |\mathbb{E}(\quad_{i:N})|}{\quad} \to \int_{\xi_1}^{\xi_2} |\Phi^{-1}(\quad)|\mathrm{d} \quad \overset{=\Phi^{-1}(q)}{=\!=\!=} \int_{\Phi^{-1}(\xi_1)}^{\Phi^{-1}(\xi_2)} | \quad |\varphi(\quad)\mathrm{d} \quad = \int_{\Phi^{-1}(\xi_1)}^{\Phi^{-1}(\xi_2)} \quad |$$

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

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

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:*

$$\mathbb{E}(\theta_{[i:N]}) = \sigma_\theta \cdot \frac{\rho - \tilde{\rho}}{\sqrt{1-\rho_X}} \cdot \mathbb{E}(\ _{i:N}) + \mu_\theta, \tag{EC.63}$$
$$\mathrm{Var}(\theta_{[i:N]}) = \sigma_\theta^2 \cdot \left(1 - \frac{(\rho-\tilde{\rho})^2}{1-\rho_X} + \frac{(\rho-\tilde{\rho})^2}{1-\rho_X} \cdot \mathrm{Var}(\ _{i:N})\right), \tag{EC.64}$$
$$\mathrm{Cov}(\theta_{[i:N]}, \theta_{[j:N]}) = \sigma_\theta^2 \cdot \left(\rho_\theta + \frac{(\rho-\tilde{\rho})^2}{1-\rho_X} \cdot \mathrm{Cov}(\ _{i:N},\ _{j:N})\right), \tag{EC.65}$$

for $\ ,\ = 1, 2, \ldots,$ *and* $\ \neq\ .$ *Here,* $\ _{1:N} \le\ _{2:N} \le \cdots \le\ _{N:N}$ *are the order statistics of* $\ _1,\ _2, \ldots,\ _N \overset{\text{IID}}{\sim} \mathcal{N}(0,1)$.

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

EC.3.

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

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

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

For any $\in A$, Fang et al. (2020, Theorem 2.1) demonstrate that, there exists a set B_v with Lebesgue measure 0 such that $\mathscr{D}_1 C(\ ,\ ) = \frac{\partial C}{\partial u}(\ ,\ )$ for $\in [0,1]\backslash B_v$. 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 $\mathscr{D}_1 C(\ ,\ ) = \frac{\partial C}{\partial u}(\ ,\ )$ for any $\in [0,1]\backslash E_1$ and $\in A$. Hence, (EC.66) holds for any $\in \Delta_X$, $\in (F_X(\ ^-), F_X(\ ))\backslash E_1$ and $\in \overline{\mathcal{R}}_\theta$.

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

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

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

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

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

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

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

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

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

Third, we define

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

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

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

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

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

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

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

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

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

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

$$\begin{aligned}&\mathbb{P}(F_X^{-1}(\ _i) \le \ , \tilde{\ }(F_X(F_X^{-1}(\ _i)),\ _i) \le \ ) = \int_0^{F_X(\ )} \mathbb{P}\left(\ _i \le \tilde{\mathscr{D}}_1 C(F_X(F_X^{-1}(\ )), F_\theta(\ ))\right) \mathrm{d}\ \\ &= \int_0^{F_X(\ )} \mathscr{D}_1 C(\ , F_\theta(\ )) \mathrm{d}\ = C(F_X(\ ), F_\theta(\ )) = \mathbb{P}(\ _i \le \ , \theta_i \le\end{aligned}$$

In addition,

$$\begin{aligned}
&\mathbb{P}\left(\tilde{}(F_X(H_{1:N}),\ _1) \le\ _1, \ldots, \tilde{}(F_X(H_{N:N}),\ _N) \le\ _N, H_1 \le \cdots \le H_N\right) \\
=&\mathbb{P}\left(\tilde{}(F_X(H_1),\ _1) \le\ _1, \ldots, \tilde{}(F_X(H_N),\ _N) \le\ _N, H_1 \le \cdots \le H_N\right) \\
=&\mathbb{P}\left(\tilde{}(F_X(F_X^{-1}(\ _1)),\ _1) \le\ _1, \ldots, \tilde{}(F_X(F_X^{-1}(\ _N)),\ _N) \le\ _N, F_X^{-1}(\ _1) \le \cdots \le F_X^{-1}(\ _N)\right) \\
=&\int_0^1 \cdots \int_0^1 \mathbf{1}_{\{\tilde{g}(F_X(F_X^{-1}(u_1)),\ _1) \le\ _1, \ldots, \tilde{g}(F_X(F_X^{-1}(u_N)),\ _N) \le\ _N\}} \cdot \\
&\qquad \mathbf{1}_{\{F_X^{-1}(u_1) \le \cdots \le F_X^{-1}(u_N)\}} \mathrm{d}\ _1 \cdots \mathrm{d}\ _N \mathrm{d}\ _1 \cdots \mathrm{d}\ _N,
\end{aligned} \tag{EC.80}$$

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

$$\mathbb{P}\left(\ _{[1:N]} \le\ _1, \ldots,\ _{[N:N]} \le\ _N\right) = \mathbb{P}\left(\tilde{}(F_X(H_{1:N}),\ _1) \le\ _1, \ldots, \tilde{}(F_X(H_{N:N}),\ _N) \le\ _N\right),$$

which proves the claim. In addition, because $H_{i:N} = F_X^{-1}(\ _{i:N})$, 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

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

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

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

and similarly, $\mathbb{P}\left((\ _{N:N},\ _N) \le\ \right) = \int_0^1 \mathscr{D}_1 C(\ , F_\theta(\ ))\ ^{N-1}\mathrm{d}\$. Therefore,

$$0 = \mathbb{P}\left(\hat{}(\ _{N:N},\ _N) \le\ \right) - \mathbb{P}($$

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

$$\mathscr{D}_1\hat{C}(\;,F_\theta(\;))=\mathscr{D}_1C(\;,F_\theta(\;)),\quad \text{almost everywhere,}\quad \;\in[0,1].$$

For any $\;\in[0,1]$, by integrating the equation above with respect to $\;$ from 0 to $\;$, according to Remark EC.1, we have $\hat{C}(\;,F_\theta(\;))=C(\;,F_\theta(\;))$. 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. □

Proof of Proposition 2. This is a corollary of Theorem 2 given Lemma EC.4. □

Proof of Theorem 3. This is a corollary of Theorem 4. □

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

Proof of "if." According to Theorem 2 and Lemma EC.4, for any fixed $\;_1,\dots,\;_m$, we have

$$\begin{aligned}
&\mathbb{P}\left(\theta_{[i_1:N]}\le\;_1,\dots,\theta_{[i_m:N]}\le\;_m\right)=\mathbb{P}\left(\;(\;_{i_1:N},\;_{i_1})\le\;_1,\dots,\;(\;_{i_m:N},\;_{i_m})\le\;_m\right)\\
&=\int_0^1\cdots\int_0^{u_3}\int_0^{u_2}\mathbb{P}\left(\;_{u_1}(\;_{i_1})\le F_\theta(\;_1),\dots,\;_{u_m}(\;_{i_m})\le F_\theta(\;_m)\right)\\
&\qquad\cdot\frac{\;_1^{i_1-1}(\;_2-\;_1)^{i_2-i_1-1}\cdots(1-\;_m)^{N-i_m}\;!}{(\;_1-1)!(\;_2-\;_1-1)!\cdots(\;-\;_m)!}\mathrm{d}\;_1\mathrm{d}\;_2\cdots\mathrm{d}\;_m\\
&=\int_0^1\cdots\int_0^{u_3}\int_0^{u_2}\prod_{k=1}^m\mathscr{D}_1C(\;_k,F_\theta(\;_k))\\
&\qquad\cdot\frac{\;_1^{i_1-1}(\;_2-\;_1)^{i_2-i_1-1}\cdots(1-\;_m)^{N-i_m}\;!}{(\;_1-1)!(\;_2-\;_1-1)!\cdots(\;-\;_m)!}\mathrm{d}\;_1\mathrm{d}\;_2\cdots\mathrm{d}\;_m\\
&=\int_0^1\cdots\int_0^{u_3}\left[\int_0^{u_2}\mathscr{D}_1C(\;_1,F_\theta(\;_1))\;_1^{i_1-1}(\;_2-\;_1)^{i_2-i_1-1}\frac{(\;_2-1)!}{(\;_1-1)!(\;_2-\;_1-1)!}\mathrm{d}\;_1\right]\\
&\qquad\cdot\prod_{k=2}^m\mathscr{D}_1C(\;_k,F_\theta(\;_k))\cdot\frac{(\;_3-\;_2)^{i_3-i_2-1}\cdots(1-\;_m)^{N-i_m}\;!}{(\;_2-1)!(\;_3-\;_2-1)!\cdots(\;-\;_m)!}\mathrm{d}\;_2\cdots\mathrm{d}\;_m.
\end{aligned}\tag{EC.83}$$

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

$$\begin{aligned}
&\int_0^{u_2}\mathscr{D}_1C(\;_1,F_\theta(\;_1))\;_1^{i_1-1}(\;_2-\;_1)^{i_2-i_1-1}\frac{(\;_2-1)!}{(\;_1-1)!(\;_2-\;_1-1)!}\mathrm{d}\;_1\\
&\overset{u_1=u_2t}{=\!=}\;_2^{i_2-1}\int_0^1\mathscr{D}_1C(\;_2\;,F_\theta(\;_1))\;^{i_1-1}(1-\;)^{i_2-i_1-1}\frac{(\;_2-1)!}{(\;_1-1)!(\;_2-\;_1-1)!}\mathrm{d}\;.
\end{aligned}\tag{EC.84}$$

Let $\;_N(\;)=\;^{i_1-1}(1-\;)^{i_2-i_1-1}\frac{(i_2-1)!}{(i_1-1)!(i_2-i_1-1)!}$, which is the density of a beta distribution. Then, we have $\int_0^1\;_N(\;)\mathrm{d}\;=1$, and $\;_N(\;)$ increases with $\;$ when $\;\in[0,\frac{i_1-1}{i_2-2}]$ and decreases when $\;\in[\frac{i_1-1}{i_2-2},1]$. Let $\;^*=\frac{\xi_1}{\xi_2}$, which satisfies that $\lim_{N\to+\infty}\frac{i_1-1}{i_2-2}=\;^*$. Because $\mathscr{D}_1C(\;,\;)$ is continuous with respect to $\;$ on $[0,1]$, it is also uniformly continuous. Thus, for any $\varepsilon>0$, there exists $\delta>0$ (which does not depend on $\;_2,\ldots,\;_m$) such that $|\mathscr{D}_1C(\;_2\;,F_\theta(\;_1)) - \mathscr{D}_1C(\;_2\;^*,F_\theta(\;_1))| < \varepsilon$ for any $\;_2 \in [0,1]$ and $|\; - \;^*| < \delta$. Therefore,

$$\begin{aligned}
&\left|\int_0^1 \mathscr{D}_1C(\;_2\;,F_\theta(\;_1))\;^{i_1-1}(1-\;)^{i_2-i_1-1}\frac{(\;_2-1)!}{(\;_1-1)!(\;_2-\;_1-1)!}\mathrm{d}\; - \mathscr{D}_1C(\;_2\;^*,F_\theta(\;_1))\right| \\
&\le \int_0^1 |\mathscr{D}_1C(\;_2\;,F_\theta(\;_1)) - \mathscr{D}_1C(\;_2\;^*,F_\theta(\;_1))|\;_N(\;)\mathrm{d}\; \\
&= \int_0^{t^*-\delta}\cdots\mathrm{d}\; + \int_{t^*-\delta}^{t^*+\delta}\cdots\mathrm{d}\; + \int_{t^*+\delta}^{1}\cdots\mathrm{d}\; = (\mathrm{I}) + (\mathrm{II}) + (\mathrm{III}).
\end{aligned} \tag{EC.85}$$

Term (I) satisfies that $(\mathrm{I}) \le 2\int_0^{t^*-\delta}\;_N(\;)\mathrm{d}\; \le 2(\;^*-\delta)\;_N(\;^*-\delta) \le 2\;_N(\;^*-\delta)$ when is sufficiently large because of $0 \le \mathscr{D}_1C \le 1$ (Remark EC.1) and the monotonicity of $\;_N(\;)$. We claim that $\lim_{N\to+\infty}\;_N(\;^*-\delta) = 0$. To prove this claim, let us consider a random variable η_N following the beta distribution with density $\;_N(\;)$. For any $\tau > 0$, by Chebyshev's inequality, we have $\mathbb{P}(|\eta_N - \;^*| > \tau) \le \mathbb{P}(|\eta_N - \mathbb{E}(\eta_N)| > \tau/2) + \mathbb{P}(|\mathbb{E}(\eta_N) - \;^*| > \tau/2) \le \mathrm{Var}(\eta_N)/(\tau/2)^2 + \mathbb{P}(|\mathbb{E}(\eta_N) - \;^*| > \tau/2)$. It is straightforward to show that $\mathrm{Var}(\eta_N) \to 0$ and $\mathbb{E}(\eta_N) \to \;^*$ as $\; \to +\infty$ and, therefore, $\mathbb{P}(|\eta_N - \;^*| > \tau) \to 0$, implying that η_N converges to $\;^*$ in probability. This demonstrates the claim, which implies that $(\mathrm{I}) \to 0$ as $\; \to +\infty$. Similarly, we have $(\mathrm{III}) \to 0$. Term (II) satisfies that $(\mathrm{II}) \le \varepsilon\int_{t^*-\delta}^{t^*+\delta}\;_N(\;)\mathrm{d}\; \le \varepsilon\int_0^1\;_N(\;)\mathrm{d}\; = \varepsilon$. Therefore, (EC.85) is less than 2ε when is sufficiently large, and we can estimate (EC.84) by

$$\begin{aligned}
&(\mathscr{D}_1C(\;_2\;^*,F_\theta(\;_1)) - 2\varepsilon)\;_2^{i_2-1} \le \\
&\qquad \int_0^{u_2} \mathscr{D}_1C(\;_1,F_\theta(\;_1))\;_1^{i_1-1}(\;_2-\;_1)^{i_2-i_1-1}\frac{(\;_2-1)!}{(\;_1-1)!(\;_2-\;_1-1)!}\mathrm{d}\;_1 \\
&\qquad\qquad \le (\mathscr{D}_1C(\;_2\;^*,F_\theta(\;_1)) + 2\varepsilon)\;_2^{i_2-1}.
\end{aligned}$$

Using similar approaches above, we can deductively derive similar bounds for the iterated integrals in (EC.83) and conclude that (EC.83) converges to $\prod_{k=1}^m \mathscr{D}_1C(\xi_k,F_\theta(\;_k))$ as increases without bound thanks to the arbitrariness of ε. In addition, we also have

$$\begin{aligned}
&\mathbb{P}(\;(\xi_1,\;_1) \le \;_1,\ldots,\;(\xi_m,\;_m) \le \;_m) = \mathbb{P}(\;_{\xi_1}(\;_1) \le F_\theta(\;_1),\ldots,\;_{\xi_m}(\;_m) \le F_\theta(\;_m)) \\
&= \mathbb{P}(\;_1 \le \mathscr{D}_1C(\xi_1,F_\theta(\;_1)),\ldots,\;_m \le \mathscr{D}_1C(\xi \quad \theta
\end{aligned}$$

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

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

For any $\ \in [0,1]$, by integrating the equation above with respect to ξ_k from 0 to , according to Remark EC.1, we have $\hat{C}(\ , F_\theta(\ )) = C(\ , F_\theta(\ ))$. 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. □

EC.3.4. Proofs for Section 4

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

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

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

EC.3.5. Proofs for Section 5

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

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

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

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

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

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

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

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

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

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

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

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

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

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 $^{\mathrm{Ga}}(\xi) > 0$. For notational simplicity, let $_\beta(\ ) \equiv \begin{cases} |\ |^\beta, & \geq 0, \\ -|\ |^\beta, & < 0, \end{cases}$ and $\equiv \frac{1}{\Phi^{-1}(\xi)} > 0$. Then,

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

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

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

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

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

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

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

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

Therefore, by Durrett (2019, Theorem A.5.3), $\ '_{\beta}(0) = 0$. With similar arguments, we can further show that $\ ''_{\beta}(0) = \frac{1-\rho^2}{\rho^2}\beta(\beta-1)$. Hence, when $\beta > 2$, the second-order Taylor expansion of $\ _{\beta}(\ )$ is given by:

$$\ _{\beta}(\ ) = 1 + \frac{1-\rho^2}{2\rho^2}\beta(\beta-1)\ ^2 + \ (\ ^2), \qquad \to 0^+. \tag{EC.89}$$

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

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

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

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

In addition, by direct calculation, we have

$$\ _1(\ ) \equiv \int_{-\infty}^{+\infty} \ _1(\ ,\ )\mathrm{d}\ = \int_{-\infty}^{+\infty}\left(1 + \frac{\sqrt{1-\rho^2}}{\rho}\ \right)\frac{\ ^{-\ ^2/2}}{\sqrt{2\pi}}\mathrm{d}\ = 1. \tag{EC.92}$$

Hence,

$$\begin{aligned}
&\mathrm{Var}\left[\ \left(\frac{\rho}{\ } + \sqrt{1-\rho^2}\ \right)\right] = \mathrm{Var}\left[\sigma\left(\frac{\rho}{\ } + \sqrt{1-\rho^2}\ \right) + \tau\ _{\beta}\left(\frac{\rho}{\ } + \sqrt{1-\rho^2}\ \right)\right] \\
&= \sigma^2(1-\rho^2) + \tau^2\mathrm{Var}\left[\ _{\beta}\left(\frac{\rho}{\ } + \sqrt{1-\rho^2}\ \right)\right] + 2\sigma\tau\mathrm{Cov}\left[\left(\frac{\rho}{\ } + \sqrt{1-\rho^2}\ \right),\ _{\beta}\left(\frac{\rho}{\ } + \sqrt{1-\rho^2}\ \right)\right] \\
&= \sigma^2(1-\rho^2) + \tau^2\left(\frac{\rho}{\ }\right)^{2\beta}\left[\tilde{\ }_{2\beta}(\ ) - \ _{\beta}(\ )^2\right] + 2\sigma\tau\left(\frac{\rho}{\ }\right)^{\beta+1}\left[\tilde{\ }_{\beta+1}(\ ) - \ _1(\ )\ _{\beta}(\ )\right] \\
&= \sigma^2(1-\rho^2) + \tau^2\frac{\rho^{2\beta}}{\ ^{2\beta-2}}\left[\frac{1-\rho^2}{\rho^2}\beta^2 + \ (1)\right] + 2\sigma\tau\frac{\rho^{\beta+1}}{\ ^{\beta-1}}\left[\frac{1-\rho^2}{\rho^2}\beta + \ (1)\right], \qquad \to 0^+,
\end{aligned} \tag{EC.93}$$

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

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

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

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 $+1$

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 -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 $\ ,\ = 1, 2, \ldots,\ $ and $\ \neq\ $, the (,)-entry of A is given by (EC.58), and the (,)-entry of A is given by (EC.61); the (,)-entry of B is given by (EC.59), and the (,)-entry of B is given by (EC.62).

Let $\boldsymbol{\eta} = (\mathbb{E}(\ _{1:N}), \mathbb{E}(\ _{2:N}), \ldots, \mathbb{E}(\ _{N:N}))^\top$, 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,

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

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 $\ ,\ = 1, 2, \ldots,\ $. 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,

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

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

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

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

We first prove the upper bound of $\mathrm{IR}^*_{\mathrm{T}\quad\mathrm{Va}}$. Denote by $\lambda\quad(A)$ the smallest eigenvalue of A, and Lemma EC.7 implies that $\lambda\quad(A) \geq \min_{i=1,2,\ldots,N}\left[A_{ii} - \sum_{j\neq i}|A_{ij}|\right]$, where A_{ij} is the (,)-entry of A, , $= 1,2,\ldots,$. 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 $= 1,2,\ldots,$, we have

$$A_{ii} - \sum_{j\neq i}|A_{ij}| = \sigma_\theta^2 \cdot \left[1-(\overline{\rho}^2+\sigma_\rho^2)+(\overline{\rho}^2+\sigma_\rho^2)\cdot \mathrm{Var}(\quad_{i:N})\right] - \sum_{j\neq i}\sigma_\theta^2\cdot(\overline{\rho}^2+\sigma_\rho^2)\cdot|\mathrm{Cov}(\quad_{i:N},\quad_{j:N})|$$

$$= \sigma_\theta^2 \cdot \Bigg[1-(\overline{\rho}^2+\sigma_\rho^2)+(^-$$

Hence, the optimal information ratio satisfies

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

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

$$(\mathrm{IR}^*_{\mathrm{D} \quad \mathrm{d} \quad \mathrm{c}})^2 = \frac{(\rho-\tilde{\rho})^2}{1-\rho_X} \cdot \sigma_\theta^2 \cdot \boldsymbol{\eta}^\top \quad {}^{-1} \boldsymbol{\eta} + \mu_\theta^2 \mathbf{1}^\top \quad {}^{-1} \mathbf{1} - \frac{\sigma_\theta^2 \cdot \rho_\theta \cdot \mu_\theta^2 \cdot (\mathbf{1}^\top \quad {}^{-1} \mathbf{1})^2}{1 + \sigma_\theta^2 \cdot \rho_\theta \cdot \mathbf{1}^\top \quad {}^{-1} \mathbf{1}}$$
$$= \frac{(\rho-\tilde{\rho})^2}{1-\rho_X} \cdot \sigma_\theta^2 \cdot \boldsymbol{\eta}^\top \quad {}^{-1} \boldsymbol{\eta} + \frac{\mu_\theta^2}{1/(\mathbf{1}^\top \quad {}^{-1} \mathbf{1}) + \sigma_\theta^2 \cdot \rho_\theta}. \quad \text{(EC.101)}$$

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

We first prove the upper bound of $\mathrm{IR}^*_{\mathrm{D} \quad \mathrm{d} \quad \mathrm{c}}$. Denote by $\lambda \quad (\quad)$ the smallest eigenvalue of $\quad$, and Lemma EC.7 implies that $\lambda \quad (\quad) \geq \min_{i=1,2,\ldots,N} \left[\quad {}_{ii} - \sum_{j\neq i} | \quad {}_{ij}| \right]$, where $\quad {}_{ii}$ and $\quad {}_{ij}$ are given by (EC.99). In addition, by Parts (iv) and (v) of Lemma EC.5 and (EC.22), for any $\quad = 1, 2, \ldots, \quad$, we have

$$\quad {}_{ii} - \sum_{j \neq i} | \quad {}_{ij}| = \sigma_\theta^2 \cdot \left[1 - \rho_\theta - \frac{(\rho-\tilde{\rho})^2}{1-\rho_X} + \frac{(\rho-\tilde{\rho})^2}{1-\rho_X} \cdot \mathrm{Var}(\quad {}_{i:N}) \right] - \sum_{j\neq i} \sigma_\theta^2 \cdot \frac{(\rho-\tilde{\rho})^2}{1-\rho_X} |\mathrm{Cov}(\quad {}_{i:N}, \quad {}_{j:N})|$$
$$= \sigma_\theta^2 \cdot \left[1 - \rho_\theta - \frac{(\rho-\tilde{\rho})^2}{1-\rho_X} + \frac{(\rho-\tilde{\rho})^2}{1-\rho_X} \cdot \mathrm{Var}(\quad {}_{i:N}) - \frac{(\rho-\tilde{\rho})^2}{1-\rho_X} \cdot \mathrm{Cov}\left(\quad {}_{i:N}, \sum_{j\neq i} \quad {}_{j:N} \right) \right]$$
$$= \sigma_\theta^2 \left[1 - \rho_\theta - 2 \cdot \frac{(\rho-\tilde{\rho})^2}{1-\rho_X} \cdot (1 - \mathrm{Var}(\quad {}_{i:N})) \right] \geq \sigma_\theta^2 \left[1 - \rho_\theta - 2 \cdot \frac{(\rho-\tilde{\rho})^2}{1-\rho_X} \right]. \quad \text{(EC.102)}$$

Hence,

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

Combining this with (EC.101), we have

$$(\mathrm{IR}^*_{\mathrm{D} \quad \mathrm{d} \quad \mathrm{c}})^2 = \frac{(\rho-\tilde{\rho})^2}{1-\rho_X} \cdot \sigma_\theta^2 \cdot \boldsymbol{\eta}^\top \quad {}^{-1} \boldsymbol{\eta} + \frac{\mu_\theta^2}{1/(\mathbf{1}^\top \quad {}^{-1} \mathbf{1}) + \sigma_\theta^2 \cdot \rho_\theta}$$
$$\leq \frac{(\rho-\tilde{\rho})^2}{1-\rho_X} \cdot \sigma_\theta^2 \cdot \frac{1}{\lambda \quad (\quad)} \boldsymbol{\eta}^\top \quad \quad {}^\top \boldsymbol{\eta} + \frac{\mu_\theta^2}{\lambda \quad (\quad)/(\mathbf{1}^\top \quad \quad {}^\top \mathbf{1}) + \sigma_\theta^2 \cdot \rho_\theta}$$
$$= \frac{(\rho-\tilde{\rho})^2}{1-\rho_X} \cdot \sigma_\theta^2 \cdot \frac{1}{\lambda \quad (\quad)} \quad (\quad) + \frac{\mu_\theta^2}{\lambda \quad (\quad)/ \quad + \sigma_\theta^2 \cdot \rho_\theta}$$
$$\leq \frac{(\rho-\tilde{\rho})^2 \quad (\quad)}{(1-\rho_\theta)(1-\rho_X) - 2(\rho-\tilde{\rho})^2} + \frac{\mu_\theta^2/\sigma_\theta^2}{[(1-\rho_\theta)(1-\rho_X) - 2(\rho-\tilde{\rho})^2]/[(1-\rho_X) \quad] + \rho_\theta}, \quad \text{(EC.104)}$$

which proves the upper bound.

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

$$(\mathrm{IR}^*_{\mathrm{D} \quad \mathrm{d} \quad \mathrm{c}})^2 \ge \frac{(\rho - \tilde{\rho})^2 \ (\ )}{(1-\rho_\theta)(1-\rho_X)} + \frac{\mu_\theta^2/\sigma_\theta^2}{(1-\rho_\theta)/\ +\rho_\theta},$$

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 $\ _{1:N} \le \ _{2:N} \le \cdots \le \ _{N:N}$ be the order statistics of $\ _1, \ _2, \ldots, \ _N \overset{\mathrm{IID}}{\sim} \mathcal{N}(0,1)$. 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 $\mathbb{E}(\theta_{[i:N]}) = \sigma_\theta \cdot \rho \cdot \mathbb{E}(\ _{i:N})$. Thus, the optimal weight to maximize the IR, $\bar{\boldsymbol{w}}^* \equiv (0,\ldots,0,\ ^*_{\lfloor N\xi_1 \rfloor},\ ^*_{\lfloor N\xi_1+1 \rfloor},\ldots,\ ^*_{\lfloor N\xi_2 \rfloor},0,\ldots,0)$, can be approximated by

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

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

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

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

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

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 $\mathbb{E}(\theta_{[i:N]}) = \sigma_\theta \cdot \rho \cdot \mathbb{E}(\ _{i:N})$. Hence, the information ratio of the equal-weighted portfolio can be approximated by

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

$$= \frac{\rho}{\sqrt{1-\rho^2}} \cdot \frac{1}{\sqrt{\lfloor \quad \xi_2 \rfloor - \lfloor \quad \xi_1 \rfloor + 1}} \cdot \sum_{i=\lfloor N\xi_1 \rfloor}^{\lfloor N\xi_2 \rfloor} \mathbb{E}(\quad _{i:N}) \approx \frac{\rho}{\sqrt{1-\rho^2}} \cdot \frac{\sqrt{\quad}}{\sqrt{\xi_2 - \xi_1}} \cdot \sum_{i=\lfloor N\xi_1 \rfloor}^{\lfloor N\xi_2 \rfloor} \frac{\mathbb{E}(\quad _{i:N})}{}$$

$$\to \frac{\rho \cdot \sqrt{\quad}}{\sqrt{1-\rho^2}} \cdot \frac{\varphi\big(\Phi^{-1}(\xi_1)\big) - \varphi\big(\Phi^{-1}(\xi_2)\big)}{\sqrt{\xi_2 - \xi_1}},$$

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

$$\bar{\ } = \mathbb{E}\left[\frac{\sum_{i=\lfloor N\xi_1 \rfloor}^{\lfloor N\xi_2 \rfloor} \frac{1}{\lfloor N\xi_2 \rfloor - \lfloor N\xi_1 \rfloor + 1} \quad _{i:N}}{\sum_{i=\lfloor N\xi_1 \rfloor}^{\lfloor N\xi_2 \rfloor} \frac{1}{\lfloor N\xi_2 \rfloor - \lfloor N\xi_1 \rfloor + 1}} \right] = \frac{\sum_{i=\lfloor N\xi_1 \rfloor}^{\lfloor N\xi_2 \rfloor} \mathbb{E}(\quad _{i:N})}{\lfloor \quad \xi_2 \rfloor - \lfloor \quad \xi_1 \rfloor + 1} = \mu_X + \sigma_X \cdot \frac{\sum_{i=\lfloor N\xi_1 \rfloor}^{\lfloor N\xi_2 \rfloor} \mathbb{E}(\quad _{i:N})}{\lfloor \quad \xi_2 \rfloor - \lfloor \quad \xi_1 \rfloor + 1}$$

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

$$
\begin{aligned}
&= \sum_{\pi=(i_1,\dots,i_N)\in\mathcal{P}} \ _{i_1,\dots,i_N}\mathbb{P}(\ _1(\pi,\ \ _1,\ _1)\le\ _1,\dots,\ \ _N(\pi,\ \ _N,\ _N)\le\ _N|\Pi=\pi)\\
&= \sum_{\pi=(i_1,\dots,i_N)\in\mathcal{P}} \ _{i_1,\dots,i_N}\cdot\\
&\qquad \int\cdots\int \mathbb{P}(\ _1(\pi,\ _1,\ _1)\le\ _1,\dots,\ \ _N(\pi,\ _N,\ _N)\le\ _N|\Pi=\pi,\ \ _1=\ _1,\dots,\ \ _N=\ _N)\\
&\qquad\qquad \cdot\ \ _{i_1,i_2,\dots,i_N}(\ _1,\ _2,\dots,\ _N)\mathrm{d}\ _1\cdots\mathrm{d}\ _N\\
&= \sum_{\pi=(i_1,\dots,i_N)\in\mathcal{P}} \int\cdots\int \mathbb{P}(\ _1(\pi,\ _1,\ _1)\le\ _1,\dots,\ \ _N(\pi,\ _N,\ _N)\le\ _N)\\
&\qquad\qquad \cdot\prod_{k=1}^{N}\ _{X,i_k}(\ _k)\cdot\mathbf{1}_{\{\ _1\le\cdots\le\ _N\}}\mathrm{d}\ _1\cdots\mathrm{d}\ _N\\
&= \sum_{\pi=(i_1,\dots,i_N)\in\mathcal{P}} \int\cdots\int \prod_{k=1}^{N}\mathbb{P}(\ _k(\pi,\ _k,\ _k)\le\ _k)\cdot\prod_{k=1}^{N}\ _{X,i_k}(\ _k)\cdot\mathbf{1}_{\{\ _1\le\cdots\le\ _N\}}\mathrm{d}\ _1\cdots\mathrm{d}\ _N. \qquad \text{(EC.106)}
\end{aligned}
$$

By the definition of $\ _k$,

$$
\begin{aligned}
\mathbb{P}(\ _k(\pi,\ _k,\ _k)\le\ _k) &= \mathbb{P}\left(F^{-1}_{\theta,i_k}\circ\ ^{i_k}_{F_{X,i_k}(\ _k)}(\ _k)\le\ _k\right)=\mathbb{P}\left(\ ^{i_k}_{F_{X,i_k}(\ _k)}(\ _k)\le F_{\theta,i_k}(\ _k)\right)\\
&= \mathbb{P}\left(\ _k\le\frac{\partial C_{i_k}}{\partial\ }\left(F_{X,i_k}(\ _k),F_{\theta,i_k}(\ _k)\right)\right)=\frac{\partial C_{i_k}}{\partial\ }\left(F_{X,i_k}(\ _k),F_{\theta,i_k}(\ _k)\right),
\end{aligned}
$$

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

$$
\begin{aligned}
\mathbb{P}\left(\theta_{i_k}\le\ _k|\ \ _{i_k}=\ _k\right) &= \mathbb{P}\left(F_{\theta,i_k}(\theta_{i_k})\le F_{\theta,i_k}(\ _k)\big|F_{X,i_k}(\ _{i_k})=F_{X,i_k}(\ _k)\right)\\
&= \frac{\partial C_{i_k}}{\partial\ }\left(F_{X,i_k}(\ _k),F_{\theta,i_k}(\ _k)\right)=\mathbb{P}(\ _k(\pi,\ _k,\ _k)\le\ _k).
\end{aligned}
$$

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

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

$$
\begin{aligned}
&\mathbb{E}(\theta_{[i:N]})=\mathbb{E}(\ _i(\Pi,\ \ _i,\ _i))\\
&= \sum_{\pi=(j_1,j_2,\dots,j_N)} \ _{j_1,j_2,\dots,j_N}\int_0^1\int_{-\infty}^{+\infty}\Bigg[F^{-1}_{\theta,j_i}\circ\ ^{j_i}_{F_{X,j_i}(\ )}(\ )\cdot\frac{1}{\ _{j_1,j_2,\dots,j_N}}\ _{X,j_i}(\ )\\
&\qquad\cdot\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\prod_{s=1,s\neq i}^{N}\ _{X,j_s}(\ _s)\mathbf{1}_{\{\ _1\le\cdots\le\ _{i-1}\le\ \le\ _{i+1}\le\cdots\le\ _N\}}\mathrm{d}\ _1\cdots\mathrm{d}\ _{i-1}\mathrm{d}\ _{i+1}\mathrm{d}\ _N\Bigg]\mathrm{d}\ \mathrm{d}\\
&= \sum_{k=1}^{N}\int_0^1\int_{-\infty}^{+\infty}\Bigg[F^{-1}_{\theta,j_i}\circ\ ^{j_i}_{F_{X,j_i}(\ )}(\ )\ \ _{X,j_i}(\ )\cdot\sum_{\pi=(j_1,j_2,\dots,j_N),j_i=k}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\\
&\qquad\prod_{s=1,s\neq i}^{N}\ _{X,j_s}(\ _s)\mathbf{1}_{\{\ _1\le\cdots\le\ _{i-1}\le\ \le\ _{i+1}\le\cdots\le\ _N\}}\mathrm{d}\ _1\cdots\mathrm{d}\ _{i-1}\mathrm{d}\ _{i+1}\mathrm{d}\ _N\Bigg]\mathrm{d}\ \mathrm{d}\\
&= \sum_{k=1}^{N}\int_0^1\int_{-\infty}^{+\infty}\left[F^{-1}_{\theta,j_i}\circ\ ^{j_i}_{F_{X,j_i}(\ )}(\ )\ \ _{X,j_i}(\ )\cdot H^i_k(\ )\right]\mathrm{d}\ \mathrm{d}\ =\sum_{k=1}^{N}\int_0^1\int_0^1\tilde{\ }_k(\ ,\ )H^i_k(\ )\mathrm{d}\ \mathrm{d}\ ,
\end{aligned}
$$

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

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 $\tilde{C}(\ ,\ )$ is stochastically increasing if and only if, for any $\in [0,1]$, $\tilde{C}(\ ,\ )$ is a concave function of . Therefore, by Lemma EC.3, the linearly interpolating copula C is also stochastically increasing.

By Theorem 2, for $= 1, 2, \ldots,\ $, $\theta_{[i:N]} \overset{d}{=} (\ _{i:N},\ _i)$, where is defined as (15). According to the definition of stochastically increasing, $\mathscr{D}_1 C(\ ,\ )$ is non-increasing with . We claim that this implies that the function $(\ ,\ ) = F_\theta^{-1} \circ\ _u(\ )$ is non-decreasing with . To prove this claim, consider $0 \leq\ _1 <\ _2 \leq 1$, then for any fixed , we have $\mathscr{D}_1 C(\ _1,\ ) \geq \mathscr{D}_1 C(\ _2,\ )$. Because $_{u_1}(\ ) = \inf\{\ : \mathscr{D}_1 C(\ _1,\ ) \geq\ \}$ and $_{u_2}(\ ) = \inf\{\ : \mathscr{D}_1 C(\ _2,\ ) \geq\ \}$, we immediately have $_{u_1}(\ ) \leq\ _{u_2}(\ )$ for any fixed . Further because F_θ is non-decreasing, $(\ _1,\ ) \leq (\ _2,\ )$ holds. This proves the claim.

Through the representation $\theta_{[i:N]} \overset{d}{=} ($

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 $\mathscr{D}_1 C(\ ,\ ) = \begin{cases} 1, & \geq\ , \\ 0, & <\ , \end{cases}$ and therefore $_u(\ ) =$ for $\in (0,1]$. This implies that

$$(\ (\ _{1:N},\ _1),\ldots,\ (\ _{N:N},\ _N)) = \left(F_\theta^{-1}(\ _{1:N}),\ldots,F_\theta^{-1}(\ _{N:N})\right) \stackrel{d}{=} (\theta_{1:N},\ldots,\theta_{N:N}).$$

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

$$(\ (\ _{1:N},\ _1)$$

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

$$
\begin{aligned}
&\mathbb{P}\left(\mu_\theta - \theta_{[N:N]} \le \quad {}_1, \ldots, \mu_\theta - \theta_{[1:N]} \le \quad {}_N\right) \\
&= \int_0^1 \cdots \int_0^1 \prod_{i=1}^N \left[1 - \frac{\partial C}{\partial \;}(\quad {}_i, F_\theta(\mu_\theta - \quad {}_i))\right] \quad !\mathbf{1}_{\{u_1 > \cdots > u_N\}} \mathrm{d} \quad {}_1 \cdots \mathrm{d} \quad {}_N \\
&= \int_0^1 \cdots \int_0^1 \prod_{i=1}^N \left[1 - \frac{\partial C}{\partial \;}(1 - \quad {}_i, F_\theta(\mu_\theta - \quad {}_i))\right] \quad !\mathbf{1}_{\{u_1 < \cdots < u_N\}} \mathrm{d} \quad {}_1 \cdots \mathrm{d} \quad {}_N,
\end{aligned}
\tag{EC.109}
$$

where the last equality replaces $\quad {}_i$ with $1 - \quad {}_i$.

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

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 ϕ_γ^{G} is $(\phi_\gamma^{\mathrm{G}})'(\;) = -\exp(- \; ^{1/\gamma}) \; ^{1/\gamma - 1}/\gamma$, whose inverse function $(\phi_\gamma^{\mathrm{G}})'^{-1}$ 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

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

where $_{i,t}$ and $_{i,t}$ are the weight and return of stock in the portfolio in year , 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 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.