Operations Research

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

Optimal Impact Portfolios with General Dependence and Marginals

Andrew W. Lo, Lan Wu, Ruixun Zhang, Chaoyi Zhao

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

OPERATIONS RESEARCH
Articles in Advance, pp. 1–15
ISSN 0030-364X (print), ISSN 1526-5463 (online)

Contextual Areas

Optimal Impact Portfolios with General Dependence and Marginals

Andrew W. Lo,[a,b,c,d] **Lan Wu,**[e,f] **Ruixun Zhang,**[e,f,g,h,*] **Chaoyi Zhao**[e,*]

[a] MIT Sloan School of Management, Cambridge, Massachusetts 02142; [b] Laboratory for Financial Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142; [c] Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142; [d] Santa Fe Institute, Santa Fe, New Mexico 87501; [e] School of Mathematical Sciences, Peking University, Beijing 100871, China; [f] Laboratory for Mathematical Economics and Quantitative Finance, Peking University, Beijing 100871, China; [g] Center for Statistical Science, Peking University, Beijing 100871, China; [h] National Engineering Laboratory for Big Data Analysis and Applications, Peking University, Beijing 100871, China

*Corresponding author

Contact: alo@mit.edu (AWL); lwu@pku.edu.cn (LW); zhangruixun@pku.edu.cn, https://orcid.org/0000-0002-7670-8393 (RZ); zhaochaoyi@pku.edu.cn, https://orcid.org/0000-0002-5921-3612 (CZ)

Received: July 23, 2023
Revised: November 21, 2023; February 5, 2024
Accepted: February 12, 2024
Published Online in Articles in Advance: March 20, 2024

Area of Review: Financial Engineering

https://doi.org/10.1287/opre.2023.0400

Abstract. We develop a mathematical framework for constructing optimal impact portfolios and quantifying their financial performance by characterizing the returns of impact-ranked assets using induced order statistics and copulas. The distribution of induced order statistics can be represented by a mixture of order statistics and uniformly distributed random variables, where the mixture function is determined by the dependence structure between residual returns and impact factors—characterized by copulas—and the marginal distribution of residual returns. This representation theorem allows us to explicitly and efficiently compute optimal portfolio weights under any copula. This framework provides a systematic approach for constructing and quantifying the performance of optimal impact portfolios with arbitrary dependence structures and return distributions.

Funding: Research funding from the China National Key R&D Program [Grant 2022YFA1007900], the China National Natural Science Foundation [Grants 12271013, 72342004], the Fundamental Research Funds for the Central Universities (Peking University), and the MIT Laboratory for Financial Engineering is gratefully acknowledged.
Supplemental Material: The online appendix is available at https://doi.org/10.1287/opre.2023.0400.

Keywords: impact investing • environmental, social, and governance (ESG) • induced order statistics • copula • representation theorem • portfolio theory

1. Introduction

Impact investing, broadly defined as investments that consider not only financial objectives but also other goals that support certain social priorities and agendas, has drawn an increasing amount of attention in recent years. This form of investing has been given other names, such as "sustainable" and "green" investing, and examples include divestment from "sin stocks," "socially responsible" investing (SRI), the use of environmental, social, and governance (ESG) investment criteria, and private equity funds seeking social impact alongside financial returns.[1]

The construction of impact portfolios typically involves ranking assets based on a measure of impact, such as a company's ESG score, its amount of carbon emissions, or the prospect of developing a disease-modifying drug. Popular methods include negative screening, imposing filters so that certain companies are excluded from the investable universe; positive screening, selecting companies for high values of certain attributes; relative weighting, over- or under-weighting companies within an industry based on a certain measure; and full ESG factor integration in which ESG information is combined with other fundamental and technical factors of a company to improve the investment process.[2]

Despite the rapid growth in popularity and assets under the management of impact investing, it is unclear whether they are adding or removing value from an investor's point of view.[3] In addition, there are ongoing controversies regarding whether ESG portfolios and other impact investing products may be violations of the fiduciary duties of the managers of those products.[4] These challenges call for a general framework for constructing optimal impact portfolios and for measuring and disclosing the financial impact of impact investing.

1.1. Main Results

In this article, we develop a framework to quantify the distribution of asset returns after ranking, screening, and weighting based on a measure of impact. We demonstrate how to construct optimal impact portfolios, and explicitly quantify their performance in terms of the well-known information ratio (see, for example, Grinold and Kahn 2019). Our framework allows for arbitrary marginal distributions of the asset returns and the impact factor, as well as arbitrary dependence structures between the two, characterized by copulas.

We formalize impact investing as the sorting and selection within an investment universe of N assets based on an impact factor, X_i, for asset i, so that higher values of X_i correspond to greater impact, for example, lower carbon emissions, higher ESG score, etc. We consider a linear multifactor model for asset returns and focus on the residual returns in excess of all known factors, denoted by $\boldsymbol{\theta} \equiv (\theta_1, \theta_2, \ldots, \theta_N)^\top$. The impact on investment performance is therefore determined by the joint distribution of the vector $\boldsymbol{X} \equiv (X_1, X_2, \ldots, X_N)^\top$ of impact factors with the residual returns, $\boldsymbol{\theta}$.

Our fundamental tool to quantify the distribution of $\boldsymbol{\theta}$ after ranking based on $\boldsymbol{X}$ is the *induced order statistic*, which uses random variables that are ranked not by their own values ($\boldsymbol{\theta}$ in our case) but by the values of other random variables ($\boldsymbol{X}$ in our case). However, the existing literature on induced order statistics often relies on strong distributional assumptions, and their properties under general distributions are rarely discussed.[5]

Our main result is a *representation theorem*, which shows that, for a *general* bivariate distribution of $(X_i, \theta_i)^\top$, the distribution of induced order statistics of θ_i ranked by X_i can be represented by a mixture of order statistics and uniformly distributed random noise, and the mixture function is determined by the copula of X_i and θ_i and the marginal distribution of θ_i. These results completely characterize the distribution—both in finite sample and asymptotically—of induced order statistics and allow for the efficient computation of optimal portfolio weights. To the best of our knowledge, we are the first to provide results for induced order statistics under general copulas and marginal distributions.

When the distribution of $(X_i, \theta_i)^\top$ is discontinuous, our construction proof shows that the representation holds if and only if the copula is "linearly interpolating" in the undefined regions given by Sklar's (1959) theorem.[6] This generality allows us to derive weights for optimal impact portfolios when the impact factor, $\boldsymbol{X}$, represents discrete rating values, which is often the case in practice, and when residual returns, $\boldsymbol{\theta}$, include jumps or take discrete values in high-frequency data.

The representation theorem has several important applications. To understand the impact of general dependence structures between the residual return, $\boldsymbol{\theta}$, and the impact factor, $\boldsymbol{X}$, we apply the representation theorem under different copulas. In particular, we can quantify the impact of asymmetric tail dependence between asset returns and the impact factor—modeled by the Archimedean copula—which is commonly observed in many impact investing contexts.[7] We find that stronger tail dependence implies larger weights for assets on that tail in optimal impact portfolios.

In addition, we use the representation theorem to study the impact of the marginal distribution of residual returns on impact portfolios. It is well documented that stock returns are skewed and heavy-tailed. We show that a larger positive (negative) skewness leads to smaller weights for higher (lower) impact-ranked assets. Heavy-tailed returns may lead to nonmonotonic weights for the optimal impact portfolio, because of the extreme levels of risk from assets with the most extreme impact values.

1.2. Related Literature

Our work contributes to a rapidly growing literature on quantitative models for green finance. For example, Blasberg et al. (2021) propose the carbon default swap to manage the exposure to transition risk. De Angelis et al. (2023) model the effect of impact investors on mitigating climate risks by raising the cost of capital of the most carbon-intensive companies. Aïd and Biagini (2023) discuss the impact of regulations on the carbon emission market. Gobet and Lage (2023) model the optimal transition path of credit portfolios. Our work focuses on impact investing from the perspective of portfolio construction. The literature on green portfolio construction and sustainable asset pricing includes, for example, Pástor et al. (2021), Pedersen et al. (2021), Sorensen et al. (2021), and Flora and Tankov (2023). Although this literature incorporates green constraints into the traditional asset-pricing framework, the interaction between these constraints and asset returns is not explicitly modeled.

Our results extend the recent work of Lo and Zhang (2023), which models the dependence between impact factor and asset returns when they are jointly normally distributed with a constant correlation. However, as demonstrated empirically in Online Appendix EC.2, impact factors and returns do not follow normal distributions in practice. Although the joint normality assumption yields simple analytical results, it is unclear how impact portfolios and their performances are affected by deviations from normality (e.g., skewed and heavy-tailed distributions) and asymmetric dependencies between asset returns and impact factors. In this article, we provide a complete characterization of these interactions and a framework to construct impact portfolios that account for them. Empirical results in Online Appendix EC.2 show that allowing for general dependence and marginal distributions can significantly improve the financial performance of impact portfolios.

As mentioned above, we contribute to the theory of induced order statistics by characterizing their distributions under general dependence structures. These statistics, whose term was first coined by Bhattacharya (1974), are also referred to as *concomitants* of the order statistics (David 1973).[8] Lo and MacKinlay (1990) apply these same statistical tools to quantify data-snooping biases in testing financial asset-pricing models. To our knowledge, we are the first to combine induced order statistics with copulas, and the latter are widely used in credit risk models (Giesecke 2003, 2004; Cont and Kan 2011; Filiz et al. 2012; Cont and Minca 2013; Brigo et al. 2014). In this article, we derive the distribution of induced order statistics under arbitrary copulas, and provide a novel application in the context of impact investing.

We present the main results under the assumption that $(X_i, \theta_i)^\top$ is independent and identically distributed (IID) across assets (Assumption 1) for three reasons. First, the analytical results under the IID assumption yield considerable intuition that is mathematically clear yet fundamental to describing the distribution of induced order statistics and the performance of impact portfolios. Second, the main results can be generalized to allow for cross-sectional heterogeneity with more mathematical complexity (see Online Appendix EC.1.3.2), but both the representation theorem and its associated intuition carry over. Finally, our characterization of induced order statistics under general marginal distributions and dependence is new even under this IID assumption, which may yield applications in other domains and contexts beyond impact investing.

In fact, the theory of induced order statistics for bivariate random variables with arbitrary copulas can potentially be applied much more broadly to contexts that involve ranking bivariate pairs by one covariate. This includes portfolio selection based on hundreds of new factors and anomalies in the "Factor Zoo" discussed in the recent literature (Harvey et al. 2016, Feng et al. 2020, Hou et al. 2020), as well as the pricing and portfolio construction of credit bonds, credit default swaps (CDS), CDS indices, and so on, because investors may form portfolios by ranking these assets based on their credit ratings.

1.3. Outline

The remainder of this article is organized as follows. We describe our framework in Section 2. Section 3 shows the representation theorem. We then discuss the impact of general dependence structures in Section 4 and marginal distributions in Section 5. We conclude in Section 6, and provide additional technical details, empirical studies, and proofs in the online appendix.

2. The Framework

We consider a world of N assets whose returns in excess of the risk-free rate satisfy the following linear multifactor model:

$$r_{it} = \alpha_i + \beta_{i1}\Lambda_{1t} + \beta_{i2}\Lambda_{2t} + \cdots + \beta_{iK}\Lambda_{Kt} + \varepsilon_{it},$$
$$i = 1, 2, \ldots, N; \quad t = 1, 2, \ldots, T; \tag{1}$$

where Λ_{kt} is the k-th factor return (in excess of the risk-free rate), $k = 1, 2, \ldots, K$, α_i and β_{ik} are the excess return and factor betas, respectively, and ε_{it} is the idiosyncratic return component independent of all factors. The ε_{it}'s are zero-mean random variables that are IID both across assets i and over time t. To allow for potential mispricings, we assume that α_i are random variables that are IID across assets i, and they follow a nondegenerate distribution.[9]

We define the following quantity to be the *residual return* of asset i at time t, for $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$:

$$\theta_{it} \equiv \alpha_i + \varepsilon_{it}. \tag{2}$$

Therefore, given a specific time t, the unconditional distribution of θ_{it} is IID across asset i. Given a specific asset i, the unconditional distribution of θ_{it} is identically distributed but dependent over time t because of the common term, α_i.

Under suitable restrictions on $\{\alpha_i, \beta_{ik}\}$ and the definitions of the factor returns $\{\Lambda_k\}$, the linear multifactor model in (1) is consistent with a number of asset-pricing models, such as the Capital Asset Pricing Model (CAPM) (Sharpe 1964, Lintner 1965), the Intertemporal CAPM (Merton 1973), the Arbitrage Pricing Theory (APT) (Ross 1976), and the Fama-French multifactor models (Fama and French 1993). In particular, all of these asset-pricing models imply that $\alpha_i = 0$, and that the expected returns of the assets are simply the sum of the risk-free rate plus all the risk premia multiplied by the asset's corresponding risk exposures or β_{ik}'s.

However, to measure the impact of impact investing, we take no position as to whether any particular asset-pricing model holds. We allow for the possibility of superior investment performance of individual securities, but also include the conventional case of equilibrium or no-arbitrage pricing if we set the mean and variance of the alphas to zero. The implications of our model are broadly applicable to an equilibrium asset-pricing setup where "alpha" may be reinterpreted as omitted factors which investors are either unaware of or unable to access as easily as portfolio managers (see Lo and Zhang 2023, section 4.4).

2.1. Impact Factors and Induced Order Statistics

Impact investors rank assets according to a specific impact variable,[10] and form portfolios based on the impact variable. We assume that there exists a specific impact factor for security i at time t (e.g., the ESG score), X_{it}, which may be correlated with the residual returns of the i-th asset, θ_{it}. Although θ_{it} is not observable at time t, its corresponding impact factor, X_{it}, is observable. Let $\boldsymbol{X}_t = (X_{1t}, \ldots, X_{Nt})^\top$ and $\boldsymbol{\theta}_t = (\theta_{1t}, \ldots, \theta_{Nt})^\top$. In particular, we assume the following:

Assumption 1. *For a specific time t,*

$$\begin{pmatrix} X_{1t} \\ \theta_{1t} \end{pmatrix}, \begin{pmatrix} X_{2t} \\ \theta_{2t} \end{pmatrix}, \ldots, \begin{pmatrix} X_{Nt} \\ \theta_{Nt} \end{pmatrix} \overset{\text{IID}}{\sim} F(\cdot, \cdot). \tag{3}$$

Assumption 1 states that the impact factors and residual returns are IID across different assets. Independence across assets is justifiable because we are modeling residual returns. The assumption of identical distributions allows for mathematical tractability and implies that the relationship between X_{it} and θ_{it} is homogeneous. In our main article, we employ this IID assumption to derive concise theoretical results and useful mathematical intuition. However, the IID assumption can be relaxed to model cross-sectional heterogeneity (such as industry effects) by mixtures of homogeneous subgroups. We relax this assumption in Online Appendix EC.1.3.2. Moreover, because the majority of this article discusses the static case, we omit the subscript t for notational simplicity except when considering a dynamic model in Online Appendix EC.1.2.3.

The bivariate vectors, $(X_i, \theta_i)^\top, i = 1, 2, \ldots, N$, are ranked according to their first components (e.g., ESG scores), X_i:

$$\begin{pmatrix} X_{1:N} \\ \theta_{[1:N]} \end{pmatrix}, \begin{pmatrix} X_{2:N} \\ \theta_{[2:N]} \end{pmatrix}, \ldots, \begin{pmatrix} X_{N:N} \\ \theta_{[N:N]} \end{pmatrix}, \tag{4}$$

where $X_{1:N} \le X_{2:N} \le \cdots \le X_{N:N}$ are the order statistics of $\boldsymbol{X}$. The notation $\theta_{[i:N]}$ represents the i-th *induced order statistic* of $\boldsymbol{\theta}$ induced by another variable, $\boldsymbol{X}$. Hereinafter, we use the notation

$$\boldsymbol{\theta}_{[X]} = (\theta_{[1:N]}, \theta_{[2:N]}, \ldots, \theta_{[N:N]})^\top \tag{5}$$

to denote the vector of induced order statistics of $\boldsymbol{\theta}$ ranked by values of $\boldsymbol{X}$.

The framework in (1)–(4) is different from the impact investing model proposed in Lo and Zhang (2023) in several important ways. First, we generalize Lo and Zhang (2023) by considering the interaction between the impact factor and residual returns, $\boldsymbol{\theta}$, instead of the alpha as in Lo and Zhang (2023). This is also consistent with the framework of the fundamental law of active management in Grinold (1989). Second, Lo and Zhang (2023) consider the case where $(X_i, \theta_i)^\top$ follows a bivariate normal distribution, which implies that the dependence between the impact factor and asset return is described by a single correlation coefficient. Here, we consider a much richer set of dependence structures modeled by copulas (see Section 4). Third, the general bivariate distribution F in (3) allows for the analysis of different marginal distributions of X_i and θ_i. In particular, asset returns in practice include important deviations from normality such as nonzero skewness and heavy-tailedness which have important implications for impact portfolio construction, as shown in Section 5. Our empirical results in Online Appendix EC.2 demonstrate that allowing for general marginal distributions and dependence structures between X_i and θ_i can greatly enhance the performance of impact portfolios.

2.2. Impact Portfolio Construction

Impact investing essentially involves selecting securities based on the impact factor, $\boldsymbol{X}$. We consider a portfolio based on the N assets, and $\boldsymbol{w} = (w_1, w_2, \ldots, w_N)^\top$ represents the vector of asset weights for the N assets ranked by $\boldsymbol{X}$. In other words, we hold a fraction of our portfolio, w_i, in the asset with the i-th smallest impact factor, $X_{i:N}$. The residual return of the portfolio is therefore

$$\theta_p \equiv w_1 \theta_{[1:N]} + w_2 \theta_{[2:N]} + \cdots + w_N \theta_{[N:N]} = \boldsymbol{w}^\top \boldsymbol{\theta}_{[X]}. \tag{6}$$

Equation (6) shows that the residual return of the portfolio, θ_p, is determined by the distribution of $\boldsymbol{\theta}_{[X]}$ and the choice of $\boldsymbol{w}$. In addition, impact investors use the distribution of $\boldsymbol{\theta}_{[X]}$ to determine the portfolio weights, $\boldsymbol{w}$. Therefore, our goal is to characterize the distribution of $\boldsymbol{\theta}_{[X]}$, given the joint distribution F in Assumption 1.

To quantify the performance of the impact portfolio, let $\boldsymbol{\mu}$ and Σ be the expectation and covariance matrix of the residual returns of the N impact-ranked assets:

$$\boldsymbol{\mu} \equiv \mathbb{E}(\boldsymbol{\theta}_{[X]}), \quad \Sigma \equiv \mathrm{Cov}(\boldsymbol{\theta}_{[X]}). \tag{7}$$

Then, the expectation and variance of the portfolio's residual return are

$$\mathbb{E}(\theta_p) = \boldsymbol{w}^\top \boldsymbol{\mu}, \quad \mathrm{Var}(\theta_p) = \boldsymbol{w}^\top \Sigma \boldsymbol{w}. \tag{8}$$

The optimal portfolio weights and their corresponding performance metrics can be derived from standard portfolio theory. For completeness, we explicitly provide these constructions based on the moments of residual returns given by (7), which serves as the foundation for our results in subsequent sections. We consider impact investors who use either the information ratio or the mean-variance utility as the objective function to form portfolios. The following proposition from standard portfolio theory characterizes the optimal portfolios with these objectives.

Proposition 1 (Optimal Portfolio). *Under the multifactor model of (1), if investors construct portfolios based on N assets with frictionless borrowing and lending at the risk-free rate, and if they maximize the information ratio, $\mathrm{IR} = \mathbb{E}(\theta_p)/\sqrt{\mathrm{Var}(\theta_p)}$, or the mean-variance utility, $\mathbb{E}(\theta_p) - 0.5\lambda \mathrm{Var}(\theta_p)$, with a constant risk-aversion* *parameter $\lambda > 0$, the optimal portfolio weights and the optimal information ratio will be given by*

$$\boldsymbol{w}^* \propto \Sigma^{-1}\boldsymbol{\mu}, \quad \textit{and} \quad \mathrm{IR}^* = \sqrt{\boldsymbol{\mu}^\top \Sigma^{-1}\boldsymbol{\mu}}, \tag{9}$$

where $\boldsymbol{\mu}$ and Σ are given by (7).

Proposition 1 characterizes the optimal portfolio weights. This is useful for both relative weighting strategies, when companies are reweighted based on the rank of a certain measure,[11] and exclusionary investing, when filters are imposed to exclude certain companies from the investable universe. We discuss the trade-off between performance and impact for exclusionary investing in more detail in Online Appendix EC.1.2.6, and provide an extension to accommodate investors who care about both active and passive returns in Online Appendix EC.1.1. In this article, without loss of generality, we set the proportional constant to be one, that is, $\boldsymbol{w}^* = \Sigma^{-1}\boldsymbol{\mu}$.

One advantage of our framework is that one needs to estimate a much smaller number of parameters in portfolio construction. In Markowitz portfolio theory, one needs to estimate $N(N+1)/2$ parameters for the covariance matrix (Sharpe 1963). In comparison, our framework circumvents this issue because both $\boldsymbol{\mu}$ and Σ depend solely on the bivariate distribution F defined in (3). With certain distributional assumptions for F, one only needs to estimate a small number of parameters of F. This leads to more robust impact portfolios in practice, as demonstrated empirically by Lo et al. (2022).

In this article, we focus on the residual return of the portfolio rather than its raw return for the following reasons. First, institutional investors usually assess a portfolio's *relative* performance in comparison with benchmark portfolios (such as an index fund) or baseline asset-pricing models (such as the Fama-French factor models), rather than a portfolio's raw returns. This is the typical target of interest for ESG and sustainable investment products.[12] Second, more broadly, the literature on active investing also focuses on the residual returns to derive active investment performance metrics such as the information ratio.[13] Our choice follows both strands of the literature. Third, although residual returns of a portfolio are not directly realizable by holding individual assets, they are realizable when using index funds and factor exchange-traded funds to neutralize a portfolio's exposure to existing asset-pricing factors. Investors can realize the excess returns in this case.

3. The Distribution of Induced Order Statistics

The key to constructing impact portfolios using Proposition 1 lies in the characterization of the distribution of induced order statistics, $\boldsymbol{\theta}_{[X]}$, and, in particular, their first and second moments. In this section, we employ copulas to derive a representation theorem for the distribution of induced order statistics as a mixture of order statistics of uniform random variables and uniform random noise. The mixture is determined by the dependence structure between residual returns and impact factors, which is characterized by copulas, as well as the marginal distribution of residual returns. We present an exact finite-sample representation theorem in Section 3.1, and provide an asymptotic version in Section 3.2.

3.1. The Representation Theorem

Sklar's (1959) well-known theorem decomposes the joint distribution of an arbitrary pair of random variables—$(X_i, \theta_i)^\top$ in our case—into a *copula* function that characterizes their dependence structure, and two marginal distributions. To be specific, we denote the marginal distribution functions of X_i and θ_i by F_X and F_θ, respectively. Sklar's theorem states that there exists a copula $C(\cdot,\cdot)$ such that the bivariate distribution function of $(X_i, \theta_i)^\top$, F, can be represented as

$$F(x,y) = C(F_X(x), F_\theta(y)), \quad \forall x, y \in \mathbb{R}. \tag{10}$$

If F_X and F_θ are continuous, C is unique; otherwise, C is uniquely determined on $\overline{\mathcal{R}}_X \times \overline{\mathcal{R}}_\theta$, where $\overline{\mathcal{R}}_X$ and $\overline{\mathcal{R}}_\theta$ are the closures of the ranges of F_X and F_θ, respectively. In this article, we refer to C as a copula of F when it satisfies (10).

Next, we present our main representation theorem that explicitly shows the relationship between the distribution of $\boldsymbol{\theta}_{[X]}$ and the three components of F—F_X, F_θ, and C—under a general F.

3.1.1. Representation Theorem Under Smoothness Condition.

We first present the key result under a smoothness condition of F to build intuition, and then provide the general representation theorem without any smoothness assumptions.

Theorem 1 (The Representation Theorem). *Under* Assumption 1, *if both F_X and F_θ are continuous functions, and C is a copula of F with a density, we have*

$$\begin{aligned}&(\theta_{[1:N]}, \theta_{[2:N]}, \ldots, \theta_{[N:N]})\\ &\overset{d}{=} (g(U_{1:N}, V_1), g(U_{2:N}, V_2), \ldots, g(U_{N:N}, V_N)),\end{aligned} \tag{11}$$

where "$\overset{d}{=}$" denotes equality in distribution. Function g is defined as

$$g(u,w) \equiv F_\theta^{-1} \circ g_u(w), \tag{12}$$

where $g_u(w)$ is the inverse function of $v \longmapsto \frac{\partial C}{\partial u}(u,v)$,[14] *and "$\circ$" represents function composition. In addition, $U_1, U_2 \ldots, U_N$ and $V_1, V_2 \ldots, V_N$ are mutually independent and IID random variables, each following* Uniform(0,1), *and $U_{1:N} \le U_{2:N} \le \cdots \le U_{N:N}$ are the order statistics of $U_1, U_2 \ldots, U_N$.*

Theorem 1 provides an explicit representation for the distribution of induced order statistics, $\boldsymbol{\theta}_{[X]}$, by a mixture of order statistics, $U_{i:N}$, and IID uniform noise, V_i, for any smooth bivariate distribution, F. The mixture function given by (12), $g(\cdot,\cdot)$, is constructed using F_θ and C but does not depend on F_X. This theorem explicitly establishes a relationship between the distribution of $\boldsymbol{\theta}_{[X]}$ and the joint distribution of $\boldsymbol{X}$ and $\boldsymbol{\theta}$.

The representation given by Theorem 1 requires the smoothness of F, specifically the continuity of F_X and F_θ, which are assumptions also adopted by prior studies (Lo and Zhang 2023), and the differentiability of C, which are satisfied by many commonly used copulas. However, from both the mathematical and practical perspectives, we are still interested in whether this representation holds for a more general F.

3.1.2. Representation Theorem for General Distributions.

To extend the representation theorem to general distributions F, we require a few technical results. First, when F_θ is discontinuous, we need to specify the definition of the inverse function, $F_\theta^{-1}(\cdot)$, in the mixture function (12). Second, when C is nondifferentiable, the function $g_u(\cdot)$ in (12), which is the inverse of $v \longmapsto \frac{\partial C}{\partial u}(u,v)$, is not well defined. We adopt the concept of modified partial Dini derivative first proposed by Fang et al. (2020), instead of the traditional definition of derivatives, to address this issue (see Online Appendix EC.1.3.1 for further details).

Finally, when F_X is discontinuous, we demonstrate that the representation holds only when the copula of F is "linearly interpolating" in an undetermined region of the copula. To be precise, according to Sklar's theorem, the copula of F satisfying (10) is uniquely determined on $\overline{\mathcal{R}}_X \times \overline{\mathcal{R}}_\theta$. For the undetermined regions, we define a notion of linear interpolation as follows.

Definition 1 (Linearly Interpolating Copula). Let $\overline{\mathcal{R}}_X^c = [0,1]\backslash\overline{\mathcal{R}}_X$. A copula $C(u, v)$ is linearly interpolating on $\overline{\mathcal{R}}_X^c \times \overline{\mathcal{R}}_\theta$ with respect to u if

$$C(u,v) = \frac{F_X(d) - u}{F_X(d) - F_X(d^-)} C(F_X(d^-), v) + \frac{u - F_X(d^-)}{F_X(d) - F_X(d^-)} C(F_X(d), v) \tag{13}$$

holds for any $(u,v) \in (F_X(d^-), F_X(d)) \times \overline{\mathcal{R}}_\theta$ and $d \in \Delta_X$, where Δ_X is the set of all discontinuity points of F_X, and $F_X(d^-) = \lim_{x \to d^-} F_X(x)$.[15]

In Definition 1, we say a copula is linearly interpolating on $\overline{\mathcal{R}}_X^c \times \overline{\mathcal{R}}_\theta$ with respect to u when it is a linear function of $u \in \overline{\mathcal{R}}_X^c$ given any value of $v \in \overline{\mathcal{R}}_\theta$. Online Appendix EC.1.3.1 provides the intuition behind this concept.

With these technical results in hand, we can proceed to the representation theorem of the distribution of $\boldsymbol{\theta}_{[X]}$ for a very general joint distribution function, F.

Theorem 2 (Representation Theorem for General F). *Under* Assumption 1, *we have*

$$\begin{aligned}&(\theta_{[1:N]}, \theta_{[2:N]}, \ldots, \theta_{[N:N]}) \\ &\overset{d}{=} (g(U_{1:N}, V_1), g(U_{2:N}, V_2), \ldots, g(U_{N:N}, V_N))\end{aligned} \tag{14}$$

for any $N \geq 1$ ***if and only if*** *the copula $C(u, v)$ is linearly interpolating on $\overline{\mathcal{R}}_X^c \times \overline{\mathcal{R}}_\theta$ with respect to u. Here, the function g is defined as*

$$g(u,w) \equiv F_\theta^{-1} \circ g_u(w), \tag{15}$$

where $g_u(w)$ is the inverse function of $v \longmapsto \mathscr{D}_1 C(u,v)$, and $\mathscr{D}_1 C(u,v)$ is defined by (EC.30) *in Online Appendix EC.1.3.1. Other notation is defined as in Theorem 1.*

Theorem 2 gives a necessary and sufficient condition for the representation—the copula, $C(u, v)$, must be linearly interpolating on $\overline{\mathcal{R}}_X^c \times \overline{\mathcal{R}}_\theta$ with respect to u. Proposition EC.9 in Online Appendix EC.1.3.1 provides details of the existence and uniqueness of this linearly interpolating copula, as well as how to construct it. Note that, when F_X is continuous, we have $\overline{\mathcal{R}}_X = [0,1]$ and $\overline{\mathcal{R}}_X^c = \emptyset$. Therefore, Theorem 1 can be viewed as a special case of Theorem 2 where no interpolation is needed for the copula on $\overline{\mathcal{R}}_X^c \times \overline{\mathcal{R}}_\theta$.

3.1.3. Implications.

Representation Theorems 1 and 2 have several important implications.

First, they demonstrate that the distribution of induced order statistics, $\boldsymbol{\theta}_{[X]}$, can always be represented as a mixture of order statistics, $U_{i:N}$, and IID uniformly distributed random variables, V_i. In addition, the mixture function, g, is explicitly given by (15). This not only provides insight into the source of the randomness of $\boldsymbol{\theta}_{[X]}$ but also allows us to analyze the distribution of induced order statistics using the many established tools for order statistics.

Second, the representation theorems allow us to calculate the moments of $\boldsymbol{\theta}_{[X]}$, which is required for constructing optimal impact portfolios. By specifying the dependence structure between $\boldsymbol{X}$ and $\boldsymbol{\theta}$, along with their marginal distributions, we can construct a linearly interpolating copula and then use Theorem 2 to efficiently compute the moments of $\boldsymbol{\theta}_{[X]}$ through numerical integration. In particular, the first two moments of $\boldsymbol{\theta}_{[X]}$ are given by the following proposition, which shows that the computational complexity of computing the first two moments of one asset (pair) does not scale with the number of assets N:

Proposition 2. *Under the assumptions of Theorem 2, if the copula $C(u, v)$ is linearly interpolating on $\overline{\mathcal{R}}_X^c \times \overline{\mathcal{R}}_\theta$ with respect to u, we have*

$$\mathbb{E}(\theta_{[i:N]}) = \int_0^1 \int_0^1 g(u,w) \cdot \frac{N! u^{i-1}(1-u)^{N-i}}{(i-1)!(N-i)!} \mathrm{d}u\mathrm{d}w,$$

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

$$\mathbb{E}(\theta_{[i:N]}\theta_{[j:N]}) = \int_0^1 \int_0^1 \int_0^1 \int_u^1 g(u,w) \cdot g(z,s) \cdot \frac{N! u^{i-1}(z-u)^{j-i-1}(1-z)^{N-j}}{(i-1)!(j-i-1)!(N-j)!} \mathrm{d}z\mathrm{d}u\mathrm{d}w\mathrm{d}s,$$

for $i, j = 1, 2, \ldots, N$ and $i < j$, where function g is defined by (15).

Third, the representation theorems have a broad range of potential practical applications because of their ability to characterize the distribution of $\boldsymbol{\theta}_{[X]}$ under any bivariate distribution between the impact factors and the residual returns. For example, we allow the impact factors, $\boldsymbol{X}$, to take discrete rating values.[16] The residual returns, $\boldsymbol{\theta}$, can also be discontinuous, which can describe prices with jumps, daily price limits, and high-frequency data with discrete price levels.[17] We also allow nondifferentiable dependence structures, C, between $\boldsymbol{X}$ and $\boldsymbol{\theta}$.[18]

Fourth, the representation theorems assert that the distribution of $\boldsymbol{\theta}_{[X]}$ is typically determined by the dependence structure, that is, the copula C, and the marginal distribution of residual returns, $\boldsymbol{\theta}$, but *not* the marginal distribution of the impact factors, $\boldsymbol{X}$. This is because the mixture function, g, given by (15), depends solely on C and F_θ. The specific form of F_X does not affect the representation, and only the discontinuity points of F_X are necessary to determine the interpolation of C (see Definition 1 in Online Appendix EC.1.3). This implies that impact investors who follow Proposition 1 to construct optimal impact portfolios should only focus on the *rank* of impact factors, rather than their specific distribution. It is well known that considerable noise exists in impact measurements such as the ESG score (Berg et al. 2022). The fact that portfolio weights depend only on the *rank* suggests that our methodology is more robust against noise and outliers in impact measurements.

We generalize the representation theorems to allow for cross-sectional heterogeneity in Online Appendix EC.1.3.2 and highlight the similarities and differences in terms of the implications above. We also use the generalized representation theorem to model markets with a mixture of heterogeneous subgroups and show how that affects the optimal impact portfolio.

3.2. Asymptotic Version

Theorems 1 and 2 characterize the exact finite-sample distribution of $\boldsymbol{\theta}_{[X]}$ for N assets. In practice, the universe of investable assets can be very large, in which case we can approximate the distribution of $\boldsymbol{\theta}_{[X]}$ as N increases without bound. As in Section 3.1, we first present the results for a smooth F and then the general version.

Theorem 3. *Under Assumption 1, assume that both F_X and F_θ are continuous functions, C is a copula of F with a density, and $\frac{\partial C}{\partial u}(u,v)$ is a continuous function of u. For any fixed $m \geq 1$, any constants $\xi_1 < \xi_2 < \cdots < \xi_m$, and any sequence $1 < i_1(N) < i_2(N) < \cdots < i_m(N) < N$ such that $i_k(N)/N \to \xi_k \in (0,1)$ for $k = 1, 2, \ldots, m$ as $N \to +\infty$, we have*

$$(\theta_{[i_1(N):N]}, \theta_{[i_2(N):N]}, \ldots, \theta_{[i_m(N):N]}) \xrightarrow{d} (g(\xi_1, V_1), g(\xi_2, V_2), \ldots, g(\xi_m, V_m)). \quad (16)$$

Here, "$\xrightarrow{d}$" denotes convergence in distribution, and random variables V_i and function g are defined as in Theorem 1.

Theorem 3 is the asymptotic counterpart of Theorem 1. We follow both the theory of order statistics (David and Nagaraja 2003) and that of impact investing (Lo and Zhang 2023) to consider asymptotic convergence where the number of investable assets, N, grows, and assets $i_1(N), i_2(N), \ldots, i_m(N)$ converge to assets ranking at $100\xi_1, 100\xi_2, \ldots, 100\xi_m$ percentiles among all assets in the universe. Practitioners who invest in a large number of assets or target assets ranking at specific percentiles can use this theorem to guide their impact investing strategies.

The following result provides the asymptotic version of the representation theorem for general F.

Theorem 4. *Under Assumption 1, let C be a copula of F and $\mathfrak{D}_1 C(u,v)$ be a continuous function of u. We have*

$$(\theta_{[i_1(N):N]}, \theta_{[i_2(N):N]}, \ldots, \theta_{[i_m(N):N]}) \xrightarrow{d} (g(\xi_1, V_1), g(\xi_2, V_2), \ldots, g(\xi_m, V_m)) \quad (17)$$

*for any fixed $m \geq 1$, any constants $\xi_1 < \xi_2 < \cdots < \xi_m$, and any sequence $1 < i_1(N) < i_2(N) < \cdots < i_m(N) < N$ such that $i_k(N)/N \to \xi_k \in (0,1)$ for $k = 1, 2, \ldots, m$ as $N \to +\infty$, **if and only if** the copula $C(u, v)$ is linearly interpolating on $\overline{\mathcal{R}}_X^c \times \overline{\mathcal{R}}_\theta$ with respect to u. Here, "$\xrightarrow{d}$" denotes convergence in distribution, and random variables V_i and function g are defined as in Theorem 2.*

Compared with Theorems 1 and 2, the asymptotic version has the following implications.

First, similar to Theorem 2, the linearly interpolating copula is necessary and sufficient for the asymptotic representation. One can easily construct the linearly interpolating copula and calculate the moments of the asymptotic distribution numerically using Theorem 4.

Second, unlike the exact finite-sample version, the asymptotic version reveals that the asymptotic distribution of induced order statistics is no longer a mixture of order statistics and random noise. Instead, it becomes a mixture of asset quantiles, ξ_i, and uniform noise, V_i. The mixture function, g, remains the same as in the finite-sample version and is determined by the marginal distribution of $\boldsymbol{\theta}$ and the dependence structure, C.

Third, because ξ_i in (17) are constants, the randomness of the asymptotic distribution of $(\theta_{[i_1(N):N]}, \theta_{[i_2(N):N]}, \ldots, \theta_{[i_m(N):N]})$ arises solely from V_i. Therefore, the induced order statistics $\theta_{[i_1(N):N]}, \theta_{[i_2(N):N]}, \ldots, \theta_{[i_m(N):N]}$ are asymptotically mutually independent. This highlights an advantage of our framework compared with the standard mean-variance portfolio construction—the estimation of the inverse of Σ is more stable when the number of assets N is large. Proposition 1 shows that the optimal weight depends on Σ^{-1}, which can be estimated stably in our framework because the asymptotic independence of the induced order statistics implies that Σ is approximately diagonal.

Finally, as with the finite-sample version, the asymptotic version also shows that the asymptotic distribution of $\boldsymbol{\theta}_{[X]}$ is typically determined by C and F_θ, but not F_X. This provides a road map for investigating the distribution of induced order statistics, and therefore the optimal impact portfolios. In Sections 4 and 5, we study how different copulas and marginal distributions of $\boldsymbol{\theta}$ affect the optimal impact portfolio weights and their corresponding returns, respectively. In Online Appendix EC.1.2, we discuss the special case where $(X_i, \theta_i)^\top$ is jointly normally distributed. Most of the results in this article are derived based on the powerful representation theorems, Theorems 1 and 4.

4. General Dependence via Copulas

In this section, we consider general dependence structures between $\boldsymbol{X}$ and $\boldsymbol{\theta}$. In particular, based on our representation theorems, we demonstrate how the moments of $\boldsymbol{\theta}_{[X]}$ depend on the characteristics of the copula. We focus on specific copulas, including the Gaussian and Archimedean copula (which encompasses the Clayton copula and Gumbel copula), to model tail dependence. The fundamental copulas studied by Yang et al. (2006) and the elliptical copula are discussed in Online Appendices EC.1.4.3 and EC.1.4.4, respectively.

4.1. Gaussian Copula

The Gaussian copula, which is constructed from multivariate normal distributions, is one of the most widely used copulas in the literature (Cont and Kan 2011).

Definition 2 (Gaussian Copula). The bivariate Gaussian copula with parameter $\rho \in (-1, 1)$ is defined as

$$C_\rho^{\mathrm{Ga}}(u, v) \equiv \boldsymbol{\Phi}_\rho(\Phi^{-1}(u), \Phi^{-1}(v)), \quad u, v \in [0, 1],$$

where Φ is the distribution function of $\mathcal{N}(0, 1)$, and $\boldsymbol{\Phi}_\rho$ is the bivariate distribution function of a bivariate normal distribution with mean 0, variance 1, and correlation ρ.

The parameter ρ captures the degree of dependence between the impact factor and the residual return.

The following result characterizes the distribution of $\boldsymbol{\theta}_{[X]}$ under the Gaussian copula using Theorems 1 and 2.

Proposition 3. *Under Assumption 1, if F_X is continuous and the copula of F is a Gaussian copula with parameter $\rho \in (-1, 1)$, we have*

$$\begin{aligned}(\theta_{[1:N]}, \ldots, \theta_{[N:N]}) &\\ \stackrel{d}{=} \Big(&F_\theta^{-1} \circ \Phi(\rho Y_{1:N} + \sqrt{1-\rho^2} Z_1), \ldots, \\ &F_\theta^{-1} \circ \Phi(\rho Y_{N:N} + \sqrt{1-\rho^2} Z_N)\Big),\end{aligned} \tag{18}$$

where $Z_1, Z_2, \ldots, Z_N \overset{\mathrm{IID}}{\sim} \mathcal{N}(0, 1)$, $Y_{1:N} \le Y_{2:N} \le \cdots \le Y_{N:N}$ are the order statistics of $Y_1, Y_2, \ldots, Y_N \overset{\mathrm{IID}}{\sim} \mathcal{N}(0, 1)$, and random variables $\{Y_i\}_{i=1}^N$ and $\{Z_i\}_{i=1}^N$ are mutually independent.

Proposition 3 shows that, under the Gaussian copula, the distribution of $\theta_{[i:N]}$ is determined by $\rho Y_{i:N} + \sqrt{1-\rho^2} Z_i$, which is a *linear* mixture of order statistics of standard normal random variables and IID standard normal noise. Proposition 3 also implies that the distribution of $\theta_{[i:N]}$ is linked to $\rho Y_{i:N} + \sqrt{1-\rho^2} Z_i$ via a special function:

$$Q(x) \equiv F_\theta^{-1} \circ \Phi(x), \quad x \in \mathbb{R}. \tag{19}$$

In fact, $Q(x)$ is precisely the quantile-quantile plot (Q-Q plot) of F_θ versus the standard normal distribution, a widely used statistical tool. Financial practitioners often use the Q-Q plot to visualize the heaviness of the tails of random variables. In particular, if $\boldsymbol{\theta}$ is normally distributed with mean μ_θ and variance σ_θ^2, we have $F_\theta(y) = \Phi((y - \mu_\theta)/\sigma_\theta)$, and $Q(x) = F_\theta^{-1} \circ \Phi(x) = \mu_\theta + \sigma_\theta x$, which is a linear Q-Q plot. In this case, it is straightforward to verify that Proposition 3 reduces to the case of a bivariate normal distribution (see Proposition EC.4 in Online Appendix EC.1.2).

The following result gives the asymptotic distribution of $\boldsymbol{\theta}_{[X]}$ under the Gaussian copula, which is a corollary of the asymptotic version of Theorems 3 and 4.

Proposition 4. *Under the assumptions of Proposition 3, for any fixed $m \ge 1$, any constants $\xi_1 < \xi_2 < \cdots < \xi_m$, and any sequence $1 < i_1(N) < i_2(N) < \cdots < i_m(N) < N$ such that $i_k(N)/N \to \xi_k \in (0, 1)$ for $k = 1, 2, \ldots, m$ as $N \to +\infty$, we have*

$$\begin{aligned}(\theta_{[i_1(N):N]}, \ldots, \theta_{[i_m(N):N]}) &\\ \xrightarrow{d} \Big(&F_\theta^{-1} \circ \Phi\Big(\rho\Phi^{-1}(\xi_1) + \sqrt{1-\rho^2} Z_1\Big), \ldots, \\ &F_\theta^{-1} \circ \Phi\Big(\rho\Phi^{-1}(\xi_m) + \sqrt{1-\rho^2} Z_m\Big)\Big),\end{aligned}$$

where the notation follows that in Proposition 3.

By combining Proposition 4 and the optimal weights given by (9), the optimal weight for assets ranked at quantile ξ under the Gaussian copula can be approximated by

$$w^{\mathrm{Ga}}(\xi) \equiv \frac{\mathbb{E}\left[F_\theta^{-1} \circ \Phi\left(\rho\Phi^{-1}(\xi) + \sqrt{1-\rho^2}Z\right)\right]}{\mathrm{Var}\left[F_\theta^{-1} \circ \Phi\left(\rho\Phi^{-1}(\xi) + \sqrt{1-\rho^2}Z\right)\right]}, \quad \xi \in (0,1), \tag{20}$$

where Z is a standard normal random variable. There is no need to calculate the covariances among induced order statistics because $\theta_{[i_1(N):N]}, \theta_{[i_2(N):N]}, \ldots, \theta_{[i_m(N):N]}$ are asymptotically mutually independent. Investors can use (20), which can be easily computed numerically, to approximate the optimal weights when the total number of assets, N, is large.

4.2. Archimedean Copula

Although the Gaussian copula is widely used, it has been documented that the tail dependencies between asset returns and impact factors for both tails are asymmetric in many contexts applicable to impact investing (Bax et al. 2023), a feature that cannot be captured by the Gaussian copula. Therefore, in this section, we construct optimal portfolios under another important family of copulas—the Archimedean copula family—which has been proven to be useful for modeling tail dependence in financial data (McNeil et al. 2015).

Definition 3 (Archimedean Copula). The Archimedean copula with generator function ψ is defined as

$$C_\psi^{\mathrm{Ar}}(u,v) \equiv \phi(\psi(u) + \psi(v)), \quad u,v \in [0,1],$$

where the generator function $\psi : [0,1] \to [0,+\infty]$ is a continuous, strictly decreasing, and strictly convex function such that $\psi(1) = 0$, and its inverse $\phi \equiv \psi^{-1}$ is defined on $[0,+\infty]$.[19]

The dependence structure of an Archimedean copula is determined by its generator function, ψ, instead of the single parameter ρ of the Gaussian copula. This allows the Archimedean copula to capture a rich set of dependence structures.[20] The following proposition characterizes the distribution of $\boldsymbol{\theta}_{[X]}$ under the Archimedean copula using Theorems 1 and 2. The corresponding asymptotic result can also be derived using Theorems 3 and 4, which we omit because of space constraints.

Proposition 5. *Under Assumption 1, if F_X is continuous and the copula of F is an Archimedean copula with generator function ψ (let $\phi = \psi^{-1}$), and if the first- and second-order derivatives of ϕ, ϕ' and ϕ'', both exist on $[0,+\infty)$, we have*

$$\begin{pmatrix}\theta_{[1:N]}\\ \theta_{[2:N]}\\ \vdots \\ \theta_{[N:N]}\end{pmatrix} \stackrel{d}{=} \begin{pmatrix} F_\theta^{-1} \circ \phi[\phi'-1[\phi' \circ \phi^{-1}(U_{1:N}) \cdot V_1] - \phi^{-1}(U_{1:N})] \\ F_\theta^{-1} \circ \phi[\phi'-1[\phi' \circ \phi^{-1}(U_{2:N}) \cdot V_2] - \phi^{-1}(U_{2:N})] \\ \vdots \\ F_\theta^{-1} \circ \phi[\phi'-1[\phi' \circ \phi^{-1}(U_{N:N}) \cdot V_N] - \phi^{-1}(U_{N:N})]\end{pmatrix},$$

where the notation is the same as in Theorem 1.

Given an Archimedean copula with any generator function ψ, we can use Proposition 5 to derive the distribution of $\boldsymbol{\theta}_{[X]}$ explicitly. Like the result for the Gaussian copula, Proposition 5 demonstrates that the distribution of $\boldsymbol{\theta}_{[X]}$ under the Archimedean copula is also a mixture of order statistics, $U_{i:N}$, and IID uniform noise, V_i, although the mixture is more complicated than the linear mixture for the Gaussian copula given by Proposition 3.

Archimedean copulas are widely used for characterizing various tail dependencies. To derive optimal impact portfolios when the dependence between the impact factors and residual returns is asymmetric in its tails, we consider two special cases of the Archimedean copula: the Clayton and Gumbel copulas. The Clayton copula characterizes lower tail dependence (i.e., small values of the residual returns, $\boldsymbol{\theta}$, generally appear together with small values of the impact factors, $\boldsymbol{X}$), and the Gumbel copula characterizes upper tail dependence (i.e., large values of $\boldsymbol{\theta}$ generally appear together with large values of $\boldsymbol{X}$).[21] One can expect that the two copulas yield opposite results. For brevity, we discuss the Clayton copula below and relegate the discussion of the Gumbel copula to Online Appendix EC.1.4.2.

4.3. Clayton Copula

The Clayton copula, which models lower tail dependence, is defined as follows:

Definition 4 (Clayton Copula). The Clayton copula with parameter $\gamma \in (0,+\infty)$ is defined as an Archimedean copula with generator function

$$\psi_\gamma^{\mathrm{Cl}}(u) \equiv \frac{u^{-\gamma}-1}{\gamma}, \quad u \in [0,1]. \tag{21}$$

The Clayton copula parameterizes the Archimedean copula generated by (21), and the parameter γ measures the strength of dependence between $\boldsymbol{X}$ and $\boldsymbol{\theta}$.

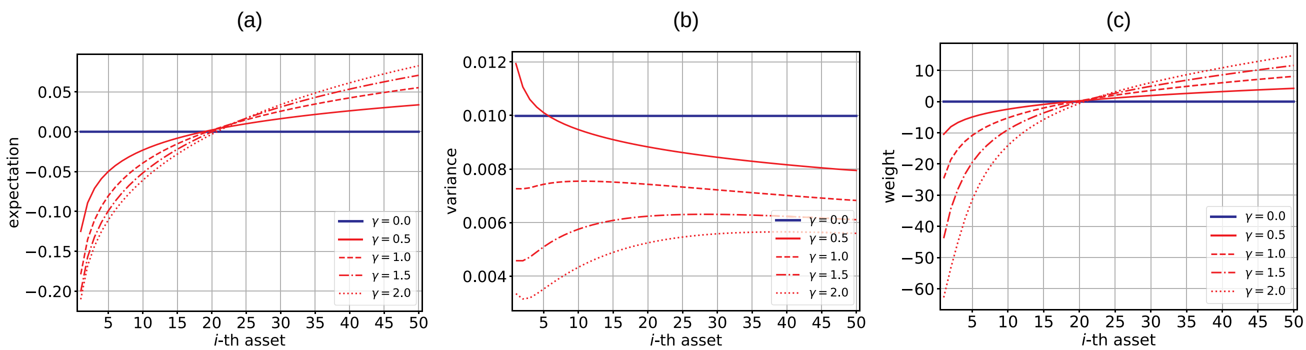
By replacing the generator function of the Archimedean copula in Proposition 5 by $\psi_\gamma^{\mathrm{Cl}}$, we can derive the distribution of its induced order statistics, $\boldsymbol{\theta}_{[X]}$ (for brevity, we write the i-th entry only):

$$\theta_{[i:N]} \stackrel{d}{=} F_\theta^{-1}[(1 + U_{i:N}^{-\gamma} V_i^{-\gamma/(\gamma+1)} - U_{i:N}^{-\gamma})^{-1/\gamma}]. \tag{22}$$

This result allows us to calculate the moments of $\boldsymbol{\theta}_{[X]}$ numerically. In addition, by taking the limits of (22) with respect to γ, we can show that the right-hand side of (22) converges to $F_\theta^{-1}(U_{i:N})$ as γ increases without bound, and converges to $F_\theta^{-1}(V_i)$ as γ approaches zero. Therefore, γ determines the relative importance of the order statistics, $U_{i:N}$, and the uniform noise, V_i, in the representation.

We use numerical examples to investigate the optimal impact portfolios under the Clayton copula using (22) and Proposition 2. Figure 1 displays the

Figure 1. (Color online) Expectations and Variances of $\boldsymbol{\theta}_{[X]}$, and Optimal Weights Assuming a Clayton Copula



Notes. The marginal distribution of $\boldsymbol{\theta}$ is $\mathcal{N}(0,\sigma_\theta^2)$ with $\sigma_\theta = 10\%$. We set $n = 50$ for illustrative purposes. (a) Expectations. (b) Variances. (c) Weights.

expectations and variances of the ranked residual returns, and the optimal weights (in terms of maximizing the information ratio given by (9)) under the Clayton copula. For illustrative purposes, we assume that $\boldsymbol{\theta}$ is normally distributed with zero mean and a 10% standard deviation, and the number of assets is $n = 50$.[22]

We observe that assets with higher impact factors have higher expectations (Figure 1(a)). The variances show different monotonicity patterns with respect to i (Figure 1(b)). Assets with higher impact factors generally have higher weights (Figure 1(c)). Meanwhile, as γ increases, the dispersion in expected residual returns across assets becomes larger (Figure 1(a)), the variances become generally smaller (Figure 1(b)), and thus the optimal weights are more dispersed across assets (Figure 1(c)).

It is worth noting that the distribution of residual returns and optimal weights shown in Figure 1 is not symmetric. In particular, the overall weights of long positions are lower than the magnitudes of weights of short positions, because of the lower expected residual returns for assets in long positions. This is consistent with the fact that the Clayton copula characterizes lower tail dependence, meaning that the dependence between the impact factor and residual returns will be stronger for assets with lower impact factors.

In summary, our findings for the Clayton copula shed light on the implications of tail dependence for impact portfolios. Generally speaking, stronger tail dependence implies larger weights for those assets that are ranked on that tail. This also reveals the importance of choosing proper dependence structures between the impact factors, $\boldsymbol{X}$, and the residual returns, $\boldsymbol{\theta}$, in practice when constructing impact portfolios.

5. General Return Distributions

In this section, we investigate how different marginal distributions of residual returns, $\boldsymbol{\theta}$, affect the construction of impact portfolios under Assumption 1. In particular, we focus on the influence of skewed and heavy-tailed residual returns. Online Appendix EC.1.5.2 provides additional discussions on the implications of symmetric distributions of $\boldsymbol{\theta}$.

5.1. Skewed Returns

It has been well documented that individual stock returns are skewed (Cont 2001, Jondeau and Rockinger 2003). In this section, we investigate the influence of a skewed $\boldsymbol{\theta}$ on the optimal weights. To derive the key result, we use the Gaussian copula and analyze the influence of a skewed $\boldsymbol{\theta}$ using the Q-Q plot, $Q(x)$, defined by (19).

We model a skewed distribution for $\boldsymbol{\theta}$ by considering a piecewise linear Q-Q plot:

$$Q_{a,b}^{\text{Skew}}(x) = (F_{\theta;a,b}^{\text{Skew}})^{-1} \circ \Phi(x) \equiv \begin{cases} ax, & x \geq 0, \\ bx, & x < 0, \end{cases} \tag{23}$$

where $a > 0$, $b > 0$, $F_{\theta;a,b}^{\text{Skew}}$ is the marginal distribution function of $\boldsymbol{\theta}$, and Φ is the distribution function of $\mathcal{N}(0,1)$. The parameters a and b control the skewness. Specifically, if $a > b$, the distribution is positively skewed (heavier right tails), whereas if $a < b$, the distribution is negatively skewed (heavier left tails). The case of $a = b$ corresponds to a normal distribution.

For $\boldsymbol{\theta}$ specified by (23), the following proposition characterizes the optimal portfolio weights at both tails.

Proposition 6. *Under Assumption 1, assume that the marginal distribution of $\boldsymbol{X}$ is continuous, the marginal distribution of $\boldsymbol{\theta}$ is given by (23), and the copula is Gaussian with parameter $\rho \in (-1,1)$. Then, $w^{\text{Ga}}(\xi)$ defined by (20) satisfies*

$$\lim_{\xi \to 1^-} \frac{w^{\text{Ga}}(\xi)}{\rho \Phi^{-1}(\xi)/[a(1-\rho^2)]} = 1,$$

$$\lim_{\xi \to 0^+} \frac{w^{\text{Ga}}(\xi)}{\rho \Phi^{-1}(\xi)/[b(1-\rho^2)]} = 1.$$

Proposition 6 demonstrates that the optimal weights on assets ranked at quantile ξ can be approximated by

$\rho\Phi^{-1}(\xi)/[a(1-\rho^2)]$ on the right tail ($\xi \to 1^-$), and by $\rho\Phi^{-1}(\xi)/[b(1-\rho^2)]$ on the left tail ($\xi \to 0^+$). This proves the key result that, under a skewed marginal distribution of residual returns, a larger positive skewness ($a > b$) results in smaller weights for top-ranking assets based on the impact factor ($|\rho\Phi^{-1}(\xi)|/[a(1-\rho^2)] < |\rho\Phi^{-1}(\xi)|/[b(1-\rho^2)]$), whereas a larger negative skewness ($a < b$) leads to smaller weights for bottom-ranking assets ($|\rho\Phi^{-1}(\xi)|/[a(1-\rho^2)] > |\rho\Phi^{-1}(\xi)|/[b(1-\rho^2)]$). In other words, positive skewness reduces the weights of top-ranked assets, and negative skewness reduces the weights of bottom-ranked assets. Online Appendix EC.1.5.1 provides additional numerical examples to further illustrate this effect.

5.2. Heavy-Tailed Returns

It is widely recognized that the empirical distribution of financial asset returns exhibits a higher peak and heavier tail compared with the normal distribution (Cont 2001, Kou and Peng 2016). Here we consider the influence of these characteristics of $\boldsymbol{\theta}$.

To derive the key result, we begin by considering a special distribution of $\boldsymbol{\theta}$ with heavy tails:

$$Q^{\text{HeavyTail}}_{\sigma,\tau,\beta}(x) = (F^{\text{HeavyTail}}_{\theta;\sigma,\tau,\beta})^{-1} \circ \Phi(x) \equiv \begin{cases} \sigma x + \tau |x|^{\beta}, & x \geq 0, \\ \sigma x - \tau |x|^{\beta}, & x < 0, \end{cases} \tag{24}$$

where $\sigma > 0$, $\tau > 0$, $\beta \geq 1$, $F^{\text{HeavyTail}}_{\theta;\sigma,\tau,\beta}$ is the marginal distribution function of $\boldsymbol{\theta}$, and Φ is the distribution function of $\mathcal{N}(0,1)$. Parameter β controls the heaviness of the tails, and a larger value of β implies a distribution with a higher peak and heavier tails. Parameters σ and τ control the balance between the linear term, x, and the heavy-tailedness term, $|x|^{\beta}$.

The following proposition characterizes the optimal portfolio weights when the residual returns, $\boldsymbol{\theta}$, have heavy tails, specifically for $\beta > 2$.

Proposition 7. *Under Assumption 1, assume that the marginal distribution of $\boldsymbol{X}$ is continuous, the marginal distribution of $\boldsymbol{\theta}$ is given by (24) with $\beta > 2$, and the copula is Gaussian with parameter $\rho \in (0,1)$. Then, $w^{\text{Ga}}(\xi)$ defined by (20) satisfies*

$$w^{\text{Ga}}(0.5) = 0; \qquad w^{\text{Ga}}(\xi) > 0,\ \xi \in (0.5,1); \qquad \lim_{\xi \to 1^-} w^{\text{Ga}}(\xi) = 0;$$
$$w^{\text{Ga}}(\xi) < 0,\ \xi \in (0,0.5); \qquad \lim_{\xi \to 0^+} w^{\text{Ga}}(\xi) = 0.$$

If $\rho \in (-1,0)$, the signs of the inequalities above are reversed.

Proposition 7 reveals an interesting result that, when the tails of the marginal distribution of residual returns are heavy enough ($\beta > 2$), the optimal weights are no longer monotonic. In particular, according to Proposition 7, assets with the most extreme values of impact factors (i.e., the highest or lowest) should have optimal weights of nearly zero. This explains why impact investors may not put the most money on assets with the highest or lowest impact factors, but prefer to invest more in nonextreme assets.

The intuition behind the nearly zero optimal weights when ξ approaches zero and one is that the variance of the impact returns increases without bound and, in particular, at a higher rate than the expectation. This is demonstrated in the proof of Proposition 7 in Online Appendix EC.3.5. In other words, the high risk from assets with extreme impact values leads to their low weights in the optimal allocation. This is also illustrated in the following numerical example based on the scaled-*t* distribution.

Definition 5 (Scaled-*t* Distribution). A random variable θ follows a scaled-*t* distribution with parameters $(\text{df}, \sigma_\theta)$ if $\theta \overset{d}{=} \sigma_\theta \cdot S/\sqrt{\text{Var}(S)}$, where $\sigma_\theta > 0$, and the random variable S follows Student's *t*-distribution with degree of freedom df. We denote this by $\theta \sim \text{Scaled-}t(\text{df}, \sigma_\theta)$.

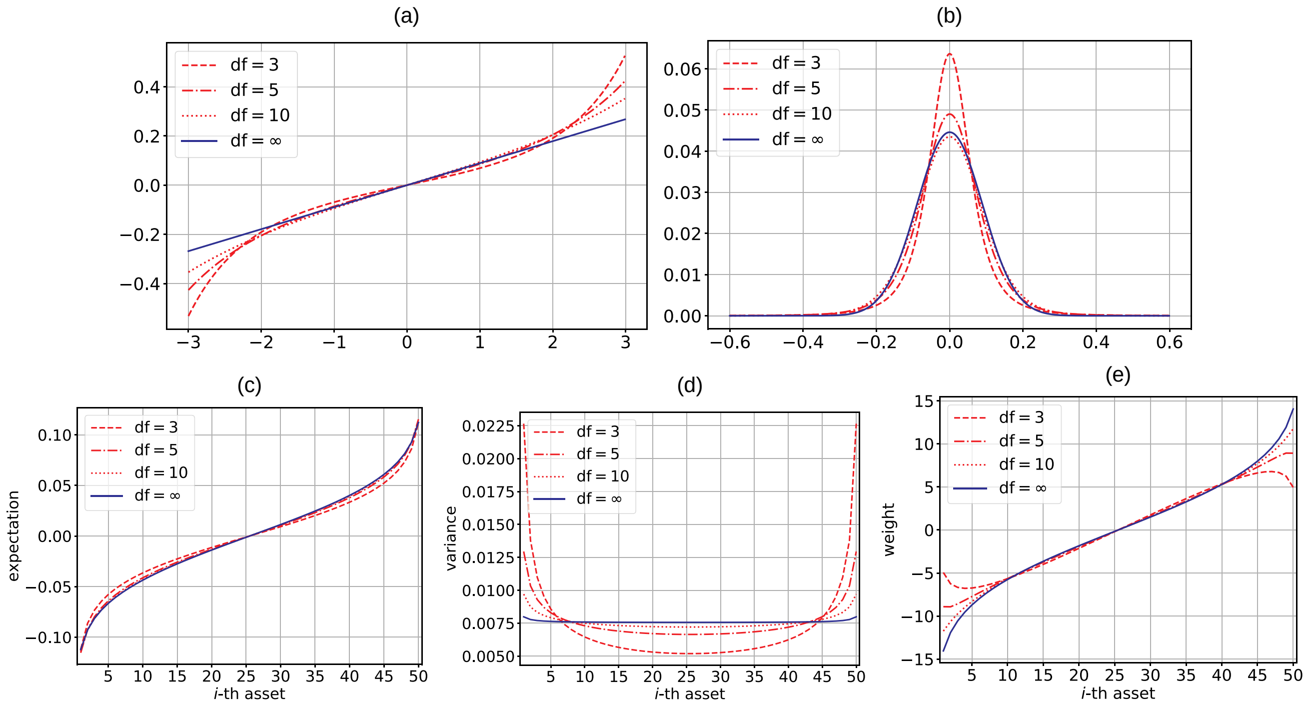
Figure 2, (a) and (b), shows the Q-Q plot and the density function of the scaled-*t* distribution, respectively. Figure 2(a) shows that the Q-Q plot deviates from a straight line as the degrees of freedom, df, decrease. Figure 2(b) implies that the degrees of freedom control the kurtosis of the distribution, with smaller values corresponding to higher peaks and heavier tails. In particular, as the degrees of freedom increase without bound, by definition, the scaled-*t* distribution converges to the normal distribution.

Figure 2, (c) and (d), displays the expectations and variances of the induced order statistics, $\boldsymbol{\theta}_{[X]}$, and Figure 2(e) contains the optimal weights in a market with $n = 50$, scaled-*t* distributed $\boldsymbol{\theta}$, and Gaussian copula with parameter $\rho = 50\%$. Figure 2(e) highlights that the optimal weights are no longer monotonic. When the distribution of $\boldsymbol{\theta}$ has a higher peak and a heavier tail, the optimal weights for assets with extreme impact factors are lower, which confirms the results given by Proposition 7. In addition, the decrease in the magnitudes of weights for assets ranked at the two extremes is driven by their extremely high variances (see Figure 2(d)). In other words, impact investors may not want to put the highest weights on assets with the highest or lowest impact factors because of their high risk.

6. Conclusion

In this article, we develop a framework for constructing optimal impact portfolios and analyzing their performance. Our results apply to any joint distribution of impact factors and residual returns, making them broadly applicable to a wide range of contexts. We develop significant extensions of the theory of induced order statistics, with which we are able to characterize the distribution of residual returns of individual assets ranked by the impact factor.

Figure 2. (Color online) Q-Q Plots, Density Functions, Expectations, Variances, and Weights



Notes. Panels (a) and (b) are the Q-Q plot and density function of scaled-$t(\mathrm{df}, \sigma_\theta)$. Panels (c)–(e) are expectations, variances of $\boldsymbol{\theta}_{[X]}$, and the optimal weights under Assumption 1. The copula is Gaussian with parameter ρ, and the marginal distribution of $\boldsymbol{\theta}$ is scaled-$t(\mathrm{df}, \sigma_\theta)$. We set $n = 50$, $\rho = 50\%$, and $\sigma_\theta = 10\%$ for illustrative purposes. (a) Q-Q plots. (b) Density functions. (c) Expectations. (d) Variances. (e) Weights.

By modeling the dependence between impact factors and residual returns using copulas, we derive representation theorems for the distribution of impact-ranked residual returns, both in finite samples and asymptotically. These representation theorems completely characterize the distribution as a mixture of order statistics and uniformly distributed random noise, and the mixture function is determined by the copula and the marginal distribution of residual returns, but does not depend on the marginal distribution of impact factors. Our representation theorems allow for discontinuous joint distributions, in which case we demonstrate that the representation holds if and only if the copula is linearly interpolating.

In practice, impact investing involves nonnormal dependence structures and nonnormal return distributions. In particular, we apply our representation theorems to Gaussian and Archimedean copulas and investigate the corresponding distribution of impact returns. We also study how tail dependence affects the construction of optimal impact portfolios, and find that stronger tail dependence implies larger weights for assets that are ranked on that tail in optimal impact portfolios.

We also demonstrate several important implications for the optimal impact portfolio weights when residual returns are skewed or exhibit heavy tails. Larger positive skewness leads to smaller weights for higher impact-ranked assets, and larger negative skewness leads to smaller weights for lower impact-ranked assets. In addition, because of the high risk on heavy tails, the optimal impact portfolio should not put the largest weights on assets with the highest or the lowest impact factors. Rigorous theoretical justifications are provided for these heuristics to construct impact portfolios.

Overall, our results completely characterize how to construct optimal impact portfolios for arbitrary joint distributions of impact factor and returns, and provide an efficient methodology to compute their portfolio weights numerically.

Our framework provides a toolkit for practitioners to construct impact portfolios and quantify their performance based on real data. As demonstrated in Online Appendix EC.2, impact investors can first estimate an appropriate copula and a specific marginal distribution for residual returns, and then use our analytical results to form optimal impact portfolios. This allows impact investors to achieve higher risk-adjusted returns than those for impact portfolios constructed using simpler heuristics such as negative or positive screening.

More broadly, our framework can be regarded as an extension of standard portfolio theory by considering the rank rather than the value of a specific factor, leading to portfolios with more robust performance in practice. Beyond impact investing, the mathematics of our framework also applies more broadly to any factor, such as the momentum, size, and value factors, or the hundreds of new factors in the "Factor Zoo" discussed in the recent literature (Cochrane 2011, Feng et al. 2020).

Our results also yield several other potentially useful applications. For example, the copula is widely used in credit risk models (Cont and Kan 2011, Filiz et al. 2012, Cont and Minca 2013). Credit bond investors may form portfolios by ranking bonds based on their credit ratings. They can choose a copula to characterize the dependence structure between credit ratings and bond returns, use our framework to derive the distribution of credit-ranked returns of the bonds, and thus construct optimal bond portfolios. Our theory may also be used in the pricing and portfolio construction of credit derivatives, such as CDS, CDS indices, and so on. In general, our framework can be applied whenever there exists an observable factor (e.g., impact measures or credit ratings) that can influence another target variable of interest (e.g., asset returns).

Acknowledgments

The authors gratefully acknowledge Nan Chen, Paul Glasserman, Matheus Grasselli, Xin Guo, Ruediger Kiesel, Ruodu Wang, Jingping Yang, Shushang Zhu, an anonymous editor, two anonymous referees, and seminar and conference participants at the 2023 SIAM Conference on Financial Mathematics and Engineering, the 2023 Energy Finance Italia 8, the 2022 INFORMS Annual Meetings, the 11th Annual Conference of Financial Engineering and Risk Management Branch of the OR Society of China, the 2022 China Society for Industrial and Applied Mathematics (CSIAM) Annual Conference, the 19th Chinese Finance Annual Meeting, UC Berkeley, Suzhou University, and Sun Yat-sen University for very helpful comments and discussion.

Endnotes

[1] Impact investing has also been used to refer specifically to these investments (Barber et al. 2021), but we use the term more broadly in this article.

[2] In a 2017 survey, Eccles et al. (2017) find that the two most popular strategies are exclusionary screening (i.e., negative screening, 47% of surveyed investors) and best-in-class selection (i.e., relative weighting, 37%). See also Roselle (2016), Amel-Zadeh and Serafeim (2018), and Cappucci (2018).

[3] Several studies show that investments with ESG considerations may sacrifice returns in stock markets (Hong and Kacperczyk 2009, Pástor et al. 2022), bond markets (Baker et al. 2022), and venture capital funds (Barber et al. 2021), whereas others (Bansal et al. 2022, Lo et al. 2022) suggest that ESG is associated with higher returns, at least under certain market conditions.

[4] For example, on August 4, 2022, a letter signed by the attorneys general of 19 states in the United States was sent to BlackRock—one of the world's largest institutional asset managers—expressing concern over its ESG policies and how they may violate multiple state laws and fiduciary duties (https://www.texasattorneygeneral.gov/sites/default/files/images/executive-management/BlackRock%20Letter.pdf, accessed December 15, 2022).

[5] See, for example, David (1973) and Bhattacharya (1974).

[6] Sklar's theorem maintains that the copula between X_i and θ_i is uniquely determined only on the Cartesian product of the closures of the ranges of the marginal distribution functions, but undefined in other regions.

[7] See, for example, Nofsinger and Varma (2014), Lööf et al. (2022), and Bax et al. (2023).

[8] For the literature on induced order statistics, see, for example, Yang (1977), Kim and David (1990), Balakrishnan (1993), Lee and Viana (1999), and Wang and Nagaraja (2009).

[9] The assumption that α_i is random is somewhat unconventional, so a few clarifying remarks are in order. This assumption was first used in Lo and MacKinlay (1990) to represent cross-sectional estimation errors of intercepts from CAPM regressions. In the current context, we interpret the randomness in α_i as a measure of uncertainty regarding the degree of mispricings of assets in our investment universe, similar to Pástor and Stambaugh (1999) and Lo and Zhang (2023). This uncertainty can be interpreted from a Bayesian perspective as the degree of conviction that mispricings exist in the cross-section.

[10] Examples include negative screening, positive screening, and relative weighting (or "best-in-class selection"). Investors use the *rank* information but not necessarily the *exact values* of $\boldsymbol{X}$ to form portfolios. This is a weaker assumption compared with the framework of Grinold (1989), and is particularly relevant in the context of impact investing, because many impact measurements (such as ESG) are known to be very noisy (Berg et al. 2022).

[11] Even if investors include all assets in the portfolio, their weights are determined by the rank of the impact factor. Because the rank of each asset is random, their returns are also random. Therefore, we need to quantify the distribution of the induced order statistics in order to quantify the returns of the sorted assets and their weights in the optimal impact portfolio.

[12] See, for example, Bolton and Kacperczyk (2021, 2023), Pástor et al. (2021, 2022), Pedersen et al. (2021), and Lo et al. (2022).

[13] See, for example, Treynor and Black (1973), Grinold and Kahn (1999, 2019), and Ding and Martin (2017).

[14] The function $g_u(w)$ corresponds to the inverse Rosenblatt transform (Rosenblatt 1952), which is widely used for sampling random vectors from copula. See, for example, Nelsen (2007, section 2.9).

[15] Because F_X is nondecreasing, Δ_X must be at most countably infinite and $\overline{\mathcal{R}}_X^c$ can always be represented as $\overline{\mathcal{R}}_X^c = \cup_{d\in\Delta_X}(F_X(d^-), F_X(d))$.

[16] For example, the MSCI ESG Ratings gives categorical ratings (CCC, B, BB, BBB, A, AA, and AAA) for companies. Similarly, bond credit ratings are almost always discrete.

[17] For example, Kou et al. (2017) and Cai and Yang (2018) investigate jumps in equity markets, Chen and Kou (2009) propose a jump model for credit risk, and Cai and Kou (2012) use jump diffusion models to price options. In addition, the U.S. equity index market has a market-wide 20% circuit breaker, and the Chinese stock market has $\pm 10\%$ price limits on individual stocks. And for high-frequency data, most markets have a minimum tick size of, for example, one cent.

[18] For example, the comonotonicity copula and the countermonotonicity copula discussed in Online Appendix EC.1.4.3 are not differentiable.

[19] Some authors refer to the inverse function ψ^{-1} as the generator function of the Archimedean copula, as in McNeil et al. (2015). Here we largely follow the definition given by Nelsen (2007).

[20] Nelsen (2007, section 4) provides a list of Archimedean copulas with over 20 types of generator functions.

[21] Formally, the lower tail dependence parameter, λ_L, and the upper tail dependence parameter, λ_U, for a copula C are defined as

$$\lambda_L \equiv \lim_{u\to 0^+} \frac{C(u,u)}{u}, \quad \lambda_U \equiv 2 - \lim_{u\to 1^-} \frac{1-C(u,u)}{1-u}.$$

In particular, the Clayton copula with parameter γ has $\lambda_L = 2^{-1/\gamma}$ and $\lambda_U = 0$, and the Gumbel copula with parameter γ has $\lambda_L = 0$ and $\lambda_U = 2 - 2^{1/\gamma}$. See Nelsen (2007, section 5.4).

[22] The first two moments shown in Figure 1 are computed numerically using Proposition 2. We perform the numerical integrations using the `scipy.integrate` package in Python 3.7 on a laptop with an Intel(R) Core(TM) i7-9750H CPU @ 2.60 GHz. The computation times are 14.77, 15.37, and 505.59 seconds for expectations, variances, and covariances, respectively. When the number of assets is high, one may skip the calculation of covariances in practice to further reduce computation time because they are very close to zero (see, for example, Theorems 3 and 4).

References

Aïd R, Biagini S (2023) Optimal dynamic regulation of carbon emissions market. *Math. Finance* 33(1):80–115.

Amel-Zadeh A, Serafeim G (2018) Why and how investors use ESG information: Evidence from a global survey. *Financial Anal. J.* 74(3):87–103.

Baker M, Bergstresser D, Serafeim G, Wurgler J (2022) The pricing and ownership of US green bonds. *Annu. Rev. Financial Econom.* 14:415–437.

Balakrishnan N (1993) Multivariate normal distribution and multivariate order statistics induced by ordering linear combinations. *Statist. Probab. Lett.* 17(5):343–350.

Bansal R, Wu D, Yaron A (2022) Socially responsible investing in good and bad times. *Rev. Financial Stud.* 35(4):2067–2099.

Barber BM, Morse A, Yasuda A (2021) Impact investing. *J. Financial Econom.* 139(1):162–185.

Bax K, Sahin Ö, Czado C, Paterlini S (2023) ESG, risk, and (tail) dependence. *Internat. Rev. Financial Anal.* 87:102513.

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

Bhattacharya P (1974) Convergence of sample paths of normalized sums of induced order statistics. *Ann. Statist.* 2(5):1034–1039.

Blasberg A, Kiesel R, Taschini L (2021) Carbon default swap—Disentangling the exposure to carbon risk through CDS. Preprint, submitted May 31, https://dx.doi.org/10.2139/ssrn.3856993.

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

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

Brigo D, Capponi A, Pallavicini A (2014) Arbitrage-free bilateral counterparty risk valuation under collateralization and application to credit default swaps. *Math. Finance* 24(1):125–146.

Cai N, Kou S (2012) Pricing Asian options under a hyper-exponential jump diffusion model. *Oper. Res.* 60(1):64–77.

Cai N, Yang X (2018) International reserve management: A drift-switching reflected jump-diffusion model. *Math. Finance* 28(1): 409–446.

Cappucci M (2018) The ESG integration paradox. *J. Appl. Corporate Finance* 30(2):22–28.

Chen N, Kou SG (2009) Credit spreads, optimal capital structure, and implied volatility with endogenous default and jump risk. *Math. Finance* 19(3):343–378.

Cochrane JH (2011) Presidential address: Discount rates. *J. Finance* 66(4):1047–1108.

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

Cont R, Kan YH (2011) Dynamic hedging of portfolio credit derivatives. *SIAM J. Financial Math.* 2(1):112–140.

Cont R, Minca A (2013) Recovering portfolio default intensities implied by CDO quotes. *Math. Finance* 23(1):94–121.

David H (1973) Concomitants of order statistics. *Bull. Internat. Statist. Inst.* 45(1):295–300.

David HA, Nagaraja HN (2003) *Order Statistics* (John Wiley & Sons, Hoboken, NJ).

De Angelis T, Tankov P, Zerbib OD (2023) Climate impact investing. *Management Sci.* 69(12):7669–7692.

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

Eccles RG, Kastrapeli MD, Potter SJ (2017) How to integrate ESG into investment decision-making: Results of a global survey of institutional investors. *J. Appl. Corporate Finance* 29(4):125–133.

Fama EF, French KR (1993) Common risk factors in the returns on stocks and bonds. *J. Financial Econom.* 33(1):3–56.

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

Feng G, Giglio S, Xiu D (2020) Taming the factor zoo: A test of new factors. *J. Finance* 75(3):1327–1370.

Filiz IO, Guo X, Morton J, Sturmfels B (2012) Graphical models for correlated defaults. *Math. Finance* 22(4):621–644.

Flora M, Tankov P (2023) Green investment and asset stranding under transition scenario uncertainty. *Energy Econom.* 124: 106773.

Giesecke K (2003) A simple exponential model for dependent defaults. *J. Fixed Income* 13(3):74–83.

Giesecke K (2004) Correlated default with incomplete information. *J. Banking Finance* 28(7):1521–1545.

Gobet E, Lage C (2023) Optimal ecological transition path of a credit portfolio distribution, based on multidate Monge–Kantorovich formulation. *Ann. Oper. Res.* Published ahead of print May 25, https://doi.org/10.1007/s10479-023-05385-4.

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

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

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

Harvey CR, Liu Y, Zhu H (2016) … and the cross-section of expected returns. *Rev. Financial Stud.* 29(1):5–68.

Hong H, Kacperczyk M (2009) The price of sin: The effects of social norms on markets. *J. Financial Econom.* 93(1):15–36.

Hou K, Xue C, Zhang L (2020) Replicating anomalies. *Rev. Financial Stud.* 33(5):2019–2133.

Jondeau E, Rockinger M (2003) Testing for differences in the tails of stock-market returns. *J. Empirical Finance* 10(5):559–581.

Kim SH, David HA (1990) On the dependence structure of order statistics and concomitants of order statistics. *J. Statist. Planning Inference* 24(3):363–368.

Kou S, Peng X (2016) On the measurement of economic tail risk. *Oper. Res.* 64(5):1056–1072.

Kou S, Yu C, Zhong H (2017) Jumps in equity index returns before and during the recent financial crisis: A Bayesian analysis. *Management Sci.* 63(4):988–1010.

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

Lintner J (1965) The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Rev. Econom. Statist.* 47(1):13–37.

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

Lo AW, Zhang R (2023) Quantifying the impact of impact investing. *Management Sci.* ePub ahead of print December 7, https://doi.org/10.1287/mnsc.2022.01168.

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

Lööf H, Sahamkhadam M, Stephan A (2022) Is corporate social responsibility investing a free lunch? The relationship between ESG, tail risk, and upside potential of stocks before and during the COVID-19 crisis. *Finance Res. Lett.* 46:102499.

McNeil AJ, Frey R, Embrechts P (2015) *Quantitative Risk Management: Concepts, Techniques and Tools*, revised ed. (Princeton University Press, Princeton, NJ).

Merton R (1973) An intertemporal capital asset pricing model. *Econometrica* 41(5):867–887.

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

Nofsinger J, Varma A (2014) Socially responsible funds and market crises. *J. Banking Finance* 48:180–193.

Pástor L, Stambaugh RF (1999) Costs of equity capital and model mispricing. *J. Finance* 54(1):67–121.

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

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

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

Roselle P (2016) The evolution of integrating ESG analysis into wealth management decisions. *J. Appl. Corporate Finance* 28(2): 75–79.

Rosenblatt M (1952) Remarks on a multivariate transformation. *Ann. Math. Statist.* 23(3):470–472.

Ross S (1976) The arbitrage theory of capital asset pricing. *J. Econom. Theory* 13(3):341–360.

Sharpe WF (1963) A simplified model for portfolio analysis. *Management Sci.* 9(2):277–293.

Sharpe WF (1964) Capital asset prices: A theory of market equilibrium under conditions of risk. *J. Finance* 19(3):425–442.

Sklar M (1959) Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris* 8(3):229–231.

Sorensen E, Chen M, Mussalli G (2021) The quantitative approach for sustainable investing. *J. Portfolio Management* 47(8):38–49.

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

Wang K, Nagaraja HN (2009) Concomitants of order statistics for dependent samples. *Statist. Probab. Lett.* 79(4):553–558.

Yang SS (1977) General distribution theory of the concomitants of order statistics. *Ann. Statist.* 5(5):996–1002.

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

Andrew W. Lo is the Charles E. and Susan T. Harris Professor at the MIT Sloan School of Management, director of the MIT Laboratory for Financial Engineering, and a principal investigator at the MIT Computer Science and Artificial Intelligence Laboratory. He is also an external faculty member of the Santa Fe Institute and a research associate of the National Bureau of Economic Research.

Lan Wu is a professor at the School of Mathematical Sciences at Peking University (PKU), with affiliations at the PKU Center for Statistical Science and the PKU Laboratory for Mathematical Economics and Quantitative Finance.

Ruixun Zhang is an assistant professor and Boya Young Fellow at the School of Mathematical Sciences at Peking University (PKU), with affiliations at the PKU Center for Statistical Science, the PKU Laboratory for Mathematical Economics and Quantitative Finance, and the PKU National Engineering Laboratory for Big Data Analysis and Applications.

Chaoyi Zhao is a PhD student at the School of Mathematical Sciences at Peking University.