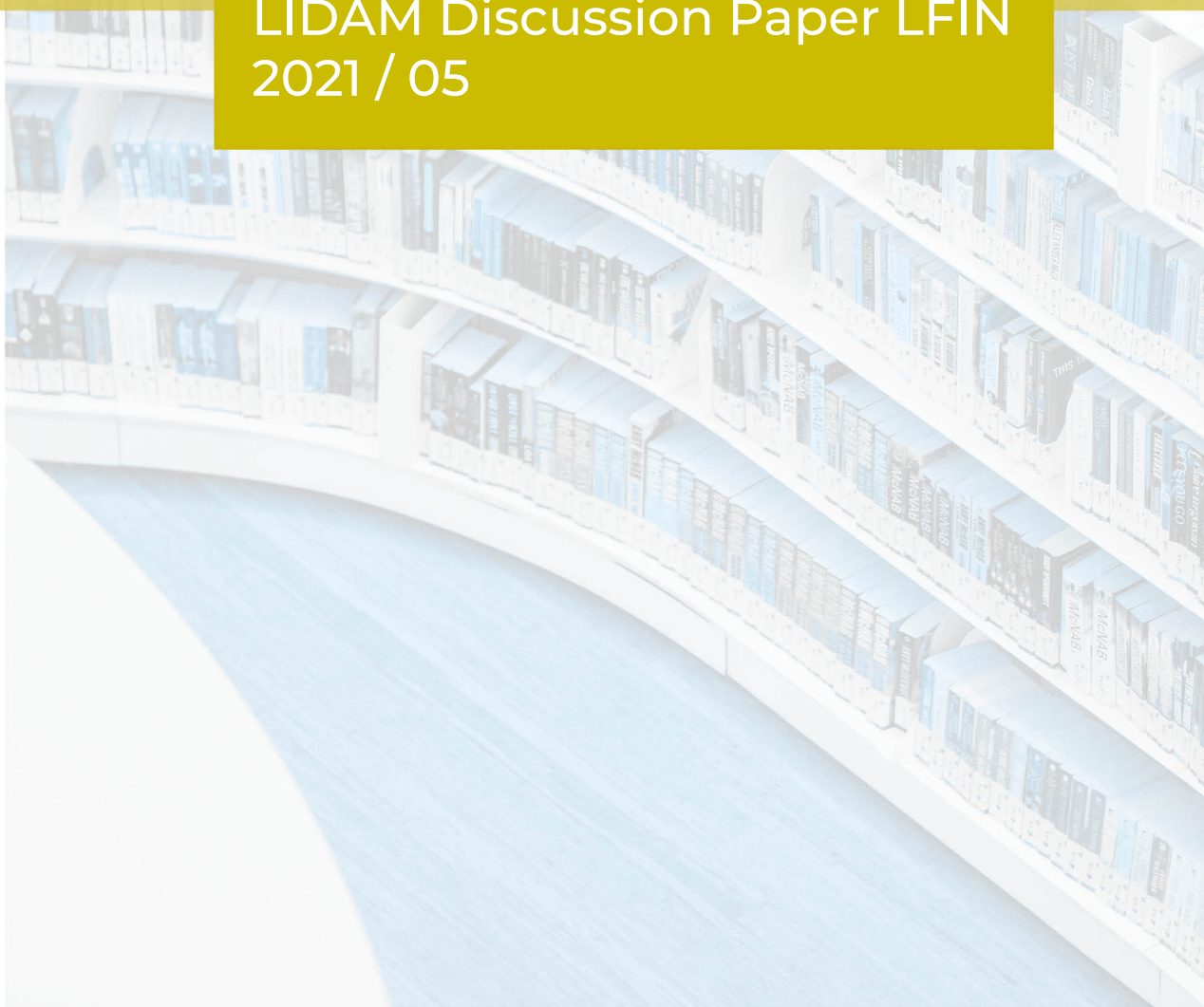
PORTFOLIO SELECTION: A TARGET-DISTRIBUTION APPROACH

Nathan Lassance, Frédéric Vrins

LIDAM Discussion Paper LFIN
2021 / 05



UCLouvain

LFIN | LIDAM
Louvain Institute of Data Analysis and Modeling in economics and statistics

LFIN

Voie du Roman Pays 34, L1.03.01 B-1348 Louvain-la-Neuve

Tel (32 10) 47 43 04

Email: lidam-library@uclouvain.be

https://uclouvain.be/en/research-institutes/lidam/lfin/publications.html

Portfolio Selection: A Target-Distribution Approach*

Nathan Lassance Frédéric Vrins

July 26, 2021

Abstract

We introduce a general framework to the portfolio-selection problem in which investors aim at targeting a *distribution* of returns, which can accommodate a wide range of preferences. The resulting optimal portfolio has a return density that is as close as possible to the target-return density. We study the theoretical properties of this approach for two classes of target distribution that allow for different first four moments. Three results that stand out are, first, that the fit to higher moments is controlled by the entropy of standardized portfolio returns when targeting a Gaussian distribution. Second, when targeting a specific Dirac-delta distribution, no norm-constrained portfolio can stochastically dominate the proposed optimal portfolio. Third, if the target-return mean and variance are located on or above the efficient frontier, the optimal portfolio is mean-variance efficient when asset returns are Gaussian. For non-Gaussian returns, the optimal portfolio may move away from the frontier to better fit the higher moments of the target distribution. The empirical analysis illustrates that the proposed framework helps the investor obtain portfolio returns in line with her preferences.

Keywords: portfolio selection, higher moments, tail risk, Kullback-Leibler divergence
JEL Classification: G11, G12

*Lassance, UCLouvain, Louvain Finance, corresponding author, nathan.lassance@uclouvain.be; Vrins, UCLouvain, Louvain Finance, frederic.vrins@uclouvain.be. We gratefully acknowledge Kris Boudt, Victor DeMiguel, Roberto Renò, Guofu Zhou, and conference participants for comments and discussions. Part of this work was conducted while N. Lassance's research was supported by the Fonds de la Recherche Scientifique (F.R.S.-FNRS) via a grant ASP [grant number FC 17775]. The work of F. Vrins is supported by the Belgian Federal Science Policy Office via a grant ARC [grant number 18-23/089].

1 Introduction

The portfolio-selection problem aims at allocating wealth to financial assets according to some criterion. It involves two steps. First, model the investor's preferences by choosing an appropriate objective function. Second, invest in the assets of a given investment universe to maximize the objective; that is, find the optimal portfolio weights. The relevance of the investment strategy is often assessed in a third step, via an out-of-sample performance analysis. In practice, the second step requires the investor to estimate the objective function from a sample of data. For that purpose, a wide range of methods have been proposed in the literature on portfolio selection under parameter uncertainty.[1]

In this paper, we are primarily interested in the first step of the process, which we address in a general framework. As we review at the end of this section, many approaches exist to address this step. The standard approach dating from the pioneering work of Markowitz (1952) consists in taking the expectation of a given utility function, but practitioners often favour alternative objectives that yield more diversified portfolios, such as factor diversification (Lassance et al. 2021). In this paper, we introduce a new route that, we believe, is both sound from a theoretical perspective and intuitive from a practitioner's viewpoint. Specifically, we propose a framework that consists in modeling the investor's preferences by specifying a *target distribution* for the returns. Then, the objective function in the first step of the process consists in finding the portfolio weights minimizing the "distance" between the portfolio-return distribution and the investor's target-return distribution.

The proposed approach has several benefits. First, it can accommodate a wide range of investor's preferences in a flexible way. Indeed, most metrics that are used in the literature to design and evaluate portfolio strategies—moments, lower partial moments, risk-return ratios, quantiles, etc.— can directly be obtained from the underlying distribution.[2] Those metrics can be reflected in the investor's objective by choosing a target distribution complying with them. For example, in the paper, we consider two classes of target distributions

[1] Some popular approaches in that area are shrinkage estimation (Jorion 1986, Ledoit and Wolf 2004), portfolio combination (Kan and Zhou 2007, Kan et al. 2021), portfolio-weight constraints (Jagannathan and Ma 2003, DeMiguel et al. 2009), sparse estimation (Goto and Xu 2015, Ao et al. 2019), and factor models (De Nard et al. 2019, Lassance and Vrins 2021b).

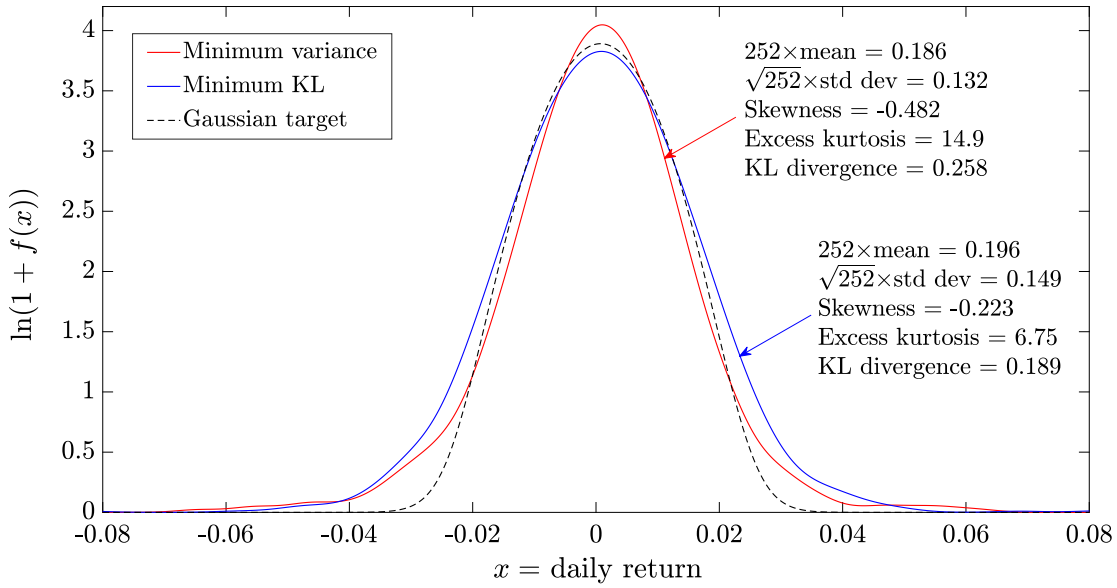
[2] See Cogneau and Hubner (2014) for a comprehensive review of portfolio performance metrics.

that can accommodate different values of mean, variance, skewness, and kurtosis. Second, in practice, the return distribution is often the key information that stakeholders care about, either directly (via histograms built from historical time-series) or indirectly (via statistics). Most investors are likely to be able to identify the type of return distribution they would like to obtain (e.g., for risk-averse investors, high mean, low variance, positive asymmetry, thin tails), and those they would like to avoid. Desirable distributions can also be identified from the extensive experimental evidence on higher-order risk attitudes of economic agents; see, for example, Deck and Schlesinger (2010), Trautmann and van de Kuilen (2018). Therefore, specifying the objective of the investment strategy via the shape of the desired return distribution is both flexible and intuitive. Third, finding the portfolio that best fits a target distribution has the advantage over a strategy that would feature constraints on a few moments that all moments are embedded in one go; there is no need to choose which moments to ignore, and how to aggregate the selected ones into a single objective function.

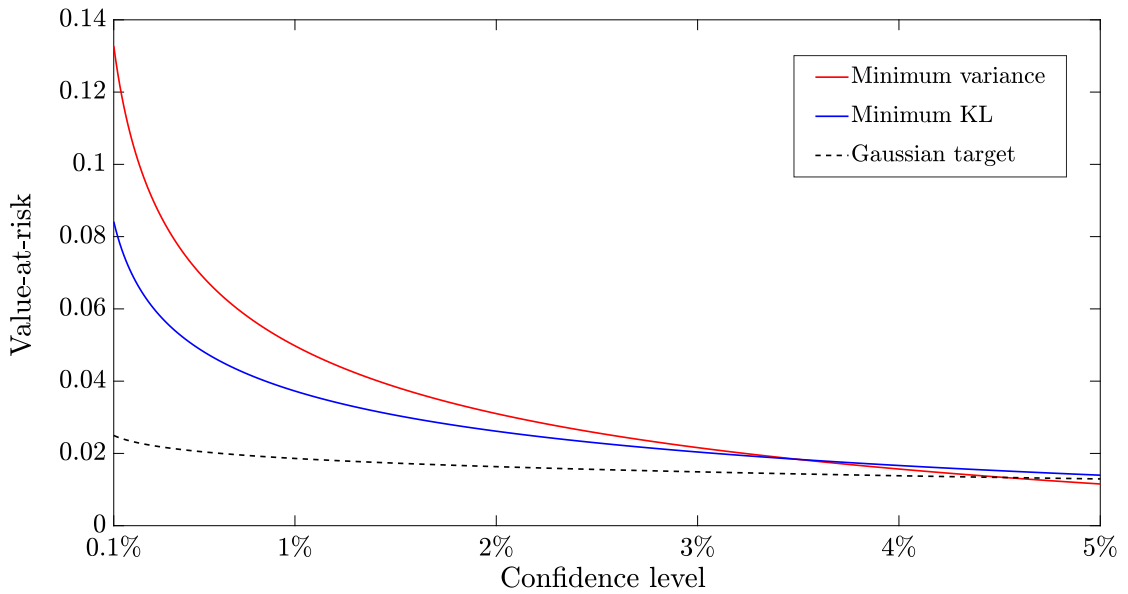
Although different strategies can be designed by relying on our method, we provide one possible application in Figure 1 in guise of illustration. The purpose here is to improve the higher moments of a portfolio obtained from a base strategy. We use daily returns on the 25 portfolios of stocks sorted on size and book-to-market spanning January 1980 to March 2021. In this example, the investor considers as base strategy the global-minimum-variance (GMV) portfolio, whose return density is depicted in red. While this portfolio has the smallest achievable variance in-sample, it suffers from a large negative skewness (-0.482) and positive excess kurtosis (14.9). To obtain a portfolio with improved higher moments, the standard approach would be to adjust the minimum-variance objective, penalizing for such higher-moment characteristics. In our framework, a similar goal can be achieved by targeting a return distribution whose mean and variance correspond to those of the GMV portfolio, but that exhibits better higher moments. This can by achieved by choosing, for example, a *Gaussian* target, whose skewness and excess kurtosis are zero and, thus, are more appealing that those of GMV. The Gaussian target distribution is depicted in dashed black in Figure 1a. Adopting our strategy, the investor looks for the portfolio whose return distribution is as close as possible to the target according to an appropriate "distance measure" between densities; here, we use the well-known Kullback-Leibler (KL) divergence. This

Figure 1: Targeting a Gaussian return distribution

a) Comparison of density functions



b) Comparison of value-at-risk



Notes. The two figures are constructed from daily returns on the 25 portfolios of stocks sorted on size and book-to-market spanning January 1980 to March 2021. In Figure a), the return density of the global-minimum-variance portfolio is depicted in solid red, and the Gaussian density with the same mean and variance in dashed black. The return density of the minimum-KL portfolio, which minimizes the Kullback-Leibler divergence with respect to the Gaussian density, is depicted in solid blue. We plot $\ln(1+f(x))$, where $f(x)$ is the density function, to make the tails more visible. In Figure b), we depict the value-at-risk of the three density functions for confidence levels between 0.1% and 5%. The value-at-risk is computed via the Cornish-Fisher expansion.

strategy amounts to enhancing the higher moments of GMV by shrinking the GMV portfolio towards a portfolio whose return has similar first two moments, but more attractive skewness and kurtosis. By construction, the return density of the optimal portfolio (in blue) is closer to the Gaussian density (in black) relative to GMV (in red) (KL divergence of 0.189 versus 0.258). In particular, the return mean and variance of the optimal portfolio are close to the return mean and variance of GMV but, in addition, the optimal portfolio benefits from a skewness of -0.223 and an excess kurtosis of 6.75 that are closer to the higher moments of the Gaussian. This result comes from a compromise between all the moments and, in particular, at the cost of a higher variance than that of GMV. Nevertheless, the return distribution of the optimal portfolio enjoys a reduced tail risk as shown by the value-at-risk in Figure 1b; for example, the 0.1% value-at-risk is nearly reduced by 40% relative to GMV.

The first step of the portfolio-selection process—modeling the investor's preferences via an appropriate objective function—is an important problem to address because it lacks a unifying approach. Consider first a world in which investors have mean-variance preferences as in the pioneering approach of Markowitz (1952), which is an appropriate modeling approach when asset returns are distributed according to a Gaussian or elliptical distribution (Chamberlain 1983, Schuhmacher et al. 2021). Life is theoretically easy in such a world: the optimal portfolio is always located on the mean-variance efficient (MVE) frontier, and the objective function unequivocally consists in minimizing the variance for a given mean return. Similarly, we prove that our proposed optimal portfolio is also located on the MVE frontier when asset returns are Gaussian if the target-return mean and variance are located on or above the MVE frontier. However, this simple world is unrealistic because it is well-known that in most cases asset returns are not distributed according to a Gaussian or elliptical distribution; see, for example, the large literature that allows for stochastic volatility and jumps (Heston 1993, Das and Uppal 2004). Consequently, mean-variance preferences and the minimum-variance objective function are no longer appropriate because investors care for other distributional properties than just the first two moments; see Jean (1971), Scott and Horvath (1980) for early studies on higher-moment portfolio analysis, and Ang et al. (2006), Kelly and Jiang (2014) for evidence of downside-risk preferences in the cross-section of stock returns. As Kadan and Liu (2014 p.134) put it: "What looks like an attractive in-

vestment strategy when focusing on the first two moments, can easily become less appealing when considering higher moments and disaster risk." In this more complicated world, as we review at the end of this section, many different approaches have been proposed to capture preferences beyond the mean and variance, which indicates that investors lack a general framework to translate their preferences into an optimal portfolio. Our goal is to propose such a framework and analyze its theoretical and empirical properties.

Our contribution is threefold. First, we introduce a general and flexible framework in which the investor's preferences are captured by a target-return distribution. The investor's optimal portfolio is then obtained by minimizing the conditional Kullback-Leibler (CKL) divergence between the portfolio-return and the target-return density functions, and is called the minimum-CKL (MCKL) portfolio. The CKL divergence provides a novel extension to the plain KL divergence (Kullback and Leibler 1951). The two measures coincide when the KL divergence is well-defined but, contrary to the KL divergence, the CKL divergence is also defined when the portfolio-return support set is not included in that of the target return, which allows us to consider a broader set of target densities. To operationalize this general approach, we study two classes of target distributions. On the one hand, the generalized-normal distribution (Nadarajah 2005), which extends the Gaussian distribution via an additional parameter that controls the kurtosis. On the other hand, the shifted-log-normal distribution (Yuan 1933), which extends the Gaussian distribution via an additional parameter that controls the skewness. We propose to match the mean and variance of the target distribution to those of a reference portfolio located on the MVE frontier, which ensures that the first two moments can not be dominated. The additional parameter of the target distribution then controls for preferences beyond mean and variance. In doing so, the MCKL portfolio balances between the optimal mean-variance tradeoff of the reference portfolio and the desired higher-moment properties embedded in the target distribution.

Second, we investigate the theoretical properties of the proposed framework. We begin by deriving closed-form expressions for the objective functions related to the generalized-normal and shifted-log-normal target distributions, respectively. We illustrate that the objective functions are more sensitive to the variance than to the mean and the higher moments, which is desirable from an estimation-risk perspective (Martellini and Ziemann 2010). We

then consider the special case in which the target distribution is Gaussian, as in Figure 1. We show that the corresponding objective function decomposes as a sum of three terms. The first two terms measure the fit to the target-return mean and variance, the fit to the variance being relatively more important. The third term is the entropy of the standardized portfolio return and shrinks the higher moments toward those of a Gaussian distribution, which is desirable in terms of downside risk. We also study the special case of the Dirac-delta target distribution and show that the MCKL portfolio aims at minimizing the euclidean distance to the target-return mean and variance, which is achieved for a particular mean-variance portfolio. Moreover, for a specific choice of the Dirac-delta mean, no norm-constrained portfolio can dominate the MCKL portfolio in terms of stochastic dominance of the third order (Whitmore 1970, Post and Kopa 2016), which represents the preferences of investors with increasingly concave utility functions. Finally, we study the classical setting in which asset returns are assumed Gaussian and demonstrate that the objective functions associated with the two classes of target distributions have a unique local (and so global) optimum in the mean-variance space. One important consequence of this result is that, if the target-return mean and variance are located on or above the mean-variance efficient frontier, then the MCKL portfolio is always mean-variance efficient, which is consistent with Markowitz's theory. Otherwise, if asset returns are not Gaussian, the MCKL portfolio may move away from the efficient frontier so as to better fit the higher moments of the target distribution.

Third, we provide finite-sample estimates for the considered objective functions, and use them to study the in-sample properties and out-of-sample performance of the MCKL portfolio, using daily returns from three datasets of characteristic-sorted equity portfolios. We consider the proposed setting whereby the investor has identified her desired MVE portfolio, but wishes to improve the higher moments of returns of this portfolio. Taking the sample GMV portfolio and the robust mean-variance portfolio of Kan et al. (2021) as MVE portfolios, our empirical results show that the MCKL portfolio trades off between mean-variance efficiency on the one hand, and the improved higher moments of the target distribution on the other hand. Moreover, relative to the chosen MVE portfolio, the MCKL portfolio yields a smaller disaster risk according to the probability of a three-sigma loss considered in Gormsen and Jensen (2020). Also, the tradeoff between mean-variance efficiency and improved higher

moments is more satisfying relative to that of the higher-moment approaches of Briec et al. (2007), Martellini and Ziemann (2010), and Lassance et al. (2021).

Our work contributes to the literature on portfolio selection with preferences beyond mean and variance, which encompasses many different approaches. A first approach in this literature consists in maximizing the expectation of another utility function than the quadratic utility. Most commonly used are cubic and quartic utility functions obtained as Taylor-series expansions of power or exponential utility functions (Jondeau and Rockinger 2006, Guidolin and Timmermann 2008, Harvey et al. 2010, Martellini and Ziemann 2010). However, the third and fourth moments have a negligible impact on optimal portfolios when using such utility functions (Levy and Markowitz 1979, Kroll et al 1984, Markowitz 2014), which has led researchers to investigate other utility functions that better capture actual preferences such as those capturing disappointment aversion (Routledge and Zin 2010, Dahlquist et al. 2017). A second approach relies on stochastic-dominance performance indices, accounting for all distributional moments. In particular, Kadan and Liu (2014) show that the performance measures introduced by Aumann and Serrano (2008) and Foster and Hart (2009) capture higher moments in a desirable manner, and that the impact of higher moments on performance is significant and not vanishing unlike in the first approach. A third approach departs from the expected-utility framework and imposes higher-moment constraints besides the objective function; see Athayde and Flores (2004), Briec et al. (2007) for two popular examples. A fourth approach relies on the minimization of tail risk criteria such as the value-at-risk (Basak and Shapiro 2001, Alexander and Baptista 2004), expected shortfall (Rockafellar and Uryasev 2000), lower partial moment (Price et al. 1982, Jarrow and Zhao 2006), and drawdown (Alexander and Baptista 2006, Van Hemert et al. 2020). A fifth approach optimizes mean-risk ratios that can be seen as generalizations of the Sharpe ratio that account for asymmetry and fat tails in returns. Popular ones are the Sortino ratio (Sortino and van der Meer 1991), the Omega ratio (Keating and Shadwick 2002), and the more general Kappa ratio (Kaplan and Knowles 2004). Finally, a sixth approach involves improving higher moments indirectly—i.e., without including them in the optimization program—via robust optimization (Kim et al. 2014) or factor diversification (Lassance et al. 2021). The main distinguishable feature of our approach compared to the above papers is that we model

preferences via a distribution of returns, instead of via a few statistics of the latter.

Our work is also related to a number of other approaches. Bera and Park (2008) minimize the KL divergence between the portfolio weights and target weights specified by the investor, subject to moment constraints. Our approach is different because we minimize the KL (or CKL) divergence between the portfolio-return distribution and a specified target distribution. Chalabi and Würtz (2012) study the mathematical aspects of a similar problem via a dual representation of the optimization program. In this paper, we mainly focus on the choice of target distribution and the theoretical and empirical properties of MCKL portfolios. Finally, our method is conceptually related to existing approaches in the derivatives pricing literature. As an example, the cost-efficient portfolio is defined as the lowest-cost dynamic portfolio that replicates a given payoff distribution at some future time horizon, scenario by scenario (Dybvig 1988, Bernard et al. 2014). Another example is that of optimal positioning, which consists in finding a buy-and-hold portfolio of derivatives that best replicates the terminal wealth of a given optimal dynamic asset-allocation rule (Haugh and Lo 2001). Our paper differs from those approaches because we aim at finding the static portfolio whose return distribution approaches a given target-return distribution, and we do not consider derivatives products in the investment universe.

The rest of the paper is organized as follows. Section 2 introduces notation and reviews useful properties of mean-variance portfolios. Section 3 presents the general formulation for the MCKL portfolio, and Section 4 studies the theoretical properties of this portfolio for several choices of target-return distribution. Section 5 explains how to estimate the considered objective functions, and Section 6 contains our empirical results. Section 7 concludes. The internet appendix contains proofs of all results and additional material.

2 Notation and mean-variance portfolios

Let $R \in \mathbb{R}^N$ be the random (column) vector of asset returns, which is assumed static over time. We denote by μ and $\mathbf{\Sigma}$ the mean and covariance matrix of R, respectively, where $\mathbf{\Sigma}$ is

assumed of full rank. The investor searches for an optimal portfolio w, which belongs to

$$\mathcal{W} := \{w \in \mathbb{R}^N \mid \iota' w = 1\}, \tag{1}$$

where ι is the N-dimensional vector of ones. The return of this portfolio is $P = P(w) := w'R$, which admits a density f_P that is non-zero in Ω_P, the support set of P. The first four portfolio-return moments, assumed finite, are as follows: $\mu_P := w'\mu$ the mean, $\sigma_P^2 := w'\Sigma w$ the variance, $\zeta_P := \mathbb{E}\left[P^{\star 3}\right]$ the skewness, and $\kappa_P := \mathbb{E}\left[P^{\star 4}\right] - 3$ the excess kurtosis, where

$$P^\star := \frac{P - \mu_P}{\sigma_P} = \frac{P(w) - w'\mu}{\sqrt{w'\Sigma w}} \tag{2}$$

is the standardized portfolio return. The central moments are denoted $m_{k,P}$ for $k \geq 3$. Finally, we denote the set of attainable portfolio-return mean and volatility by

$$\mathcal{MV} := \{(\mu_0, \sigma_0) \in \mathbb{R} \times \mathbb{R}^+ \mid \exists w \in \mathcal{W} : w'\mu = \mu_0, w'\mathbf{\Sigma} w = \sigma_0^2\}. \tag{3}$$

In the next definition, we review several useful properties of mean-variance portfolios.

Definition 1 (Properties of mean-variance portfolios)**.** *The following holds:*

1. *The global-minimum-variance (GMV) portfolio is given by*

$$w_{GMV} := \frac{\mathbf{\Sigma}^{-1}\iota}{\iota'\mathbf{\Sigma}^{-1}\iota}, \tag{4}$$

   with mean return $\mu_g := \iota'\mathbf{\Sigma}^{-1}\mu(\iota'\mathbf{\Sigma}^{-1}\iota)^{-1}$, variance $\sigma_g^2 := (\iota'\mathbf{\Sigma}^{-1}\iota)^{-1}$, and Sharpe ratio $\psi_g := \mu_g/\sigma_g = \iota'\mathbf{\Sigma}^{-1}\mu(\iota'\mathbf{\Sigma}^{-1}\iota)^{-1/2}$.

2. *The maximum-Sharpe-ratio portfolio is given by*

$$w_{MSR} := \frac{\mathbf{\Sigma}^{-1}\mu}{\iota'\mathbf{\Sigma}^{-1}\mu}, \tag{5}$$

   with Sharpe ratio $\psi := \sqrt{\mu'\mathbf{\Sigma}^{-1}\mu}$.

3. *The mean-variance portfolio with mean return μ_0 is given by*

$$w_{\mu_0} := w_{GMV} + \frac{\mu_g(\mu_0 - \mu_g)}{\sigma_g^2(\psi^2 - \psi_g^2)}(w_{MSR} - w_{GMV}), \tag{6}$$

and is mean-variance efficient if and only if $\mu_0 \geq \mu_g$.

4. *The mean-variance efficient (MVE) frontier is given by the upper branch of the mean-variance hyperbola:*

$$\sigma_{EF}(\mu_0) := \sqrt{\sigma_g^2 + \frac{(\mu_0 - \mu_g)^2}{\psi^2 - \psi_g^2}}, \quad \mu_0 \geq \mu_g. \tag{7}$$

3 General framework

In this section, we discuss the choice of divergence measure, define the optimal portfolio, and propose to fix the target-return mean and variance via a reference portfolio on the MVE frontier. We keep the choice of target-return distribution unspecified; specific choices are analyzed in Section 4 for which explicit analytical results will be given.

3.1 Choice of divergence measure

Consider a target-return density f_T that corresponds to a random variable T and captures the investor's preferences. The optimal portfolio is defined as the portfolio whose return density f_P is as close as possible to f_T. To that end, we first review the Kullback-Leibler (KL) divergence between f_P and f_T. We then introduce a generalization of the KL, called *conditional KL* (CKL), which allows us to deal with a broader set of target densities.

3.1.1 Kullback-Leibler divergence

The KL divergence between f_P and f_T requires f_P to be absolutely continuous with respect to f_T. This means that $f_T(x) = 0$ implies $f_P(x) = 0$ for all x or, equivalently, that the support set of f_P is included in that of f_T.

Definition 2 (Kullback-Leibler divergence)**.** *Let $\Omega_P \subseteq \Omega_T$. Then, the KL divergence between f_P and f_T is defined as*

$$\langle f_P | f_T \rangle_1 := \mathbb{E}\left[\ln \frac{f_P(P)}{f_T(P)}\right] = \int_{\Omega_P} f_P(x) \ln \frac{f_P(x)}{f_T(x)} \mathrm{d}x. \tag{8}$$

The KL divergence is also known as relative entropy and was introduced by Kullback and Leibler (1951). It is a measure of divergence because of the property that $\langle f_P | f_T \rangle_1 \geq 0$ for all f_P, f_T and $\langle f_P | f_T \rangle_1 = 0$ if and only if $f_P = f_T$ almost everywhere.

Although other divergence measures exist (see Basseville 2013 for a review), the KL divergence is the standard choice and has been used in many financial applications; see Jiang et al. (2018) and references therein. There are three main reasons why we also rely on the KL divergence. First, it has a natural interpretation as the increase of entropy, and thus the amount of lost information, when approximating f_P by f_T. Second, the KL divergence captures higher moments and, thus, measures distribution fit in a more general manner than via the first two moments only. Third, the KL divergence leads to an objective function that is mathematically tractable when the target-return density belongs to the exponential family, as considered in Section 4.[3] Indeed, the KL divergence admits the decomposition

$$-\langle f_P | f_T \rangle_1 = \mathbb{E}[\ln f_T(P)] + H[P], \tag{9}$$

where

$$H[P] := -\mathbb{E}[\ln f_P(P)] \tag{10}$$

is the portfolio-return entropy (Shannon 1948), and the first term in the right-hand side of (9) simplifies when f_T has an exponential expression.

The KL divergence quantifies the divergence between the densities f_P and f_T in a special way. Specifically, the decomposition (9) yields the following interpretation. The first term is the so-called cross-entropy, and can be thought of as $\lim_{h\to\infty} \frac{1}{h} \sum_{i=1}^{h} \ln f_T(P_i)$ where $P_i \sim f_P$

[3]Note that the KL divergence is not symmetric: $\langle f_P | f_T \rangle_1 \neq \langle f_T | f_P \rangle_1$, in general. The Gaussian target-return density studied in Section 4.3.1 clarifies why the direction $\langle f_P | f_T \rangle_1$ is more appropriate: minimizing the other direction amounts to fitting only the first two moments of the target return, whereas our chosen direction also controls for the fit to the higher moments of the target return.

i.i.d., which is the theoretical log-likelihood that $P \sim f_T$ (Cardoso 1997). Maximizing the cross-entropy with respect to f_P tends to yield a Dirac delta centered at the mode of T. The second term is the portfolio-return entropy, satisfying $H[P] = H[P^\star] + \ln \sigma_P$, where $H[P^\star]$ is maximum if $P^\star \sim \mathcal{N}(0,1)$ whenever $\Omega_P = \mathbb{R}$ (Cover and Thomas 2006). Thus, the entropy acts as a shift-invariant penalty of the log-likelihood, shrinking f_P away from a Dirac delta centered at the mode of f_T by rewarding widespread Gaussian distributions.

3.1.2 Conditional Kullback-Leibler divergence

A severe restriction of the KL divergence is that it is not defined when $\Omega_P \not\subseteq \Omega_T$, a condition that depends on the chosen target density. For example, the SLN target distribution proposed in Section 4.2 is defined on $\Omega_T = \{x \in \mathbb{R} | x > \delta\}$, and Ω_P is not contained in that set when $\Omega_P = \mathbb{R}$, e.g., when P is Gaussian. To account for such cases, we introduce an extension of the KL divergence, that we call conditional KL (CKL) divergence. To the best of our knowledge, this extension was not analyzed previously.

Definition 3 (Conditional KL divergence)**.** *Let $p := \mathbb{P}(P \notin \Omega_T)$. Then, the CKL divergence between f_P and f_T is defined as $\langle f_P | f_T \rangle_2 := +\infty$ if $p = 1$ and*

$$\langle f_P | f_T \rangle_2 := \mathbb{E}\left[\ln \frac{f_P(P)}{f_T(P)} \middle| P \in \Omega_T \right] + \frac{p}{1-p} \tag{11}$$

otherwise.

The next proposition highlights three important properties of the CKL divergence; namely, it is always positive, it corresponds to the KL divergence when the condition $\Omega_P \subseteq \Omega_T$ for the KL to exist holds, and it is infinite when f_P and f_T do not overlap.

Proposition 1 (Properties of CKL divergence)**.** *The CKL divergence $\langle f_P | f_T \rangle_2$ satisfies the following properties:*

1. *$\langle f_P | f_T \rangle_2 \geq 0$ for all f_P, f_T. Moreover, $\langle f_P | f_T \rangle_2 = 0$ if $f_P = f_T$ almost everywhere and, otherwise, $\langle f_P | f_T \rangle_2 > 0$ whenever (i) $p = 1$, (ii) $\Omega_P \subseteq \Omega_T$, or (iii) $\Omega_T \subset \Omega_P$ and $p > 0$.*

2. *$\langle f_P | f_T \rangle_2 = \langle f_P | f_T \rangle_1$ if $\Omega_P \subseteq \Omega_T$.*

3. $\langle f_P | f_T \rangle_2 = +\infty$ *if* $\Omega_P \cap \Omega_T = \emptyset$.

Minimizing the CKL divergence amounts to fit the two densities on $\Omega_T \cap \Omega_P$, but also to minimize the probability $\mathbb{P}(P \notin \Omega_T) = p$, ensuring that the support sets tend to coincide. The term $p/(1-p)$ in (11) guarantees the non-negativity of the CKL divergence. Note that the CKL divergence can vanish even when f_P and f_T do not agree almost everywhere, in the case where Ω_P and Ω_T partially overlap each other contrary to the conditions in part 1 of Proposition 1. For instance, $\langle f_P | f_T \rangle_2 = 0$ when $f_P(x) = f_T(x) e^{-p/(1-p)}$ for all $x \in \Omega_T \cap \Omega_P$.[4]

Finally, the CKL divergence admits a similar decomposition to that in (9):

$$-\langle f_P | f_T \rangle_2 = \mathbb{E}[\ln f_T(P) | P \in \Omega_T] + H[P | P \in \Omega_T] - \frac{p}{1-p}, \tag{12}$$

where

$$H[P | P \in \Omega_T] := -\mathbb{E}[\ln f_P(P) | P \in \Omega_T]. \tag{13}$$

3.2 Definition of optimal portfolio

Considering that the CKL divergence generalizes the KL divergence, we define the optimal portfolio as the portfolio that, given a choice of target-return density f_T, minimizes the CKL divergence between f_P and f_T.

Definition 4 (Minimum-CKL portfolio)**.** *The MCKL portfolio minimizes the conditional KL divergence between the portfolio-return density* f_P *and the target-return density* f_T*:*

$$w_{f_T} := \underset{w \in \mathcal{W}}{\operatorname{argmax}}\ \mathcal{C}(w; f_T) \quad \textit{with} \quad \mathcal{C}(w; f_T) := -\langle f_{P(w)} | f_T \rangle_2. \tag{14}$$

We rely on a maximization formulation because it makes some formulas more natural. In the sequel, equalities in different specifications of $\mathcal{C}(w; f_T)$ are considered to hold up to additive and strictly positive multiplicative constants independent of w, since those do not affect the maximizer w_{f_T}. In Section 4, we analyze properties of the MCKL portfolio for the generalized-normal and shifted-log-normal target distributions.

[4] $\langle f_P | f_T \rangle_2 = 0$ if $\mathbb{E}\left[\ln \frac{f_T(P)}{f_P(P)} 1_{\{P \in \Omega_T\}}\right] = p$, which happens when $\ln \frac{f_T(x)}{f_P(x)} = p/(1-p)$ for all $x \in \Omega_T \cap \Omega_P$. Isolating $f_P(x)$ proves the result.

3.3 Reference-portfolio approach

In the empirical analysis, we propose to match the target-return mean and variance to those of a reference portfolio located on the MVE frontier. Although the investor is free to choose her target distribution differently, this seems a reasonable choice given the mean-variance dominance of MVE portfolios. Specifically, given an MVE reference portfolio w_0 satisfying $\mu_0 \geq \mu_g$, $\sigma_0 = \sigma_{EF}(\mu_0)$, we set

$$\left(\mu_T, \sigma_T^2\right) \leftarrow (\mu_0, \sigma_0^2) = (w_0'\mu, w_0'\boldsymbol{\Sigma} w_0)\,, \tag{15}$$

which ensures that $(\mu_T, \sigma_T) \in \mathcal{MV}$; that is, they are attainable.

We underline that this approach does not imply that investors have mean-variance preferences, because fitting the whole target-return distribution does not amount to just fitting the first two moments. Instead, the MCKL portfolio will balance between the optimal mean-variance tradeoff of the reference portfolio and the desired higher-moment properties reflected by the chosen target distribution.

4 Choosing the target-return distribution

In this section, we propose and analyze two classes of target-return distribution. First, the generalized-normal (GN) distribution in Section 4.1, which is a three-parameter symmetric distribution that allows for varying kurtosis. Second, the shifted-log-normal (SLN) distribution in Section 4.2, which is a three-parameter distribution that allows for varying skewness. These two choices are appealing because (i) they are parsimonious (only three parameters), (ii) they allow for a wide range of skewness or kurtosis, (iii) they belong to the exponential family, which makes the CKL divergence more tractable analytically, and (iv) they allow for an easy calibration of the target-return parameters, e.g., by moment matching. Then, in Section 4.3, we analyze the Gaussian and Dirac-delta target distributions, which are special cases of the GN and SLN distributions. Finally, in Section 4.4, we prove that when asset returns are Gaussian and the target-return mean and variance are located on or above the MVE frontier, then the MCKL portfolio is always MVE.

4.1 The generalized-normal target distribution

We consider first the GN distribution of Nadarajah (2005), which is a special case of the skewed generalized t distribution of Theodossiou (1998). The GN distribution belongs to the exponential family, is symmetric, and is defined via three parameters that control for the mean, variance, and fat tails. Thus, this distribution can be used to capture the preferences of investors who aim at avoiding asymmetry and fat tails in their returns beyond targeting a specific mean and variance.

Definition 5 (Generalized-normal target)**.** *The target return follows a GN distribution, $T \sim \mathcal{GN}(\alpha, \beta, \gamma)$, if its density function is given by*

$$f_T(x; \alpha, \beta, \gamma) := \frac{2^{-\frac{\gamma+1}{\gamma}}\gamma}{\beta\Gamma(1/\gamma)} \exp\left(-\frac{1}{2}\left(\frac{|x-\alpha|}{\beta}\right)^{\gamma}\right), \tag{16}$$

where $\alpha \in \mathbb{R}$ and $\beta, \gamma > 0$. The first four moments of T are given by

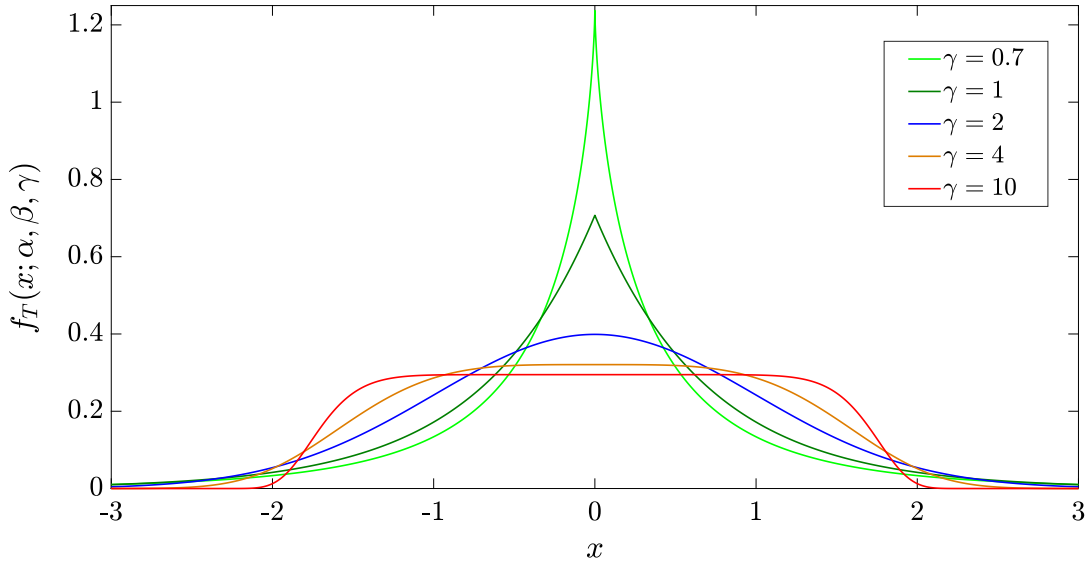
$$\mu_T = \alpha, \quad \sigma_T^2 = \beta^2 4^{1/\gamma}\frac{\Gamma(3/\gamma)}{\Gamma(1/\gamma)}, \quad \zeta_T = 0, \quad \text{and} \quad \kappa_T = \frac{\Gamma(5/\gamma)\Gamma(1/\gamma)}{\Gamma(3/\gamma)^2} - 3. \tag{17}$$

Observe that the GN distribution is symmetric and that its excess kurtosis is negative for $\gamma > 2$. The minimum excess kurtosis, obtained at the limit $\gamma \to \infty$, is -1.2. The GN distribution retrieves the Laplace distribution for $\gamma = 1$, the Gaussian distribution for $\gamma = 2$, and the Uniform distribution on the interval $[\alpha - \beta, \alpha + \beta]$ for $\gamma \to \infty$. In Figure 2, we depict the zero-mean unit-variance GN density for different values of γ, and observe that the smaller the γ the more heavy-tailed the distribution. Finally, moment-matching can easily be done for the mean and variance from (17), and the γ matching a specified excess kurtosis κ_T, or a specified quantile for example, can be found numerically.

In the next proposition, we characterize the objective function in (14) associated with the GN target distribution. Note that $\Omega_T = \mathbb{R}$ and thus necessarily $\Omega_P \subseteq \Omega_T$, so that the CKL divergence reduces to the KL divergence as established in Proposition 1.

Proposition 2 (MCKL portfolio with GN target)**.** *Let $T \sim \mathcal{GN}(\alpha, \beta, \gamma)$. Then, the MCKL*

Figure 2: Density of the generalized-normal distribution



Notes. The figure depicts the generalized-normal density in (16) with mean zero, variance one, and several values of γ that controls the kurtosis via (17). We recover the Gaussian density when $\gamma = 2$.

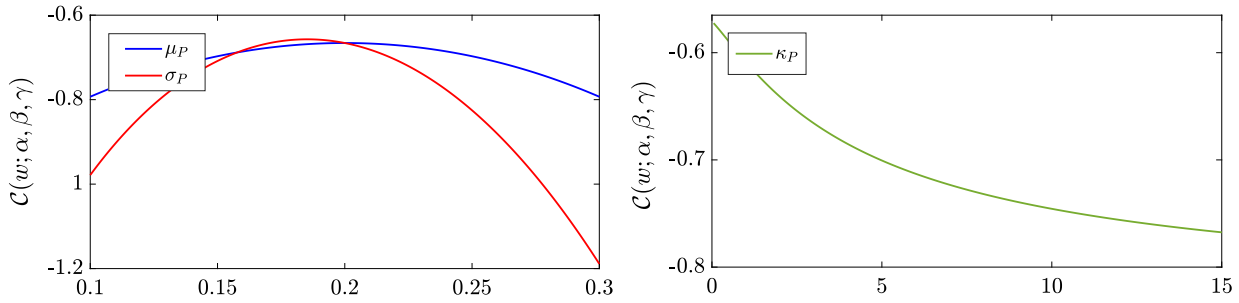
portfolio $w_{f_T} = w_{\alpha,\beta,\gamma}$ *is obtained by maximizing*

$$\mathcal{C}\left(w;\alpha,\beta,\gamma\right) = H[P^\star] + \ln \sigma_P - \frac{1}{2\beta^\gamma}\mathbb{E}\left[|P-\alpha|^\gamma\right]. \tag{18}$$

The first and last terms in this expression can not be further simplified without additional assumption on the distribution of P. We analyze the special case of Gaussian asset returns in Section 4.4 and the Gaussian target distribution, corresponding to $\gamma = 2$, in Section 4.3.1.

In Figure 3, we evaluate the moment sensitivity of the objective function $\mathcal{C}(w;\alpha,\beta,\gamma)$. We consider $(\mu_T, \sigma_T, \kappa_T) = (0.2, 0.2, -0.5)$ $(\gamma = 2.78)$ and depict how the objective function evolves as we vary each moment $(\mu_P, \sigma_P, \kappa_P)$, assuming a Student t distribution for P. When varying κ_P, we take $(\mu_P, \sigma_P) = (\mu_T, \sigma_T)$ and, when varying μ_P or σ_P, we take $\kappa_P = 3$ because κ_T is not achievable for the Student t distribution. The figure shows that the objective function is more sensitive to the standard deviation than to the mean and the kurtosis. This is a desirable property because the standard deviation is subject to less estimation risk than the mean return and higher moments (Martellini and Ziemann 2010). In Section 4.3.1, we show theoretically that the KL divergence is more sensitive to the standard deviation than

Figure 3: Sensitivity to moments for the generalized-normal target distribution



Notes. The figure depicts how the objective function associated with the generalized-normal target distribution, $\mathcal{C}(w;\alpha,\beta,\gamma)$ in (18), evolves as we vary the portfolio-return mean μ_P, standard deviation σ_P, and kurtosis κ_P, assuming a Student t distribution for P. We match the parameters of the target distribution to the moments $(\mu_T,\sigma_T,\kappa_T) = (0.2, 0.2, -0.5)$. When varying κ_P, we take $(\mu_P,\sigma_P) = (\mu_T,\sigma_T)$ and, when varying μ_P or σ_P, we take $\mu_P = \mu_T$ or $\sigma_P = \sigma_T$ and $\kappa_P = 3$ because κ_T is not achievable for the Student t.

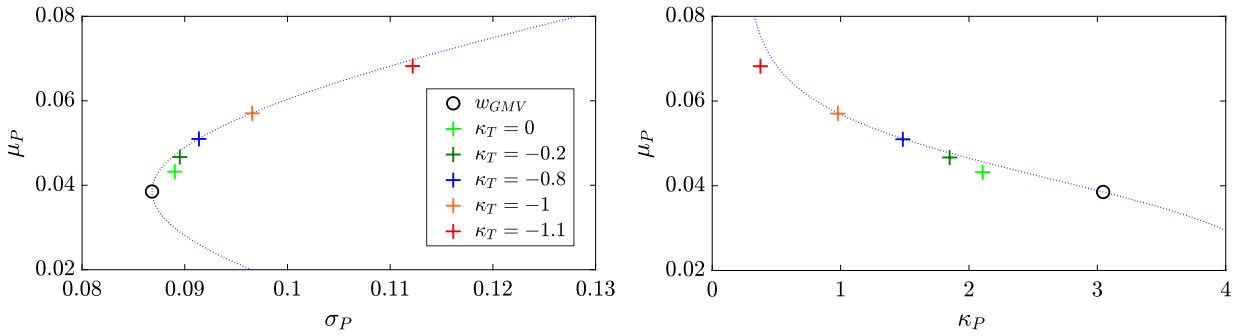
the mean for the Gaussian target. We conclude this section with an illustrative example.

Example 1. Consider $N = 3$ independent Student t asset returns with mean $(0.02, 0.07, 0.15)$, volatility $(0.10, 0.20, 0.30)$, and excess kurtosis $(8, 4, 0)$. As proposed in Section 3.3, we match the mean and variance of the GN target distribution to those of an MVE reference portfolio; namely, the GMV portfolio: $w_0 = w_{GMV} = (0.74, 0.18, 0.08)'$. The parameter γ is calibrated to an excess kurtosis κ_T between 0 and -1.1, which improves upon the excess kurtosis of GMV that equals 3.05. We observe that the lower the target kurtosis κ_T, the more the MCKL portfolio invests in the third asset that has the lowest kurtosis: $w_3 = 0.08$ for $\kappa_T = 0$ versus $w_3 = 0.30$ for $\kappa_T = -1.1$. In Figure 4, we depict the mean-variance-kurtosis tradeoff for the GMV portfolio (black circle), MCKL portfolios (colored crosses), and a set of mean-variance portfolios (dotted blue curve). We see that the smaller the κ_T, the more the MCKL portfolio departs from the GMV reference portfolio so as to improve the kurtosis.

4.2 The shifted-log-normal target distribution

Whereas the GN target distribution is symmetric, we now turn to an investor who targets an asymmetric distribution. For that purpose, we consider the SLN distribution; see Yuan (1933), Cohen (1951) for early studies, and Kaplan and Knowles (2004), Kadan and Liu (2014, Sec. 4.2.1) for financial illustrations. Contrary to the popular skew-normal distribution

Figure 4: Optimal portfolio for the generalized-normal target distribution



Notes. The figure depicts the results for the setting in Example 1. We consider $N = 3$ independent Student t asset returns with mean $(0.02, 0.07, 0.15)$, volatility $(0.10, 0.20, 0.30)$, and excess kurtosis $(8, 4, 0)$. We consider a generalized-normal target-return distribution, $T \sim \mathcal{GN}(\alpha, \beta, \gamma)$, whose parameters are calibrated by moment-matching considering the return mean and variance of the GMV portfolio, $w_{GMV} = (0.74, 0.18, 0.08)'$, and an excess kurtosis κ_T between 0 and -1.1. The two plots compare the mean-variance and mean-kurtosis tradeoff of the minimum-CKL and GMV portfolios, $w_{\alpha,\beta,\gamma}$ and w_{GMV}, as a function of κ_T. The blue dotted curves show the tradeoff for the mean-variance portfolios with mean return $\mu_P \in [0.02, 0.08]$.

of Azzalini (1985) and other similar distributions, the SLN distribution admits any value of skewness and belongs to the exponential family.

Definition 6 (Shifted-log-normal target)**.** *The target return follows an SLN distribution, $T \sim \mathcal{SLN}(\alpha, \beta, \delta)$, if its density is a log-normal density shifted by δ:*

$$f_T(x; \alpha, \beta, \delta) := \frac{1}{(x-\delta)\beta\sqrt{2\pi}} \exp\left(-\frac{1}{2\beta^2}\big(\ln(x-\delta) - \alpha\big)^2\right), \quad x > \delta, \tag{19}$$

where $\alpha, \delta \in \mathbb{R}$ and $\beta > 0$. The first three moments of T are given by

$$\begin{aligned} \mu_T &= \exp\left(\alpha + \beta^2/2\right) + \delta, \\ \sigma_T^2 &= g(\beta)\exp(2\alpha + \beta^2), \\ \zeta_T &= (g(\beta) + 3)\sqrt{g(\beta)}, \end{aligned} \tag{20}$$

where $g(\beta) := \exp(\beta^2) - 1$.

Note that the skewness ζ_T is always positive, but can be made negative by considering the density of $-T$ as target. The investor can calibrate the SLN parameters (α, β, δ) via moment-matching using results by Yuan (1933). In our notation, and with further development for

the formula of β, the parameters (α, β, δ) matching the moments $(\mu_T, \sigma_T^2, \zeta_T)$ are given by[5]

$$\begin{aligned}
&\beta = \sqrt{\ln(A + A^{-1} - 1)}, \quad A = \left(1 + \frac{\zeta_T^2}{2} + \sqrt{\frac{\zeta_T^4}{4} + \zeta_T^2}\right)^{1/3}, \\
&\alpha = \frac{1}{2} \ln\left(\frac{\sigma_T^2}{\exp(\beta^2) g(\beta)}\right), \\
&\delta = \mu_T - \sigma_T g(\beta)^{-1/2}.
\end{aligned} \tag{21}$$

Remark 1. From (21), it is easy to show that β, $g(\beta)$, and δ are increasing with the skewness ζ_T, while α decreases with ζ_T.

In the next proposition, we show that the Gaussian distribution is a special case of the SLN distribution when the SLN skewness tends to zero for a fixed mean and variance. This is a new result to the best of our knowledge.

Proposition 3 (Gaussian as a special case of SLN)**.** *Let (μ_T, σ_T) be fixed. Then,*

$$\lim_{\zeta_T \to 0^+} f_T(x; \alpha, \beta, \delta) = \lim_{\beta \to 0^+} f_T(x; \alpha, \beta, \delta) = \frac{1}{\sigma_T} \phi\left(\frac{x - \mu_T}{\sigma_T}\right),$$

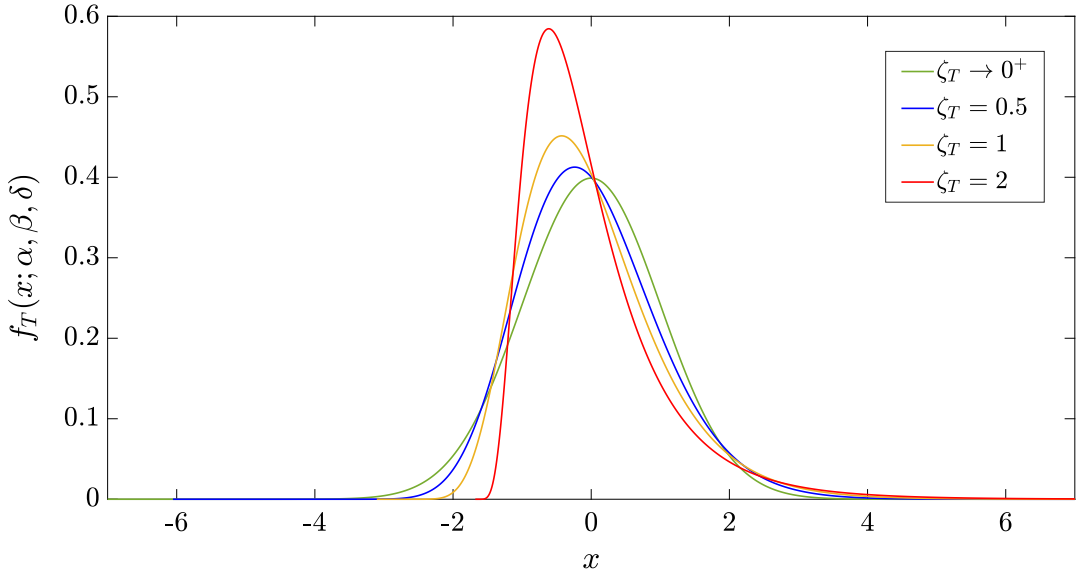
where ϕ is the standard Gaussian density.

In Figure 5, we depict the zero-mean unit-variance SLN density for different skewness ζ_T, and observe indeed that the distribution tends to a Gaussian as $\zeta_T \to 0^+$.

In the next proposition, we characterize the objective function in (14) associated with the SLN target distribution. Note that because $\Omega_T = \{x \in \mathbb{R} | x > \delta\}$, one cannot rely on the KL to quantify the divergence between f_P and f_T if, e.g., P is Gaussian. The CKL divergence, however, is suitable. Nevertheless, as we discuss below, the probability $p = \mathbb{P}(P \notin \Omega_T)$ is typically very small and, thus, we can obtain a very accurate approximation to the CKL-based objective function by assuming that $p = 0$.

[5]While the parameters (α, δ) are fitted to the mean and variance of a reference portfolio as in (15), the parameter β can be chosen in other ways than via moment-matching. For example, the quantile at confidence level ε of $T \sim \mathcal{SLN}(\alpha, \beta, \gamma)$ is $q_T(\varepsilon) = \exp(\alpha + \beta \Phi^{-1}(\varepsilon)) + \delta$, where Φ is the standard Gaussian cdf. Thus, β can be matched to a given quantile as $\beta = (\ln(q_T(\varepsilon) - \delta) - \alpha)/\Phi^{-1}(\varepsilon)$.

Figure 5: Density of the shifted-log-normal distribution



Notes. The figure depicts the shifted-log-normal density in (19) with mean zero, variance one, and skewness ζ_T between 0 and 2. As shown in Proposition 3, we recover the Gaussian density as $\zeta_T \to 0^+$.

Proposition 4 (MCKL portfolio with SLN target)**.** *Let* $T \sim \mathcal{SLN}(\alpha,\beta,\delta)$, $P_\delta := P - \delta$, $m(X) := \mathbb{E}[X|X \geq 0]$, $m_k(X) := \mathbb{E}\big[(X - m(X))^k | X \geq 0\big]$, *and* $\psi_k(X) = m_k(X)/m(X)^k$. *Then, the following holds concerning the MCKL portfolio* $w_{f_T} = w_{\alpha,\beta,\delta}$*:*

1. $w_{\alpha,\beta,\delta}$ *is obtained by maximizing*

$$\begin{aligned}\mathcal{C}(w;\alpha,\beta,\delta) = H[P|P \geq \delta] - \mathbb{E}[\ln P_\delta | P_\delta \geq 0] \\ - \frac{1}{2\beta^2}\left(\mathbb{V}[\ln P_\delta | P_\delta \geq 0] + (\mathbb{E}[\ln P_\delta | P_\delta \geq 0] - \alpha)^2\right) - \frac{p}{1-p},\end{aligned} \tag{22}$$

where the conditional expectation and variance of $\ln P_\delta$ *have second-order Taylor-series expansions equal to*

$$\mathbb{E}[\ln P_\delta | P_\delta \geq 0] \approx \ln(m(P_\delta)) - \psi_2(P_\delta)/2, \tag{23}$$

$$\mathbb{V}[\ln P_\delta | P_\delta \geq 0] \approx \psi_2(P_\delta) - \psi_3(P_\delta) + \psi_4(P_\delta)/4 - \frac{\psi_2(P_\delta)}{4m(P_\delta)^2}. \tag{24}$$

2. *When* $p = \mathbb{P}(P < \delta)$ *is small and* $\mu_P > \delta$*, the objective function* $\mathcal{C}(w;\alpha,\beta,\delta)$ *can be*

approximated as

$$\begin{aligned}\widetilde{\mathcal{C}}(w;\alpha,\beta,\delta) := H[P^\star] + \ln\sigma_P - \widetilde{\mathbb{E}}[\ln P_\delta | P_\delta \geq 0] \\ - \frac{1}{2\beta^2}\left(\widetilde{\mathbb{V}}[\ln P_\delta | P_\delta \geq 0] + (\widetilde{\mathbb{E}}[\ln P_\delta | P_\delta \geq 0] - \alpha)^2\right),\end{aligned} \tag{25}$$

where

$$\widetilde{\mathbb{E}}[\ln P_\delta | P_\delta \geq 0] := \ln(\mu_P - \delta) - \frac{\sigma_P^2}{2(\mu_P-\delta)^2}, \tag{26}$$

$$\widetilde{\mathbb{V}}[\ln P_\delta | P_\delta \geq 0] := \frac{\sigma_P^2}{(\mu_P-\delta)^2} - \frac{m_{3,P}}{(\mu_P-\delta)^3} + \frac{m_{4,P}-\sigma_P^4}{4(\mu_P-\delta)^4}. \tag{27}$$

The approximate MCKL portfolio that maximizes $\widetilde{\mathcal{C}}(w;\alpha,\beta,\delta)$ *is denoted* $\widetilde{w}_{\alpha,\beta,\delta}$.

3. *Let* $(\mu_P,\sigma_P) = (\mu_T,\sigma_T)$ *be fixed. Then, the skewness* ζ_P *and excess kurtosis* κ_P *maximizing* $\widetilde{\mathcal{C}}(w;\alpha,\beta,\delta)$ *are given by*

$$\zeta_P = 3\beta^{-2}g(\beta)^{3/2}, \tag{28}$$

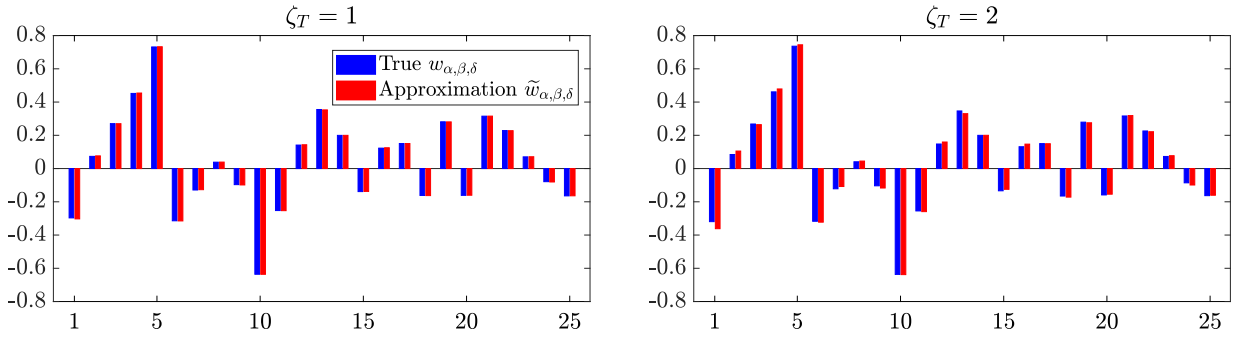
$$\kappa_P = -3\beta^{-2}g(\beta)^2 < 0, \tag{29}$$

provided they are attainable, where ζ_P *is close to* ζ_T *from* (21).[6]

Several comments are in order. First, the approximations introduced in part 2 of Proposition 4—Taylor-series expansion and small p—are accurate when the skewness ζ_T of the SLN target distribution is not too extreme. Indeed, the smaller the ζ_T, the smaller the δ and the more accurate the approximations, which become exact when $\zeta_T \to 0^+$, corresponding to a Gaussian target. Specifically, δ is typically many standard deviations below the target-return mean: the quantity $(\mu_T-\delta)/\sigma_T = g(\beta)^{-1/2}$, which decreases with the skewness ζ_T from Remark 1, varies between 30 and 3.1 for ζ_T between 0.1 and 1. In Section IA.2 of the internet appendix, we consider the special case in which P is Gaussian and we derive closed-form expressions for the true objective function, the one approximated via a Taylor-series expansion, and the one approximated via a Taylor-series expansion plus assuming p is small.

[6]Specifically, taking β as a function of ζ_T in (21), the first-order Taylor-series expansion of ζ_P around $\zeta_T = 0$ is $\zeta_P = \zeta_T$. This first-order approximation stays accurate even far from $\zeta_T = 0$; for example, applying formula (28), $\zeta_P = (1.016, 2.090)$ when $\zeta_T = (1,2)$.

Figure 6: Accuracy of approximation to the MCKL portfolio when targeting a shifted-log-normal distribution



Notes. The figure depicts the portfolio weights maximizing the true objective function for the shifted-log-normal target-return distribution (in blue), corresponding to $\mathcal{C}(w;\alpha,\beta,\delta)$ in (22), compared to those maximizing the approximation of the objective function via a Taylor-series expansion and assuming that the probability $p = \mathbb{P}(P \leq \delta)$ is small (in red), corresponding to $\widetilde{\mathcal{C}}(w;\alpha,\beta,\delta)$ in (25). We consider the special case in which the portfolio return P has a Gaussian distribution, for which we derived closed-form expressions for the two objective functions in Section IA.2 of the internet appendix. The mean and covariance matrix of asset returns are calibrated to a dataset of daily returns on the 25 portfolios of stocks sorted on size and book-to-market spanning January 1980 to March 2021. The return mean and variance of the target are matched to those of the global-minimum-variance portfolio, and the target-return skewness equals $\zeta_T = 1$ (left plot) and $\zeta_T = 2$ (right plot).

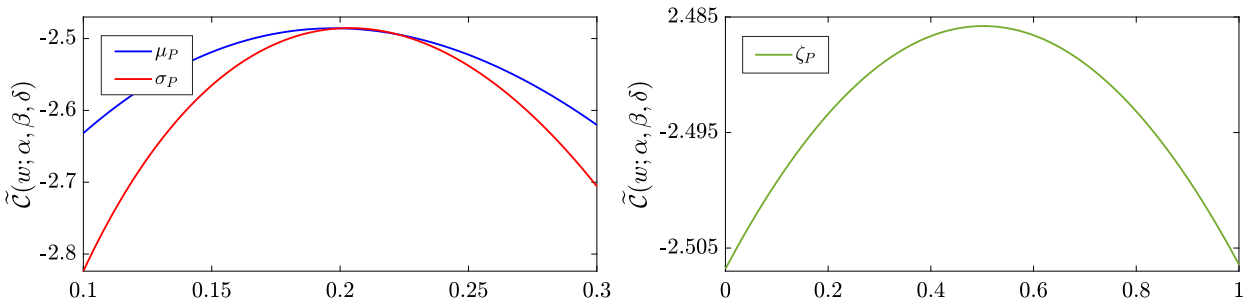
The latter typically holds in practice because, for $P \sim \mathcal{N}(\mu_T, \sigma_T)$,

$$p = \mathbb{P}(P < \delta) = \Phi\left(\frac{\delta - \mu_T}{\sigma_T}\right) = 1 - \Phi(g(\beta)^{-1/2}),$$

which, as discussed above, is small when ζ_T is not too large (e.g., $p < 0.1\%$ when $\zeta_T \leq 1$). We find that the Taylor approximation is accurate as long as ζ_T is not too large, and that the small-p approximation introduces a very negligible error. In Figure 6, we use these closed-form expressions to compare the MCKL portfolio maximizing the true objective function $\mathcal{C}(w;\alpha,\beta,\delta)$ in (22) to the MCKL portfolio maximizing the approximation $\widetilde{\mathcal{C}}(w;\alpha,\beta,\delta)$ in (25). To do so, we calibrate μ and $\mathbf{\Sigma}$ to a dataset of daily returns on the 25 portfolios of stocks sorted on size and book-to-market spanning January 1980 to March 2021, and we consider GMV as a reference portfolio and a large target-return skewness of $\zeta_T = 1$ and 2. The figure shows that the two MCKL portfolios are nearly identical, indicating the validity of our proposed approximation. In the sequel, we always rely on the objective function $\widetilde{\mathcal{C}}(w;\alpha,\beta,\delta)$, which is a simple function of the portfolio-return first four moments and entropy.

Second, part 3 of the proposition indicates that targeting an SLN distribution with skew-

Figure 7: Sensitivity to moments for the shifted-log-normal target distribution



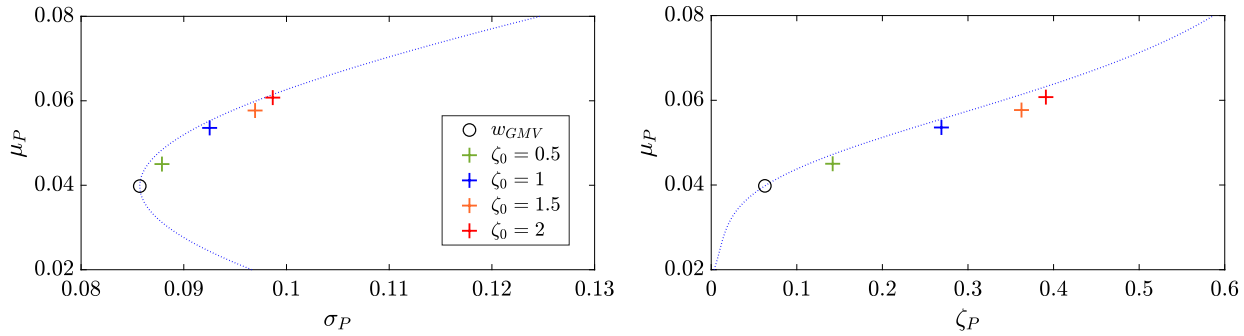
Notes. The figure depicts how the objective function $\widetilde{\mathcal{C}}(w;\alpha,\beta,\delta)$ associated with the shifted-log-normal target distribution evolves as we vary the portfolio-return mean μ_P, standard deviation σ_P, and skewness ζ_P. We match the parameters of the target distribution to the moments $(\mu_T,\sigma_T,\zeta_T) = (0.2, 0.2, 0.5)$. We vary each moment (μ_P,σ_P,ζ_P) individually while taking the other two equal to the target-return moments and fixing a portfolio-return excess kurtosis of $\kappa_P = 3$.

ness ζ_T indeed amounts to partially fitting ζ_T, but also to aim for a negative excess kurtosis, which together reduce tail risk. In practice, there is no reason to have $(\mu_P,\sigma_P) = (\mu_T,\sigma_T)$ even when considering the MCKL portfolio because $\widetilde{w}_{\alpha,\beta,\delta}$ trades off between the various moments. In Section 4.4, we derive the moments (μ_P,σ_P) of the MCKL portfolio when asset returns are Gaussian.

Third, we evaluate the moment sensitivity of the objective function $\widetilde{\mathcal{C}}(w;\alpha,\beta,\delta)$. We consider $(\mu_T,\sigma_T,\zeta_T) = (0.2, 0.2, 0.5)$ and depict in Figure 7 how the objective function evolves as we vary each moment (μ_P,σ_P,ζ_P) individually while taking the other two equal to the target-return moments and fixing $\kappa_P = 3$. The figure shows that the objective function is most sensitive to the standard deviation, then the mean, and then the skewness. As noted in Section 4.1, this is desirable in terms of estimation risk.

Example 2. Consider $N = 3$ independent asset returns with the same mean and variance as in Example 1, skewness $(0, 0.5, 1)$, and excess kurtosis $(1, 2, 3)$. We match the mean and variance of the SLN target distribution with those of GMV as a reference portfolio via (21), whereas the skewness ζ_T varies between 0.5 and 2, which improves upon the skewness of the GMV portfolio that equals 0.06. We observe that the higher the target skewness ζ_T, the more the MCKL portfolio invests in the third asset that has the largest skewness: $w_3 = 0.09$ for $\zeta_T = 0.5$ versus $w_3 = 0.19$ for $\zeta_T = 1$. In Figure 8, we depict the mean-variance-skewness tradeoff of the GMV and MCKL portfolios. We see that the higher the ζ_T, the more the

Figure 8: Optimal portfolio for the shifted-log-normal target distribution



Notes. The figure depicts the results for the setting in Example 2. We consider $N = 3$ independent asset returns with mean $(0.02, 0.07, 0.15)$, volatility $(0.10, 0.20, 0.30)$, skewness $(0, 0.5, 1)$, and excess kurtosis $(1, 2, 3)$. We take as a reference portfolio the GMV portfolio: $w_0 = w_{GMV} = (0.74, 0.18, 0.08)'$. We consider a shifted-log-normal target-return distribution, $T \sim \mathcal{SLN}(\alpha, \beta, \delta)$, whose parameters are calibrated by moment-matching considering the return mean and variance of the GMV portfolio, $w_{GMV} = (0.74, 0.18, 0.08)'$, and a skewness ζ_T between 0.5 and 2. The two plots depict the mean-variance-skewness tradeoff of the MCKL and GMV portfolios as a function of ζ_T. The blue dotted curves depict the tradeoff for the mean-variance portfolios with mean return $\mu_P \in [0.02, 0.08]$.

MCKL portfolio departs from the GMV reference portfolio so as to improve the skewness.

4.3 The Gaussian and Dirac-delta target distributions

We now study two special cases. First, the Gaussian target distribution, which is obtained as an SLN distribution with $\beta \to 0^+$ and a GN distribution with $\gamma = 2$. Second, the Dirac-delta target distribution, which is obtained as a Gaussian distribution with volatility $\sigma_T \to 0^+$.

4.3.1 The Gaussian target distribution

Consider a Gaussian target-return distribution, $T \sim \mathcal{N}(\alpha, \beta)$, with density

$$f_T(x; \alpha, \beta) = \frac{1}{\beta}\phi\left(\frac{x - \alpha}{\beta}\right). \tag{30}$$

We assume that $\beta > 0$; the case $\beta \to 0^+$ is addressed in Section 4.3.2. This first special case represents an appealing target because in practice the higher moments of returns are typically negatively skewed and leptokurtic, and thus, are less desirable than those of the Gaussian distribution. As we show in the next proposition, the resulting objective function

admits an insightful decomposition into three separate terms that measure the fit to the target-return mean, variance, and higher moments, respectively.

Proposition 5 (MCKL portfolio with Gaussian target)**.** *Let $T \sim \mathcal{N}(\alpha, \beta)$. Then, the MCKL portfolio $w_{f_T} = w_{\alpha,\beta}$ is obtained by maximizing*

$$\mathcal{C}(w; \alpha, \beta) = -\frac{1}{2}\left(\frac{\mu_P - \alpha}{\beta}\right)^2 + \frac{1}{2}\left(\ln \sigma_P^2 - \frac{\sigma_P^2}{\beta^2}\right) + H[P^\star]. \tag{31}$$

This decomposition highlights a number of interesting properties of the CKL divergence, which corresponds to the KL divergence in this case. First, the first two terms in (31) measure the fit to the target-return mean and variance, respectively. The first one is maximized when $\mu_P = \alpha$, and the second one when $\sigma_P = \beta$. Therefore, when employing the reference-portfolio approach proposed in Section 3.3, the sum of the first two terms is maximized for $w = w_0$.

Second, the KL divergence penalizes a misfit to the standard deviation β more strongly than a misfit to the mean α. To see this, observe that maximizing the first two terms in (31) amounts to minimizing the squared ℓ_2 distance between (μ_P, σ_P) and (α, β) plus an extra penalty on the variance of the form $2\beta(\sigma_P - \beta \ln \sigma_P)$ that is minimized for $\sigma_P = \beta$. This is also visible in the top two plots of Figure 9, described in Example 3, which show that the objective function is more sensitive to the variance than the mean. This may be desirable in practice given the large estimation errors in sample mean returns (Merton 1980).

Third, the additional entropy term, $H[P^\star]$, is invariant to μ_P and σ_P. It is a well-known result from information theory that, for a fixed variance, the Gaussian distribution has the largest entropy among all other distributions defined on the support set $\mathbb{R}$ (Cover and Thomas 2006). In other words, $H[P^\star]$ is maximized when $P^\star \sim \mathcal{N}(0, 1)$. Therefore, the KL objective function does not merely try to fit (μ_P, σ_P) to (μ_T, σ_T) but also to produce portfolio returns whose higher moments are close to those of the Gaussian distribution. This can be seen from the second-order Gram-Charlier expansion of the entropy (Hyvärinen et al. 2001):

$$H[P^\star] \approx H\big[\mathcal{N}(0,1)\big] - \frac{\zeta_P^2}{12} - \frac{\kappa_P^2}{48}. \tag{32}$$

The right-hand side of (32) is maximized when the skewness and excess kurtosis of the port-

folio return equal zero. Note that other papers use entropy as a higher-moment criterion for portfolio selection. Vermorken et al. (2012) and Flores et al. (2017) construct a diversification index based on entropy and show that its maximization improves skewness and kurtosis. Lassance and Vrins (2021) propose to minimize the Rényi entropy (a generalization of the Shannon entropy $H[P]$) of portfolio returns and show that it accounts for kurtosis. The way entropy appears in the objective function (31) differs from those papers because it is the standardized entropy, whose maximization shrinks skewness and excess kurtosis to zero.

In the next example, we illustrate the result in Proposition 5.

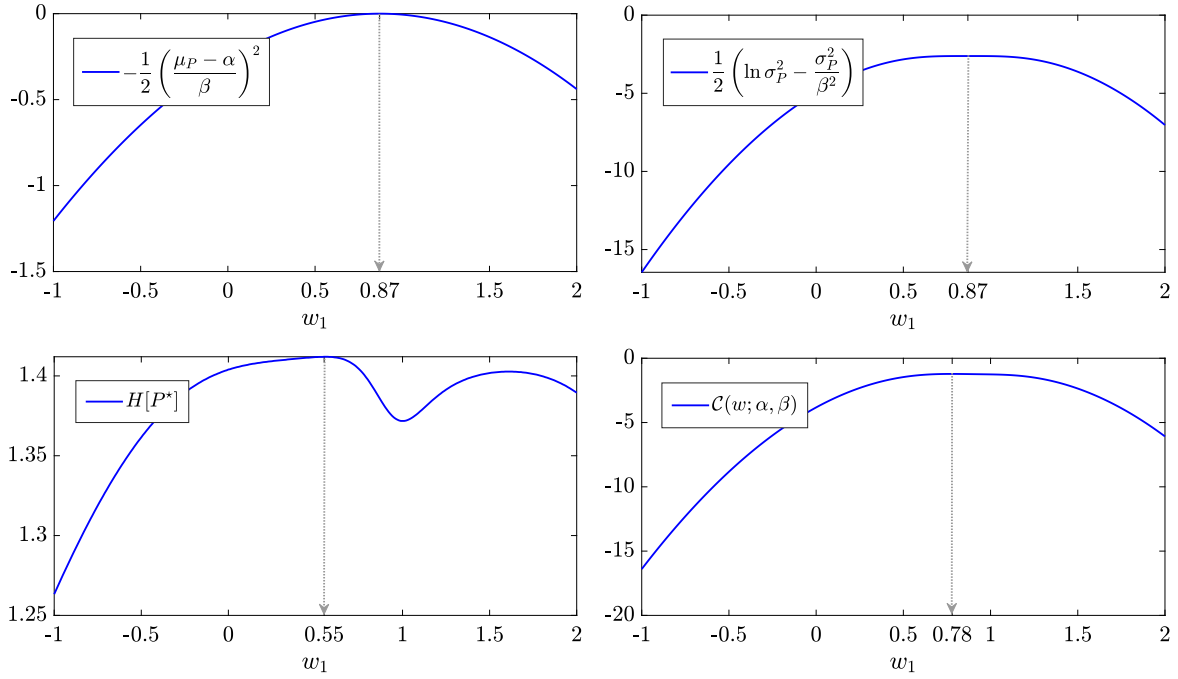
Example 3. Consider $N = 2$ independent Student t asset returns with the following parameters: $R_1 \sim t(\mu_1 = 0.05, \sigma_1 = 0.10, \nu_1 = 5)$ and $R_2 \sim t(\mu_2 = 0.15, \sigma_2 = 0.30, \nu_2 = 10)$. The corresponding excess kurtosis are $\kappa_{R_1} = 6$ and $\kappa_{R_2} = 1$, respectively. The investor targets a Gaussian distribution with (α, β) matching the return mean and variance of GMV as a reference portfolio, $w_0 = w_{GMV} = (0.87, 0.13)'$. In Figure 9, we depict the three terms in the objective function (31) as a function of the weight on the first asset, w_1. The MCKL portfolio would coincide with the GMV portfolio if we ignored the entropy term. However, the GMV portfolio is concentrated on the first asset, which has the largest return kurtosis and a low entropy $H[P^\star]$. In contrast, the portfolio maximizing the entropy $H[P^\star]$, $w = (0.55, 0.45)'$, is less exposed to the first asset. As a result, the MCKL portfolio $w_{\alpha,\beta} = (0.78, 0.22)'$ balances between the GMV and maximum-entropy portfolios and achieves a smaller excess kurtosis (2.66 versus 4.57 for GMV) for only a small increase in volatility (12.48% versus 12.05%).

4.3.2 The Dirac-delta target distribution

As a second special case, we investigate the Dirac-delta distribution, which, as we show, allows us to recover the MVE frontier via a novel parametrization and is related to the notion of stochastic dominance.

The decomposition of the KL divergence in Proposition 5 allows us to study the limit case in which the investor targets a distribution with zero variance; that is, $\beta = \sigma_T \to 0^+$. This corresponds to a Dirac-delta density centered in α, denoted by δ_α. From (31), it is useful

Figure 9: Decomposition of the objective function for the Gaussian target distribution



Notes. The figure depicts the three terms composing the objective function in the case of a Gaussian target-return distribution, as derived in Proposition 5. The setting is the one of Example 3, where we consider a portfolio of two independent assets following a Student t distribution: $X_1 \sim t(\mu_1 = 0.05, \sigma_1 = 0.10, \nu_1 = 5)$ and $X_2 \sim t(\mu_2 = 0.15, \sigma_2 = 0.30, \nu_2 = 10)$. The mean and variance of the Gaussian target return, $T \sim \mathcal{N}(\alpha, \beta)$, match those of the GMV portfolio $w_{GMV} = (0.87, 0.13)'$. The curves are drawn as a function of w_1, the portfolio weight on the first asset. The grey arrows indicate the location of the maximum for each curve.

to observe that multiplying $\mathcal{C}(w;\alpha,\beta)$ by β^2 gives the new objective function

$$\breve{\mathcal{C}}(w;\alpha,\beta) := \beta^2\mathcal{C}(w;\alpha,\beta) = -\frac{1}{2}\left((\mu_P-\alpha)^2+\sigma_P^2\right) + \beta^2 H[P]. \tag{33}$$

Clearly, maximizing $\breve{\mathcal{C}}(w;\alpha,\beta)$ amounts to maximizing $\mathcal{C}(w;\alpha,\beta)$ for all $\beta > 0$, but $\breve{\mathcal{C}}(w;\alpha,\beta)$ does not diverge as $\beta \to 0^+$. To deal with the Dirac-delta target, we define the objective function

$$\mathcal{C}(w;\alpha) = \mathcal{C}(w;\delta_\alpha) := \lim_{\beta\to 0^+} \breve{\mathcal{C}}(w;\alpha,\beta) = -(\mu_P-\alpha)^2 - \sigma_P^2, \tag{34}$$

leading to the MCKL portfolio $w_{\delta_\alpha} = w_\alpha$. The objective function $\mathcal{C}(w;\alpha)$ can be interpreted as the ℓ_2 distance between (μ_P, σ_P) and $(\alpha, 0)$, and also as the mean square error $\mathbb{E}\left[(P-\alpha)^2\right]$. In the next proposition, we derive the analytical solution for w_α, show that it is located on the MVE frontier if $\alpha \geq \mu_g$, and establish that, for a specific Dirac-delta mean, it cannot be dominated by any norm-constrained portfolio in terms of stochastic dominance of the third

order (Whitmore 1970, Post and Kopa 2016).

Proposition 6 (MCKL portfolio with Dirac-delta target)**.** *The MCKL portfolio w_α satisfies the following:*

1. *w_α is a mean-variance portfolio with mean return given by*

$$w_\alpha' \mu = \mu_g + (\alpha - \mu_g) \frac{\psi^2 - \psi_g^2}{1 + \psi^2 - \psi_g^2}. \tag{35}$$

2. *$w_\alpha = w_{GMV}$ if $\alpha = \mu_g$ and is mean-variance efficient if and only if $\alpha \geq \mu_g$.*

3. *Let the asset returns be bounded, $-\infty < r_{\min,i} \leq R_i \leq r_{\max,i} < \infty$ for all i. Let $q \geq 1$ and $K \geq N^{(1-q)/q}$. Then, w_α associated with the target mean*

$$\alpha = \max_{\substack{w \in \mathcal{W} \\ ||w||_q \leq K}} \sum_{i:w_i<0} w_i r_{\min,i} + \sum_{i:w_i \geq 0} w_i r_{\max,i} \tag{36}$$

cannot be dominated by any portfolio $w \in \mathcal{W}$ satisfying $||w||_q \leq K$ in the sense of third-order stochastic dominance.[7]

Part 1 of Proposition 6 shows that, when the Dirac-delta mean $\alpha \geq \mu_g$, the mean return of the MCKL portfolio w_α is located in between μ_g and α, which is because the objective (34) amounts to fitting the mean return α and minimizing the portfolio-return variance σ_P^2.

As shown in part 3 of Proposition 6 using results in Le Courtois and Xu (2020), this relation complies with the notion of stochastic dominance. Specifically, for α chosen according to (36), no q-norm-constrained portfolio can dominate the MCKL portfolio in the sense of third-order stochastic dominance. This means that no q-norm-constrained portfolio is preferred over the MCKL portfolio by all investors with an increasingly concave utility function, which represents a positive preference for mean and skewness and a negative preference for variance. This is true in particular for the popular ℓ_1-norm ($q = 1$) and ℓ_2-norm-constrained ($q = 2$) mean-variance portfolios; see, for instance, DeMiguel et al. (2009), Ao et al. (2019). Note that the result holds as well for stochastic dominance of the first and second order since they represent a wider class of utility functions; see the textbook of Levy (2016).

[7] A portfolio w^i dominates a portfolio w^j in terms of third-order stochastic dominance if $\mathbb{E}[U(w^i)] \geq \mathbb{E}[U(w^j)]$ for all utility functions U satisfying $U' \geq 0$, $U'' \leq 0$, and $U''' \geq 0$.

4.4 Gaussian asset returns and mean-variance efficiency

In this section, we analyze the properties of the MCKL portfolio associated with the GN and SLN target distributions in the classical setting in which the asset returns are assumed Gaussian. We first derive the moments of the Gaussian distribution minimizing the CKL divergence, which may not be attainable. Using this result, we then show that when the target-return mean and variance are located on or above the MVE frontier, the MCKL portfolio is always located on the MVE frontier, with specific bounds on the mean return when targeting a Gaussian distribution.

In the next proposition, we begin by deriving the moments of the Gaussian distribution minimizing the CKL divergence with respect to the SLN and GN target distributions.

Proposition 7 (Gaussian distributions with minimum CKL)**.** *The Gaussian distributions with minimum CKL divergence satisfy the following:*

1. *The moments of the Gaussian distribution minimizing the CKL divergence with respect to the density of* $T \sim \mathcal{GN}(\alpha, \beta, \gamma)$ *satisfy*

$$(\mu_{GN}^{\star}, \sigma_{GN}^{\star}) = (\mu_T, C_1(\gamma)\sigma_T)\,, \tag{37}$$

where

$$C_1(\gamma) := \sqrt{\frac{\Gamma(1/\gamma)}{2\Gamma(3/\gamma)}} \left(\frac{\Gamma\left(\frac{\gamma+1}{2}\right)\gamma}{\pi^{1/2}} \right)^{-1/\gamma} \leq 1, \tag{38}$$

with equality if and only if $\gamma = 2$.

2. *The moments of the Gaussian distribution minimizing the CKL divergence with respect to the density of* $T \sim \mathcal{SLN}(\alpha, \beta, \delta)$ *according to the approximation* (25) *satisfy*

$$(\mu_{SLN}^{\star}, \sigma_{SLN}^{\star}) := (\mu_T, C_2(\beta)\sigma_T), \tag{39}$$

where

$$C_2(\beta) := \beta g(\beta)^{-1/2} \leq 1, \tag{40}$$

with equality if and only if $\beta \to 0^+$.

We conclude from this result that if there exist portfolios $w \in \mathcal{W}$ corresponding to the optimal Gaussian distributions in (37) and (39), they will coincide with the MCKL portfolios $w_{\alpha,\beta,\gamma}$ and $\widetilde{w}_{\alpha,\beta,\delta}$, respectively. This is formalized in the next corollary.

Corollary 1. *Let the asset returns be multivariate Gaussian, $R \sim \mathcal{N}(\mu, \mathbf{\Sigma})$. Then, the MCKL portfolio $w_{\alpha,\beta,\gamma}$ satisfies $P \sim \mathcal{N}(\mu_T, \sigma^\star_{GN})$ provided that $(\mu_T, \sigma^\star_{GN}) \in \mathcal{MV}$. Similarly, the MCKL portfolio $\widetilde{w}_{\alpha,\beta,\delta}$ satisfies $P \sim \mathcal{N}(\mu_T, \sigma^\star_{SLN})$ provided that $(\mu_T, \sigma^\star_{SLN}) \in \mathcal{MV}$.*

Notice that Corollary 1 remains silent about what are the properties of the MCKL portfolio associated with Gaussian asset returns when choosing a GN or SLN target distribution whose first two moments are located on or above the MVE frontier. Indeed, in such a case, neither $(\mu_T, \sigma^\star_{SLN})$ nor $(\mu_T, \sigma^\star_{GN})$ belong to $\mathcal{MV}$ (except for a Gaussian target and (μ_T, σ_T) on the MVE frontier), which means that no portfolio $w \in \mathcal{W}$ delivers those moments. Choosing the target-return mean and variance on the MVE frontier is in line with the reference-portfolio approach in Section 3.3. A classical example in which the investor may be drawn to choose the target-return mean and variance above the MVE frontier is when she restricts to investing in only a subset of all considered assets (e.g., after a filtering based on ESG scores), and wishes to target the performance of the optimal portfolio computed from all assets. The next proposition guarantees that the MCKL portfolio remains MVE in such a case, for any GN and SLN target distribution.

Proposition 8 (MCKL portfolio with Gaussian asset returns)**.** *Let $R \sim \mathcal{N}(\mu, \mathbf{\Sigma})$ and the target-return mean and variance be located on or above the mean-variance efficient frontier: $\mu_T \geq \mu_g$ and $\sigma_T \leq \sigma_{EF}(\mu_T)$. Then, the MCKL portfolios $w_{\alpha,\beta,\gamma}$, $\widetilde{w}_{\alpha,\beta,\delta}$, and $w_{\alpha,\beta}$ satisfy the following:*

1. *$w_{\alpha,\beta,\gamma}$, $\widetilde{w}_{\alpha,\beta,\delta}$, and $w_{\alpha,\beta}$ are mean-variance efficient.*

2. *Focusing on $w_{\alpha,\beta}$ associated with a Gaussian target, the mean return $\mu^\star_P = w'_{\alpha,\beta}\mu$ solves*

$$(\mu^\star_P - \mu_g)(\sigma^{\star 2}_P - \beta^2) = (\alpha - \mu^\star_P)(\psi^2 - \psi^2_g)\sigma^{\star 2}_P \quad \textit{with} \quad \sigma^\star_P = \sigma_{EF}(\mu^\star_P). \tag{41}$$

Moreover, if μ_{ℓ_2} is the mean return of the mean-variance efficient portfolio minimizing the ℓ_2 distance to (α, β), we have the following:

(i) $\mu_P^\star = \mu_g$ *if* $\alpha = \mu_g$,

(ii) $\sigma_{EF}^{-1}(\beta) \leq \mu_P^\star \leq \mu_{\ell_2} \leq \alpha$ *if* $\alpha > \mu_g$, *where* $\sigma_{EF}^{-1}(\beta)$ *has to be replaced by* μ_g *if* $\beta < \sigma_g$, *and with equalities if and only if* $\beta = \sigma_{EF}(\alpha)$.

Several comments are in order. First, Proposition 8 shows that we recover portfolios located on the MVE frontier when asset returns are Gaussian, which is a desirable result. The core of the proof consists in showing that the CKL divergence has no other stationary point in the mean-variance plane than the global optimum in Proposition 7, which is desirable from an optimization perspective. This implies geometrically that the MCKL portfolio is MVE because the global optimum is located above the MVE frontier when so is (μ_T, σ_T).

Second, similar results to those in Proposition 8 hold by symmetry when choosing (μ_T, σ_T) below the lower branch of the mean-variance hyperbola. However, such a choice is less sensible for the investor.

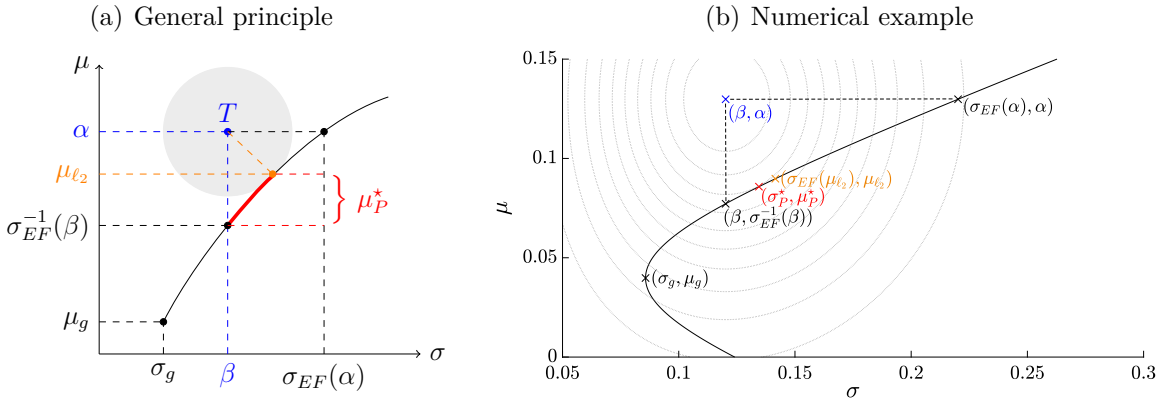
Third, the results in Proposition 8 hold when asset returns are Gaussian. For non-Gaussian returns, the MCKL portfolio will no longer be MVE if the CKL divergence deems that portfolios outside of the MVE frontier provide a better balance between fitting the target-return mean-variance and higher moments, as in Figures 4 and 8.

Fourth, when targeting a Gaussian distribution, we have more specific bounds on the mean return of the MCKL portfolio, $\mu_P^\star$, as shown in part 2 of Proposition 8. In particular, the CKL divergence induces a larger fit to the target variance than to the target mean relative to the ℓ_2 distance: $\mu_P^\star \leq \mu_{\ell_2}$. This phenomenon is depicted in Figure 10a.

We conclude this section with an illustrative example of Proposition 8.

Example 4. Consider $N = 3$ Gaussian asset returns with the same mean and covariance matrix as in Examples 1 and 2. We consider the Gaussian target, the GN target with $\gamma = 4$ (excess kurtosis $\kappa_T = -0.81$), and the SLN target with a skewness $\zeta_T = 1$. The target-return mean and variance are above the MVE frontier: $\mu_T = \mu_g + 0.09$, $\sigma_T = \sigma_{EF}(\mu_T) - 0.1$. In Figure 10b, we depict contour lines of the objective function for the Gaussian target, which cross the MVE frontier for the first time at the MCKL portfolio whose mean return $\mu_P^\star$ is below the minimum-ℓ_2-distance mean return μ_{ℓ_2}. Then, in Figure 11, we depict contour lines of the objective function for the GN and SLN target distributions. As established in

Figure 10: Mean-variance efficiency of the optimal portfolio for Gaussian asset returns and a Gaussian target



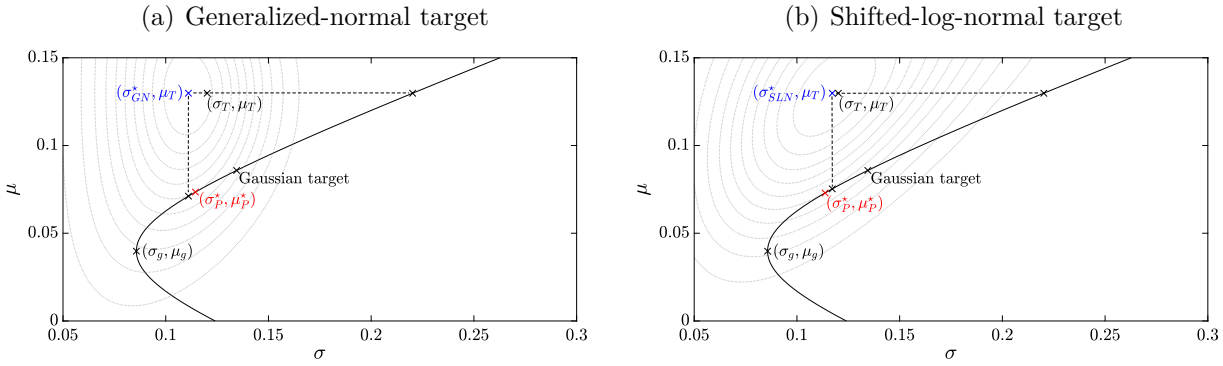
Notes. The figure illustrates the theoretical result in part 2 of Proposition 8, which considers the case in which asset returns are Gaussian and the investor targets a Gaussian distribution $\mathcal{N}(\alpha, \beta)$. In this figure, (α, β) are not attainable and $\alpha > \mu_g$, which corresponds to part 2(ii) of the proposition. In that case, the MCKL portfolio is a mean-variance efficient portfolio whose mean return $\mu_P^\star$ satisfies $\sigma_{EF}^{-1}(\beta) < \mu_P^\star < \mu_{\ell_2} < \alpha$, where μ_{ℓ_2} is the mean return of the mean-variance efficient portfolio minimizing the ℓ_2 distance to (α, β). This result is illustrated in a general fashion in plot (a), and for a numerical example in plot (b) with $N = 3$ independent asset returns with mean $(0.02, 0.07, 0.15)$, volatility $(0.10, 0.20, 0.30)$, and target-return moments $\alpha = \mu_g + 0.09$, $\beta = \sigma_{EF}(\alpha) - 0.1$. The contour lines are displayed for $\exp(4\mathcal{C}(w; \alpha, \beta))$ to reduce the flatness of the objective function $\mathcal{C}(w; \alpha, \beta)$ in (31) around the optimum.

part 1 Proposition 8, the MCKL portfolio is MVE for both targets. Contrary to the Gaussian target, the optimal volatility $\sigma_P^\star$ can be smaller than the global optimum; see Figure 11b. The contour lines have a different shape than those for the Gaussian target and induce a smaller return variance for the MCKL portfolio.

5 Estimation method

In this section, we show how to estimate the objective functions associated with the GN, SLN, and Gaussian target distributions considered in Section 4. For that purpose, we consider that the investor has collected a sample of historical asset-return observations $R_{i,t}$, with $i = 1, \ldots, N$ and $t = 1, \ldots, h$. We denote R_t the asset-return vector at time t and, for a given portfolio-weight vector w, the portfolio return at time t is denoted $P_t = w' R_t$. For all considered target distributions, we first estimate the MVE reference portfolio w_0 via $\hat{w}_0$ and

Figure 11: Mean-variance efficiency of the optimal portfolio for Gaussian asset returns and a generalized-normal and shifted-log-normal target



Notes. The figure illustrates the theoretical result in part 1 of Proposition 8, which considers the case in which asset returns are Gaussian and the investor targets a generalized-normal or a shifted-log-normal distribution whose mean and volatility (μ_T, σ_T) are above the mean-variance efficient frontier. In that case, the MCKL portfolio is a mean-variance efficient portfolio whose mean return $\mu^\star_P$ is given at the first point for which the contour lines of the objective function cross the mean-variance efficient frontier. The figure considers $N = 3$ independent asset returns with mean $(0.02, 0.07, 0.15)$ and volatility $(0.10, 0.20, 0.30)$, and a target-return mean $\mu_T = \mu_g + 0.09$ and volatility $\sigma_T = \sigma_{EF}(\mu_T) - 0.1$. The contour lines are displayed for $\exp(4\mathcal{C})$, where $\mathcal{C}$ is the objective function, to reduce the flatness around the optimum.

recover its return mean and variance,

$$\hat{\mu}_0 = \hat{w}_0' \hat{\mu} \quad \text{and} \quad \hat{\sigma}_0^2 = \hat{w}_0' \widehat{\mathbf{\Sigma}} \hat{w}_0, \tag{42}$$

where $\hat{\mu} := h^{-1} \sum_{t=1}^{h} R_t$ and $\widehat{\mathbf{\Sigma}} := (h-1)^{-1} \sum_{t=1}^{h} (R_t - \hat{\mu})(R_t - \hat{\mu})'$.

5.1 Shifted-log-normal target distribution

To estimate the objective function $\widetilde{\mathcal{C}}(w; \alpha, \beta, \delta)$ in part 2 of Proposition 4 associated with the SLN target distribution, we calibrate the parameters (α, δ) to $(\hat{\mu}_0, \hat{\sigma}_0)$ in (42),

$$\hat{\alpha} = \frac{1}{2} \ln \left(\frac{\hat{\sigma}_0^2}{\exp(\beta^2) g(\beta)} \right) \quad \text{and} \quad \hat{\delta} = \hat{\mu}_0 - \hat{\sigma}_0 g(\beta)^{-1/2}, \tag{43}$$

where β is fixed exogenously by the investor, e.g., to match a desired skewness ζ_T in (21). Then, using the property $H[P^\star] = H[P] - \ln \sigma_P$, the estimated objective function depends

on β and $\hat{w}_0$ as

$$\begin{aligned}
&\widehat{\mathcal{C}}(w;\beta,\hat{w}_0) = \widehat{H}[P] - \widehat{\mathbb{E}}[\ln P_\delta | P_\delta \geq 0] - \frac{1}{2\beta^2}\left(\left(\widehat{\mathbb{E}}[\ln P_\delta | P_\delta \geq 0] - \hat{\alpha}\right)^2 + \widehat{\mathbb{V}}[\ln P_\delta | P_\delta \geq 0]\right), \\
&\widehat{\mathbb{E}}[\ln P_\delta | P_\delta \geq 0] = \ln(\hat{\mu}_P - \hat{\delta}) - \frac{\hat{\sigma}_P^2}{2(\hat{\mu}_P - \hat{\delta})^2}, \\
&\widehat{\mathbb{V}}[\ln P_\delta | P_\delta \geq 0] = \frac{\hat{\sigma}_P^2}{(\hat{\mu}_P - \hat{\delta})^2} - \frac{\hat{m}_{3,P}}{(\hat{\mu}_P - \hat{\delta})^3} + \frac{\hat{m}_{4,P} - \hat{\sigma}_P^4}{4(\hat{\mu}_P - \hat{\delta})^4}.
\end{aligned} \tag{44}$$

The estimate (44) involves estimating the portfolio-return first four moments and entropy.

Concerning the moments, we rely on the classical unbiased sample estimates (Kendall and Stuart 1963):

$$\begin{aligned}
&\hat{\mu}_P = \frac{1}{h}\sum_{t=1}^{h} P_t, \quad \hat{\sigma}_P^2 = \frac{1}{h-1}\sum_{t=1}^{h}(P_t - \hat{\mu}_P)^2, \quad \hat{m}_{3,P} = \frac{h}{(h-1)(h-2)}\sum_{t=1}^{h}(P_t - \hat{\mu}_P)^3, \\
&\hat{m}_{4,P} = \frac{1}{h^2 - 3h + 3}\left(\frac{h^2}{h-1}\sum_{t=1}^{h}(P_t - \hat{\mu}_P)^4 - 3(2h-3)\hat{\sigma}_P^4\right).
\end{aligned} \tag{45}$$

We also consider the robust shrinkage estimators of the variance, skewness, and kurtosis in Ledoit and Wolf (2004), Martellini and Ziemann (2010), and Boudt et al. (2018). However, because we use daily return data in the empirical analysis of Section 6, we find that they perform as well as the sample estimates in (45).

Concerning the entropy $H[P]$, various estimators exist in the literature; see Beirlant et al. (1997) for a review. We rely on the m-spacings estimator because it is both robust and accurate; see Vasicek (1976), van Es (1992), and Learned-Miller and Fisher (2003).

Definition 7 (Entropy estimator)**.** *The m-spacings estimator of $H[P]$ is given by*

$$\widehat{H}[P] = \frac{1}{h-m}\sum_{t=1}^{h-m}\ln\left(\frac{h+1}{m}\left(P_{(t+m:h)} - P_{(t:h)}\right)\right) + \ln m - \psi(m), \tag{46}$$

where $P_{(1:h)} \leq P_{(2:h)} \leq \cdots \leq P_{(h:h)}$ are the order statistics, $m \in [1, h-1]$ is an integer, $\psi(m) := \Gamma'(m)/\Gamma(m)$ is the digamma function, and $\ln m - \psi(m)$ is a bias-correction term.

Increasing the parameter m improves the robustness to outliers and decreases the variance of $\widehat{H}[P]$. As shown by van Es (1992), this estimator is consistent under the two conditions

$m, h \to \infty$ and $m/h \to 0$. Therefore, m is commonly taken as a power function of h, and we fix $m = \left[h^{2/3}\right]$ as in Lassance and Vrins (2021a).

5.2 Generalized-normal target distribution

To estimate the objective function $\mathcal{C}(w; \alpha, \beta, \gamma)$ in Proposition 2 associated with the GN target distribution, we calibrate the parameters (α, β) to $(\hat{\mu}_0, \hat{\sigma}_0)$ in (42),

$$\hat{\alpha} = \hat{\mu}_0 \quad \text{and} \quad \hat{\beta} = \hat{\sigma}_0 2^{-1/\gamma} \sqrt{\frac{\Gamma(1/\gamma)}{\Gamma(3/\gamma)}}, \tag{47}$$

where γ is fixed exogenously by the investor, e.g., to match a desired excess kurtosis κ_T in (17). Then, the estimated objective function depends on γ and $\hat{w}_0$ as

$$\widehat{\mathcal{C}}(w; \gamma, \hat{w}_0) = \widehat{H}[P] - \frac{1}{2\hat{\beta}^\gamma} \widehat{\mathbb{E}}\left[|P - \hat{\alpha}|^\gamma\right], \tag{48}$$

where $\widehat{H}[P]$ is given by (46) and $\widehat{\mathbb{E}}\left[|P - \hat{\alpha}|^\gamma\right]$ is an estimate of $\mathbb{E}\left[|P - \hat{\alpha}|^\gamma\right]$.

To estimate the expectation $\mathbb{E}\left[|P - \hat{\alpha}|^\gamma\right]$, we rely on the Gaussian-mixture estimator of the portfolio-return density f_P, which is a popular estimation method that, as we show below, yields an analytical expression for the estimator $\widehat{\mathbb{E}}\left[|P - \hat{\alpha}|^\gamma\right]$. The Gaussian-mixture density estimator belongs to the class of finite-mixture estimators (Everitt and Hand 1981), is a generalization of the popular kernel density estimator, and can readily be implemented via the expectation-maximization algorithm (Dempster et al. 1977, Benaglia et al. 2009).

Definition 8 (Gaussian-mixture density estimator)**.** *The Gaussian-mixture estimator of f_P is given by*

$$\hat{f}_P^{GME}(x) := \sum_{k=1}^{K} \frac{\pi_k}{b_k} \phi\left(\frac{x - m_k}{b_k}\right), \tag{49}$$

where $\pi_k \in [0, 1]$ for all k and $\sum_{k=1}^{K} \pi_k = 1$. As a special case, the kernel density estimator is obtained for $\pi_k = 1/K$ for all k, $K = h$, and kernels centered on the observations P_t:

$$\widehat{f}_P^{KDE}(x) := \frac{1}{h} \sum_{t=1}^{h} \frac{1}{b_t} \phi\left(\frac{x - P_t}{b_t}\right). \tag{50}$$

Using the density estimator in (49), the expectation $\mathbb{E}\left[|P-\hat{\alpha}|^{\gamma}\right]$ is estimated as

$$\widehat{\mathbb{E}}\left[|P-\hat{\alpha}|^{\gamma}\right]=\int_{-\infty}^{+\infty}\hat{f}_{P}^{GME}(x)|x-\hat{\alpha}|^{\gamma}\mathrm{d}x, \tag{51}$$

and we provide a closed-form expression for this integral in the next proposition.

Proposition 9 (Estimator of objective function with GN target)**.** *The estimator* $\widehat{\mathbb{E}}\left[|P-\hat{\alpha}|^{\gamma}\right]$ *in* (51) *is given by*

$$\widehat{\mathbb{E}}\left[|P-\hat{\alpha}|^{\gamma}\right]=\frac{2^{\frac{\gamma-2}{2}}\Gamma\left(\frac{\gamma+1}{2}\right)}{\pi^{1/2}}\sum_{k=1}^{K}\pi_{k}b_{k}^{\gamma}\,{}_{1}F_{1}\left(-\frac{1}{2}\left(\frac{m_{k}-\hat{\alpha}}{b_{k}}\right)^{2};-\frac{\gamma}{2},\frac{1}{2}\right), \tag{52}$$

where ${}_{1}F_{1}$ *is the confluent hypergeometric function (Slater 1972):*

$${}_{1}F_{1}(x;c_{1},c_{2}):=\frac{\Gamma(c_{2})}{\Gamma(c_{1})}\sum_{n=0}^{\infty}\frac{\Gamma(c_{1}+n)}{\Gamma(c_{2}+n)}\frac{x^{n}}{n!}. \tag{53}$$

In the empirical analysis of Section 6, we rely on the kernel density estimator $\hat{f}_{P}^{KDE}$ in (50) with a constant bandwidth b based on the robust ratio-quantile-ranges approach of Zhang and Wang (2008); see Section IA.3 of the internet appendix for more details.

5.3 Gaussian target

In the special case of a Gaussian target, corresponding to $\beta\to 0^{+}$ in the SLN target or $\gamma=2$ in the GN target, the estimated objective function is obtained from Proposition 5 as

$$\widehat{\mathcal{C}}(w;\hat{w}_{0})=\widehat{H}[P]-\frac{(\hat{\mu}_{P}-\hat{\mu}_{0})^{2}+\hat{\sigma}_{P}^{2}}{2\hat{\sigma}_{0}^{2}}. \tag{54}$$

From Proposition 3, this estimate coincides with the estimate of the objective function for the SLN target in (44) as we let $\beta\to 0^{+}$. Moreover, as we show in the following proposition, it also coincides with the objective function for the GN target in (48) when $\gamma=2$ in case one relies on the kernel density estimator in (50) and a bandwidth function independent of the portfolio weights, which corresponds to the setting used in the empirical analysis.

Proposition 10 (Correspondence of estimators for GN and Gaussian targets)**.** *Let* $\hat{f}_{P}=$

$\hat{f}_P^{KDE}$ and the bandwidth b_t be independent from the portfolio w. Then, $\widehat{\mathcal{C}}(w; \gamma = 2, \hat{w}_0)$ in (48)–(52) corresponds to $\widehat{\mathcal{C}}(w; \hat{w}_0)$ in (54).

6 Empirical analysis

We now evaluate the performance of the MCKL portfolio strategy. In Section 6.1, we discuss the data we use, and in Section 6.2, the considered performance criteria. In Section 6.3, we discuss in-sample results. In Section 6.4, we present the methodology employed for the out-of-sample analysis, and we analyze the results in Section 6.5.

6.1 Data

We consider three commonly used characteristic-sorted equity portfolios from Ken French's library: (i) 25 portfolios formed on size and book-to-market (25SBTM), (ii) 25 portfolios formed on operating profitability and investment (25OPINV), and (iii) 30 industry portfolios (30IND). The time period spans January 1980 to March 2021. Characteristic-sorted portfolios reduce the size of the investment universe and thus estimation risk, which is desirable for portfolio optimization, particularly with higher moments. Finally, we consider daily returns because they improve estimation accuracy and because higher-moment risk is more prominent at the daily frequency (Martellini and Ziemann 2010).

6.2 Performance criteria

Given a time series of in-sample or out-of-sample portfolio returns $P_1, \ldots, P_\tau$, we evaluate portfolio performance in terms of moments, disaster risk, and CKL divergence.

First, we report the annualized mean $252\hat{\mu}_P$, annualized volatility $\hat{\sigma}_P\sqrt{252}$, skewness $\hat{\zeta}_P = \hat{m}_{3,P}/\hat{\sigma}_P^{3/2}$, and excess kurtosis $\hat{\kappa}_P = \hat{m}_{4,P}/\hat{\sigma}_P^2 - 3$, computed as in (45).

Second, we evaluate the disaster risk of the portfolio via the probability of a three-sigma loss as in Gormsen and Jensen (2020). Specifically, we estimate the portfolio-return density via the kernel density estimator $\hat{f}_P^{KDE}$ as in Section 5.2 and compute the probability of a

three-sigma loss as

$$\mathbb{P}(P < \hat{\mu}_P - 3\hat{\sigma}_P) = \int_{-\infty}^{\hat{\mu}_P - 3\hat{\sigma}_P} \hat{f}_P^{KDE}(x)\mathrm{d}x = \frac{1}{\tau}\sum_{t=1}^{\tau} \Phi\left(\frac{\hat{\mu}_P - 3\hat{\sigma}_P - P_t}{b}\right). \tag{55}$$

Third, we report the CKL divergence with respect to the three target-return distributions considered in this empirical analysis.[8] First, the Gaussian target studied in Section 4.3.1. Second, the GN target studied in Section 4.1 with a parameter $\gamma = 2.78$ corresponding to an excess kurtosis $\kappa_T = -0.5$. Third, the SLN target studied in Section 4.2 with a skewness $\zeta_T = 0.5$. This criterion is important to assess how well our considered objective function is optimized. Those values of target-return skewness and excess kurtosis are chosen so as to improve upon the higher moments of the Gaussian target. In unreported results, we find that the out-of-sample results discussed in Section 6.5 are not very sensitive to the specific chosen parameters; they remain similar when using a GN distribution with excess kurtosis $\kappa_T = -1$ and an SLN distribution with skewness $\zeta_T = 1$.

6.3 Discussion of in-sample results

To study the properties of the MCKL portfolio without being impacted by estimation risk, we first discuss in-sample results computed from all asset returns spanning January 1980 to March 2021. For conciseness, we restrict to the 25SBTM dataset.

We follow the estimation method of Section 5. We consider two MVE reference portfolios: the GMV portfolio and the mean-variance (MV) portfolio with a risk-aversion coefficient $\lambda = 10$. We also consider the three target distributions in Section 6.2. In Table 1, we report the in-sample performance according to the criteria in Section 6.2; the CKL divergence is reported with respect to the considered reference portfolio. The case with the GMV reference portfolio and a Gaussian target distribution corresponds to that in Figure 1.

The table shows that the MCKL portfolio trades off between the mean-variance efficiency of the reference portfolio and the improved higher moments of the target distribution. When considering GMV as a reference portfolio, the skewness is -0.482 for GMV, -0.223 for MCKL

[8] We add back the proportionality constant of the target-return density when computing the CKL divergence, which was ignored in the estimated objective functions of Section 5 because it does not affect the optimizer.

Table 1: In-sample portfolio performance for the 25SBTM dataset

	GMV reference portfolio				MV reference portfolio			
	GMV	MCKL portfolios			MV	MCKL portfolios		
		Gaussian	SLN	GN		Gaussian	SLN	GN
$252 \times \hat{\mu}_P$	0.186	0.196	0.225	0.208	0.428	0.215	0.241	0.196
$\sqrt{252} \times \hat{\sigma}_P$	0.132	0.149	0.153	0.173	0.204	0.209	0.210	0.251
$\hat{\zeta}_P$	-0.482	-0.223	0.169	-0.102	-0.045	-0.075	0.218	-0.046
$\hat{\kappa}_P$	14.89	6.754	6.810	3.100	6.655	1.868	2.873	1.332
CKL (Gaussian)	0.258	0.189	0.204	0.234	0.191	0.131	0.141	0.179
CKL (SLN)	0.343	0.250	0.219	0.286	0.211	0.141	0.127	0.208
CKL (GN)	0.443	0.281	0.290	0.235	0.489	0.317	0.358	0.175
$\mathbb{P}(P < \hat{\mu}_P - 3\hat{\sigma}_P)$	0.84%	0.63%	0.64%	0.50%	0.59%	0.42%	0.33%	0.44%

Notes. The table reports the in-sample performance for the dataset of daily returns on the 25 portfolios formed on size and book-to-market spanning January 1980 to March 2021. We consider two reference portfolios: the global-minimum-variance (GMV) portfolio and the mean-variance (MV) portfolio with a risk-aversion coefficient $\lambda = 10$. We then consider three target-return distributions whose mean and variance match those of the two reference portfolios: a Gaussian distribution, a shifted-log-normal (SLN) distribution with a skewness of 0.5, and a generalized-normal (GN) distribution with an excess kurtosis of -0.5. The minimum-CKL (MCKL) portfolios minimize the conditional Kullback-Leibler divergence with respect to these target distributions. As performance criteria, we report, in order, the mean, standard deviation, skewness, excess kurtosis, CKL divergence with respect to the three target distributions, and the probability of a three-sigma loss. See Section 6.2 for more details on the computation of the criteria.

with a Gaussian target, and 0.169 for MCKL with an SLN target. Similarly, the excess kurtosis is 14.89 for GMV, 6.754 for MCKL with a Gaussian target, and 3.100 for MCKL with a GN target. Therefore, the choice of target distribution impacts the higher moments of the MCKL portfolio in the desired way: the SLN target leads to the largest skewness and the GN target to the smallest kurtosis. As a result of the improved higher moments, the MCKL portfolio delivers a smaller probability of a three-sigma loss than that of GMV and, by construction, a smaller CKL divergence. Finally, the MCKL portfolio differs from the GMV portfolio in terms of the first two moments, and more strongly when targeting an SLN or GN distribution rather than a Gaussian distribution because they have more pronounced higher moments. Specifically, the return mean and variance of the MCKL portfolio are larger than those of the GMV portfolio.

The case of the MV reference portfolio confirms that the CKL divergence favors fitting the variance of the target distribution over the mean, in line with the results in Section 4. For example, the annualized mean and volatility of the MV portfolio are 0.428 and 0.204,

and those of the MCKL portfolio with a Gaussian target are 0.215 and 0.209, respectively. In terms of higher moments, the improvement is substantial. While the MV portfolio has a skewness and excess kurtosis of -0.045 and 6.655, the MCKL portfolio has a skewness and excess kurtosis of 0.218 and 2.873 when targeting an SLN distribution, and -0.046 and 1.332 when targeting a GN distribution.

The question now is whether those in-sample gains can also be realized out of sample when facing estimation risk. Therefore, the next two sections are devoted to the out-of-sample methodology and results.

6.4 Out-of-sample methodology

We detail the considered portfolio strategies in Sections 6.4.1 and 6.4.2, and the rolling-horizon procedure in Section 6.4.3.

6.4.1 Reference and MCKL portfolios

We fix the mean and variance of the chosen target distribution via two reference portfolios w_0. First, the sample GMV portfolio, which delivers a satisfying mean-variance performance out of sample because it ignores estimates of mean returns (DeMiguel et al. 2009). Second, the combination of the sample GMV and sample mean-variance portfolios that maximizes the expected out-of-sample utility according to Kan et al. (2021) (KWZ portfolio), using a risk-aversion coefficient $\lambda = 10$. This portfolio is designed to deliver the optimal mean-variance utility under estimation risk and, thus, is a sensible choice of reference portfolio.

For a given reference portfolio, we then find the MCKL portfolios associated with the same three target-return distributions as those in Section 6.3.

6.4.2 Higher-moment portfolios from the literature

We compare our MCKL portfolios with three other approaches put forward in the literature to improve higher moments. First, the approach that consists in optimizing a four-moment Taylor-series expansion of the investor's utility; see Jondeau and Rockinger (2006), Guidolin and Timmermann (2008), Harvey et al. (2010), and Martellini and Ziemann (2010). We

follow the latter and rely on the constant relative risk aversion (CRRA) utility assuming a constant mean-return vector μ to neutralize its large estimation risk. Thus, we compute the portfolio solving

$$\min_{w\in\mathcal{W}} \frac{\lambda}{2}\hat{\sigma}_P^2 - \frac{\lambda(\lambda+1)}{6}\hat{m}_{3,P} + \frac{\lambda(\lambda+1)(\lambda+2)}{24}\hat{m}_{4,P}, \tag{56}$$

where the estimators $\hat{\sigma}_P^2$, $\hat{m}_{3,P}$, and $\hat{m}_{4,P}$ are given by (45). To be consistent with the KWZ reference portfolio, we fix $\lambda = 10$. As discussed in Section 1, this portfolio is not expected to bring much higher-moment improvements compared to the sample GMV portfolio because, in (56), higher moments are small in magnitude compared to the variance.

Second, the approach put forward by Briec et al. (2007), which consists in finding the largest moment improvements relative to a chosen benchmark portfolio w_B. We consider both the mean-variance-skewness efficient portfolio of Briec et al. (2007) and a mean-variance-kurtosis efficient portfolio similar to Jurczenko et al. (2006). Specifically, given a benchmark portfolio w_B, we consider the two portfolios solving

$$\begin{aligned}
&\max_{\delta\in[0,1],w\in\mathcal{W}} \delta \quad \text{subject to} \quad \hat{\mu}_P \geq \hat{\mu}_B,\ \hat{\sigma}_P^2 \leq \hat{\sigma}_B^2(1-\delta),\ \hat{m}_{3,P} \leq \hat{m}_{3,B}(1+\text{sign}(\hat{m}_{3,P})\delta),\\
&\max_{\delta\in[0,1],w\in\mathcal{W}} \delta \quad \text{subject to} \quad \hat{\mu}_P \geq \hat{\mu}_B,\ \hat{\sigma}_P^2 \leq \hat{\sigma}_B^2(1-\delta),\ \hat{m}_{4,P} \leq \hat{m}_{4,B}(1-\delta),
\end{aligned} \tag{57}$$

where the sample moments of the benchmark portfolio are computed as in (45). We call those two portfolios MVS and MVK. The appeal of the approach of Briec et al. (2007) is that the maximum is guaranteed to be global and the portfolio to be efficient in terms of the considered moments. However, a drawback compared to our MCKL portfolio approach is that the optimal portfolio coincides with the benchmark portfolio w_B if w_B is mean-variance efficient and, thus, the investor cannot start from such a natural portfolio. Therefore, as in Boudt et al. (2018), we take the equally weighted portfolio as benchmark: $w_B = \iota/N$.

Third, the approach put forward by Lassance et al. (2021), which consists in diversifying the portfolio-return variance across $K \leq N$ independent components, which are defined as the maximally independent factors underlying the asset returns. We call this portfolio IC-variance-parity (ICVP). We consider the ICVP portfolio because it is conceptually related

to our MCKL portfolio when choosing a Gaussian target distribution. Indeed, as explained by Lassance et al. (2021), diversifying variance across maximally independent factors drives the higher moments toward those of the Gaussian by the central limit theorem. However, contrary to the MCKL portfolio, there is less control over which mean and variance is targeted by the ICVP portfolio. Mathematically, the ICVP portfolio is given by

$$w = \frac{\boldsymbol{V}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{R}^{\dagger\prime}\iota_K^{MV}}{\iota'\boldsymbol{V}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{R}^{\dagger\prime}\iota_K^{MV}}, \quad \iota_K^{MV} = \text{sign}(\boldsymbol{R}^{\dagger}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{V}'\iota). \tag{58}$$

Three matrices appear in this formula. $\boldsymbol{V}$ and $\boldsymbol{\Lambda}$ are the eigenvectors and eigenvalues matrices of the sample covariance matrix $\widehat{\boldsymbol{\Sigma}}$ corresponding to the first K principal components. $\boldsymbol{R}^{\dagger}$ is the rotation matrix defining the independent components. Specifically, given the standardized principal components $Y_{PC} = \boldsymbol{\Lambda}^{-1/2}\boldsymbol{V}'R$, the independent components $Y_{IC} = \boldsymbol{R}^{\dagger}Y_{PC}$ are defined such that they minimize a measure of dependence known as mutual information. As in Lassance et al. (2021), we fix the number of principal and independent components, K, via the minimum-average-partial-correlation method of Velicer (1976).

6.4.3 Rolling-horizon procedure

We evaluate the out-of-sample performance of the considered portfolios via a classical rolling-horizon procedure. At a given time t, we estimate the different portfolio strategies using the past five years of daily data, a sample size $h = 1260$. We keep these portfolio weights constant for the next three months, thus rebalancing the portfolio every day to account for asset growth, and compute the corresponding out-of-sample daily portfolio returns. After three months, we roll forward the estimation window by three months and repeat this procedure, eventually generating a time series of out-of-sample returns from which we evaluate portfolio performance using the performance criteria in Section 6.2. Concerning the CKL divergence criteria, when considering one of the four higher-moment portfolios from the literature in Section 6.4.2, we consider the GMV portfolio as a reference portfolio.

Table 2: Out-of-sample performance of portfolio strategies (continued on next page)

	$252\hat{\mu}_P$	$\sqrt{252}\hat{\sigma}_P$	$\hat{\zeta}_P$	$\hat{\kappa}_P$	CKL (Gauss)	CKL (SLN)	CKL (GN)	$\mathbb{P}(P < \hat{\mu}_P - 3\hat{\sigma}_P)$
(a) 25SBTM								
GMV reference portfolio								
GMV	0.186	0.117	-0.855	19.40	0.318	0.392	0.543	0.99%
MCKL (Gauss)	0.198	0.126	-0.429	12.71	0.251	0.304	0.437	0.86%
MCKL (SLN)	0.212	0.131	-0.454	12.26	0.238	0.292	0.413	0.82%
MCKL (GN)	0.189	0.144	-0.387	8.842	0.225	0.297	0.363	0.78%
KWZ reference portfolio								
KWZ	0.363	0.194	-0.153	4.280	0.204	0.221	0.343	0.77%
MCKL (Gauss)	0.258	0.206	-0.074	4.383	0.190	0.209	0.307	0.59%
MCKL (SLN)	0.285	0.206	-0.154	5.454	0.193	0.222	0.316	0.64%
MCKL (GN)	0.344	0.224	0.030	3.309	0.180	0.191	0.258	0.55%
Higher-moment portfolios from the literature								
CRRA	0.186	0.117	-0.854	19.12	0.316	0.389	0.540	0.99%
MVS	0.181	0.120	-0.896	17.23	0.320	0.400	0.538	1.06%
MVK	0.187	0.116	-0.878	18.25	0.314	0.386	0.537	1.02%
ICVP	0.208	0.149	-0.647	11.93	0.258	0.383	0.410	0.86%
(b) 25OPINV								
GMV reference portfolio								
GMV	0.135	0.154	-0.855	27.08	0.298	0.442	0.519	0.81%
MCKL (Gauss)	0.139	0.167	-0.634	18.84	0.250	0.391	0.425	0.71%
MCKL (SLN)	0.129	0.172	-0.710	17.64	0.244	0.406	0.404	0.77%
MCKL (GN)	0.144	0.195	-0.493	13.84	0.276	0.465	0.403	0.77%
KWZ reference portfolio								
KWZ	0.154	0.165	-0.699	23.90	0.284	0.373	0.500	0.70%
MCKL (Gauss)	0.152	0.173	-0.343	11.40	0.227	0.276	0.400	0.69%
MCKL (SLN)	0.139	0.179	-0.717	16.79	0.227	0.330	0.402	0.75%
MCKL (GN)	0.161	0.206	-0.211	8.459	0.218	0.287	0.331	0.68%
Higher-moment portfolios from the literature								
CRRA	0.134	0.154	-0.853	26.78	0.294	0.438	0.514	0.81%
MVS	0.138	0.158	-0.713	25.33	0.305	0.446	0.522	0.81%
MVK	0.135	0.154	-0.848	26.82	0.299	0.443	0.519	0.81%
ICVP	0.180	0.206	-0.550	13.56	0.323	0.564	0.457	0.78%

6.5 Discussion of out-of-sample results

In this section, we discuss the out-of-sample performance of the considered portfolio strategies following the methodology in Section 6.4. All results are reported in Table 2.

First, we observe that, in order to better match the target distribution in terms of higher moments, the MCKL portfolio needs to accept a higher volatility than that of the reference portfolio. Specifically, the volatility of the MCKL portfolio is close but generally higher than that of the reference portfolio. This is natural when taking GMV as a reference portfolio,

Table 2: Out-of-sample performance of portfolio strategies (continued)

	$252\hat{\mu}_P$	$\sqrt{252}\hat{\sigma}_P$	$\hat{\zeta}_P$	$\hat{\kappa}_P$	CKL (Gauss)	CKL (SLN)	CKL (GN)	$\mathbb{P}(P < \hat{\mu}_P - 3\hat{\sigma}_P)$
(c) 30IND								
GMV reference portfolio								
GMV	0.103	0.124	-0.989	24.60	0.288	0.421	0.509	0.78%
MCKL (Gauss)	0.087	0.137	-1.016	17.47	0.241	0.406	0.416	0.80%
MCKL (SLN)	0.092	0.140	-0.960	17.19	0.236	0.407	0.402	0.72%
MCKL (GN)	0.079	0.159	-0.753	12.03	0.257	0.461	0.380	0.70%
KWZ reference portfolio								
KWZ	0.114	0.147	-0.695	16.07	0.289	0.380	0.486	0.91%
MCKL (Gauss)	0.079	0.151	-0.860	12.58	0.241	0.345	0.413	0.81%
MCKL (SLN)	0.084	0.150	-0.857	14.02	0.224	0.330	0.392	0.74%
MCKL (GN)	0.078	0.178	-0.649	8.302	0.239	0.381	0.347	0.78%
Higher-moment portfolios from the literature								
CRRA	0.102	0.124	-1.010	24.43	0.285	0.418	0.505	0.78%
MVS	0.106	0.132	-0.731	24.03	0.292	0.432	0.511	0.84%
MVK	0.111	0.128	-1.017	23.02	0.276	0.420	0.487	0.81%
ICVP	0.111	0.167	-0.618	10.68	0.262	0.466	0.351	0.61%

Notes. The table reports the out-of-sample performance of the portfolio strategies considered in Sections 6.4.1 and 6.4.2 for the three datasets of daily returns of Section 6.1. We consider two reference portfolios: the sample global-minimum-variance (GMV) portfolio and the combination of the sample GMV and sample mean-variance portfolios that maximizes the expected out-of-sample utility according to Kan et al. (2021) (KWZ portfolio), using a risk-aversion coefficient $\lambda = 10$. We then consider three target-return distributions whose mean and variance match those of the two reference portfolios: a Gaussian distribution, a shifted-log-normal (SLN) distribution with a skewness of 0.5, and a generalized-normal (GN) distribution with an excess kurtosis of -0.5. The minimum-CKL (MCKL) portfolios minimize the conditional Kullback-Leibler divergence with respect to these target distributions. As performance criteria, we report, in order, the mean, standard deviation, skewness, excess kurtosis, CKL divergence with respect to the three target distributions (we take GMV as a reference portfolio for the higher-moment portfolios from the literature), and the probability of a three-sigma loss. See Section 6.2 for more details on the computation of the criteria.

but is also the case for KWZ. The increase in volatility is more pronounced for the SLN and GN target distributions that have more pronounced higher moments that the Gaussian target. Concerning the mean return, the MCKL portfolio can have a larger or a smaller mean return than that of the reference portfolio depending on the cases. The difference in mean return compared to the reference portfolio tends to be more pronounced than the difference in volatility, which corroborates the in-sample observation in Section 6.3 that the CKL divergence is more impacted by the fit to the volatility than the fit to the mean.

Second, the MCKL portfolio has generally more desirable higher-moment characteristics than those of the reference portfolio, which is consistent with the fact that the targeted distributions embed better skewness and kurtosis. The improvement in excess kurtosis is

more substantial than the improvement in skewness, which can be a result of the weight of the moments in the CKL divergence measure, but also because skewness is subject to larger sampling errors than kurtosis (Martellini and Ziemann 2010). To give a concrete example, when taking GMV as a reference portfolio and the GN target distribution that has an excess kurtosis of -0.5, the excess kurtosis of the MCKL portfolio is twice smaller than that of GMV for all three datasets. A consequence of those improved higher moments is that the probability of a three-sigma loss is reduced when the investor opts for a MCKL portfolio over the reference portfolio, which is desirable for investors behaving according to disappointment aversion theory (Routledge and Zin 2010, Dahlquist et al. 2017).

Third, the underlying objective of minimizing the CKL divergence with respect to the chosen target distribution is better achieved out of sample with the MCKL portfolio estimated in-sample rather than the reference portfolio estimated in-sample. Specifically, the MCKL portfolio achieves a smaller CKL divergence, for the corresponding target distribution and reference portfolio, in all cases except one. This is notable because in-sample efficiency does not automatically translate to better out-of-sample performance, particularly given the large estimation errors in higher moments. This suggests that the estimation method proposed in Section 5 is relatively robust.

Fourth, we discuss the out-of-sample performance of the considered higher-moment portfolios proposed in the literature. The CRRA portfolio delivers a nearly identical performance to that of the GMV portfolio, as expected. Thus, it is not a satisfying approach to improve higher moments. The MVS and MVK portfolios deliver a modest improvement in higher moments compared to the GMV portfolio, for a correspondingly small increase in variance, and actually increase the probability of a three-sigma loss. Such portfolios may thus be preferred by investors for whom preferences for higher moments are minor. Finally, recall that the ICVP portfolio diversifies the portfolio-return variance across maximally independent factors, which, by the central limit theorem, delivers a more Gaussian portfolio-return density. Table 2 shows that this is achieved with great success because the ICVP portfolio yields a very significant improvement in skewness and kurtosis compared to the GMV portfolio, often more significant than with the MCKL portfolio targeting a Gaussian distribution. However, a drawback of the ICVP portfolio is that it also delivers a significantly larger volatility than

that of the GMV portfolio, more so than the MCKL portfolios, which translates in a larger CKL divergence. Thus, it is more desirable to explicitly target a Gaussian distribution via a minimization of the CKL divergence than implicitly targeting it via factor diversification.

Overall, the results suggest that the minimum-CKL portfolio is worth considering as an alternative to mean-variance and higher-moment portfolios proposed in the literature.

7 Conclusion

In practice, investment managers monitor their portfolios in terms of return statistics such as the mean, dispersion, asymmetry, tail risk, or quantiles; that is, in terms of return distribution. Therefore, in this paper, we propose to capture investors' preferences via a general framework that consists in optimizing the full distribution of returns. Specifically, given a target-return distribution specified by the investor, the optimal portfolio has a return density that is as close as possible to the target-return density according to the conditional Kullback-Leibler divergence. We study the theoretical properties of the proposed method for two classes of target distribution that allow for different levels of skewness and kurtosis, while picking the mean and variance of the target distribution on the mean-variance efficient frontier. We show in particular that unless asset returns are Gaussian, the optimal portfolio naturally departs from the efficient frontier so as to better fit the desired higher-moment characteristics embedded in the target distribution. Empirically, we find that the proposed approach delivers a desirable tradeoff between mean-variance efficiency and improved higher moments, and has better properties than other competing approaches that aim at improving higher moments.

References

Alexander, G., & Baptista, A. (2004). A comparison of VaR and CVaR constraints on portfolio selection with the mean-variance model. *Management Science*, 50(9):1261–1273.

Alexander, G., & Baptista, A. (2006). Portfolio selection with a drawdown constraint. *Journal of Banking and Finance*, 30(11):3171–3189.

Ang, A., Chen, J., & Yu, H. (2006). Downside risk. *Review of Financial Studies*, 19(4):1191–1239.

Ao, M., Yingying, L., & Xinghua, Z. (2019). Approaching mean-variance efficiency for large portfolios. *Review of Financial Studies*, 32(7):2890–2919.

Athayde, G., & Flores, R. (2004). Finding a maximum skewness portfolio: A general solution to three-moments portfolio choice. *Journal of Economic Dynamics and Control*, 28(7):1335–1352.

Aumann, R., & Serrano, R. (2008). An economic index of riskiness. *Journal of Political Economy*, 116(5):810–836.

Azzalini, A. (1985). A class of distributions which includes the Normal ones. *Scandinavian Journal of Statistics*, 12(2):171–178.

Basak, S., & Shapiro, A. (2001). Value-at-risk-based risk management: Optimal policies and asset prices. *The Review of Financial Studies*, 14(2):371–405.

Basseville, M. (2013). Divergence measures for statistical data processing – An annotated bibliography. *Signal Processing*, 93(4):621–633.

Beirlant, J., Dudewicz, E., Gyor, L., & van der Meulen, E. (1997). Non-parametric entropy estimation: An overview. *International Journal of Mathematical and Statistical Sciences*, 6(1):17–39.

Benaglia, T., Chauveau, D., Hunter, D., & Young, D. (2009). mixtools: An R package for analyzing finite mixture models. *Journal of Statistical Software*, 32(6):1–29.

Bera, A., & Park, S. (2008). Optimal portfolio diversification using maximum entropy principle. *Econometric Reviews*, 27:484–512.

Bernard, C., Boyle, P., & Vanduffel, S. (2014). Explicit representation of cost-efficient strategies. *Finance*, 35(2):5-55.

Boudt, K., Cornilly, D., & Verdonck, T. (2018). A coskewness shrinkage approach for estimating the skewness of linear combinations of random variables. *Journal of Financial Econometrics*, 22:1–23.

Briec, W., Kerstens, K., & Jokung, O. (2007). Mean-variance-skewness portfolio performance gauging: A general shortage function and dual approach. *Management Science*, 53(1):135–149.

Cardoso, J. (1997). Infomax and maximum likelihood for blind source separation, *IEEE Signal Processing Letters*, 4(4):112–114.

Chalabi, Y., & Würtz, D. (2012). Portfolio optimization based on divergence measures. MPRA working paper, 43332.

Chamberlain, G. (1983). A characterization of the distributions that imply mean—variance utility functions. *Journal of Economic Theory*, 29(1):185–201.

Cogneau, P., & Hubner, G. (2014). The (more than) 100 ways to measure portfolio performance. *Journal of Performance Measurement*, 14:56–69.

Cover, T., & Thomas, J. (2006). *Elements of Information Theory* (2nd ed.). New Jersey: Wiley.

Cohen, A. (1951). Estimating parameters of logarithmic-normal distributions by maximum likelihood. *Journal of the American Statistical Association*, 46(254):206–212.

Dahlquist, M., Farago, A., & Tédongap, R. (2017). Asymmetries and portfolio choice. *Review of Financial Studies*, 30(2):667–702.

Das, S., & Uppal, R. (2004). Systemic risk and international portfolio choice. *Journal of Finance*, 59(6):2809–2834.

Deck, C., & Schlesinger, H. (2010). Exploring higher order risk effects. *Review of Economic Studies*, 77:1403–1420.

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

Dempster, A., Laird, N., & Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B (Methodological)*, 39(1):1–38.

De Nard, G., Ledoit, O., & Wolf, M. (2019). Factor models for portfolio selection in large dimensions: the good, the better and the ugly. *Journal of Financial Econometrics*, 1–22.

Dybvig, P. (1988). Inefficient dynamic portfolio strategies or how to throw away a million dollars in the stock market. *Review of Financial Studies*, 1(1):67-88.

Everitt, B., & Hand, D. (1981). *Finite Mixture Distributions*. London: Chapman and Hall.

van Es, B. (1992). Estimating functionals related to a density by a class of statistics based

on spacings. *Scandinavian Journal of Statistics*, 19(1):61–72.

Flores, Y., Bianchi, R., Drew, M., & Trück, S. (2017). The diversification delta: A different perspective. *Journal of Portfolio Management*, 43(4):112–124.

Foster, D., & Hart, S. (2009). An operational measure of riskiness. *Journal of Political Economy*, 117(5):785–814.

Gormsen, N., & Jensen, C. (2020). Higher-moment risk. Available at SSRN 3069617.

Goto, S., & Xu, Y. (2015). Improving mean variance optimization through sparse hedging restrictions. *Journal of Financial and Quantitative Analysis*, 50(6):1415–1441.

Guidolin, M., & Timmermann, A. (2008). International asset allocation under regime switching, skew, and kurtosis preferences. *Review of Financial Studies*, 21(2):889–935.

Harvey, C., Liechty, J., Liechty, M., & Müller, P. (2010). Portfolio selection with higher moments. *Quantitative Finance*, 10(5):469–485.

Haugh, M., & Lo, A. (2001) Asset allocation and derivatives. *Quantitative Finance*, 1(1):45–72.

Heston, S. (1993). A closed-form solution for options with stochastic volatility with application to bond and currency options. *Review of Financial Studies*, 6(2):327–343.

Hyvärinen, A., Karhunen, J., & Oja, E. (2001). *Independent Component Analysis*. New York: John Wiley & Sons.

Jagannathan, R., & Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *Journal of Finance*, 58(4):1651–1684.

Jarrow, R., & Zhao, F. (2006). Downside loss aversion and portfolio management. *Management Science*, 52(4):558–566.

Jean, W. (1971). The extension of portfolio analysis to three or more parameters. *Journal of Financial and Quantitative Analysis*, 6(1):505–515.

Jiang, L., Wu, K., & Zhou, G. (2018). Asymmetry in stock comovements: An entropy approach. *Journal of Financial and Quantitative Analysis*, 53(4):1479–1507.

Jondeau, E., & Rockinger, M. (2006). Optimal portfolio allocation under higher moments. *European Financial Management*, 12(1):29–55.

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

Jurczenko, E., Maillet, B., & Merlin, P. (2006). Hedge fund portfolio selection with higher-order moments: A nonparametric mean–variance–skewness–kurtosis efficient frontier. In Jurczenko, E., & Maillet, B. *Multi-moment Asset Allocation and Pricing Models*. West Sussex: Wiley.

Kadan, O., & Liu, F. (2014). Performance evaluation with high moments and disaster risk. *Journal of Financial Economics*, 113(1):131–155.

Kan, R., Wang, X., & Zhou, G. (2021). Optimal portfolio choice with estimation risk: No risk-free asset case. *Management Science*, forthcoming.

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

Kaplan, P., & Knowles, J. (2004). Kappa: A generalized downside risk-adjusted performance measure. *Journal of Performance Measurement*, 8:42–54.

Keating, C., & Shadwick, W. (2002). A universal performance measure. *Journal of Performance Measurement*, 6:59–84.

Kelly, B., & Jiang, H. (2014). Tail risk and asset prices. *Review of Financial Studies*, 27(10):2841–2871.

Kendall, M., & Stuart, A. (1963). *The Advanced Theory of Statistics* (vol. 1, 2nd ed). London: Charles Griffin.

Kim, W., Fabozzi, F., Cheridito, P., & Fox, C. (2014). Controlling portfolio skewness and kurtosis without directly optimizing third and fourth moments. *Economics Letters*, 122:154–158.

Kroll, Y., Levy, H., & Markowitz, H. (1984). Mean-variance versus direct utility maximization. *Journal of Finance*, 39(1):47–61.

Kullback, S., & Leibler, R. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86.

Lassance, N., DeMiguel, V., & Vrins, F. (2021). Optimal portfolio diversification via independent component analysis. *Operations Research*, forthcoming.

Lassance, N., & Vrins, F. (2021a). Minimum Rényi entropy portfolios. *Annals of Operations Research*, 299(1):23–46.

Lassance, N., & Vrins, F. (2021b). Portfolio selection with parsimonious higher comoments

estimation. *Journal of Banking and Finance*, 126(9):106–115.

Learned-Miller, E., & Fisher, J. (2003). ICA using spacings estimates of entropy. *Journal of Machine Learning Research*, 4(7-8):1271–1295.

Le Courtois, O., & Xu, X. (2020). A complete ranking of risky prospects consistent with stochastic dominance. Available at SSRN 3729152.

Ledoit, O., & Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2):365–411.

Levy, H. (2016). *Stochastic Dominance: Investment Decision Making Under Uncertainty* (3rd ed.). Switzerland: Springer.

Levy, H., & Markowitz, H. (1979). Approximating expected utility by a function of mean and variance. *American Economic Review*, 69:308–317.

Markowitz, H. (1952). Portfolio selection. *Journal of Finance*, 7(1):77–91.

Markowitz, H. (2014). Mean–variance approximations to expected utility. *European Journal of Operational Research*, 234(2):346–355.

Martellini, L., & Ziemann, V. (2010). Improved estimates of higher-order comoments and implications for portfolio selection. *Review of Financial Studies*, 23(4):1467–1502.

Merton, R. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics*, 8(4):323–361.

Nadarajah, S. (2005). A generalized normal distribution. *Journal of Applied Statistics*, 32(7):685–694.

Post, T., & Kopa, M. (2016). Portfolio choice based on third-degree stochastic dominance. *Management Science*, 63(10):3381–3392.

Price, K., Price, B., & Nantell, T. (1982). Variance and lower partial moment measures of systematic risk: Some analytical and empirical results. *Journal of Finance*, 37(3):843–855.

Rockafellar, R., & Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of Risk*, 2(3):21–41.

Routledge, B., & Zin, S. (2010). Generalized disappointment aversion and asset prices. *Journal of Finance*, 65(4):1303–1332.

Schuhmacher, F., Kohrs, H., & Auer, B. (2021). Justifying mean-variance portfolio selection when asset returns are skewed. *Management Science*, forthcoming.

Scott, R., & Horvath, P. (1980). On the direction of preference for moments of higher order than the variance. *Journal of Finance*, 35(4):915–919.

Shannon, C. (1948). A mathematical theory of communication. *Bell Systems Technical Journal*, 27:379–423 and 623–656.

Slater, L. (1972). Confluent hypergeometric functions. In Abramowitz, M., & Stegun, I. A. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York: Dover Publications.

Sortino, F., & van der Meer, R. (1991). Downside risk. *Journal of Portfolio Management*, 17(4):27–31.

Theodossiou, P. (1998). Financial data and the skewed generalized t distribution. *Management Science*, 44(12):1650–1661.

Trautmann, S., & van de Kuilen, G. (2018). Higher order risk attitudes: A review of experimental evidence. *European Economic Review*, 103:108–124.

Van Hemert, O., Ganz, M., Harvey, C., Rattray, S., Martin, E., & Yawitch, D. (2020). Drawdowns. *Journal of Portfolio Management*, 46(8):34–50.

Vasicek, O. (1976). A test for normality based on sample entropy. *Journal of the Royal Statistical Society Series B*, 38(1):54–59.

Velicer, W. (1976) Determining the number of components from the matrix of partial correlations. *Psychometrika*, 41(3):321–327.

Vermorken, M., Medda, F., & Schroder, T. (2012). The diversification delta: A higher-moment measure for portfolio diversification. *Journal of Portfolio Management*, 39(1):67–74.

Whitmore, G. (1970). Third-degree stochastic dominance. *American Economic Review*, 60(3):457–459.

Yuan, P. (1933). On the logarithmic frequency distribution and the semi-logarithmic correlation surface. *The Annals of Mathematical Statistics*, 4(1):30–74.

Zhang, J., & Wang, X. (2008). Robust normal reference bandwidth for kernel density estimation. *Statistica Neerlandica*, 63(1):13–23.

Internet Appendix to

Portfolio Selection:
A Target-Distribution Approach

This internet appendix to the main paper contains three sections. Section IA.1 provides the proofs of all results in the paper. Section IA.2 studies the CKL divergence with a shifted-log-normal target distribution and a Gaussian portfolio return. Section IA.3 explains the bandwidth-selection method.

IA.1 Proofs of results

Proof of Proposition 1

Part 1. The non-negativity of the regular KL divergence is direct from Jensen's inequality; see Cover and Thomas (2006). Dealing with the CKL divergence is not much more involved but requires other tools. The $p = 1$ case is obvious, by definition. Assuming $p < 1$, the conditional expectation exists, and the negative CKL divergence is given by

$$-\langle f_P|f_T\rangle_2 = \frac{1}{1-p}\left(\mathbb{E}\left[\ln\frac{f_T(P)}{f_P(P)}1_{\{P\in\Omega_T\}}\right] - p\right).$$

Using $\ln x \leq x - 1$ for all $x > 0$, the expectation becomes

$$\begin{aligned}
\mathbb{E}\left[\ln\frac{f_T(P)}{f_P(P)}1_{\{P\in\Omega_T\}}\right] &= \int_{\Omega_T\cap\Omega_P} f_P(x)\ln\frac{f_T(x)}{f_P(x)}\mathrm{d}x \\
&\leq \int_{\Omega_T\cap\Omega_P} f_T(x)\mathrm{d}x - \int_{\Omega_T\cap\Omega_P} f_P(x)\mathrm{d}x \\
&= \int_{\Omega_T\cap\Omega_P} f_T(x)\mathrm{d}x - \int_{\Omega_T} f_P(x)\mathrm{d}x \\
&\leq \int_{\Omega_T} f_T(x)\mathrm{d}x - \int_{\Omega_T} f_P(x)\mathrm{d}x \\
&= 1 - \mathbb{P}(P\in\Omega_T) = p.
\end{aligned} \tag{IA1}$$

Therefore, we have $-\langle f_P|f_T\rangle_2 \leq 0$ and thus $\langle f_P|f_T\rangle_2 \geq 0$.

Next, we have that $\langle f_P|f_T\rangle_2 = 0$ if $f_P = f_T$ almost everywhere because, in this case, their integral on Ω_T is 1, so that $p = \mathbb{P}(P \notin \Omega_T) = 0$ and the CKL divergence reduces to the KL divergence. The latter is known to vanish when $f_P = f_T$ almost everywhere.

Finally, we turn to the three conditions under which $\langle f_P|f_T\rangle_2 > 0$ when f_P does not agree with f_T almost everywhere. The first condition, $p = 1$, holds because, then, $\langle f_P|f_T\rangle_2 = +\infty$

by definition. The second condition, $\Omega_P \subseteq \Omega_T$, corresponds to the regular case of the KL divergence. Let us now prove the third condition, $\Omega_T \subset \Omega_P$ and $p > 0$. To that end, note that because $\ln x - x + 1$ has a unique maximum in $x = 1$, the first inequality in (IA1) is strict as soon as $f_T(x)/f_P(x) \neq 1$ on a set of strictly positive f_P-measure. This shows that $\langle f_P | f_T \rangle_2 > 0$ if there exists a set $\mathcal{S} \subseteq \Omega_T \cap \Omega_P$ satisfying $\int_{\mathcal{S}} f_T(x)\mathrm{d}x \neq \int_{\mathcal{S}} f_P(x)\mathrm{d}x$ and $\int_{\mathcal{S}} f_P(x)\mathrm{d}x > 0$. To see that there exists such a set when $p > 0$ and $\Omega_T \subset \Omega_P$, take $\mathcal{S} = \Omega_T$. Clearly, $\mathcal{S} \subseteq \Omega_T \cap \Omega_P$ and $1 - p = \mathbb{P}(P \in \Omega_T) = \int_{\mathcal{S}} f_P(x)\mathrm{d}x \in (0, 1)$. However, $\int_{\mathcal{S}} f_T(x)\mathrm{d}x = \int_{\Omega_T} f_T(x)\mathrm{d}x = 1$. This shows that $\langle f_P | f_T \rangle_2 > 0$ when $\Omega_T \subset \Omega_P$ and $p > 0$.

Part 2. When $\Omega_P \subseteq \Omega_T$, the conditioning event $P \in \Omega_T$ holds with probability one, so that $p = 0$. Noting that in this case the conditional expectation coincides with the unconditional one, and the CKL divergence simplifies to

$$\langle f_P | f_T \rangle_2 = \mathbb{E}\left[\ln \frac{f_P(P)}{f_T(P)}\right] = \langle f_P | f_T \rangle_1.$$

Part 3. The condition $\Omega_T \cap \Omega_P = \emptyset$ implies $p = \mathbb{P}(P \notin \Omega_T) = 1$, and the claim follows from the definition of the CKL divergence with $p = 1$.

Proof of Proposition 2

Because $\Omega_T = \mathbb{R}$ for $T \sim \mathcal{GN}(\alpha, \beta, \gamma)$, we have $\Omega_P \subseteq \Omega_T$ and, as shown in Proposition 1, the CKL divergence reduces to the KL divergence. Applying the decomposition of the KL divergence in (9) to the GN target density $f_T(x; \alpha, \beta, \gamma)$ in (16) yields

$$\mathcal{C}(w; \alpha, \beta, \gamma) = H[P] + \ln \frac{2^{-\frac{\gamma+1}{\gamma}}\gamma}{\beta\Gamma(1/\gamma)} - \frac{1}{2\beta^\gamma}\mathbb{E}\left[|P - \alpha|^\gamma\right],$$

where $H[P] = H[P^\star] + \ln \sigma_P$. Removing the constant middle term completes the proof.

Proof of Proposition 3

After replacing α and δ by their moment-matching expressions in (21), proving the proposition amounts to showing that

$$\lim_{\beta\to 0^+} \frac{\exp\left(-\frac{1}{2\beta^2}\left(\frac{\beta^2}{2}+\ln(1+zg(\beta)^{1/2}/\sigma_T)\right)^2\right)}{\beta\sqrt{2\pi}\,(z+\sigma_T g(\beta)^{-1/2})} = \frac{1}{\sigma_T\sqrt{2\pi}}\exp\left(-\frac{z^2}{2\sigma_T^2}\right), \tag{IA2}$$

where $z := x-\mu_T$. As we now show, this limit holds because the numerator tends to $\exp(-z^2/(2\sigma_T^2))$ and the denominator to $\sigma_T\sqrt{2\pi}$.

We begin with the denominator of (IA2). We need to show that

$$\lim_{\beta\to 0^+} \beta\left(z+\frac{\sigma_T}{\sqrt{\exp(\beta^2)-1}}\right) = \sigma_T,$$

which is equivalent to showing that

$$\lim_{\beta\to 0^+} \frac{\beta}{\sqrt{\exp(\beta^2)-1}} = 1.$$

This last limit holds because, by application of L'Hospital rule,

$$\lim_{\beta\to 0^+} \frac{\beta}{\sqrt{\exp(\beta^2)-1}} = \left(\lim_{\beta\to 0^+}\frac{\exp(\beta^2)-1}{\beta^2}\right)^{-1/2} = \left(\lim_{\beta\to 0^+}\exp(\beta^2)\right)^{-1/2} = 1.$$

We turn next to the numerator of (IA2). We need to show that

$$\lim_{\beta\to 0^+}\frac{1}{\beta^2}\left(\frac{\beta^2}{2}\ln\left(1+\frac{z}{\sigma_T}g(\beta)^{1/2}\right)\right)^2 = \frac{z^2}{\sigma_T^2}.$$

After applying L'Hospital once and some simplifications, this limit simplifies to

$$\frac{z}{\sigma_T}\lim_{\beta\to 0^+}\left(\frac{\exp(\beta^2)\left(\frac{\beta^2}{2}+\ln\left(1+zg(\beta)^{1/2}/\sigma_T\right)\right)}{g(\beta)^{1/2}+zg(\beta)/\sigma_T}\right).$$

Applying L'Hospital once more and multiplying the numerator and denominator by $g(\beta)^{1/2}$, we can evaluate the limit and find that it is indeed equal to z^2/σ_T^2.

Proof of Proposition 4

Part 1. The support set of the SLN density is $\Omega_T = (\delta, \infty]$. Applying the decomposition of the CKL divergence in (12) to the SLN target density $f_T(x; \alpha, \beta, \delta)$ in (19) yields

$$\mathcal{C}(w; \alpha, \beta, \delta) = \mathbb{E}[\ln f_T(P)|P_\delta > 0] + H[P|P_\delta > 0] - \frac{p}{1-p},$$

where the conditional cross-entropy, ignoring the normalization constant, simplifies to

$$\begin{aligned}\mathbb{E}[\ln f_T(P)|P \geq \delta] &= -\mathbb{E}[\ln P_\delta|P_\delta \geq 0] - \frac{1}{2\beta^2}\mathbb{E}[(\ln P_\delta) - \alpha)^2|P_\delta \geq 0] \\ &= -\mathbb{E}[\ln P_\delta|P_\delta \geq 0] - \frac{1}{2\beta^2}\left(\mathbb{V}[\ln P_\delta|P_\delta \geq 0] + (\mathbb{E}[\ln P_\delta|P_\delta \geq 0] - \alpha)^2\right).\end{aligned}$$

A second-order Taylor-series expansion of $\ln x$ then yields the approximations in (23)–(24) for $\mathbb{E}[\ln P_\delta|P_\delta \geq 0]$ and $\mathbb{V}[\ln P_\delta|P_\delta \geq 0]$ by considering the Taylor-series expansion for a general transform $Z = g(X)$ around μ_X:

$$\begin{aligned}\mu_Z &\approx g(\mu_X) + \frac{g''(\mu_X)}{2}\sigma_X^2, \\ \sigma_Z^2 &\approx (g'(\mu_X))^2\sigma_X^2 + g'(\mu_X)g''(\mu_X)m_3(X) + \frac{1}{4}(g''(\mu_X))^2(m_{4,X} - \sigma_X^4).\end{aligned}$$

Part 2. When $p = 0$, we have $H[P|P_\delta \geq 0] = H[P]$, $p/(1-p) = 0$, $m(P_\delta) = \mu_P - \delta$, $m_k(P_\delta) = m_{k,P}$, and $\psi_k(P_\delta) = m_{k,P}/(\mu_P - \delta)^k$. Plugging those expressions into (22)–(23)–(24) proves the result.

Part 3. From the Gram-Charlier expansion of the entropy in (32), we have that $H[P^\star]$ depends on the skewness via the term $\zeta_P^2/12$. Then, because ζ_P only appears in $H[P^\star]$ and $\widetilde{\mathbb{V}}[\ln P_\delta|P_\delta \geq 0]$, $\widetilde{\mathcal{C}}(w; \alpha, \beta, \delta)$ depends on ζ_P as

$$\widetilde{\mathcal{C}}(w; \alpha, \beta, \delta) = \frac{\zeta_P\sigma_P^3}{2\beta^2(\mu_P - \delta)^3} - \frac{\zeta_P^2}{12} + \text{constant}, \tag{IA3}$$

where the constant is independent from ζ_P. Moreover, because we assume $(\mu_P, \sigma_P) =$

(μ_T, σ_T), we have from (21) that $\mu_P - \delta = \sigma_P g(\beta)^{1/2}$. Thus, (IA3) becomes

$$\widetilde{\mathcal{C}}(w; \alpha, \beta, \delta) = \frac{\zeta_P g(\beta)^{3/2}}{2\beta^2} - \frac{\zeta_P^2}{12} + \text{constant}, \tag{IA4}$$

The optimal skewness ζ_P maximizing (IA4) is easily found to be given by (28), provided it is attainable. The result for the kurtosis is found in a similar way from the fact that $H[P^\star]$ depends on the excess kurtosis via the term $\kappa^2/48$ in (32). Therefore, when $(\mu_P, \sigma_P) = (\mu_T, \sigma_T)$, $\widetilde{\mathcal{C}}(w; \alpha, \beta, \delta)$ depends on κ_P as

$$\widetilde{\mathcal{C}}(w; \alpha, \beta, \delta) = -\frac{(\kappa_P + 2) g(\beta)^2}{8\beta^2} - \frac{\kappa_P^2}{48} + \text{constant}, \tag{IA5}$$

The optimal excess kurtosis ζ_P maximizing (IA5) is easily found to be given by (29), provided it is attainable.

Proof of Proposition 5

Plugging $\gamma = 2$ in the objective function (18) gives

$$\mathcal{C}(w; \alpha, \beta, 2) = H[P^\star] + \ln \sigma_P - \frac{1}{2\beta^2} \mathbb{E}\left[(P - \alpha)^2\right] = H[P^\star] + \ln \sigma_P - \frac{1}{2}\left(\frac{\mu_P - \alpha}{\beta}\right)^2 - \frac{\sigma_P^2}{2\beta^2},$$

which corresponds indeed to (31).

Proof of Proposition 6

Part 1. The Lagrangian associated with the objective function $\mathcal{C}(w; \alpha)$ is given by

$$\mathcal{L} = -(w'\mu - \alpha)^2 - w'\boldsymbol{\Sigma} w + \tau(w'\iota - 1),$$

where τ is the Lagrange multiplier associated with the full-investment constraint $w'\iota = 1$. Define $\boldsymbol{A} := (\boldsymbol{\Sigma} + \mu\mu')^{-1}$. Then, setting the derivative of $\mathcal{L}$ with respect to w' equal to zero yields, after some developments,

$$w = \boldsymbol{A}^{-1}\left(\alpha\mu - \frac{\tau}{2}\iota\right). \tag{IA6}$$

After replacing (IA6) in the constraint $w'\iota = 1$, the Lagrange multiplier τ is given by

$$\tau = \frac{2\big(\alpha\iota'\boldsymbol{A}^{-1}\mu - 1\big)}{\iota'\boldsymbol{A}^{-1}\iota}. \tag{IA7}$$

Combining (IA6) and (IA7) results in the following solution for w_α:

$$w_\alpha = \boldsymbol{A}\left(\alpha\mu - \frac{(\alpha\iota'\boldsymbol{A}\mu - 1)\,\iota}{\iota'\boldsymbol{A}\iota}\right). \tag{IA8}$$

From the updating formula for the inverse matrix (Greene 1993 p.25), we have that

$$\boldsymbol{A} = \boldsymbol{\Sigma}^{-1} - \frac{1}{1+\psi^2}\boldsymbol{\Sigma}^{-1}\mu\mu'\boldsymbol{\Sigma}^{-1}.$$

As a result, we find that

$$\begin{aligned}
\boldsymbol{A}\mu &= \frac{\mu_g}{\sigma_g^2(\psi^2+1)}w_{MSR}, \\
\boldsymbol{A}\iota &= \frac{w_{GMV}}{\sigma_g^2} - \frac{w_{MSR}}{\psi^2+1}\left(\frac{\mu_g}{\sigma_g^2}\right)^2,
\end{aligned}$$

and plugging these two equalities in (IA8) results in the following formula for w_α after simplifications:

$$w_\alpha = w_{GMV}\left[\frac{(\psi^2+1)\sigma_g^2 - \alpha\mu_g}{(\psi^2+1)\sigma_g^2 - \mu_g^2}\right] + w_{MSR}\left[\frac{\mu_g(\mu_g-\alpha)}{\mu_g^2 - (\psi^2+1)\sigma_g^2}\right]. \tag{IA9}$$

Being a combination of w_{GMV} and w_{MSR}, w_α is a mean-variance portfolio, whose mean return $w_\alpha'\mu$ in (35) is obtained by noting that $w_{MSR}'\mu = \psi^2\sigma_g^2/\mu_g$.

Part 2. $w_\alpha = w_{GMV}$ if $\alpha = \mu_g$ because, from (35), $w_\alpha'\mu = \mu_g$ if $\alpha = \mu_g$. Moreover, w_α is MVE if and only if $w_\alpha'\mu \geq \mu_g$, which, from (35), happens when $\alpha \geq \mu_g$.

Part 3. Le Courtois and Xu (2020) define the deficit scale of a portfolio with return P bounded by $p_{\max}$ as

$$DS_P = \mathbb{E}\left[\frac{(p_{\max}-P)^2}{p_{\max}^2}\right] = \frac{(p_{\max}-\mu_P)^2}{p_{\max}^2} + \frac{\sigma_P^2}{p_{\max}^2}. \tag{IA10}$$

In their proposition 2.2, they show that given two investments with return P_1, P_2 in $[p_{\min}, p_{\max}]$, if DS_{P_1} is larger than DS_{P_2}, then P_1 cannot dominate P_2 in terms of third-order stochastic dominance. The proposition follows because the deficit scale in (IA10) is a scaled version of the Dirac-delta objective function $\mathcal{C}(w;\alpha)$ in (34) with $\alpha = p_{\max}$. Indeed, any portfolio $w \in \mathcal{W}$ satisfying $||w||_q \leq K$ has a return bounded above by α in (36), and bounded below by the minimum with respect to $w \in \mathcal{W}$ and $||w||_q \leq K$ of $\sum_{i:w_i<0} w_i r_{\max,i} + \sum_{i:w_i\geq 0} w_i r_{\min,i}$, and has a smaller $\mathcal{C}(w;\alpha)$, and thus a larger deficit scale, than the MCKL portfolio maximizing $\mathcal{C}(w;\alpha)$ with α in (36). Finally, we need $q \geq 1$ for $||w||_q$ to be a norm, and we need $K \geq N^{(1-q)/q}$ because the minimum norm $||w||_q$ achievable by any portfolio $w \in \mathcal{W}$ is achieved by the equally weighted portfolio with a norm $||\iota/N||_q = N^{(1-q)/q}$.

Proof of Proposition 7

As a preliminary useful result, note that $H[P] = H[P^\star] + \ln \sigma_P$ and, when $P^\star \sim \mathcal{N}(0,1)$, $H[P^\star] = \frac{1}{2}\ln 2\pi e$ is a constant that can be removed from the objective function.

Part 1. Let $Z \sim \mathcal{N}(\mu_Z, \sigma_Z)$ be a Gaussian random variable and $k > -1$ a real number. Then, the kth absolute moment of Z is given by (Winkelbauer 2014)

$$\mathbb{E}\left[|Z|^k\right] = \frac{\sigma_Z^k 2^{k/2}\Gamma\left(\frac{k+1}{2}\right)}{\pi^{1/2}} {}_1F_1\left(-\frac{1}{2}\left(\frac{\mu_Z}{\sigma_Z}\right)^2; -\frac{k}{2}, \frac{1}{2}\right), \tag{IA11}$$

where the function ${}_1F_1$ is defined in (53). Using this property with $k = 1$, the objective function in (18) is given by the following function of μ_P and σ_P when $P \sim \mathcal{N}(\mu_P, \sigma_P)$:

$$\mathcal{C}_{GN}(\mu_P, \sigma_P) := \ln \sigma_P - \frac{2^{\frac{\gamma-2}{2}}\Gamma\left(\frac{\gamma+1}{2}\right)}{\pi^{1/2}}\left(\frac{\sigma_P}{\beta}\right)^\gamma {}_1F_1\left(-\frac{1}{2}\left(\frac{\mu_P-\alpha}{\sigma_P}\right)^2; -\frac{\gamma}{2}, \frac{1}{2}\right). \tag{IA12}$$

To prove the result, we show that $\frac{\partial \mathcal{C}_{GN}(\mu_P,\sigma_P)}{\partial \mu_P} = 0$ if and only if $\mu_P = \alpha$, and then that $\frac{\partial \mathcal{C}_{GN}(\mu_P,\sigma_P)}{\partial \sigma_P}\big|_{\mu_P=\alpha} = 0$ if and only if $\sigma_P = C_1(\gamma)\sigma_T$. The first result is verified because

$$\frac{\partial_1 F_1\left(-\frac{1}{2}\left(\frac{\mu_P-\alpha}{\sigma_P}\right)^2; -\frac{\gamma}{2}, \frac{1}{2}\right)}{\partial \mu_P} = \gamma {}_1F_1\left(-\frac{1}{2}\left(\frac{\mu_P-\alpha}{\sigma_P}\right)^2; 1-\frac{\gamma}{2}, \frac{3}{2}\right)\left(\frac{\mu_P-\alpha}{\sigma_P^2}\right) = 0$$

if and only if $\mu_P = \alpha$. This is the unique zero because ${}_1F_1(-x^2; 1-\gamma/2, 3/2) > 0$ for all $x \in \mathbb{R}$ and $\gamma > 0$. Then, because ${}_1F_1\left(0; -\frac{\gamma}{2}, \frac{1}{2}\right) = 1$ for all γ, the function $\mathcal{C}_{GN}(\mu_P, \sigma_P)$ in (IA12) becomes, for $\mu_P = \alpha$,

$$\mathcal{C}_{GN}(\mu_P, \sigma_P) = \ln \sigma_P - \frac{2^{\frac{\gamma-2}{2}}\Gamma\left(\frac{\gamma+1}{2}\right)}{\pi^{1/2}}\left(\frac{\sigma_P}{\beta}\right)^{\gamma}. \tag{IA13}$$

From (IA13), it is straightforward to check that the unique solution to $\frac{\partial \mathcal{C}_{GN}(\mu_P,\sigma_P)}{\partial \sigma_P}\Big|_{\mu_P=\alpha} = 0$ is $\sigma_P = C_1(\gamma)\sigma_T = \sigma^{\star}_{GN}$. Given the unicity of the zeros to the partial derivatives, the optimum $(\mu_P, \sigma_P) = (\mu_T, \sigma^{\star}_{GN})$ is global and there is no other local optimum. Finally, $C_1(\gamma) \leq 1$ with equality if and only if $\gamma = 2$ because $\lim_{\gamma\to 0^+} C_1(\gamma) = \lim_{\gamma\to+\infty} C_1(\gamma) = 0$, and $C_1'(\gamma) > 0$ for $\gamma < 2$, $C_1'(\gamma) < 0$ for $\gamma > 2$.

Part 2. Noting that $\alpha = \ln(\mu_T - \delta) - \beta^2/2$, the objective function in (25) is given by the following function of (μ_P, μ_P) when $(\zeta_P, \kappa_P) = (0, 0)$:

$$\begin{aligned}\mathcal{C}_{SLN}(\mu_P, \sigma_P) := \ln \sigma_P - \ln(\mu_P - \delta) + \frac{\sigma_P^2}{2(\mu_P-\delta)^2} \\ - \frac{1}{2\beta^2}\left[\frac{\sigma_P^2}{(\mu_P-\delta)^2} + \frac{\sigma_P^4}{2(\mu_P-\delta)^4} + \left(\ln\left(\frac{\mu_P-\delta}{\mu_T-\delta}\right) - \frac{\sigma_P^2}{2(\mu_P-\delta)^2} + \frac{\beta^2}{2}\right)^2\right].\end{aligned} \tag{IA14}$$

To prove the result, we show that

$$\left.\frac{\partial \mathcal{C}_{SLN}(\mu_P, \sigma_P)}{\partial \sigma_P^2}\right|_{\mu_P=\mu_T} = 0 \text{ for } \sigma_P = C_2(\beta)\sigma_T \tag{IA15}$$

and

$$\left.\frac{\partial \mathcal{C}_{SLN}(\mu_P, \sigma_P)}{\partial \mu_P}\right|_{\sigma_P=C_2(\beta)\sigma_T} = 0 \text{ for } \mu_P = \mu_T. \tag{IA16}$$

In the proof of part 1 of Proposition 8, it is shown that $\mathcal{C}_{SLN}(\mu_P, \sigma_P)$ has no other local optima and, thus, that this optimum is global.

First, we prove Equation (IA15). Noting that $\mu_T - \delta = \sigma_T g(\beta)^{-1/2}$, when $\mu_P = \mu_T$ the objective function (IA14) becomes, after removing terms independent from σ_P,

$$\mathcal{C}_{SLN}(\mu_P, \sigma_P) = \frac{\sigma_P^4}{\sigma_T^4}\left(-\frac{3g(\beta)^2}{8\beta^2}\right) + \frac{\sigma_P^2}{\sigma_T^2}\left(g(\beta)\left(\frac{3}{4} - \frac{1}{2\beta^2}\right)\right) + \frac{1}{2}\ln \sigma_P^2.$$

Setting the differential with respect to σ_P^2 equal to zero, the optimal σ_P^2 must solve

$$\frac{\sigma_P^4}{\sigma_T^4}\left(\frac{3g(\beta)^2}{4\beta^2}\right)+\frac{\sigma_P^2}{\sigma_T^2}\left(g(\beta)\left(\frac{1}{2\beta^2}-\frac{3}{4}\right)\right)-\frac{1}{2}=0.$$

Of the two solutions to this second-degree polynomial, there is always only one positive solution, which is equal after simplifications to $\sigma_P = C_2(\beta)\sigma_T$.

Second, we prove Equation (IA16). Setting the differential of $\mathcal{C}_{SLN}(\mu_P, \sigma_P)$ in (IA14) with respect to μ_P equal to zero, we need to prove that

$$\begin{aligned}&\frac{1}{\mu_P-\delta}+\frac{\sigma_P^2}{(\mu_P-\delta)^3}+\frac{1}{\beta^2}\left(\ln(\mu_P-\delta)-\alpha-\frac{\sigma_P^2}{2(\mu_P-\delta)^2}\right)\left(\frac{1}{\mu_P-\delta}+\frac{\sigma_P^2}{(\mu_P-\delta)^3}\right)\\&-\frac{1}{\beta^2}\left(\frac{\sigma_P^2}{(\mu_P-\delta)^3}+\frac{\sigma_P^4}{(\mu_P-\delta)^5}\right)=0\end{aligned}$$

for $\mu_P = \mu_T$ if $\sigma_P = C_2(\beta)\sigma_T$. Setting $\mu_P = \mu_T$ and further simplifications result in the equality

$$\frac{\sigma_P^4}{\sigma_T^4}\left(\frac{3g(\beta)^2}{2\beta^2}\right)+\frac{\sigma_P^2}{\sigma_T^2}\left(g(\beta)\left(\frac{3}{2\beta^2}-\frac{3}{2}\right)\right)-\frac{3}{2}=0,$$

which must hold for $\sigma_P = C_2(\beta)\sigma_T$. Solving for σ_P^2 and keeping the only positive solution to the polynomial results in

$$\sigma_P=\sigma_T\sqrt{\frac{\beta^2-1+\sqrt{\beta^4+2\beta^2+1}}{2g(\beta)}}=C_2(\beta)\sigma_T=\sigma_{SLN}^{\star}.$$

Finally, $C_2(\beta) = 1$ if and only if $\beta \to 0^+$ from the proof of Proposition 3 and because, for $\beta > 0$, $C_2(\beta) < 1$ since $\exp(\beta^2) - 1 > \beta^2$ for all $\beta > 0$.

Proof of Corollary 1

The proof is direct because $P = w'R$ has a Gaussian distribution for any w when $R \sim \mathcal{N}(\mu, \mathbf{\Sigma})$ and, as shown in Proposition 7, the optimal mean and variance of $P \sim \mathcal{N}(\mu_P, \sigma_P)$ are $(\mu_P, \sigma_P) = (\mu_T, \sigma_{GN}^{\star})$ for the GN target distribution, and $(\mu_P, \sigma_P) = (\mu_T, \sigma_{SLN}^{\star})$ for the SLN target distribution, when those are attainable.

Proof of Proposition 8

Part 1. The strategy of the proof of part 1 of the proposition is to show that the partial derivatives of the MCKL objective function with respect to μ_P and σ_P vanish simultaneously at a single point in the plane $(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}^+$, which, of course, coincides with the maxima $(\mu_T, \sigma^\star_{GN})$ for the GN target distribution, and $(\mu_T, \sigma^\star_{SLN})$ for the SLN target distribution, in Proposition 7. This implies that there is no other stationary point than the global optimum and, thus, that starting from any point of the (μ, σ) plane, there always exists a geodesic connecting the point to the global optimum along which the MCKL objective function increases. This, in turn, shows that if (μ_T, σ_T) is chosen above the MVE frontier (implying that so is the global optimum because $\sigma^\star_{GN}, \sigma^\star_{SLN} \leq \sigma_T$), then the maximum of the MCKL objective function in the restriction of the (μ, σ) plane associated with the set of attainable portfolio-return mean and volatility, $\mathcal{MV}$, is reached on the MVE frontier.

For the MCKL portfolio $w_{\alpha,\beta,\gamma}$ associated with the GN target distribution, it is demonstrated in the proof of part 1 of Proposition 7 that $(\mu_T, \sigma^\star_{GN})$ is the unique local (and thus global) optimum of the objective function $\mathcal{C}_{GN}(\mu_P, \sigma_P)$. This also proves the case $w_{\alpha,\beta}$ corresponding to the Gaussian target distribution since it is a GN target with $\gamma = 2$.

We turn next to the MCKL portfolio $\widetilde{w}_{\alpha,\beta,\delta}$ associated with the SLN target distribution. To simplify the formulas, we introduce the notation

$$\nu_P := \frac{\sigma_P}{(\mu_P - \delta)} > 0$$

given the assumption in part 2 of Proposition 4 that $\mu_P > \delta$. First, we derive the unique zero of $\frac{\partial \mathcal{C}_{SLN}(\mu_P, \sigma_P)}{\partial \sigma_P^2}$ for a fixed μ_P, where $\mathcal{C}_{SLN}(\mu_P, \sigma_P)$ is defined in (IA14). After some algebra, this amounts to finding the roots of a second-degree polynomial in ν_P^2:

$$\nu_P^4 \left[-\frac{3}{4\beta^2} \right] + \nu_P^2 \left[\frac{1}{2\beta^2} \left(\frac{3\beta^2}{2} - 1 + \ln\left(\frac{\mu_P - \delta}{\mu_T - \delta} \right) \right) \right] + \frac{1}{2} = 0.$$

There is a unique positive root to this polynomial, which is given by

$$\nu_P^2 = \frac{1}{3}\left[\frac{3\beta^2}{2} - 1 + \ln\left(\frac{\mu_P - \delta}{\mu_T - \delta}\right) + \sqrt{6\beta^2 + \left(\frac{3\beta^2}{2} - 1 + \ln\left(\frac{\mu_P - \delta}{\mu_T - \delta}\right)\right)^2}\right]. \tag{IA17}$$

Second, we turn to the derivative with respect to μ_P for a fixed σ_P. Because $\mu_P > \delta$, the derivative of $\mathcal{C}_{SLN}(\mu_P, \sigma_P)$ with respect to $(\mu_P - \delta)^2$ has the same sign as that with respect to μ_P, and we find after some algebra that

$$-2\beta^2(\mu_P - \delta)^2 \frac{\partial \mathcal{C}_{SLN}(\mu_P, \sigma_P)}{\partial(\mu_P - \delta)^2} = (1 + \nu_P^2)\left(\frac{3}{2}(\beta^2 - \nu_P^2) + \ln\left(\frac{\mu_P - \delta}{\mu_T - \delta}\right)\right). \tag{IA18}$$

This means that the derivative has a unique zero. Indeed, $\frac{\partial \mathcal{C}_{SLN}(\mu_P, \sigma_P)}{\partial(\mu_P - \delta)^2} = 0$ when

$$\frac{3}{2}(\beta^2 - \nu^2) + \ln\left(\frac{\mu_P - \delta}{\mu_T - \delta}\right) = 0 \Leftrightarrow -\frac{3\sigma_P^2}{(\mu_P - \delta)^2} + 2\ln(\mu_P - \delta) = 2\ln(\mu_T - \delta) - 3\beta^2.$$

The left-hand side of this equality is continuous, spans $\mathbb{R}$ as a function of $\mu_P > \delta$, and is strictly increasing with μ_P. Therefore, it equals the right-hand for a single value of μ_P and the zero is unique.

Third, we show that the two partial derivatives of $\mathcal{C}_{SLN}(\mu_P, \sigma_P)$ vanish simultaneously at a single point, corresponding to the global optimum $(\mu_P, \sigma_P) = (\mu_T, \sigma^\star_{SLN})$. As explained above, this then proves part 1 of the proposition. From (IA18), the optimal μ_P for a given σ_P is such that $\ln\left(\frac{\mu_P - \delta}{\mu_T - \delta}\right) = \frac{3}{2}(\nu_P^2 - \beta^2)$. Inserting this condition into the optimal σ_P^2 in (IA17) means that the two partial derivatives vanish simultaneously when

$$\left(\frac{3\nu_P^2}{2} + 1\right)^2 = 6\beta^2 + \left(\frac{3\nu_P^2}{2} - 1\right)^2 \Leftrightarrow 6\nu_P^2 = 6\beta^2 \Leftrightarrow \nu_P = \beta$$

because both ν_P and β are positive. This unique solution coincides with the global optimum $(\mu_P, \sigma_P) = (\mu_T, \sigma^\star_{SLN})$ because $\nu_P = \beta$ in that case.

Part 2. When the asset returns are Gaussian, as noted in the proof of Proposition 7, the standardized entropy $H[P^\star]$ is a constant that can be removed from the objective function.

Thus, the Lagrangian function associated with the objective function in (31) is given by

$$\mathcal{L} = -\frac{1}{2\beta^2}\big((w'\mu - \alpha)^2 + w'\boldsymbol{\Sigma}w\big) + \frac{1}{2}\ln w'\boldsymbol{\Sigma}w + \tau(w'\iota - 1),$$

where τ is the Lagrange multiplier associated with the full-investment constraint $w'\iota = 1$. Setting the derivative of $\mathcal{L}$ with respect to w' equal to zero yields, after some developments,

$$w = \frac{w'\boldsymbol{\Sigma}w}{\beta^2 - w'\boldsymbol{\Sigma}w}\left(\frac{\mu_g(w'\mu - \alpha)}{\sigma_g^2}w_{MSR} - \frac{\tau\beta^2}{\sigma_g^2}w_{GMV}\right). \tag{IA19}$$

Replacing (IA19) in the condition $w'\iota = 1$ gives the following value for the Lagrange multiplier τ:

$$\tau = \frac{\sigma_g^2}{\beta^2}\left(\frac{\mu_g(w'\mu - \alpha)}{\sigma_g^2} + 1 - \frac{\beta^2}{w'\boldsymbol{\Sigma}w}\right). \tag{IA20}$$

Combining (IA19) and (IA20), and denoting $(\mu_P^\star, \sigma_P^{\star 2})$ the optimal MCKL portfolio-return mean and variance, results in the following solution for the MCKL portfolio $w_{\alpha,\beta}$:

$$w_{\alpha,\beta} = w_{GMV} + \frac{\sigma_P^{\star 2}\mu_g(\mu_P^\star - \alpha)}{\sigma_g^2(\beta^2 - \sigma_P^{\star 2})}(w_{MSR} - w_{GMV}), \tag{IA21}$$

which, as a special case of part 1 of the proposition, is an MVE portfolio when $\mu_T \geq \mu_g$ and $\sigma_T \leq \sigma_{EF}(\mu_T)$, i.e., $\sigma_P^\star = \sigma_{EF}(\mu_P^\star)$. To find the optimal mean return $\mu_P^\star$, one must solve $w'_{\alpha,\beta}\mu = \mu_P^\star$, which from (IA21) reduces to solving (41). Note that the division by $\beta^2 - \sigma_P^{\star 2}$ in (IA21) is not problematic. When $(\alpha, \beta) \notin \mathcal{MV}$, i.e., (α, β) is above the MVE frontier, then $\sigma_P^\star \neq \beta$ because of the presence of the first term in (31) that favors having a larger mean return than $\sigma_{EF}^{-1}(\beta)$. When $(\alpha, \beta) \in \mathcal{MV}$, i.e., (α, β) is on the MVE frontier, then the limit of $(\mu_P^\star - \alpha)/(\beta^2 - \sigma_P^{\star 2})$ in (IA21) as $\mu_P \to \alpha$ is necessarily such that $w_{\alpha,\beta}$ in an MVE portfolio with mean return $\mu_P = \alpha$.

Part 2(i). We now prove that $\mu_P^\star = \mu_g$ if $\alpha = \mu_g$. This is obvious when $\beta = \sigma_{EF}(\alpha) = \sigma_g$ because $(\alpha, \beta) \in \mathcal{MV}$ in such a case. Consider now the case $\beta < \sigma_g$ and denote $\beta^2 = \sigma_g^2/c$ with $c > 1$, in which case $(\alpha, \beta) \notin \mathcal{MV}$. Denote the MCKL objective function as a function of (μ_P, σ_P) as

$$\mathcal{C}_N(\mu_P, \sigma_P) := -\frac{1}{2}\left(\frac{\mu_P - \alpha}{\beta}\right)^2 + \frac{1}{2}\left(\ln\sigma_P^2 - \frac{\sigma_P^2}{\beta^2}\right). \tag{IA22}$$

When $(\mu_P, \sigma_P) = (\mu_g, \sigma_g)$, we easily find that

$$\mathcal{C}_N(\mu_P, \sigma_P) = \frac{1}{2}(\ln \sigma_g^2 - c). \tag{IA23}$$

Let us now show that choosing an MVE portfolio with $\mu_P > \mu_g$ will necessarily result in a smaller value for the objective function. Specifically, consider $\mu_P = \mu_g + \epsilon$, $\epsilon > 0$, and thus $\sigma_P^2 = \sigma_{EF}^2(\mu_P) = \sigma_g^2 + \epsilon^2/(\psi^2 - \psi_g^2)$. In that case, the objective function becomes

$$\begin{aligned}\mathcal{C}_N(\mu_P, \sigma_P) &= \frac{1}{2}\left(-\frac{c\epsilon^2}{\sigma_g^2}\left(1 + \frac{1}{\psi^2 - \psi_g^2}\right) + \ln\left(\sigma_g^2 + \frac{\epsilon^2}{\psi^2 - \psi_g^2}\right) - c\right) \\ &\leq \frac{1}{2}\left(-\frac{c\epsilon^2}{\sigma_g^2}\left(1 + \frac{1}{\psi^2 - \psi_g^2}\right) + \ln \sigma_g^2 + \frac{\epsilon^2}{\sigma_g^2(\psi^2 - \psi_g^2)} - c\right),\end{aligned} \tag{IA24}$$

from the inequality $\ln(x + a) \leq \ln x + a/x$. Comparing (IA23) and (IA24), we have that $\mu_P = \mu_g$ is optimal if

$$-\frac{c\epsilon^2}{\sigma_g^2}\left(1 + \frac{1}{\psi^2 - \psi_g^2}\right) + \frac{\epsilon^2}{\sigma_g^2(\psi^2 - \psi_g^2)} < 0 \quad \text{for} \quad \epsilon^2 > 0,$$

which holds because the inequality reduces to $-c + (1 - c)/(\psi^2 - \psi_g^2) < 0$ and the left-hand side is indeed negative because $c > 1$.

Part 2(ii). We now prove that $\sigma_{EF}^{-1}(\beta) \leq \mu_P^\star \leq \mu_{\ell_2} \leq \alpha$ if $\alpha > \mu_g$. It is clear that the statement holds with equalities when $\beta = \sigma_{EF}(\alpha)$ because $(\alpha, \beta) \in \mathcal{MV}$ in such a case. Therefore, we focus on proving that the statement holds with strict inequalities when $\beta < \sigma_{EF}(\alpha)$. We proceed in two steps (Figure 10 provides a graphical representation of those inequalities).

First, we prove that $\mu_P^\star > \sigma_{EF}^{-1}(\beta)$ if $\alpha > \mu_g$. Because $w_{\alpha,\beta}$ is an MVE portfolio, we have $\mu_P^\star \geq \mu_g$ and, because $\alpha > \mu_g$, we have from the first term in (IA22) that $\mu_P^\star > \mu_g$. Therefore, from (41), we must have $\text{sign}(\sigma_P^\star - \beta) = \text{sign}(\alpha - \mu_P^\star)$. Because, as we show below, $\mu_P^\star < \alpha$, we conclude that $\sigma_P^\star > \beta$ and $\mu_P^\star > \sigma_{EF}^{-1}(\beta)$. Note that $\sigma_{EF}^{-1}(\beta)$ does not exist when $\beta < \sigma_g$ and, in that case, μ_g remains a strict lower bound for $\mu_P^\star$.

Second, we prove that $\mu_P^\star < \mu_{\ell_2} < \alpha$ if $\alpha > \mu_g$. Recall first that we deal with the case where the target is not in the feasible set, $(\alpha, \beta) \notin \mathcal{MV}$. Thus, $\beta < \sigma_{EF}(\alpha)$. The mean return

of the MVE portfolio minimizing the ℓ_2 distance to (α, β) solves

$$\mu_{\ell_2} = \min_{\mu_P} (\mu_P - \alpha)^2 + (\sigma_{EF}(\mu_P) - \beta)^2. \tag{IA25}$$

Observe from the geometry of the problem that μ_{ℓ_2} is unique. Setting the derivative of the ℓ_2 distance in the right-hand side of (IA25) equal to zero, μ_{ℓ_2} is thus the unique value of μ satisfying

$$f(\mu) := \frac{\alpha - \mu}{\mu - \mu_g}(\psi^2 - \psi_g^2)\sigma_{EF}(\mu) = \sigma_{EF}(\mu) - \beta := h(\mu).$$

Moreover, $\mu_{\ell_2} > \mu_g$ (because μ_{ℓ_2} sits on the MVE frontier when $\alpha > \mu_g$) and, from the concavity of the MVE frontier, we have $\mu_{\ell_2} < \alpha$ and $\mu_{\ell_2} > \sigma_{EF}^{-1}(\beta)$ if $\beta \geq \sigma_g$. From (41), we find after some algebra that the mean return $\mu_P^\star$ of the MCKL portfolio is the value of μ solving

$$f(\mu) = h(\mu)\frac{\sigma_{EF}(\mu) + \beta}{\sigma_{EF}(\mu)} =: l(\mu).$$

Because $\beta > 0$, we necessarily have $\mu_P^\star < \mu_{\ell_2}$. To see this, recall that $f(\mu) - h(\mu)$ has a unique root $\mu_{\ell_2} < \alpha$ and that $h(\alpha) > 0 = f(\alpha)$, which shows that $h(\mu) > f(\mu)$ for all $\mu > \mu_{\ell_2}$. Let us now prove that l dominates h on (μ_{ℓ_2}, ∞), which amounts to show that $h(\mu) > 0$ for all $\mu > \mu_{\ell_2}$. Observe that $h(\mu) > 0$ for all μ if $\beta < \sigma_g$. If $\beta \geq \sigma_g$, $\sigma_{EF}^{-1}(\beta)$ exists. Using $\mu_{\ell_2} > \sigma_{EF}^{-1}(\beta)$ and the increasing property of h,

$$h(\mu_{\ell_2}) > h(\sigma_{EF}^{-1}(\beta)) = 0.$$

Because h is increasing, we find that $h(\mu) > 0$ for all $\mu > \mu_{\ell_2}$, leading to $l(\mu) > h(\mu)$ on the right of μ_{ℓ_2}. We have thus shown that on (μ_{ℓ_2}, ∞), l dominates h and h dominates f, and conclude that l dominates f on the right of μ_{ℓ_2}. This proves that the root of $f(\mu) - l(\mu)$ must satisfy $\mu_P^\star < \mu_{\ell_2}$.

Proof of Proposition 9

To prove the proposition, we need to evaluate the integral in (51). By definition of $\hat{f}_P^{GME}$ in (49) and applying Fubini's theorem, this integral becomes

$$\begin{aligned}\int_{-\infty}^{+\infty} \sum_{k=1}^{K} |x-\hat{\alpha}|^\gamma \frac{\pi_k}{b_k}\phi\left(\frac{x-m_k}{b_k}\right) \mathrm{d}x &= \sum_{k=1}^{K} \int_{-\infty}^{+\infty} |x-\hat{\alpha}|^\gamma \frac{\pi_k}{b_k}\phi\left(\frac{x-m_k}{b_k}\right) \mathrm{d}x \\ &= \sum_{k=1}^{K} \pi_k \mathbb{E}\left[|Y-\hat{\alpha}|^\gamma\right] \quad \text{where} \quad Y \sim \mathcal{N}(m_k, b_k) \\ &= \sum_{k=1}^{K} \pi_k \mathbb{E}\left[|Z|^\gamma\right] \quad \text{where} \quad Z \sim \mathcal{N}(m_k - \hat{\alpha}, h_k).\end{aligned}$$

The proposition follows from the formula for the absolute moment of a Gaussian random variable in (IA11).

Proof of Proposition 10

From the properties of the ${}_1F_1$ function in (53), we have that ${}_1F_1(x;-1,1/2) = 1-2x$. Thus, the estimated expectation in (52) becomes, for $\gamma = 2$,

$$\widehat{\mathbb{E}}\left[(P-\hat{\alpha})^2\right] = \sum_{k=1}^{K} \pi_k b_k^2 \left(1 + \left(\frac{m_k - \hat{\alpha}}{b_k}\right)^2\right).$$

Consequently, the estimated objective function $\widehat{\mathcal{C}}(w;\gamma,\hat{w}_0)$ in (48) becomes

$$\widehat{\mathcal{C}}(w;\gamma=2,\hat{w}_0) = \widehat{H}[P] - \frac{1}{2\hat{\beta}^2} \sum_{k=1}^{K} \pi_k \left((m_k-\hat{\alpha})^2 + b_k^2\right).$$

Considering the kernel density estimator $\hat{f}_P^{KDE}$, this simplifies to

$$\widehat{\mathcal{C}}(w;\gamma=2,\hat{w}_0) = \widehat{H}[P] - \frac{1}{2\hat{\beta}^2}\left(\frac{1}{h}\sum_{t=1}^{N}(P_t-\hat{\alpha})^2 + \frac{1}{h}\sum_{t=1}^{N} b_t^2\right). \tag{IA26}$$

For b_t independent from w, the second sum in (IA26) is a constant that can be removed from the objective function. Moreover, $\frac{1}{h}\sum_{t=1}^{N}(P_t-\hat{\alpha})^2$ is the sample estimator of the expectation $\mathbb{E}[(P-\hat{\alpha})^2] = (\mu_P-\hat{\alpha})^2 + \sigma_P^2$. Therefore, noting that $\hat{\beta} = \hat{\sigma}_0$ when $\gamma = 2$, (IA26) simplifies

indeed to the objective function corresponding to the Gaussian target in (54).

IA.2 CKL divergence with shifted-log-normal target

In this section, we assume that the portfolio return P is Gaussian, $P \sim \mathcal{N}(\mu_P, \sigma_P)$, and we study the CKL divergence corresponding to an SLN target; that is, the objective function

$$\mathcal{C}(w; \alpha, \beta, \delta) = \mathbb{E}[\ln f_T(P) | P \in \Omega_T] + H[P | P \in \Omega_T] - \frac{p}{1-p} \tag{IA27}$$

with f_T defined in (19). Note that, in this case, $p = \mathbb{P}(P < \delta) = \Phi\left(\frac{\delta - \mu_P}{\sigma_P}\right)$, where Φ is the standard Gaussian cdf. We first begin with the conditional entropy, $H[P|P \in \Omega_T]$, then study the conditional cross-entropy $\mathbb{E}[\ln f_T(P)|P \in \Omega_T]$, explain how the objective function simplifies under the approximation $p = 0$, and finally we compare the different objective functions via a numerical example. It is useful to define the function

$$h(p) := \phi(z_p)/(1-p), \tag{IA28}$$

where ϕ is the standard Gaussian pdf and $z_p := \Phi^{-1}(p)$ is its p-th quantile.

IA.1 Conditional entropy

The support set is $\Omega_T = [\delta, \infty)$ and the conditional entropy becomes $H[P|P \in \Omega_T] = H_\delta[P]/(1-p)$ with

$$H_\delta[P] := -\int_\delta^\infty f_P(x) \ln f_P(x) \mathrm{d}x = H[P] + \int_{-\infty}^\delta f_P(x) \ln f_P(x) \mathrm{d}x.$$

Using $f_P(x) = \phi((x-\mu_P)/\sigma_P)/\sigma_P$, the integral simplifies to

$$\begin{aligned}
\int_{-\infty}^\delta f_P(x) \ln f_P(x) \mathrm{d}x &= \int_{-\infty}^\delta \frac{1}{\sigma_P} \phi\left(\frac{x-\mu_P}{\sigma_P}\right) \ln\left(\frac{1}{\sigma_P}\phi\left(\frac{x-\mu_P}{\sigma_P}\right)\right) \mathrm{d}x \\
&= \int_{-\infty}^{\frac{\delta-\mu_P}{\sigma_P}} \phi(y) \ln\left(\frac{1}{\sigma_P}\phi(y)\right) \mathrm{d}y \\
&= I(z_p) - p \ln \sigma_P,
\end{aligned}$$

where

$$I(a) := \int_{-\infty}^{a} \phi(x) \ln \phi(x) \mathrm{d}x = \frac{a\phi(a) - \Phi(a)(1 + \ln(2\pi))}{2}.$$

Noting that $H[P] = H[P^\star] + \ln \sigma_P$ with $H[P^\star] = (1 + \ln(2\pi))/2$ and $\Phi(z_p) = p$, $H_\delta[P]$ simplifies to

$$H_\delta[P] = H[P] + I(z_p) - p \ln \sigma_P = (1-p)H[P] + \frac{z_p}{2}\phi(z_p).$$

Note that $z_p\phi(z_p) \to 0$ when $p \to 0^+$. Finally, the conditional entropy is equal to

$$H[P|P \geq \delta] = \frac{H_\delta[P]}{1-p} = H[P] + \frac{z_p}{2}h(z_p). \tag{IA29}$$

IA.2 Conditional cross-entropy

We now turn to the conditional cross-entropy term in (IA27), $\mathbb{E}[\ln f_T(P)|P \in \Omega_T]$. We study first the true value and, second, the Taylor-series approximation.

IA.2.1 True value

Ignoring the proportionality constant $1/(\beta\sqrt{2\pi})$ in (19), the conditional cross-entropy is equal to

$$\begin{aligned}
\mathbb{E}[\ln f_T(P)|P \geq \delta] &= -\mathbb{E}[\ln P_\delta | P_\delta \geq 0] - \frac{1}{2\beta^2}\mathbb{E}[(\ln P_\delta) - \alpha)^2 | P_\delta \geq 0] \\
&= -\mathbb{E}[\ln P_\delta | P_\delta \geq 0] - \frac{1}{2\beta^2}\left(\mathbb{E}[(\ln P_\delta)^2 | P_\delta \geq 0] - 2\alpha\mathbb{E}[\ln P_\delta | P_\delta \geq 0] + \alpha^2\right) \\
&= -\frac{1}{2\beta^2}\left(\mathbb{E}[(\ln P_\delta)^2 | P_\delta \geq 0] + 2(\beta^2 - \alpha)\mathbb{E}[\ln P_\delta | P_\delta \geq 0] + \alpha^2\right),
\end{aligned} \tag{IA30}$$

where

$$\mathbb{E}[(\ln P_\delta)^k | P_\delta \geq 0] = \frac{1}{1-p}\int_\delta^\infty f_P(x) \left(\ln(x-\delta)\right)^k \mathrm{d}x = \frac{1}{1-p}\int_0^\infty f_P(x+\delta) \left(\ln x\right)^k \mathrm{d}x. \tag{IA31}$$

This integral does not simplify further for $f_P(x) = \phi((x-\mu_P)/\sigma_P)/\sigma_P$, but can be evaluated numerically.

IA.2.2 Taylor-series approximation

Let us write the conditional cross-entropy in (IA30) as

$$\mathbb{E}[\ln f_T(P)|P \geq \delta] = -\mathbb{E}[\ln P_\delta|P_\delta \geq 0] - \frac{1}{2\beta^2}\left(\mathbb{V}[\ln P_\delta|P_\delta \geq 0] + (\mathbb{E}[\ln P_\delta|P_\delta \geq 0] - \alpha)^2\right). \tag{IA32}$$

A second-order Taylor-series expansion of $\ln x$ yields the approximations in (23)–(24):

$$\begin{aligned} \mathbb{E}[\ln P_\delta|P_\delta \geq 0] &\approx \ln(m(P_\delta)) - \psi_2(P_\delta)/2, \\ \mathbb{V}[\ln P_\delta|P_\delta \geq 0] &\approx \psi_2(P_\delta) - \psi_3(P_\delta) + \psi_4(P_\delta)/4 - \frac{\psi_2(P_\delta)}{4m(P_\delta)^2}. \end{aligned} \tag{IA33}$$

Moreover, when $P \sim \mathcal{N}(\mu_P, \sigma_P)$, we have from Pender (2015) that

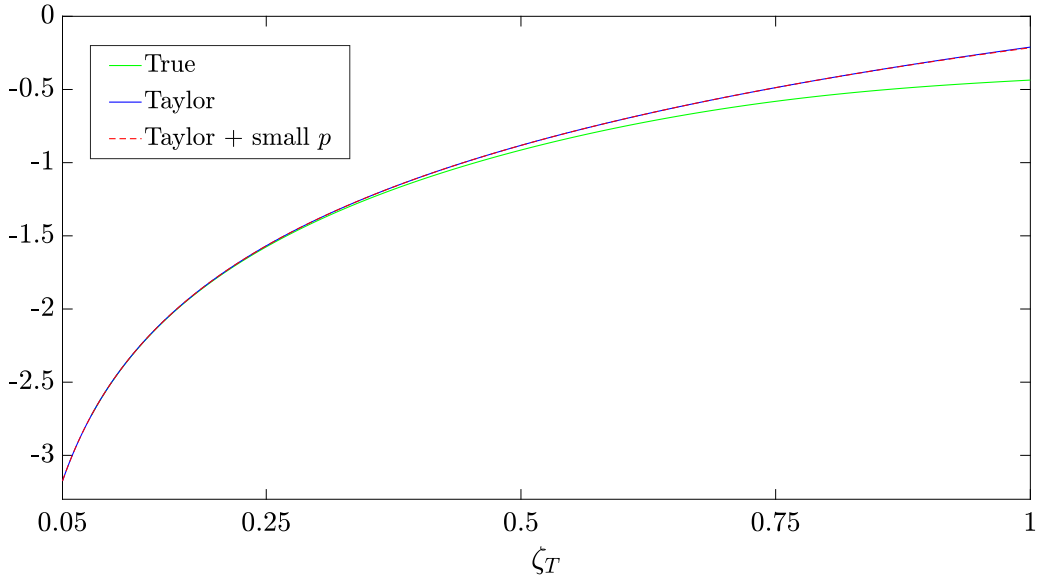
$$\begin{aligned} m(P_\delta) &= (\mu_P - \delta) + \sigma_P h(p), \\ \frac{m_2(P_\delta)}{\sigma_P^2} &= 1 + z_p h(p) - h^2(p), \\ \frac{m_3(P_\delta)}{\sigma_P^3} &= (z_P^2 - 1)h(p) - 3z_p h^2(p) + 2h^3(p), \\ \frac{m_4(P_\delta)}{\sigma_P^4} &= 3 + 12z_p h^3(p) - 4(z_p^2 - 1)h^2(p) - 3z_p h^2(p) - 6h^4(p) + (z_p^3 - 3z_p)h(p). \end{aligned} \tag{IA34}$$

IA.3 Approximation for $p = 0$

As discussed in Section 4.2, the probability $p = \mathbb{P}(P \leq \delta)$ is typically very small. If we plug in $p = 0$ in the objective function (IA27) with the conditional entropy given by (IA29) and the Taylor-series approximation to the conditional cross-entropy given by (IA32)–(IA33)–(IA34), we obtain the approximation to the objective function $\widetilde{\mathcal{C}}(w; \alpha, \beta, \delta)$ in (25), which for P being Gaussian is equal to

$$\widetilde{\mathcal{C}}(w; \alpha, \beta, \delta) = H[P] - \widetilde{\mathbb{E}}[\ln P_\delta|P_\delta \geq 0] - \frac{1}{2\beta^2}\left(\widetilde{\mathbb{V}}[\ln P_\delta|P_\delta \geq 0] + (\widetilde{\mathbb{E}}[\ln P_\delta|P_\delta \geq 0] - \alpha)^2\right) \tag{IA35}$$

Figure IA.1: Accuracy of approximation to the objective function when targeting a shifted-log-normal distribution



Notes. The figure depicts the true objective function corresponding to the shifted-log-normal target-return distribution (solid green), its approximation via a Taylor-series expansion (solid blue), and its approximation via a Taylor-series expansion plus assuming that the probability $p = \mathbb{P}(P \leq \delta)$ is small (dotted red). We consider the special case in which the portfolio return P has a Gaussian distribution, for which we derive closed-form expressions in Section IA.2. We fix $(\mu_P, \sigma_P) = (\mu_T, \sigma_T) = (0.1, 0.2)$, and the target-return skewness ζ_T varies between 0.05 and 1.

with

$$\widetilde{\mathbb{E}}[\ln P_\delta | P_\delta \geq 0] = \ln(\mu_P - \delta) - \frac{\sigma_P^2}{2(\mu_P - \delta)^2}, \tag{IA36}$$

$$\widetilde{\mathbb{V}}[\ln P_\delta | P_\delta \geq 0] = \frac{\sigma_P^2}{(\mu_P - \delta)^2} + \frac{\sigma_P^4}{2(\mu_P - \delta)^4}. \tag{IA37}$$

IA.4 Numerical comparison

In Figure IA.1, we set $(\mu_P, \sigma_P) = (\mu_T, \sigma_T) = (0.1, 0.2)$ and depict, as a function of the target-return skewness ζ_T between 0.05 and 1, the true objective function, the one approximated via a Taylor-series expansion, and the one approximated via a Taylor-series expansion plus assuming $p = 0$, for P being Gaussian as considered above. We find that the Taylor-series approximation is accurate as long as ζ_T is not too large, and that the small-p approximation introduces a very negligible error. Note that when $\zeta_T \to 0^+$, which corresponds to targeting a Gaussian distribution, the three objective functions correspond to one another.

IA.3 Choice of kernel bandwidth

One common choice to set the kernel bandwidth b is Silverman (1986)'s rule-of-thumb. However, it is well-known that this bandwidth tends to oversmooth the density. Another class of bandwidths are those found via cross-validation; see Rudemo (1982). While more accurate, such bandwidths are known to lack robustness and be highly variable, thus often undersmoothing the density. We rely instead on the approach of Zhang and Wang (2008), who propose a bandwidth based on quantile ratios that is both robust and accurate. This bandwidth is defined as

$$b = \left(\frac{4}{3h}\right)^{1/5} \widehat{Q}_{0.3}, \tag{IA38}$$

where h is the sample size and $\widehat{Q}_{0.3}$ is the 30% quantile of the ratio quantile ranges

$$\widehat{\mathrm{RQR}}_t = \frac{P_{(\underline{t+m}:h)} - P_{(\underline{t-m}:h)}}{\Phi^{-1}(q_t) - \Phi^{-1}(p_t)},\ q_t = \frac{\underline{t+m} - 0.5}{h},\ p_t = \frac{\underline{t-m} - 0.5}{h},\ t = 1, \ldots, h, \tag{IA39}$$

where $m = \left[h^{1/2}\right]$, $\underline{x} = x$ if $1 \leq x \leq h$, 1 if $x < 1$, and h if $x > h$. To avoid that the bandwidth b varies with the portfolio weights w, which could unpredictably affect the obtained optimal portfolio, we compute the bandwidth in (IA38)–(IA39) based on the equally weighted portfolio with sample returns $P_t = \iota' R_t / N$.

References

Cover, T., & Thomas, J. (2006). *Elements of Information Theory* (2nd ed.). New Jersey: Wiley.

Greene, W. (1993). *Econometric Analysis*. New York: Macmillan.

Pender, J. (2015). The truncated normal distribution: Applications to queues with impatient customers. *Operations Research Letters*, 43(1):40–45.

Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics*, 9(2):65–78.

Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman & Hall.

Winkelbauer, A. (2014). Moments and absolute moments of the normal distribution. arXiv:1209.4340v2.

Zhang, J., & Wang, X. (2008). Robust normal reference bandwidth for kernel density estimation. *Statistica Neerlandica*, 63(1):13–23.