Quantitative Finance

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/rquf20

Myopic robust index tracking with Bregman divergence

S. Penev, P. V. Shevchenko & W. Wu

To cite this article: S. Penev, P. V. Shevchenko & W. Wu (2022) Myopic robust index tracking with Bregman divergence, Quantitative Finance, 22:2, 289-302, DOI: 10.1080/14697688.2021.1950918

To link to this article: https://doi.org/10.1080/14697688.2021.1950918

Published online: 23 Jul 2021.

Submit your article to this journal

Article views: 372

View related articles

View Crossmark data

Citing articles: 2 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=rquf20

Quantitative Finance, 2022
Vol. 22, No. 2, 289–302, https://doi.org/10.1080/14697688.2021.1950918

Routledge
Taylor & Francis Group
Check for updates

Myopic robust index tracking with Bregman divergence

S. PENEV†*, P. V. SHEVCHENKO‡§ and W. WU†

†School of Mathematics and Statistics, UNSW Sydney, Sydney 2052, Australia
‡Department of Actuarial Studies and Business Analytics, Macquarie University, Sydney 2109, Australia
§Center for Econometrics and Business Analytics, St. Petersburg State University, Saint Petersburg, Russia

(*Received 6 September 2020; accepted 25 June 2021; published online 23 July 2021*)

Index tracking is a popular form of asset management. Typically, a quadratic function is used to define the tracking error of a portfolio and the look back approach is applied to solve the index tracking problem. We argue that a forward looking approach is more suitable, whereby the tracking error is expressed as an expectation of a function of the difference between the returns of the index and of the portfolio. We also assume that there is model uncertainty in the distribution of the assets, hence a robust version of the optimization problem needs to be adopted. We use Bregman divergence in describing the deviation between the nominal and actual (true) distribution of the components of the index. In this scenario, we derive the optimal robust index tracking portfolio in a semi-analytical form as a solution of a system of nonlinear equations. Several numerical results are presented that allow us to compare the performance of this robust portfolio with the optimal non-robust portfolio. We show that, especially during market downturns, the robust portfolio can be very advantageous.

Keywords: Index tracking; Robust index tracking; Bregman divergence; Kullback–Leibler divergence

JEL Classifications: G11, D81

1. Introduction

A popular form of passive asset management is so-called index tracking (see discussions, for example, in Beasley *et al.* (2003), Gaivoronski *et al.* (2005) and Andriosopoulos and Nomikos (2014)). Essentially, it means that the fund manager (or the investor) tries to replicate the performance of an index either through its value or its return (see Strub and Baumann (2018) and the references therein). In a frictionless and liquid market, a full replication portfolio (i.e. by holding exactly the same composition as the index) obviously yields the best tracking performance. This has already been discussed in many past studies, e.g. Beasley *et al.* (2003) and Strub and Baumann (2018). However, if transaction costs are considered or some of the components of the index are illiquid (see Maginn *et al.* (2007)), then a full replication portfolio does not necessarily deliver the best performance. This is due to the fact that a full replication portfolio will involve high transaction costs and because buying and selling illiquid assets will be difficult. Thus, to replicate the index, the fund manager may choose a tracking portfolio with only a subset of representative assets (see Guastaroba and Speranza (2012), de Paulo *et al.* (2016) and Strub and Baumann (2018) among others). It is worth noting that, in general, the assets in the tracking portfolio do not have to be the components of the index as long as they exhibit good similarity with the index (see, e.g., Andriosopoulos and Nomikos (2014)).

To satisfactorily solve the index tracking problem or the related enhanced tracking problem (which tracks the index as well as outperforms it), predominantly the look back approach has been used in the literature, see, e.g. Beasley *et al.* (2003), Montfort *et al.* (2008), Chiam *et al.* (2013), Guastaroba *et al.* (2016), and Strub and Baumann (2018). This approach relies on the assumption that a portfolio that has tracked the index well in the past will also demonstrate good tracking performance in the future. The past realizations of the return (or value) of the index and the return (or value) of the tracking portfolio are collected and the tracking error is defined as a function of the difference between the index and the tracking portfolio. A quadratic function is often used to define such a tracking error.

The above approach may lead to poor performance if the future differs vastly from the past. Another way to solve the index tracking problem is by adopting the forward looking

*Corresponding author. Email: S.Penev@unsw.edu.au

© 2021 Informa UK Limited, trading as Taylor & Francis Group

approach. This is based on defining the tracking error as the expectation of a function of the difference between the return of the index and the return of the tracking portfolio (see, e.g., Meade and Salkin (1990) and de Paulo *et al.* (2016)). The expectation is calculated by using the joint distribution of the index and of the tracking portfolio. A reliable estimation of the joint distribution is required in order to guarantee the good performance of this approach. If uncertainty in the distribution of the assets is present, a robust version of the approach needs to be taken.

From the onset of this paper, we should point out that index tracking has not received the attention it deserves until recently, but the situation has since changed dramatically. Many aspects of index tracking currently attract the attention of the research community.

An interesting recent research is directed towards the attempt to unify the two tasks, selecting the 'optimal subset' and determining the 'best' weights in a single step. This approach is attractive, especially if the index contains a large number of stocks, while one is interested in having a (very) small number of stocks in the portfolio (Benidis *et al.* 2018). In other words, the goal in this case is to select a sparse index tracking portfolio. While we are also interested in developing such types of approaches, they are based on different optimization techniques (e.g. sparse non-convex optimization (Benidis *et al.* 2018), mixed-integer programming Canakgoz and Beasley (2009), Filippi *et al.* (2016)), and we will not discuss these approaches in this paper.

It may be possible to use quite sophisticated statistical multivariate models and numerical approximation techniques to better mimic the underlying joint distribution of the stock returns and the index returns, and our theory allows for such models to be used as a nominal distribution in index tracking. However, our point of view in this paper is that even then a robustness approach is worth using, as it has the potential to safeguard against deviations from the nominal model.

Gnägi and Strub (2020) give a review of recent approaches and examine critically the gaps in the literature on the enhanced index-tracking problem. They also propose novel matheuristics (combining metaheuristics with mathematical programming) numerical algorithms to solve index-tracking problems with very large indices (e.g. with more than 9000 constituents). Among the many constraints they consider, is the achievement of a given target excess return in the future. Because future outcomes are uncertain, they end up requiring a minimum expected excess return which may not deliver the target in an out-of-sample period. Their paper does not deal with the robustness-type penalties that are of main interest in our paper.

As mentioned, the most recent literature focuses on the look back approach and a great effort has been made to model transaction costs and other sophisticated restrictions on the tracking portfolio such as choosing which components of the index to include. The aim of this paper is, however, to develop a *robust* version of index tracking. As far as the authors are aware, this has not been done in the past. When done, the concern has mainly been about robustness with respect to parameter uncertainty rather than robustness about distributional uncertainty (Fabozzi 2007). One possible exception is Lejeune (2012) where a minimax game theoretic interpretation of the robust forward looking approach is discussed. However, in this paper, only parameter uncertainty is considered, that is, the author assumes that the excess returns are imperfectly known, but belong to a class of distributions characterized by an ellipsoidal distributional set. In contrast, we consider the uncertainty in the joint distribution of the index and the tracking portfolio, and find the optimal way to track the index under the worst case distribution. Our approach could be compared to the recently popular distributional robustness methodology. A survey of this methodology, with a focus on actuarial applications, can be found in Blanchet *et al.* (2019). To the best of our knowledge, we have not seen an application of this methodology to the index tracking problem.

The model uncertainty in our paper is measured through a special form of Bregman divergence (see Bregman (1967)). The notion of Bregman divergence is quite general. It has been used as a means to measure the pairwise dissimilarity between matrices (Penev and Prvan 2016), between vectors (Banerjee *et al.* 2005), and also between functions (Goh and Dey 2014, Penev and Naito 2018). In the latter case, Goh and Dey (2014) call it a functional Bregman divergence. A precise definition is given in Section 3.

The classical Kullback–Leibler (KL) divergence also belongs to the class of Bregman divergences. It can be used as a benchmark. However, KL divergence is not appropriate to handle heavy tailed distributions which are commonly observed in financial asset returns (Dey and Juneja 2010, Poczos and Schneider 2011). Hence we choose another family of Bregman divergences whereby the convex function in the divergence's definition in Section 3 has a stronger polynomial growth than the one that is used in the case of KL. Yet we are using a family of Bregman divergences that is parameterized by one positive parameter λ only and is such that it allows us to recover the KL divergence in the limit when $\lambda \to 0$. In this way we can study the effect of the stronger robustification achieved when λ moves away from the zero value. As we point out in detail in Remark 3.3, this goal is indeed achieved by our choice of the function $F_\lambda(\cdot)$ in the specification of the Bregman divergence.

Our first contribution is the derivation of a semi-closed form of the worst case distribution for the chosen Bregman divergence. The second contribution is the derivation of the optimal index tracking portfolio in a semi-analytical form. The derivation of such a robust index tracking portfolio was the main goal of our work. Third, we investigate the performance of our proposed robust tracking portfolio in a short numerical study aiming to demonstrate the robustness effect. Our next contribution is related to extending our approach to deal with a variety of loss functions that are suitable to measure the quality of an index tracking portfolio. Finally, we also include a real data example in order to illustrate the out-of-sample performance of our method.

The structure of this paper is outlined below. In Section 2, we formulate the index tracking problem and present the look back approach and the forward looking approach through a simple example. In Section 3, we formulate the robust index tracking problem and derive the robust index tracking portfolio. In Section 4, we extend our model to tackle enhanced index tracking. In Section 5, we present our numerical study. Section 6 concludes and outlines avenues for further research.

2. Myopic index tracking

In this section, we formulate the index tracking problem and compare the look back approach and the forward looking approach through a simple example. Consider an index that consists of $\ell > 1$ risky assets. A fund manager is interested in constructing a tracking portfolio that contains $d < \ell$ risky assets that may not necessarily belong to the index. The aim is to replicate the return of the index over a fixed investment period. In particular, since the return of the index is available for analysis after each investment period, we focus on a short term one-period (called 'myopic') index tracking.

Within this period, we denote by $\boldsymbol{r}$ the random vector of returns of the risky assets $\boldsymbol{r} = (r^1, \ldots, r^d)^\intercal$ included in the portfolio. Hence r^i, $i = 1, \ldots, d$ denotes the return of the ith individual asset over the intended investment period. The return of the index over this investment period is denoted as b. We then define $\boldsymbol{R} = \boldsymbol{1} + \boldsymbol{r}$, and $B = 1 + b$.

Throughout the paper, we assume that all random quantities are defined on a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with the sample space Ω, the σ-algebra $\mathcal{F}$, and the probability measure $\mathbb{P}$, where the σ-algebra $\mathcal{F} = \sigma(\boldsymbol{r}, B)$.

In addition, we assume that short selling is permitted.

At the beginning of the investment period, the fund manager re-balances the portfolio with a control $\boldsymbol{u}$, where $\boldsymbol{u}^i \in \mathbb{R}$, $i = 1, \ldots, d$, denotes the proportional allocation of the wealth of the investor into the ith asset. Let $\mathcal{U}$ be the set of admissible strategies $\boldsymbol{u} \in \mathcal{U}$ such that

$$U = \left\{ \boldsymbol{u} \in \mathbb{R}^d : \boldsymbol{1}^\intercal \boldsymbol{u} = 1 \right\}.$$

Two approaches exist to track an index. The look back approach finds the optimal portfolio based on the historical data. Let $\boldsymbol{r}_{-n}$ and $b_{-n}, n = T, (T-1), \ldots, 1$, denote the return of the risky assets and the return of the index, respectively, on the nth day before today. Define $\boldsymbol{R}_{-n} = \boldsymbol{1} + \boldsymbol{r}_{-n}$, and $B_{-n} = 1 + b_{-n}$. Under this approach, the optimal control $\boldsymbol{u}$ can be obtained by solving

$$\mathcal{V} = \inf_{\boldsymbol{u} \in \mathcal{U}} \frac{1}{T} \sum_{n=1}^{T} \left(\boldsymbol{R}_{-n}^\intercal \boldsymbol{u} - B_{-n} \right)^2. \tag{1}$$

The joint distribution of the index and of the tracking portfolio is taken to be the empirical distribution.

In contrast, the forward looking approach finds the optimal portfolio by solving

$$\mathcal{V} = \inf_{\boldsymbol{u} \in \mathcal{U}} \mathbb{E} \left((\boldsymbol{R}^\intercal \boldsymbol{u} - B)^2 \right). \tag{2}$$

Under this approach, the actual distribution can be assumed to be essentially arbitrary.

It is easy to see that the forward looking approach relies on the future estimation of the actual distribution. If the empirical distribution delivers a good estimate of the future, then the look back approach is equivalent to the forward looking approach. However, this is not the case if another assumption is made about the actual distribution. Thus, unless the empirical distribution represents a reliable estimate, the two approaches yield different outcomes in general. In addition, even though the empirical distribution may represent a good estimate of the future, there is always an uncertainty in the estimation of the actual distribution. It is obvious that the 'look back' is just a specific way to perform 'forward looking' by using an empirical distribution of risk factors.

3. Myopic robust index tracking

To model the uncertainty associated with the estimation of the actual distribution, we use the Bregman divergence.

3.1. The Bregman divergence

The following definition of a Bregman divergence is taken from Penev and Naito (2018).

DEFINITION 3.1 *Given a strictly convex function $F : A \to \mathbb{R}$, where $A \subset \mathbb{R}^d$ is a convex set, the local Bregman divergence between two points $X \in A$ and $Y \in A$ is defined as*

$$d_F(X, Y) = F(Y) - F(X) - \nabla F(X)(Y - X).$$

The above definition can also be applied point wise for positive density functions f, g defined on a common domain. The point wise application means that in this case $d = 1$, ∇ means a simple derivative F' and we interpret locally, for a fixed t

$$d_F(f(t), g(t)) = F(g(t)) - F(f(t)) - F'(f(t))(g(t) - f(t)).$$

Using this localized divergence measure at the point t, we then define the global (or also called functional) Bregman divergence between the densities f and g:

$$D_{Breg}(f, g) := \int d_F(f(t), g(t)) v(t)\, \mathrm{d}t, \tag{3}$$

where v is some non-negative weight function.

In the definition of the Bregman divergence, any strictly convex function $F(\cdot)$ could be chosen. A specific choice has been suggested in the paper (Penev and Naito 2018), which we also take on board here. We take the strictly convex function $F_\lambda(z) : (0, \infty) \to \mathbb{R}^1$ to be

$$F_\lambda(z) = \frac{z^{\lambda+1} - (\lambda+1)z}{\lambda}, \quad \text{for a fixed } \lambda > 0.$$

Then, for two densities f and g of the d-dimensional argument x, it is easy to see that

$$\begin{aligned} &d_F(f(x), g(x)) \\ &\quad = F_\lambda(g(x)) - F_\lambda(f(x)) - < \nabla F_\lambda(f(x))^T, g(x) - f(x) > \\ &\quad = \frac{g(x)^{\lambda+1}}{\lambda} - \frac{(\lambda+1) g(x) f(x)^\lambda}{\lambda} + f(x)^{\lambda+1}. \end{aligned}$$

It is worth noting that our special form of Bregman divergence is closely related to the so-called Tsallis divergence and α-divergence (see, e.g. Poczos and Schneider (2011), Cichocki and Amari (2018)). α-divergence is closely related to the special case of our Bregman divergence. It differs by a re-parameterization on which both divergences are defined. It is quite a laborious task to discuss and compare many different types of divergences that exist in the literature. Such a comparison is done in the monograph (Amari 2016) and in Dey and Juneja (2010). The reader can also consult p. 187 in Amari and Cichotski (2010) where our current use of Bregman divergence is discussed under the name 'β-divergence'. It contains both Kullback–Leibler and Itakura–Saito divergences and has found earlier applications in classical robustness and in machine learning. Similar comparisons are also done on p. 9 of Dey and Juneja (2010) which demonstrates that the polynomial divergence of which our chosen Bregman divergence is based is a strictly monotone increasing function of the Rényi entropy. Hence minimizing the one or the other delivers essentially the same result.

In this paper, we prefer to work with the Bregman divergence in its *functional form* (3) when describing an ambiguity set around certain nominal probability distribution. Earlier approaches have described ambiguity sets using support and moment information or structural properties such as symmetry, unimodality, etc. We suggest that in the setting of index tracking, the uncertainty around the nominal model is difficult to specify accurately enough and for this reason considering a 'full' functional-type neighbourhood is more suitable to start with. It is remarkable that the objective function for the robust index tracking problem in this functional setting is tractable and allows us to derive the optimal robust portfolio in a semi-analytical form as a solution of a system of nonlinear equations.

Since our purpose is to robustify the inference, it is *essential* for us to use the Bregman divergence (in particular, in its specific form of β- divergence). In Mihoko and Eguchi (2002), the β-divergence is used (which is essentially the same as the density power divergence of Basu *et al.* (1998)) and it is explicitly demonstrated that using the global β-divergence leads to parameter estimators that are robust in the classical sense i.e. their influence function is bounded.

Now, by choosing the following weight function:

$$v(x) = \frac{f(x)}{f(x)^{\lambda+1}},$$

we have constructed the following functional Bregman divergence:

$$\begin{aligned} D_{Breg}(f,g) &:= \int_{\mathbb{R}^d} d_F(f(x), g(x)) v(x)\, \mathrm{d}x \\ &= \int \left(\frac{g^{\lambda+1}(x)}{\lambda} - \frac{(\lambda+1) g(x) f^{\lambda}(x)}{\lambda} + f^{\lambda+1}(x) \right) \\ &\quad \times \frac{f(x)}{f(x)^{\lambda+1}}\, \mathrm{d}x \\ &= \int \left(\frac{1}{\lambda}\mathcal{E}^{\lambda+1} - \frac{\lambda+1}{\lambda}\mathcal{E} + 1 \right) f(x)\, \mathrm{d}x \\ &= \mathbb{E}\left(\frac{1}{\lambda}\mathcal{E}^{\lambda+1} - \frac{\lambda+1}{\lambda}\mathcal{E} + 1 \right) \\ &= \mathbb{E}\left(G(\mathcal{E})\right), \end{aligned} \tag{4}$$

where $G(\mathcal{E}) = \frac{1}{\lambda}\mathcal{E}^{\lambda+1} - \frac{\lambda+1}{\lambda}\mathcal{E} + 1$.

From now on, we will always denote the nominal distribution's density by f. Expectations in this paper are always supposed to be taken with respect to the nominal distribution f. Then there will be a one-to-one correspondence between the pair $(f(x), g(x))$ and the pair $(f(x), \mathcal{E}(x))$ where

$$\mathcal{E}(x) = \frac{g(x)}{f(x)}.$$

Hence by slight abuse of notation, we will shorten the notation by replacing $D_{Breg}(f,g)$ with $D_{Breg}(\mathcal{E})$.

Remark 3.1 We comment on our choice of the weight function $v(x) = \frac{f(x)}{f(x)^{\lambda+1}}$ in the definition of the functional Bregman divergence. Its usefulness goes much beyond the obvious fact that it leads to a simple expression for $D_{Breg}(\mathcal{E})$. In Vemuri *et al.* (2011), it is argued that re-scaling at each x the usual Bregman divergence can bring about intrinsic robustification. The way the Bregman divergence is re-scaled in Vemuri *et al.* (2011) leads to the so-called Total Bregman Divergence, which is too computationally heavy to be implemented in our case of functional Bregman divergence. However, the main idea of using a re-scaling that is inversely dependent on the derivative of the $F_\lambda(\cdot)$ function can be applied also in our case and suggests the above choice of $v(x)$. Our numerical experiments support this choice.

Remark 3.2 One essential advantage of the special form of the Bregman divergence we use is that by just one parameter ($\lambda > 0$) we can control the deviation from the well-known KL divergence $\mathbb{E}(\mathcal{E}\log(\mathcal{E}))$. The latter is obtained as a limiting case when $\lambda \to 0$. Indeed, as $\lambda \to 0$, we see that

$$\frac{z^{1+\lambda} - (\lambda+1)z}{\lambda} \to z\log(z) - z.$$

This limit result then yields

$$\begin{aligned} &\mathbb{E}\left(\frac{1}{\lambda}\mathcal{E}^{\lambda+1} - \frac{\lambda+1}{\lambda}\mathcal{E} + 1 \right) \\ &\quad \to \mathbb{E}\left(\frac{g(x)}{f(x)} \log\left(\frac{g(x)}{f(x)}\right) - \frac{g(x)}{f(x)} + 1 \right) \\ &\quad = \mathbb{E}\left(\frac{g(x)}{f(x)} \log\left(\frac{g(x)}{f(x)}\right) \right) \\ &\quad = \mathbb{E}\left(\mathcal{E}\log(\mathcal{E})\right). \end{aligned}$$

The value of $\lambda > 0$ parameterizes a whole class of Bregman divergences and dictates the extent to which the robust method differs from the non-robust method (in Basu *et al.* (1998) it is called an algorithmic parameter). By varying the value of $\lambda > 0$, we can achieve a compromise between robustness and efficiency, as is standard in robust statistic setting. Larger values of λ correspond to a stronger emphasis on robustness. These would be useful when there is a belief that the divergence between the nominal distribution and the actual distribution of the returns might be large.

In a special case where both the nominal distribution and the actual distribution are multivariate normal, the above Bregman divergence can be calculated in a closed form.

Example Multivariate Normal Suppose that the nominal distribution is a d-dimensional multivariate normal distribution $N_d(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ and an actual distribution is another d-dimensional multivariate normal distribution $N_d(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$. Then, the Bregman divergence as defined above can be calculated in closed form as:

$$
\begin{aligned}
D_{Breg}(\mathcal{E}) = \frac{1}{\lambda}\left(\frac{\left(\det(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2^{-1})\right)^{\lambda+1}\det(\tilde{\boldsymbol{\Sigma}}_\lambda)}{\det(\boldsymbol{\Sigma}_1)}\right)^{\frac{1}{2}} \\
\times \exp\left(-\frac{\lambda+1}{2}\boldsymbol{\mu}_2^\intercal\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}_2\right. \\
\left. + \frac{\lambda}{2}\boldsymbol{\mu}_1^\intercal\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1 + \frac{1}{2}\tilde{\boldsymbol{\mu}}_\lambda^\intercal\tilde{\boldsymbol{\Sigma}}_\lambda^{-1}\tilde{\boldsymbol{\mu}}_\lambda\right) - \frac{1}{\lambda},
\end{aligned} \tag{5}
$$

provided $\tilde{\boldsymbol{\Sigma}}_\lambda$ is positive definite, where

$$
\begin{aligned}
\tilde{\boldsymbol{\Sigma}}_\lambda &= \left((\lambda+1)\boldsymbol{\Sigma}_2^{-1} - \lambda\boldsymbol{\Sigma}_1^{-1}\right)^{-1}, \\
\tilde{\boldsymbol{\mu}}_\lambda &= \tilde{\boldsymbol{\Sigma}}_\lambda\left((\lambda+1)\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}_2 - \lambda\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1\right).
\end{aligned}
$$

If $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$, the above formula simplifies to

$$
\begin{aligned}
D_{Breg}(\mathcal{E}) \\
= \frac{1}{\lambda}\exp\left(\frac{\lambda(\lambda+1)}{2}(\boldsymbol{\mu}_2-\boldsymbol{\mu}_1)^\intercal\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_2-\boldsymbol{\mu}_1)\right) - \frac{1}{\lambda}.
\end{aligned} \tag{6}
$$

It is worth noting that if we set $\lambda \to 0$ in (6) we get

$$
\frac{1}{2}(\mu_2 - \boldsymbol{\mu}_1)^\intercal\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1),
$$

i.e. one half of the squared Mahalanobis distance between the multivariate normal distributions $N_d(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$ and $N_d(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$.

3.2. Robust index tracking

To perform index tracking in a robust way, we need to consider perturbations of the nominal distribution of the index. These perturbed distributions can be contained inside a ball of certain radius around the nominal distribution. To this end, let us construct a Bregman divergence ball. Suppose that f, the so-called nominal distribution, is the joint density of the ℓ assets in the index, and g is the density of a perturbation of f, we denote by S_f the set of all functions $\mathcal{E} : \mathbb{R}^d \to \mathbb{R}$ representable in the form $\mathcal{E}(x) = \frac{g(x)}{f(x)}$, where g is a density. A Bregman divergence ball around of radius $\eta > 0$ around f is defined as

$$
\mathcal{B} := \mathcal{B}(\eta) = \{g : \mathcal{E} \in S_f \text{ and } D_{Breg}(\mathcal{E}) \leq \eta\}. \tag{7}
$$

We stress again that all moments throughout this paper are defined with respect to the nominal distribution. Then the robust version of the control problem (2) is defined as

$$
V = \sup_{\boldsymbol{u}\in\mathcal{U}}\inf_{\mathcal{E}\in\mathcal{B}} J(\mathcal{E}, \boldsymbol{u}), \tag{8}
$$

where

$$
J(\mathcal{E}, \boldsymbol{u}) = \mathbb{E}\left(-\mathcal{E}\left(\boldsymbol{R}^\intercal\boldsymbol{u} - B\right)^2\right),
$$

In the next section, we will derive a semi-analytical form of the optimal portfolio under the constructed Bregman divergence.

3.3. Robust optimal portfolio under Bregman divergence

The robust optimal index tracking portfolio can be obtained by applying the following result.

THEOREM 3.1 *For a fixed, small enough $\lambda > 0$, if there exist $\alpha^* > 0$, β^* and θ^* such that*

$$
\begin{aligned}
\theta^*\mathbf{1} &= \mathbb{E}\left(\frac{\partial H}{\partial \boldsymbol{u}}\left(\frac{\lambda}{\lambda+1}\left(\frac{-\beta^* - H(\boldsymbol{u})}{\alpha^*}\right)+1\right)^{\frac{1}{\lambda}}\right), \\
\mathbf{1}^\intercal\boldsymbol{u} &= 1, \\
\mathbb{E}\left(G(\mathcal{E}^*)\right) &= \eta, \\
\mathbb{E}\left(\mathcal{E}^*\right) &= 1,
\end{aligned}
$$

where

$$
\begin{aligned}
H(\boldsymbol{u}) &= -\left(\boldsymbol{R}^\intercal\boldsymbol{u} - B\right)^2, \\
\mathcal{E}^* &= \left(\frac{\lambda}{\lambda+1}\left(\frac{-\beta^* - H(\boldsymbol{u})}{\alpha^*}\right)+1\right)^{\frac{1}{\lambda}},
\end{aligned}
$$

then $\boldsymbol{u}$ is an optimal index tracking portfolio.

Proof Using the definition of $H(\boldsymbol{u})$ we see that (8) becomes

$$
\sup_{\boldsymbol{u}\in\mathcal{U}}\inf_{\mathcal{E}\in\mathcal{B}} \mathbb{E}\left(\mathcal{E}H(\boldsymbol{u})\right). \tag{9}
$$

To solve the inner optimization problem, we first write down the Lagrangian. For a fixed $\alpha \geq 0$ and a $\beta \in \mathbb{R}$, the Lagrangian is

$$
L^{inner}(\mathcal{E}, \boldsymbol{u}) = \mathbb{E}\left(\mathcal{E}H(\boldsymbol{u}) + \alpha(G(\mathcal{E}) - \eta) + \beta(\mathcal{E} - 1)\right),
$$

where

$$
G(\mathcal{E}) = \frac{1}{\lambda}\mathcal{E}^{\lambda+1} - \frac{\lambda+1}{\lambda}\mathcal{E} + 1.
$$

Differentiating inside the expectation and setting the result equal to zero yields:

$$
0 = H(\boldsymbol{u}) + \alpha G'(\mathcal{E}) + \beta. \tag{10}
$$

Solving this equation, we obtain

$$
\mathcal{E}^* = (G')^{-1}\left(\frac{-\beta - H(\boldsymbol{u})}{\alpha}\right), \tag{11}
$$

provided that $\alpha \neq 0$.

Next, we verify that this is indeed an optimal solution. The proof follows similarly to Glasserman and Xu (2014, proposition 2.3) and Dey and Juneja (2010, Theorem 2). The idea is to show that along any feasible direction the value of the Lagrangian cannot be optimized any further.

Choose an arbitrary $\mathcal{E}$. For $t \in [0, 1]$, define $\hat{\mathcal{E}} = t\mathcal{E} + (1-t)\mathcal{E}^*$, then we have

$$\begin{aligned} L^{inner}(\hat{\mathcal{E}}, \boldsymbol{u}) &= \mathbb{E}\left(H(\boldsymbol{u})\hat{\mathcal{E}} + \alpha(G(\hat{\mathcal{E}}) - \eta) + \beta(\hat{\mathcal{E}} - 1)\right) \\ &= \mathbb{E}\left(H(\boldsymbol{u})\left(t\mathcal{E} + (1-t)\mathcal{E}^*\right)\right. \\ &\quad + \alpha\left(G\left(t\mathcal{E} + (1-t)\mathcal{E}^*\right) - \eta\right) \\ &\quad \left. + \beta\left(\left(t\mathcal{E} + (1-t)\mathcal{E}^*\right) - 1\right)\right). \end{aligned}$$

If we consider $L^{inner}(\hat{\mathcal{E}}, \boldsymbol{u})$ as a function of t, and define

$$\begin{aligned} K(t) = \mathbb{E}&\left(H(\boldsymbol{u})\left(t\mathcal{E} + (1-t)\mathcal{E}^*\right)\right. \\ &+ \alpha\left(G\left(t\mathcal{E} + (1-t)\mathcal{E}^*\right) - \eta\right) \\ &\left. + \beta\left(\left(t\mathcal{E} + (1-t)\mathcal{E}^*\right) - 1\right)\right), \end{aligned}$$

it is then easy to calculate

$$K'(t) = \mathbb{E}\left(\left(H(\boldsymbol{u}) + \alpha G'\left(t\mathcal{E} + (1-t)\mathcal{E}^*\right) + \beta\right)\left(\mathcal{E} - \mathcal{E}^*\right)\right).$$

This implies then

$$K'(0) = \mathbb{E}\left(\left(H(\boldsymbol{u}) + \alpha G'(\mathcal{E}^*) + \beta\right)\left(\mathcal{E} - \mathcal{E}^*\right)\right) = 0,$$

since $\mathcal{E}^*$ satisfies (10). In addition, we know that L^{inner} is convex in its first argument, thus K is convex in t which implies $t = 0$ is an optimal solution. Because $\mathcal{E}$ is arbitrary, we cannot improve the value of the objective along any feasible direction from $\mathcal{E}^*$. This concludes that $\mathcal{E}^*$ is an optimal solution.

Next, we notice that the set

$$\{\mathcal{E} : \mathbb{E}\left(G(\mathcal{E})\right) < \eta\}$$

is not empty. By Theorem 2.1. in Ben-Tal *et al.* (1988), strong duality holds. This implies (see, e.g., pp. 242–243 in Boyd and Vandenberghe (2004)) that the optimal solution $\mathcal{E}^*$ and its corresponding α satisfies the following system:

$$\begin{aligned} \alpha\mathbb{E}\left(G(\mathcal{E}^*) - \eta\right) &= 0, \\ \mathbb{E}\left(G(\mathcal{E}^*)\right) &\leq \eta, \\ \alpha &> 0. \end{aligned}$$

We denote the solution α of this system as α^*. Thus we obtain

$$\mathcal{E}^* = \left(\frac{\lambda}{\lambda+1}\left(\frac{-\beta^* - H(\boldsymbol{u})}{\alpha^*}\right) + 1\right)^{\frac{1}{\lambda}}, \tag{12}$$

where β^* is the solution of $\mathbb{E}(\mathcal{E}^*) = 1$. With an appropriate choice of (small) λ, we can always achieve that $\mathcal{E}^*$ in (12) is well-defined and positive. Indeed, for the limiting case $\lambda = 0$ (corresponding to the KL divergence) we have

$$G(\mathcal{E}) = \mathcal{E}\log\mathcal{E} - \mathcal{E} + 1, G'(\mathcal{E}) = \log(\mathcal{E})$$

and the solution $\mathcal{E}^* = e^{\frac{-\beta^* - H(\boldsymbol{u})}{\alpha^*}}$ clearly being positive. As in Proposition 2.3 of Glasserman and Xu (2014) we can argue that by continuity we can find a set $[0, \lambda_0]$ such that $\mathcal{E}^* = (\frac{\lambda}{\lambda+1}(\frac{-\beta^* - H(\boldsymbol{u})}{\alpha^*}) + 1)^{\frac{1}{\lambda}}$ will stay positive for any $\lambda \in [0, \lambda_0]$.

Using (12), we end up with the optimization problem

$$\sup_{\boldsymbol{u}\in\mathcal{U}} J(\mathcal{E}^*, \boldsymbol{u}). \tag{13}$$

In Appendix A.2, we demonstrate that the solution to this optimization problem satisfies the system of equations in the statement of Theorem 3.1. ■

As mentioned in Remark 3.2, as $\lambda \to 0$, the chosen Bregman divergence converges to the KL divergence. Indeed, as $\lambda \to 0$, the system in Theorem 3.1 also converges to the corresponding system of the KL divergence. This is summarized in the following result.

Remark 3.3 We have assumed existence of $\alpha^* > 0$ in Theorem 3.1 and this is exploited in the resulting presentation of $\mathcal{E}^*$. It is clear, however, that the case $\alpha = 0$ should safely be excluded. Indeed if α was zero then the restriction about g belonging to the ball of radius η around f is ignored. Then $L^{inner}(\mathcal{E}, \boldsymbol{u})$ implies to minimize $\mathbb{E}(\mathcal{E}H(\boldsymbol{u}))$, where $H(\boldsymbol{u})$ is a negative random variable, under the *only* restriction that $\mathbb{E}(\mathcal{E}) = 1$. This is equivalent to ask to minimize $\mathbb{E}_g H(\boldsymbol{u})$ where $\mathbb{E}_g$ stands for calculating the expected value under the 'arbitrary' actual distribution. This problem does not have a solution since for any specified g^* such that $\mathbb{E}_{g^*}H(\boldsymbol{u}) = A$ we can find another $\tilde{g}$ such that $\mathbb{E}_{\tilde{g}}H(\boldsymbol{u}) < A$ as long as $\tilde{g}$ puts higher mass at the negative values of $H(\boldsymbol{u})$ with a large magnitude.

It is also important to note that for the chosen $\lambda > 0$ in Theorem 3.1 the solution of the equation system for $\alpha^* > 0, \beta^*$ and θ^* is also required to satisfy the condition $\frac{\beta^*}{\alpha^*} < 1 + \frac{1}{\lambda}$. Finding precise conditions which $\lambda > 0$ must satisfy under a general nominal distribution seems to be very difficult. However, we can state that numerically, in all examples that we have tried, we have observed that if certain λ' is found which 'works', then all values $\lambda \in (0, \lambda')$ also work, i.e. for them, respective $\alpha^* > 0, \beta^*$ and θ^* satisfying the system of equations exist with the inequality

$$\frac{\beta^*}{\alpha^*} < 1 + \frac{1}{\lambda}$$

being satisfied.

Corollary 3.2 *Suppose that there exist $\alpha^* > 0$, β^* and θ^* such that*

$$\begin{aligned} \mathbf{1}^\intercal\theta^* &= \mathbb{E}\left(\frac{\partial H}{\partial \boldsymbol{u}}\exp\left(\frac{-\beta^* - H(\boldsymbol{u})}{\alpha^*}\right)\right), \\ \mathbf{1}^\intercal\boldsymbol{u} &= 1, \\ \mathbb{E}\left(G(\mathcal{E}^*)\right) &= \eta, \\ \mathbb{E}\left(\mathcal{E}^*\right) &= 1, \end{aligned}$$

where

$$H(\boldsymbol{u}) = -\left(\boldsymbol{R}^\intercal\boldsymbol{u} - B\right)^2, \quad \text{and} \quad \mathcal{E}^* = \exp\left(\frac{-\beta^* - H(\boldsymbol{u})}{\alpha^*}\right),$$

then $\boldsymbol{u}$ is an optimal index tracking portfolio.

The proof of the Corollary is obtained by taking a limit as $\lambda \to 0$ in Theorem 3.1.

Remark 3.4 The main part of the numerical procedure of our method is the implementation of the solution of the nonlinear equation system from Theorem 3.1. We used MATLAB and applied the interior-point method for solving box-constrained nonlinear systems. Initial guesses for the solution were necessary to be chosen for the portfolio's weights, and for the parameters α, β and θ. Uniform initial weights turned out to be working fine all the time. The initial weights for the remaining parameters required more careful choosing and experimentation. The ones that worked fine for our numerical experiments were $\alpha = 0.02, \beta = 0.01$ and $\theta = -0.05$.

4. Modified myopic robust index tracking

The discussion in Section 3.3 can be extended to cover a specific form of modified myopic robust index tracking problem. In the previous section, we measured the quality of the index tracking by using the quadratic loss function since this is the typical choice in the portfolio tracking literature. This choice equally penalizes the performance of the portfolio whenever it deviates by the same magnitude irrespectively of whether the deviation is above or below the value of the index.

The main focus in de Paulo *et al.* (2016) is on formulating an optimization problem that represents a balancing of the trade-off between tracking error and excess return. A different goal may be of interest in a robust setting. Typically, in the latter setting, the goal is to safeguard against worst-case scenarios and the solution obtained reflects this goal. Hence it is expected to give superior performance, especially in a downturn market. If for various reasons the investor still remains in the market during a downturn (e.g., expecting that this downturn would be relatively short-lived, or because of limited liquidity), one would not be willing to penalize if the portfolio outperforms the index in such cases. Obviously, a more reasonable choice to replace the loss $\ell(x) = x^2$ to be used in such a situation would be based, for example, on a smooth approximation of the function $\ell_1(x) = x^2$ if $x > 0$ and 0 else. Other choices also make sense, for example $\ell_2(x) = [x]_+$. Direct utilization of these types of functions makes a lot of sense since we do not really want to penalize when the portfolio happens to outperform the index.

However, there is a technical difficulty to overcome if we want to include such type of losses in our approach. It is related to the fact that the functions ℓ_1 and ℓ_2 are not smooth at the origin. If we would like to utilize the steps as in Theorem 3.1 and show that the Hessian is negative semi-definite, we need a convex twice differentiable loss function $\ell_i(x), i = 1, 2$ to replace $\ell(x)$. Also, from a technical prospective, the gradient of H should be possible to calculate, preferably in a closed form. We suggest the function $\tilde{\ell}_1(x) = \frac{1}{\epsilon}\int_0^\infty \phi(\frac{1}{\epsilon}(x-t))\ell_1(t)\,\mathrm{d}t$ with a suitably chosen small $\epsilon > 0$ as approximation for $\ell_1(x)$. For approximation of $\ell_2(x)$, the expression $\tilde{\ell}_2(x) = x + \epsilon \log(1 + e^{-x/\epsilon})$ from the literature (see, e.g. Chen and Mangasarian (1995)) can be used and is known as 'the neural networks smooth plus function'. Using these, the function $H(\boldsymbol{u}) = -(\boldsymbol{R}^\top \boldsymbol{u} - B)^2 = -(B - \boldsymbol{R}^\top \boldsymbol{u})^2$ in Theorem 3.1 can be replaced by $\tilde{H}_1(\boldsymbol{u}) = -\tilde{\ell}_1(B - \boldsymbol{R}^\top \boldsymbol{u})$ or by $\tilde{H}_2(\boldsymbol{u}) = -\tilde{\ell}_2(B - \boldsymbol{R}^\top \boldsymbol{u})$, respectively. The corresponding gradient of H is to be replaced by the gradient of $\tilde{H}_1$ or $\tilde{H}_2$ and these are easily calculated by using the chain rule and the derivatives of one-dimensional argument for $\tilde{\ell}_1(t)$ and $\tilde{\ell}_2(t)$. Both derivatives of $\tilde{H}_1$ or $\tilde{H}_2$ w.r.t. the components of $\boldsymbol{u}$ deliver smooth approximating functions. We prefer the first approximation since its second mixed derivatives appear to be varying more smoothly around the origin for small values of ϵ. Elementary calculation of the integral gives the following approximations for the function $\tilde{\ell}_1(x)$:

$$\tilde{\ell}_1(x) = x^2\Phi\left(\frac{x}{\epsilon}\right) + 2x\epsilon\phi\left(\frac{x}{\epsilon}\right) + \epsilon^2\left[\frac{1}{2} + \frac{1}{2}Erf\left(\frac{x}{\sqrt{2}\epsilon}\right) - \frac{x}{\epsilon}\phi\left(\frac{x}{\epsilon}\right)\right] \tag{14}$$

Here $\Phi(\cdot)$ denotes the cumulative distribution function of the univariate standard normal distribution, $\phi(\cdot)$ denotes the density and the $Erf(\cdot)$ function is defined as $Erf(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\,\mathrm{d}t$. Having in mind the relationship $\frac{1}{2} + \frac{1}{2}Erf(\frac{x}{\sqrt{2}\epsilon}) = \Phi(\frac{x}{\epsilon})$, (14) simplifies further to the explicit expression

$$\tilde{\ell}_1(x) = (x^2 + \epsilon^2)\Phi\left(\frac{x}{\epsilon}\right) + x\epsilon\phi\left(\frac{x}{\epsilon}\right). \tag{15}$$

Differentiating (14) delivers the resulting approximations for the derivatives of $\tilde{\ell}_1(x)$:

$$\tilde{\ell}_1'(x) = 2x\Phi\left(\frac{x}{\epsilon}\right) + 2\epsilon\phi\left(\frac{x}{\epsilon}\right), \quad \tilde{\ell}_1''(x) = 2\Phi\left(\frac{x}{\epsilon}\right). \tag{16}$$

Of course, the approximations for the derivatives of $\tilde{\ell}_2(x)$ are:

$$\tilde{\ell}_2'(x) = \frac{1}{1 + \exp(-x/\epsilon)}, \quad \tilde{\ell}_2''(x) = \frac{1}{\epsilon}\frac{e^{x/\epsilon}}{(1 + e^{x/\epsilon})^2}.$$

5. Numerical analysis

In this section, we perform various numerical comparisons to illustrate the usefulness of our model. It is worth noting that there is some difference between the general theory (in particular, the statements in Section 3) and the way to illustrate the theory via examples. We stress that, as seen in Section 3, the main theoretical statement in Theorem 3.1 on the basis of which our numerical procedure is implemented, does not explicitly require calculation of the least favourable distribution in the Bregman ball; all that is needed is the radius η of the ball. If we wanted to illustrate the full effect of the robustification theory on a particular example, we should ideally be able to calculate the least favourable distribution and simulate from it. Determining the least favourable distribution is very difficult even if the nominal distribution was multivariate normal. On the other hand, we know that the least favourable distribution is on the surface of the ball (since the Lagrange multiplier α is not equal to zero). Hence, just for the purpose of generating illustrative examples, we have chosen a distribution that has the maximal allowed divergence from the nominal *and* is possible to deal with (e.g. in Example 5.1, we choose it to be multivariate normal with the same covariance matrix as the nominal but with a re-scaled mean). Of course, this distribution is not necessarily the least favourable though it is on the maximal allowable distance from the nominal distribution. This approach allows us to simulate our toy examples. We follow traditional approaches in the robustness literature where a contaminated neighbourhood is typically a neighbourhood of the multivariate normal distribution, e.g. Huber and Ronchetti (2009). For this reason, we choose our testing of the statements of Theorem 3.1 to involve perturbations of multivariate normal and of multivariate t. We are aware of the fact that these simulations represent a simplistic indicative study of the effect of the general statement of Theorem 3.1. A more comprehensive numerical work would be required to test our procedure by examining Bregman neighbourhoods of other nominal distributions of interest in finance. As this paper is more on the methodological side, such numerical work is beyond the scope of the presented research. We hope that the reader will still be able to get some insight about the usefulness of our approach by examining the presented numerical examples.

One more point to make is that the least favourable distribution corresponds to a pessimistic scenario that may not be the one that would actually happen hence our examples, by not directly simulating from it, hopefully still give useful insight in the performance of the robust procedure.

5.1. Performance comparison via simulation: index tracking

We first compare the performance of the robust and the non-robust portfolio in the context of index tracking. Suppose that we have an index which is made up of five assets according to the following weight vector:

$$\boldsymbol{w} = \begin{pmatrix} 0.15 \\ 0.20 \\ 0.20 \\ 0.15 \\ 0.30 \end{pmatrix}.$$

The expected return and the covariance matrix of these five assets are given by

$$\tilde{\mu} = \begin{pmatrix} 0.0025 \\ 0.0035 \\ 0.0010 \\ 0.0005 \\ 0.0045 \end{pmatrix},$$

$$\tilde{\boldsymbol{\Sigma}} = \begin{pmatrix} 0.0020 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0025 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0012 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0001 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0033 \end{pmatrix}. \quad (17)$$

We will use the first four assets to track this hypothetical index. The following example is used to demonstrate the comparison.

Example Multivariate Normal (MVN) Suppose that the nominal distribution is an ℓ-dimensional multivariate normal distribution $N_\ell(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ and an actual distribution is an ℓ-dimensional multivariate normal distribution $N_\ell(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$ which is on a Bregman divergence η from the nominal. Fix $\lambda = 0.1$, we assume $\boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_1$ and take $\eta = 0.1, 0.2, 0.5, 0.8, 1.0, 2.0, 5.0$, which results in $\boldsymbol{\mu}_2 = k\boldsymbol{\mu}_1$ for some $k(\eta) \in \mathbb{R}$.

We simulate 5, 000, 000 returns from the nominal distribution to calculate the robust portfolio. The same number of simulations is used to draw samples from the actual distribution to make a comparison between the robust and non-robust portfolios. Several measures are used to accomplish the comparison. The first measure is the number of times (in percentage) that the robust case outperforms the non-robust case. We call this measure the Beating Time (BT). The larger the BT, the more times the robust portfolio outperforms the non-robust one. The outperformance is in the sense of a lower tracking error, where the tracking error (TE) is defined as

$$TE = (\boldsymbol{u}^\intercal \boldsymbol{R} - B)^2 = (\boldsymbol{u}^\intercal (\mathbf{1} + \boldsymbol{r}) - B)^2 .$$

Here $\boldsymbol{u}$ is a control (either robust or non-robust) applied to the tracking portfolio, $\boldsymbol{r}$ denotes the return of the assets in the tracking portfolio, and B denotes the return of the index. The second measure we apply to both controls is the expected tracking error (ETE). Obviously, the smaller the expected tracking error, the better the portfolio (on average). We also compare the performance of the robust and non-robust portfolios with the index, and introduce a measure called the expected excess of index (EEI), where the excess of index (EI) is defined as

$$EI = \boldsymbol{u}^\intercal (\mathbf{1} + \boldsymbol{r}) - B.$$

Thus a negative EI indicates the portfolio is beaten by the index. The EEI is the average of EI over the number of simulations performed.

In this example, both the nominal and the actual distribution are MVN. From (6), it is easy to see that the Bregman divergence between these two distributions is given by

$$\eta = \frac{1}{\lambda} \exp\left(\frac{\lambda(\lambda+1)}{2} (1-k)^2 \boldsymbol{\mu}_1^\intercal \boldsymbol{\Sigma}_1^{-1} \boldsymbol{\mu}_1 \right) - \frac{1}{\lambda}.$$

This then implies

$$k = 1 \pm \sqrt{\frac{\log(\eta\lambda + 1)}{\frac{\lambda(\lambda+1)}{2} \boldsymbol{\mu}_1^\intercal \boldsymbol{\Sigma}_1^{-1} \boldsymbol{\mu}_1}},$$

which allows us to get the relevant k.

If we take

$$k = 1 - \sqrt{\frac{\log(\eta\lambda + 1)}{\frac{\lambda(\lambda+1)}{2} \boldsymbol{\mu}_1^\intercal \boldsymbol{\Sigma}_1^{-1} \boldsymbol{\mu}_1}},$$

we obtain table 1, and in the other case, we obtain table 2.

Table 1. Tracking performance: robust optimal solution (Bregman) versus non-robust optimal solution, $\lambda = 0.1, k < 0$.

η	BT %	ETE Robust	ETE Non-robust ($*10^{-4}$)	ETE Difference	EEI Robust	EEI Non-robust ($*10^{-4}$)	EEI Difference
0.1 ($k = -2.2158$)	50.51	3.1055	3.1055	0.0000	24.8692	24.8728	−0.0036
0.2 ($k = -3.5366$)	51.67	3.2016	3.2017	−0.0001	39.7446	39.7556	−0.0110
0.5 ($k = -6.1208$)	53.99	3.5174	3.5180	−0.0006	68.8326	68.8753	−0.0427
0.8 ($k = -7.9434$)	55.25	3.8417	3.8431	−0.0014	89.3307	89.4115	−0.0808
1.0 ($k = -8.9526$)	55.91	4.0573	4.0594	−0.0021	100.6753	100.7832	−0.1079
2.0 ($k = -12.7653$)	58.25	5.1028	5.1099	−0.0071	143.4931	143.7447	−0.2516
5.0 ($k = -19.5278$)	61.51	7.8510	7.8812	−0.0302	219.2462	219.9450	−0.6988

Table 2. Tracking performance: robust optimal solution (Bregman) versus non-robust optimal solution, $\lambda = 0.1, k > 0$.

η	BT %	ETE Robust	ETE Non-robust ($*10^{-4}$)	ETE Difference	EEI Robust	EEI Non-robust ($*10^{-4}$)	EEI Difference
0.1 ($k = 4.2158$)	52.10	3.2702	3.2702	0.0000	−47.5911	−47.5978	0.0067
0.2 ($k = 5.5366$)	54.20	3.4338	3.4340	−0.0002	−62.4633	−62.4805	0.0172
0.5 ($k = 8.1208$)	56.89	3.8817	3.8827	−0.0010	−91.5435	−91.6003	0.0568
0.8 ($k = 9.9434$)	58.00	4.2989	4.3011	−0.0022	−112.0351	−112.1365	0.1014
1.0 ($k = 10.9526$)	58.56	4.5659	4.5691	−0.0032	−123.3760	−123.5082	0.1322
2.0 ($k = 14.7653$)	60.64	5.8053	5.8149	−0.0096	−166.1784	−166.4697	0.2913
5.0 ($k = 21.5278$)	63.36	8.8956	8.9325	−0.0369	−242.6700	−241.8988	0.7712

It can be seen that in both tables 1 and 2, the robust portfolio outperforms the non-robust one when BT or ETE is used as a comparison measure. In contrast, when EEI is used, if there is a loss made, i.e. the portfolio underperforms the index, the robust portfolio safeguards and performs better. This leads to a positive difference in the last column of table 2. When there is a profit made, the opposite happens and a negative difference is recognized as shown in table 1.

Recall that the parameter λ controls the amount of robustness applied: the smaller the λ, the less robustness effect. This belief is confirmed from the results obtained in tables 3 and 4 when λ is taken to be 0.05. Attention should be directed at comparing the pairs: table 3 with table 1, and tables 4 with table 2, respectively. It becomes apparent that, when the ball radius η is small (hence no need of significant robustification), the performance is about the same no matter whether $\lambda = 0.05$

Table 3. Tracking performance: robust optimal solution (Bregman) versus non-robust optimal solution, $\lambda = 0.05, k < 0$.

η	BT %	ETE Robust	ETE Non-robust ($*10^{-4}$)	ETE Difference	EEI Robust	EEI Non-robust ($*10^{-4}$)	EEI Difference
0.1 ($k = -2.2955$)	50.63	3.1101	3.1101	0.0000	25.7668	25.7716	−0.0048
0.2 ($k = -3.6548$)	51.36	3.2124	3.2125	−0.0001	41.0720	41.0879	−0.0159
0.5 ($k = -6.3327$)	52.38	3.5506	3.5515	−0.0009	71.1951	71.2628	−0.0677
0.8 ($k = -8.2414$)	53.09	3.9019	3.9043	−0.0024	92.6362	92.7703	−0.1341
1.0 ($k = -9.3074$)	53.53	4.1379	4.1416	−0.0037	104.5987	104.7813	−0.1826
2.0 ($k = -13.4063$)	55.27	5.3099	5.3228	−0.0129	150.5243	150.9677	−0.4434
5.0 ($k = -21.0432$)	58.17	8.6066	8.6615	−0.0549	235.8081	237.0205	−1.2124

Table 4. Tracking performance: robust optimal solution (Bregman) versus non-robust optimal solution, $\lambda = 0.05, k > 0$.

η	BT %	ETE Robust	ETE Non-robust ($*10^{-4}$)	ETE Difference	EEI Robust	EEI Non-robust ($*10^{-4}$)	EEI Difference
0.1 ($k = 4.2955$)	52.52	3.2788	3.2789	−0.0001	−48.4876	−48.4966	0.0090
0.2 ($k = 5.6548$)	53.38	3.4506	3.4509	−0.0003	−63.7882	−63.8128	0.0246
0.5 ($k = 8.3327$)	54.05	3.9254	3.9270	−0.0016	−93.8985	−93.9878	0.0893
0.8 ($k = 10.2414$)	54.70	4.3738	4.3776	−0.0038	−115.3283	−115.4953	0.1670
1.0 ($k = 11.3074$)	55.10	4.6639	4.6695	−0.0056	−127.2841	−127.5062	0.2221
2.0 ($k = 15.4063$)	56.78	6.0434	6.0606	−0.0172	−173.1823	−173.6927	0.5104
5.0 ($k = 23.0432$)	59.40	9.7240	9.7904	−0.0664	−258.4161	−259.7455	1.3294

or $\lambda = 0.1$ was used. However, when η is increased to, say, 2 or 5, more robustification is required and using the higher value of $\lambda = 0.1$ proves to bring a higher percentage of BT.

5.2. *Performance comparison via simulation during market downturn*

In this section, we illustrate the effect of using the loss function $\tilde{\ell}_1(\cdot)$ from Section 4 on two examples: the first example involves the multivariate normal as a nominal distribution and the second example deals with multivariate t as a nominal distribution.

Example Nominal multivariate normal Given that all components of the chosen mean vector of the multivariate normal are positive, the single scalar multiplication with a value of $k < 1$ represents a market downturn scenario. As k is related to the radius η, a larger value of η pushes k further in the negative territory. The number of simulations used to calculate the robust portfolio and to assess the performance was kept at 1, 000, 000 as a sufficient stabilization of the results was already appearing at this number of simulations. We applied the smoothed loss function $\tilde{\ell}_1(\cdot)$ from Section 4 with $\epsilon = 0.01$ and $\lambda = 0.1$. We varied the radius η through the range 0.1, 0.2, 0.5, 0.8, 1, 2, 5 as before and registered the percentage of cases in which the Bregman-based portfolio outperformed the non-robust one. (The non-robust portfolio was defined as minimizing the same loss $\tilde{\ell}_1(x)$ but without considering a neighbourhood around the nominal distribution).

As expected, the percentage of cases in which the robust portfolio was not worse than the non-robust one was quite large. The results are presented in table 5.

Similar results to the ones presented in table 5 can be obtained when $\tilde{\ell}_2(\cdot)$ was used as a smoothed loss function. The results clearly outline the significant benefits of using the robust portfolio in a market downturn scenario. Of course, this is to be expected by the nature of the optimization problem that is solved in the robust setting. We note that in table 5 the cases where both portfolios deliver a zero value for the loss $\tilde{\ell}_1(\cdot)$ have been counted towards the percentage BT (since these cases are considered as 'not worse' for the robust portfolio). One may suggest that it might be fairer to exclude these cases from the comparison (i.e. to consider in what proportion of cases the robust portfolio delivered a truly better outcome).

Table 5. Tracking performance using the loss $\tilde{\ell}_1$: robust optimal solution versus non-robust optimal solution (MVN) $\lambda = 0.1, \epsilon = 0.01$ (cases of both losses being zero included).

η	BT %	ETE Robust	ETE Non-robust ($*10^{-4}$)	ETE Difference
0.1 ($k = -2.2158$)	81.24	58.5587	58.9210	−0.3623
0.2 ($k = -3.5366$)	84.06	52.3295	52.9541	−0.62464
0.5 ($k = -6.1208$)	88.65	41.2324	42.3339	−1.1015
0.8 ($k = -7.9434$)	91.46	34.5631	36.0577	−1.4946
1.0 ($k = -8.9526$)	92.72	31.1530	32.8128	−1.6598
2.0 ($k = -12.7653$)	96.28	20.3921	22.4897	−2.0976
5.0 ($k = -19.5278$)	99.11	8.36022	10.5283	−2.1681

Table 6. Tracking performance using the loss $\tilde{\ell}_1$: robust optimal solution versus non-robust optimal solution (MVN) $\lambda = 0.1, \epsilon = 0.01$ (cases of both losses being zero excluded).

η	BT %	ETE Robust	ETE Non-robust ($*10^{-4}$)	ETE Difference
0.1 ($k = -2.2158$)	58.34	58.5587	58.9210	−0.3623
0.2 ($k = -3.5366$)	62.00	52.3295	52.9541	−0.6246
0.5 ($k = -6.1208$)	68.65	41.2324	42.3339	−1.1015
0.8 ($k = -7.9434$)	72.98	34.5631	36.0577	−1.4946
1.0 ($k = -8.9526$)	75.20	31.1530	32.8128	−1.6598
2.0 ($k = -12.7653$)	82.69	20.3921	22.4897	−2.0976
5.0 ($k = -19.5278$)	91.66	8.36022	10.5283	−2.1681

It would be expected that this proportion would be smaller but still high enough. Indeed this expectation is confirmed by the results that are presented in table 6.

Example Multivariate t *(MVT)* Suppose now that the nominal distribution is an ℓ-dimensional multivariate t distribution $t_\ell(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \nu_1)$ and an actual distribution is taken to be an ℓ-dimensional multivariate t distribution $t_\ell(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \nu_2)$ such that $\boldsymbol{\mu}_2 = k\boldsymbol{\mu}_1$, $\boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_1$ for some $k \in \mathbb{R}$.

First, we note that a multivariate t distribution $t_l(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ has a density (see, e.g., Nadarajah and Kotz (2008, p. 99)):

$$f(\boldsymbol{x}) = \frac{\Gamma\left(\frac{l+\nu}{2}\right)}{(\pi\nu)^{\frac{l}{2}}\Gamma\left(\frac{\nu}{2}\right)|\Sigma|^{\frac{1}{2}}} \times \left(1 + \frac{1}{\nu}(\boldsymbol{x}-\boldsymbol{\mu})^{\intercal}\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right)^{-\frac{(l+\nu)}{2}}. \tag{18}$$

We remind the reader that the matrix Σ in (18) is *not* the covariance matrix of the multivariate t distribution, but the covariance matrix is defined for every $\nu > 2$ and can be expressed as $\frac{\nu}{\nu-2}\Sigma$.

We again applied the smoothed loss function $\tilde{\ell}_1(\cdot)$ from Section 4 with $\epsilon = 0.01$ and $\lambda = 0.1$. We present results below for the case of 10 degrees of freedom but we experimented with many other values for the degrees of freedom and the results follow the same pattern. We varied the values of k through the range $1, 0, -1, -2, -3, -4, -5, -8$ and calculated the resulting radius η numerically. The portfolio weights generally stabilized with fewer than the 1,000,000 simulations we performed. As before, we registered the percentage of cases in which the Bregman-based portfolio was not worse than the non-robust portfolio with respect to the loss $\ell_1(\cdot)$. (The non-robust portfolio was defined as minimizing the same loss $\ell_1(x)$ but without considering a neighbourhood around the nominal distribution.) The results are summarized in table 7 and similar results can be obtained when $\tilde{\ell}_2(\cdot)$ is used as a smooth loss function.

As before, an additional table is provided for the market downturn scenario, where we exclude 'zero' loss results where the robust and non-robust strategies performed equally. As shown in table 8, the proportion BT was again in favour of the robust portfolio.

The results of this section clearly outline the significant benefits of using the robust tracking portfolio in a market

Table 7. Tracking performance using the loss $\tilde{\ell}_1$: robust optimal solution versus non-robust optimal solution (MVT) $\lambda = 0.1, \epsilon = 0.01, \nu = 10$ (cases where both losses are zero are included).

k, η	BT %	ETE Robust	ETE Non-robust $(*10^{-4})$	ETE Difference
$k = 1$	74.0965	72.9088	72.9100	− 0.0012
$k = 0$	75.6374	67.4686	67.5311	− 0.0625
$k = -1$	77.1499	62.0974	62.4280	− 0.3306
$k = -2$	78.5679	56.8677	57.5992	− 0.7315
$k = -3$	80.1557	51.8374	53.0417	− 1.2043
$k = -4$	81.8112	47.0439	48.7512	− 1.7073
$k = -5$	83.5051	42.5242	44.7248	− 2.2006
$k = -8$	88.1117	30.7946	34.1600	− 3.3655

Table 8. Tracking performance using the loss $\tilde{\ell}_1$: robust optimal solution versus non-robust optimal solution (MVT) $\lambda = 0.1, \epsilon = 0.01, \nu = 10$ (cases of both losses being zero excluded).

k, η	BT %	ETE Robust	ETE Non-robust $(*10^{-4})$	ETE Difference
$k = 1$	50.77	72.9088	72.9100	− 0.0012
$k = 0$	51.55	67.4686	67.5311	− 0.0625
$k = -1$	52.66	62.0974	62.4280	− 0.3306
$k = -2$	53.50	56.8677	57.5992	− 0.7315
$k = -3$	54.53	51.8374	53.0417	− 1.2043
$k = -4$	55.68	47.0439	48.7512	− 1.7073
$k = -5$	56.93	42.5242	44.7248	− 2.2006
$k = -8$	60.72	30.7946	34.1600	− 3.3655

downturn scenario also for the case where the nominal distribution is heavy-tailed (such as the multivariate t with 10 degrees of freedom). When $k = 1$, the radius η is zero and the robust and non-robust strategies coincide, hence, up to a negligible numerical effect, the percentage was about 75% in table 7 and about 50% in table 8. As k starts getting smaller, the advantage of the robust approach starts popping up and is increasing monotonically when the magnitude of k increases.

5.3. *Real data example*

The simulation method offers the right vehicle to examine the merits of our methodology. This is because only via simulations (where we know the nominal, the actual model and the true radius of contamination) we can sense the effects of our methodology. However, an indirect way to investigate these effects would be to investigate the out-of-sample performance on some real data. We are not aware of another method in the literature for robust index tracking (in the sense we used it here as robustness to deviations from a nominal model), with which to compare. Hence we illustrate the performance of our method on data from the Hang Seng index (Hong Kong). This index contains 31 stocks. The paper (Beasley *et al.* 2003) contains many data sets that have been refereed to since then in other papers. In particular, Guastaroba and Speranza (2012), too, use these data sets. The description of the data sets and a way to access it can be found on page 60 of Guastaroba and Speranza (2012). We are following the same design as done in Guastaroba and Speranza (2012), and apply the same experimental setting as described in table 1 in their paper. For each stock, 291 consecutive weekly prices are provided. From these, the first 104 weekly prices are chosen as in-sample observations, i.e. the initial time period is $T = 104$. The next 52 weeks are set to be the out-of-sample (or validation) period.

Based on the testing data, we used the MATLAB procedure `stepwiseglm` to extract 12 stocks to include in our portfolio. This procedure creates a linear regression model using the stepwise regression approach by adding or removing predictors from the set of 31 stocks to model the response (i.e. the index value in our case). Admittedly, this may not be the best way to choose stocks to be included in the index tracking portfolio but it is a reasonable way of doing it and, as discussed in the introduction, we are unable to discuss rigorously the problem of selecting the 'optimal subset' selection in this paper. As a result of applying the MATLAB procedure, we obtained stocks with indices 4, 11, 12, 13, 15, 18, 21, 22, 23, 25, 26, 27 which we included in our portfolio.

Table 9 gives the estimated correlations matrix of the returns of the chosen 12 stocks (columns 1–12) and the return of the index (column 13), rounded by two decimal places.

The correlation matrix is clearly non-diagonal. It has one outstanding large eigenvalue (8.93689), with the smallest eigenvalue being 0.0051 and the remaining ones are all larger than 0.0051 but smaller than 1.

Replacing expected values by empirical averages in the main Theorem 3.1, we are able to find the optimal weighting for our stock selection:

$$0.0841, 0.1365, 0.0429, 0.0760, 0.2040, 0.0476, 0.0528,$$
$$0.0908, 0.0253, 0.0446, 0.0888, 0.1067$$

(rounded up to the fourth decimal). The values for the other parameters were: $\lambda = 0.2, \eta = 0.005$. These weights should be compared to the optimal non-robust weights:

$$0.0786, 0.1317, 0.0487, 0.0722, 0.2026, 0.0508,$$
$$0.0567, 0.0931, 0.0250, 0.0443, 0.0902, 0.1060.$$

Following the spirit of myopic index tracking, we then applied a moving window of size 104 when working on the remaining out-of-sample 52 data points. Each time, a new data point was added the earliest one was deleted when re-calculating the weights in the next time point. That is, at each time point after $t = 104$, the sliding window contained the last 104 time points when re-calculating the weights. The predicted weights form time point t were used to update the portfolio weights to be applied at time point $t + 1$. We applied the non-robust approach, too, and compared the outcomes of the two procedures. The ETE values for the robust and non-robust approach were almost identical on the testing data set: $9.9707 * 10^{-6}$ for robust versus $9.9552 * 10^{-6}$ for the non-robust but the robust approach outperformed the non-robust in the out-of-sample performance ($2.8869 * 10^{-5}$ for robust versus $2.9152 * 10^{-5}$ for non-robust). The important quantity of interest, BT, was

Table 9. Estimated correlation matrix between the returns of the chosen 12 stocks (columns 1–12) and the index (column 13).

1	0.60	0.66	0.69	0.64	0.76	0.83	0.61	0.31	0.70	0.56	0.69	0.82
0.60	1	0.63	0.64	0.51	0.65	0.67	0.67	0.32	0.59	0.47	0.57	0.77
0.66	0.63	1	0.80	0.67	0.74	0.73	0.69	0.35	0.67	0.67	0.76	0.86
0.69	0.64	0.80	1	0.70	0.81	0.75	0.73	0.33	0.73	0.68	0.78	0.89
0.64	0.51	0.67	0.70	1	0.71	0.65	0.64	0.37	0.64	0.69	0.64	0.82
0.76	0.65	0.74	0.81	0.71	1	0.72	0.70	0.27	0.77	0.64	0.74	0.88
0.83	0.67	0.73	0.75	0.65	0.72	1	0.66	0.27	0.71	0.60	0.78	0.87
0.61	0.67	0.69	0.73	0.64	0.70	0.66	1	0.35	0.69	0.58	0.69	0.84
0.31	0.32	0.35	0.33	0.37	0.27	0.27	0.35	1	0.33	0.33	0.28	0.41
0.70	0.59	0.67	0.73	0.64	0.77	0.71	0.69	0.33	1	0.54	0.73	0.83
0.56	0.47	0.67	0.68	0.69	0.64	0.60	0.58	0.33	0.54	1	0.62	0.76
0.69	0.57	0.76	0.78	0.64	0.74	0.78	0.69	0.28	0.73	0.62	1	0.86
0.82	0.77	0.86	0.89	0.82	0.88	0.87	0.84	0.41	0.83	0.76	0.86	1

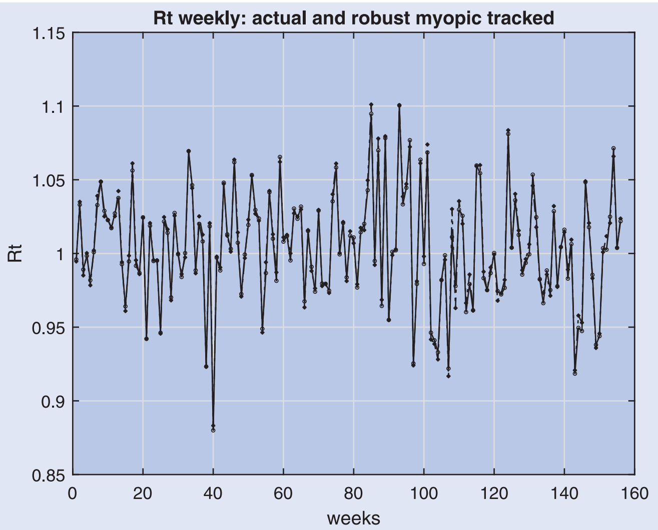

Figure 1. Time series plot of $B_t = 1 + b_t$ and of its estimation using $\boldsymbol{R_t}^{\top}\boldsymbol{u}^*$-the robust myopic index tracking method. First 104 points: learning sample, next 52 points: out-of-sample approximation. Legend: circle-observed, asterix:robust estimator.

also in favour of the robust approach again in the out-of-sample performance (27 out of 52) when the quadratic loss $\ell(x) = x^2$ was used. This increased to 42 out of 52, i.e. 80.77%, when the $\tilde{\ell}_1(x)$ of equation (15), with $\epsilon = 0.01$ was used. These outcomes are in line with the heuristic of the methodology and with the simulations results of sections 5.1 and 5.2. Figure 1 illustrates the accuracy of the approximation of the vector $B_t = 1 + b_t$ via $\boldsymbol{R_t}^{\top}\boldsymbol{u}^*$ across all 156 time points. It is perhaps not surprising that we obtain such an excellent index tracking performance as the proportion of chosen stocks (12 out of 31) is relatively high, and in the same setting as in Guastaroba and Speranza (2012), 2 years of in-sample training and a subsequent 12 months of out-of-sample testing has been applied.

6. Discussion

Various extensions of the suggested approach are of interest and are left for further research. As is to be expected, the robustness effect depends on more than one factor. The value of λ, the 'radius of contamination' η , and the model distribution, all have an effect on performance and a thorough investigation of their interplay needs to be addressed in the future. Obviously, the value of the chosen radius η of the divergence ball is strongly related to the amount of contamination around the nominal model. This information, especially in realistic financial portfolios, is difficult to access. However, our simulations lead us to believe that even with a slightly miss-specified value of η, one can still enjoy the improvement delivered by the robust procedure.

Another adjustment parameter is the λ value in the definition of the divergence. As pointed by Basu *et al.* (1998), there is no universal way of selecting it. It becomes apparent that the choices of η and λ must be inter-related. In simplistic situations, recommendations in this paper about the choice of λ are given as a way of compromise by fixing an acceptable level of efficiency loss for gaining robustness, but a thorough study of the issue is lacking. This represents an avenue for future research.

Another important question is how to measure the quality of the index tracking. We have focused on the quadratic loss function since this is the typical choice in the portfolio tracking literature. This choice equally penalizes the performance of the portfolio whenever it deviates by the same magnitude irrespectively of whether the deviation is above or below the value of the index. A more reasonable choice of loss can be based on the functions ℓ_1 and ℓ_2 discussed in Section 4. Using such a type of loss function delivers a better performance in a clear downturn market scenario, as is shown in Section 5. However, in alternative mixed scenarios, and since there is often no clear separation between market upturn and market downturn in reality, using this loss may be disadvantageous for the investor, as it may reduce their average gains. Further research in this direction will also be beneficial. A thorough comparison with the approach from Roll (1992) is on our agenda.

Finally, the numerical examples in this paper illustrated effects when the actual distribution is a maximal allowable distance from the nominal. This is not necessarily the least-favourable distribution: the least favourable distribution will never be known in practice and in the theoretical discussion we can only get it in the semi-closed form (11) as a part of an implicit solution of an equation system. Despite this, the actual distributions we used for numerical illustrations still give a good proxy for the expected effect. Of course, in our theoretical derivations, we did not need the explicit form of the least-favourable distribution and the derivations in Section 3.3 remain universally valid. The mean vector and the covariance matrix used in our simulations were selected to be close to the daily returns in the Australian share market. For daily returns, assuming multivariate normality is often appropriate. However, when the returns are collected from a longer time horizon or when the nominal distribution itself is different from the multidimensional normal, the benefits of the robust portfolio in the case of quadratic loss may be more or less spectacular depending on how heavy-tailed the nominal distribution turns out to be. In this case, explicit formulae for the divergence, such as, for example, (5), would rarely be available. This does not prevent our methodology from working; for this the required expected values under the nominal distribution in the main Theorem 3.1 should be replaced by their empirical counterparts. More numerical work could demonstrate the advantages of the methodology in such heavy-tailed cases.

As mentioned in the introduction, we do not discuss approaches to sparse index tracking portfolio selection in this paper. The combined requirement for sparsity and robustness in index tracking leads to interesting and challenging optimization problems that are left as a future research avenue.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Australian Research Council's Discovery Project funding scheme (Project DP160103 489).

References

Amari, S. *Information Geometry and Its Applications*, 2016 (Springer: Japan).
Amari, S. and Cichotski, A., Information geometry of divergence functions. *Bull. Polish Acad. Sci.: Tech. Sci.*, 2010, **58**(1), 183–195.
Andriosopoulos, K. and Nomikos, N., Performance replication of the spot energy index with optimal equity portfolio selection: Evidence from the UK, US and Brazilian markets. *Eur. J. Oper. Res.*, 2014, **234**(2), 571–582.
Banerjee, A., Merugu, S., Dhillon, I. and Ghosh, J., Clustering with Bregman divergences. *J. Mach. Learn. Res.*, 2005, **6**, 1705–1749.
Basu, A., Harris, N., Hjort, N. and Jones, M.C., Robust and efficient estimation by minimizing a density power divergence. *Biometrika*, 1998, **85**, 549–559.
Beasley, J.E., Meade, N. and Chang, T.J., An evolutionary heuristic for the index tracking problem. *Eur. J. Oper. Res.*, 2003, **148**(3), 621–643.
Benidis, K., Feng, Y. and Palomar, D., Sparse portfolios for high-dimensional financial index tracking. *IEEE. Trans. Signal. Process.*, 2018, **66**(1), 155–170.
Ben-Tal, A., Teboulle, M. and Charnes, A., The role of duality in optimization problems involving entropy functionals with applications to information theory. *J. Optim. Theory. Appl.*, 1988, **58**(2), 209–223.
Blanchet, J., Lam, H., Tang, Q. and Yuan, Z., Robust actuarial risk analysis. *N. Am. Actuar. J.*, 2019, **23**(1), 33–63.
Boyd, S. and Vandenberghe, L., *Convex Optimization*, 2004 (Cambridge University Press: Cambridge). (Seventh printing with corrections 2009).
Bregman, L.M., The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming. *USSR Comput. Math. Math. Phys.*, 1967, **7**(3), 200–217.
Canakgoz, N.A. and Beasley, J.E, Mixed-integer programming approaches for index tracking and enhanced indexation. *Eur. J. Oper. Res.*, 2009, **196**(1), 384–399.
Chen, C. and Mangasarian, O., Smoothing methods for convex inequalities and linear complementarity problems. *Math. Program.*, 1995, **71**, 51–69.
Chiam, S.C., Tan, K.C. and Mamun, A.A., Dynamic index tracking via multi-objective evolutionary algorithm. *Appl. Soft. Comput.*, 2013, **13**(7), 3392–3408.
Cichocki, A. and Amari, S., Families of alpha- beta- and gamma-divergences: Flexible and robust measures and similarities. *Entropy*, 2018, **12**, 1532–1568.
Dey, S. and Juneja, S., Entropy approach to incorporate fat tailed constraints in financial models. SSRN Electronic Journal. https://papers.ssrn.com/sol3/papers.cfm?abstract\ _id = 1647048, Tata Institute of Fundamental Research Mumbai, India (last download 27 11 2017), 2010.
de Paulo, W.L., de Oliveira, E.M. and do Valle Costa, O.L., Enhanced index tracking optimal portfolio selection. *Finance Res. Lett.*, 2016, **16**, 92–102.
Fabozzi, F.J., *Robust Portfolio Optimization and Management*, 2007 (Wiley: New Jersey).
Filippi, C., Guastaroba, G. and Speranza, M., A heuristic framework for the bi-objective enhanced index tracking problem. *Omega*, 2016, **65**, 122–137.
Gaivoronski, A.A., Krylov, A. and van der Wijst, N., Optimal portfolio selection and dynamic benchmark tracking. *Eur. J. Oper. Res.*, 2005, **163**(1), 115–131.
Glasserman, P. and Xu, X.B., Robust risk measurement and model risk. *Quant. Finance*, 2014, **14**(1), 29–58.
Gnägi, M. and Strub, O., Tracking and outperforming large stock-market indices. *Omega*, 2020, **90**, 101999.
Goh, G. and Dey, D., Bayesian model diagnostics using functional Bregman divergence. *J. Multivar. Anal.*, 2014, **124**, 371–383.
Guastaroba, G., Mansini, R., Ogryczak, W. and Speranza, M.G., Linear programming models based on Omega ratio for the enhanced index tracking problem. *Eur. J. Oper. Res.*, 2016, **251**(3), 938–956.
Guastaroba, G. and Speranza, M.G., Kernel search: An application to the index tracking problem. *Eur. J. Oper. Res.*, 2012, **217**(1), 54–68.
Huber, P. and Ronchetti, E., *Robust Statistics*, 2nd ed., 2009 (Wiley: New York).
Lejeune, M.A., Game theoretical approach for reliable enhanced indexation. *Decis. Anal.*, 2012, **9**(2), 146–155.
Maginn, J.L., Tuttle, D.L., McLeavey, D.W. and Pinto, J.E., *Managing Investment Portfolios*, 3rd ed., 2007 (John Wiley & Sons, Inc.: Hoboken, NJ).
Meade, N. and Salkin, G.R., Developing and maintaining an equity index fund. *J. Oper. Res. Soc.*, 1990, **41**(7), 599–607.
Mihoko, M. and Eguchi, S., Robust blind source separation by beta divergence. *Neural Comput.*, 2002, **14**(8), 1859–1886.
Montfort, K.V., Visser, E. and Draat, L.F.V., Index tracking by means of optimized sampling. *J. Portfolio Manage.*, 2008, **34**(2), 143–152.
Nadarajah, S. and Kotz, S., Estimation methods for the multivariate *t* distribution. *Acta. Appl. Math.*, 2008, **102**(1), 99–118.
Penev, S. and Naito, K., Locally robust methods and near-parametric asymptotics. *J. Multivariate Anal.*, 2018, **167**, 395–417.
Penev, S. and Prvan, T., Robust estimation in structural equation models using Bregman and other divergences with t-centre approach to estimate the covariance matrix. *ANZIAM J.*, 2016, **56**, 339–354. (Proceedings CTAC2014).

Poczos, B.A. and Schneider, J., On the estimation of α-divergences. *J. Maching Learning Res: Workshops Conf.*, 2011, **15**, 609–617.
Roll, R., A mean/variance analysis of tracking error. *J. Portfolio Manage.*, 1992, **18**, 13–22.
Strub, O. and Baumann, P., Optimal construction and rebalancing of index-tracking portfolios. *Eur. J. Oper. Res.*, 2018, **264**(1), 370–387.
Vemuri, B., Liu, M., Amari, S.-I. and Nielsen, F., Total Bregman divergence and its applications to DTI analysis. *IEEE Trans. Medical Imaging*, 2011, **30**, 475–483.

Appendix

A.1. Justification of (5) and (6)

Direct calculation yields:

$$\begin{aligned}
D_{Breg}(\mathcal{E}) &= \mathbb{E}\left(\frac{1}{\lambda}\mathcal{E}^{\lambda+1} - \frac{\lambda+1}{\lambda}\mathcal{E} + 1\right) \\
&= \frac{1}{\lambda}\mathbb{E}(\mathcal{E}^{\lambda+1}) - \frac{1}{\lambda} \\
&= \frac{1}{\lambda}\mathbb{E}\left(\left(\frac{\sqrt{\det(2\pi\boldsymbol{\Sigma}_1)}}{\sqrt{\det(2\pi\boldsymbol{\Sigma}_2)}}\right.\right. \\
&\quad \left.\left.\times\ \frac{\exp\left(-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{\mu}_2)^\mathsf{T}\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{X}-\boldsymbol{\mu}_2)\right)}{\exp\left(-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{\mu}_1)^\mathsf{T}\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{X}-\boldsymbol{\mu}_1)\right)}\right)^{\lambda+1}\right) \\
&\quad - \frac{1}{\lambda} \\
&= \frac{1}{\lambda}\left(\frac{\left(\det(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2^{-1})\right)^{\lambda+1}\det(\tilde{\boldsymbol{\Sigma}}_\lambda)}{\det(\boldsymbol{\Sigma}_1)}\right)^{\frac{1}{2}} \\
&\quad \times \exp\left(-\frac{\lambda+1}{2}\boldsymbol{\mu}_2^\mathsf{T}\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}_2\right. \\
&\quad \left. + \frac{\lambda}{2}\boldsymbol{\mu}_1^\mathsf{T}\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1 + \frac{1}{2}\tilde{\boldsymbol{\mu}}_\lambda^\mathsf{T}\tilde{\boldsymbol{\Sigma}}_\lambda^{-1}\tilde{\boldsymbol{\mu}}_\lambda\right) - \frac{1}{\lambda}.
\end{aligned}$$

By substituting $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$, we obtain (6).

A.2. Justification of the solution to the outer optimization problem

We notice that

$$\begin{aligned}
L^{inner}(\mathcal{E}^*,\boldsymbol{u}) &= \mathbb{E}\left(H(\boldsymbol{u})\mathcal{E}^* + \alpha^*(G(\mathcal{E}^*)-\eta) + \beta^*(\mathcal{E}^*-1)\right) \\
&= \mathbb{E}\left(H(\boldsymbol{u})\left(\frac{\lambda}{\lambda+1}\left(\frac{-\beta^*-H(\boldsymbol{u})}{\alpha^*}\right)+1\right)^{\frac{1}{\lambda}}\right) \\
&\quad + \alpha^*\mathbb{E}\left(\frac{1}{\lambda}\left(\frac{\lambda}{\lambda+1}\left(\frac{-\beta^*-H(\boldsymbol{u})}{\alpha^*}\right)+1\right)^{\frac{\lambda+1}{\lambda}}\right. \\
&\quad \left. -\frac{\lambda+1}{\lambda}\left(\frac{\lambda}{\lambda+1}\left(\frac{-\beta^*-H(\boldsymbol{u})}{\alpha^*}\right)+1\right)^{\frac{1}{\lambda}} + 1 - \eta\right) \\
&\quad + \beta^*\mathbb{E}\left(\left(\frac{\lambda}{\lambda+1}\left(\frac{-\beta^*-H(\boldsymbol{u})}{\alpha^*}\right)+1\right)^{\frac{1}{\lambda}} - 1\right).
\end{aligned}$$

We write the Lagrangian of the outer optimization problem.

$$\begin{aligned}
L^{outer}(\mathcal{E}^*,\boldsymbol{u}) &= \mathbb{E}\left(H(\boldsymbol{u})\left(\frac{\lambda}{\lambda+1}\left(\frac{-\beta^*-H(\boldsymbol{u})}{\alpha^*}\right)+1\right)^{\frac{1}{\lambda}}\right) \\
&\quad + \alpha^*\mathbb{E}\left(\frac{1}{\lambda}\left(\frac{\lambda}{\lambda+1}\left(\frac{-\beta^*-H(\boldsymbol{u})}{\alpha^*}\right)+1\right)^{\frac{\lambda+1}{\lambda}}\right. \\
&\quad \left. -\frac{\lambda+1}{\lambda}\left(\frac{\lambda}{\lambda+1}\left(\frac{-\beta^*-H(\boldsymbol{u})}{\alpha^*}\right)+1\right)^{\frac{1}{\lambda}} + 1 - \eta\right) \\
&\quad + \beta^*\mathbb{E}\left(\left(\frac{\lambda}{\lambda+1}\left(\frac{-\beta^*-H(\boldsymbol{u})}{\alpha^*}\right)+1\right)^{\frac{1}{\lambda}} - 1\right) \\
&\quad - \theta\left(\mathbf{1}^\mathsf{T}\boldsymbol{u} - 1\right).
\end{aligned}$$

The first order condition can then be obtained:

$$\mathbb{E}\left(\frac{\partial H}{\partial \boldsymbol{u}}\left(\frac{\lambda}{\lambda+1}\left(\frac{-\beta^*-H(\boldsymbol{u})}{\alpha^*}\right)+1\right)^{\frac{1}{\lambda}}\right) = \theta^*\mathbf{1},$$
$$\mathbf{1}^\mathsf{T}\boldsymbol{u} = 1. \tag{A1}$$

To check that the solution of the above equation is optimal, we calculate the Hessian. For $\boldsymbol{y} \in \mathbb{R}^n$, we see that

$$\begin{aligned}
&\boldsymbol{y}^\mathsf{T}\left(\frac{\partial^2 L^{outer}(\mathcal{E}^*,\boldsymbol{u})}{\partial \boldsymbol{u}\boldsymbol{u}^\mathsf{T}}\right)\boldsymbol{y} \\
&\quad = \mathbb{E}\left(\boldsymbol{y}^\mathsf{T}\frac{\partial^2 H}{\partial \boldsymbol{u}\boldsymbol{u}^\mathsf{T}}\boldsymbol{y}\mathcal{E}^* - \frac{1}{\alpha^*}\frac{1}{1+\lambda}\left(\boldsymbol{y}^\mathsf{T}\frac{\partial H}{\partial \boldsymbol{u}}\right)^2\left(\mathcal{E}^*\right)^{\frac{1}{1-\lambda}}\right)
\end{aligned}$$

Since $\alpha^* > 0$,

$$\boldsymbol{y}^\mathsf{T}\frac{\partial^2 H}{\partial \boldsymbol{u}\boldsymbol{u}^\mathsf{T}}\boldsymbol{y} = -2\left(\boldsymbol{y}^\mathsf{T}\boldsymbol{R}\right)^2 \le 0,$$

it is easy to see that the Hessian is negative semi-definite. As a consequence, we have verified that the solution of (A1) is optimal, and we will denote it as $\boldsymbol{u}^*$. This finishes the proof of Theorem 3.1.