Quantitative Portfolio Construction and Systematic Trading Strategies using Factor Entropy Pooling

January 2014

Attilio Meucci
Research Associate, EDHEC-Risk Institute

David Ardia
Assistant Professor of Finance, Laval University

Marcello Colasante
Consultant, SYMMYS

EDHEC BUSINESS SCHOOL

Abstract

The Entropy Pooling approach is a versatile theoretical framework to process market views and generalised stress-tests into an optimal "posterior" market distribution, which is then used for risk management and portfolio management. Entropy Pooling can be implemented non-parametrically or parametrically. The non-parametric implementation with historical scenarios is more suitable for risk management applications.

Here introduce the parametric implementation of Entropy Pooling under a factor structure, which we name Factor Entropy Pooling. The factor structure reduces the dimension of the problem and linearises the parameter space, allowing for fast computation of the posterior market distribution.

We apply Factor Entropy Pooling to two portfolio construction problems.

First, we use the Factor Entropy Pooling to construct the "implied returns", i.e. a market distribution consistent with a target optimal portfolio, such as maximum diversification/risk parity, or the CAPM equilibrium. Our approach improves on the implied returns à la Black-Litterman, and the ensuing distribution can be used as the starting point for further portfolio construction.

Second, we use Factor Entropy Pooling to construct and backtest quantitative systematic trading strategies based on ranking views, or "portfolios from sorts". Unlike standard approaches, Factor Entropy Pooling closely ties to the actual empirical data.

Fully documented code is available at symmys.com/node/160.

JEL Classification: C1, G11

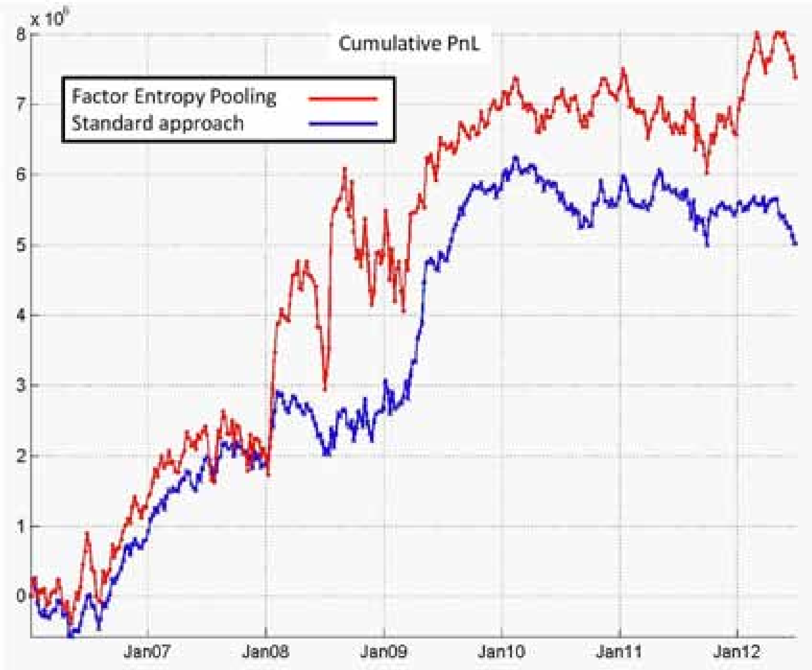
Keywords: trading signals, tactical allocation, Black-Litterman, equilibrium prior, shrinkage, risk management, Entropy Pooling, factor models, inequality views, portfolios from sorts, ranking, Kullback-Leibler.

The authors are grateful to Simon Keel for his support and feedback throughout and to Lennart Hoogerheide for his constructive comments.

1. Introduction

Processing trading signals, or, more in general, views on the market, to compute an optimal allocation is one of the main challenges in quantitative portfolio construction. Similarly, embedding stress-tests in a risk model in a statistically sound way is key to a healthy risk management process. The generalised Bayesian approach "Entropy Pooling" is a flexible framework to process views and generalised stress-tests.

Figure 1: Cumulative backtested P&L of a quantitative systematic strategy: Factor Entropy Pooling (red) and standard approach (blue).



The theoretical framework for Entropy Pooling was laid out in full generality in [Meucci, 2008]. Entropy Pooling combines an arbitrary market model, which is referred to as the "prior" and fully general views or stress-tests on the underlying market. The prior market model can be any, not necessarily normal, multivariate distribution; the views are fully general statements such as spread views, relative rankings, views on tails, correlations, volatilities, etc. The output is a distribution, referred to as the "posterior", which incorporates all the inputs and which can be used for portfolio construction and risk management. In Entropy Pooling, the posterior is obtained by warping the prior distribution so that the views are fulfilled, in such a way that the prior is minimally distorted. Specifically, the posterior distribution minimises the entropy relative to the prior, which is a natural measure of discrepancy between two distributions.

The Entropy Pooling framework can be implemented in two ways: non-parametrically, by representing the prior and the posterior in terms of scenarios and associated "Flexible Probabilities"; and parametrically, by making parametric assumptions on the prior and the posterior distributions.

The non-parametric implementation of Entropy Pooling via Flexible Probabilities was studied in the original article, and explored further with applications to tail risk in [Meucci et al., 2011], and with applications to risk management in [Meucci, 2010], [Meucci, 2012] and [Meucci, 2013]. In the non-parametric implementation, the scenarios, historical or Monte Carlo, are kept fixed, and the respective probabilities are modulated to reflect the portfolio manager's views on the market, or the risk manager's stress-tests.

1 - In this article we adopt the following notation

$\boldsymbol{m} = \mathrm{Diag}(\boldsymbol{v})$	$\bar{n}$ x $\bar{n}$ matrix of zeros, except principal diagonal which is $\bar{n}$ x 1 vector $\boldsymbol{v}$
v = diag(m)	$\bar{n}$ x1 vector equal principal diagonal of $\bar{n}$ x $\bar{n}$ matrix $\boldsymbol{m}$
$\boldsymbol{\sigma}^2$	$\bar{n}$ x $\bar{n}$ symmetric, positive (semi) definite matrix
$\boldsymbol{\sigma}_{vec} \equiv \sqrt{diag(\boldsymbol{\sigma}^2)}$	$\bar{n}$ x 1 vector of square roots of principal diagonal of $\boldsymbol{\sigma}^2$
	$\bar{n}$ x $\bar{n}$ symmetric matrix $\boldsymbol{\sigma}$ such that $\boldsymbol{\sigma}\boldsymbol{\sigma}' = \boldsymbol{\sigma}^2$
n = 1,..., $\bar{n}$	indices of market entries ($\bar{n}$ is the market dimension)
k = 1,..., $\bar{k}$	factor indices ($\bar{k}$ is the number of factors)
j = 1,..., $\bar{j}$	scenario indices ($\bar{j}$ is the number of scenarios)

The parametric implementation of Entropy Pooling was studied in the original article under special views, namely equalities on expectations and covariances. With these special types of views, the posterior was computed analytically.

In this article we discuss the parametric implementation of Entropy Pooling with fully general views, such as ranking signals, which allow us to build systematic trading strategies such as the one in Figure 1, which we detail further below. With general views it is not possibleto compute the posterior analytically. Hence, we propose an efficient numerical approach, which rests on two pillars: first, the dimension reduction of the correlation structure to a "factor" model, which increases the statistical efficiency of our estimates; second, the choice of coordinates in the reduced-dimension parameter space, which become unconstrained. We call the ensuing approach "Factor Entropy Pooling".

Factor Entropy Pooling is fast, and as such it allows us to address a variety of applications that require speed. In particular, we discuss two portfolio construction applications. The first application of Factor Entropy Pooling is the estimation of the "implied returns" consistent with a target optimal portfolio, such as maximum diversification/risk parity, or a CAPM-like equilibrium. Implied returns were first proposed in [Black and Litterman, 1990] as the starting point of a sensible mean-variance portfolio construction. The implied returns based on Factor Entropy Pooling improve on the Black-Litterman approach, by being closer to the original market data.

The second application of Factor Entropy Pooling is the construction of quantitative trading strategies based on ranking signals for alpha-generation, the so-called "portfolios from sorts". In the standard approach, discussed e.g. in [Grinold and Kahn, 1999], the expected returns of all the securities in a given market are set proportional to a given predictive signal. However, the strict proportionality assumption imposes spurious additional information in the optimisation process. [Almgren and Chriss, 2006] first addressed this issue, but their solution does not take empirical data into account. Factor Entropy Pooling effectively estimates ranking-consistent expected returns that do not impose spurious information and at the same time starts from the empirical observations.

We emphasise that Factor Entropy Pooling is not a methodology to guarantee more (risk adjusted) money, but rather a richer approach to portfolio construction, which may do worse or better ex-post, see also [Bailey et al., 2013].

The remainder of this article is organised as follows. In Section 2 we review the original general Entropy Pooling framework. In Section 3 we introduce the Factor Entropy Pooling framework. Then we show two applications of Factor Entropy Pooling. In Section 4 we use Factor Entropy Pooling to determine the implied returns of a given optimal target portfolio. In Section 5 we illustrate how to use Factor Entropy Pooling to build quantitative systematic strategies. In the appendix we report all proofs and technical details. Fully documented code is available at symmys.com/node/160.

2. Review of Entropy Pooling

This section draws from [Meucci, 2008], please refer to that publication for more details.

Entropy Pooling proceeds in three main steps. The first step of Entropy Pooling is the estimation of a "prior" distribution for a set of $\bar{n}$ risk drivers $\boldsymbol{X} \equiv (X_1, \ldots, X_{\bar{n}})'$ in the market, as represented by its probability density function (pdf), which we denote by $\underline{f}$

$$\boldsymbol{X} \sim \underline{f}. \tag{1}$$

The risk drivers are any set of random variables that fully determine the securities P&L, such as interest rates, implied volatility surfaces, etc.

The second step of Entropy Pooling is expressing the views or stress-tests $\mathcal{V}$. These are statements on expectations, correlations, tail risk conditions, etc. that possibly contradict the prior, and yet we want to embed in our risk management or allocation. For instance, the prior could represent a regular regime in the markets, and the views/stress-test can be a regime where some of the correlations, or all of them, increase substantially. Therefore, views and stress-tests $\mathcal{V}$ are constraints on the yet to be defined posterior distribution of the market.

We denote that a generic distribution f satisfies these constraints as follows

$$f \in \mathcal{V}. \tag{2}$$

Since the views possibly contradict the prior, the prior distribution (1) does not satisfy the views ($\underline{f} \notin \mathcal{V}$) and we need to search for a new, suitable distribution, the "posterior" distribution. The third step of Entropy Pooling is the computation of the posterior distribution $\bar{f}$ for the risk drivers, which incorporates the views or stress-tests $\mathcal{V}$. To compute the posterior, first we introduce the relative entropy, a measure of the similarity of a distribution f with respect to a reference distribution, in our case the prior $\underline{f}$

$$\mathcal{E}(f \| \underline{f}) \equiv \int f(\boldsymbol{x}) \ln(f(\boldsymbol{x}) / \underline{f}(\boldsymbol{x})) d\boldsymbol{x}. \tag{3}$$

Then we define the posterior $\bar{f}$ as the one distribution which is the most similar to the prior $\underline{f}$, but at the same time, unlike in general the prior, satisfies the views $\mathcal{V}$. Therefore, we define the posterior as follows

$$\bar{f} \equiv \underset{f \in \mathcal{V}}{\operatorname{argmin}} \, \mathcal{E}(f \| \underline{f}). \tag{4}$$

The posterior distribution $\bar{f}$ is then used as input to an optimiser to compute the optimal portfolios that incorporate the views $\mathcal{V}$, or to compute summary statistics that reflect the stress-tests $\mathcal{V}$ for risk management purposes. Finally, a confidence level in the views can be added, by computing a confidence-weighted mixture of the prior and the posterior.

As mentioned in the introduction, Entropy Pooling can be implemented in two ways: the non-parametric and the parametric approach.

In the non-parametric approach the prior $\underline{f}$ is represented in terms of a large number $\bar{\jmath}$ of joint scenarios for the risk drivers $\boldsymbol{x}^{(j)} \equiv (x_1^{(j)}, \ldots, x_{\bar{n}}^{(j)})'$ and the associated probabilities $\{\boldsymbol{x}^{(j)}; \underline{p}^{(j)}\}_{j=1}^{\bar{\jmath}}$. Then the posterior (4) is represented by the same scenarios with a new set of probabilities $\{\boldsymbol{x}^{(j)}; \bar{p}^{(j)}\}_{j=1}^{\bar{\jmath}}$. Organising all the $\bar{\jmath}$ probabilities in a vector $\boldsymbol{p} \equiv (p^{(1)}, \ldots, p^{(\bar{\jmath})})$, the posterior probabilities are defined as follows

$$\bar{\boldsymbol{p}} \equiv \underset{\boldsymbol{p} \in \mathcal{V}}{\operatorname{argmin}} \, \mathcal{E}(\boldsymbol{p} \| \underline{\boldsymbol{p}}), \tag{5}$$

where with minor abuse we let $\mathcal{E}(\boldsymbol{p} \| \underline{\boldsymbol{p}}) \equiv \sum_{j=1}^{\bar{\jmath}} p^{(j)} \ln(p^{(j)} / \underline{p}^{(j)})$ denote the discrete counterpart of the relative entropy (3). As it turns out, for several types of views the optimisation (5) can be transformed in an instance of linear programming with a low number of variables, and thus it can be efficiently solved numerically.

In the parametric approach, all the distributions belong to a given parametric class, i.e. $f \equiv f_{\boldsymbol{\theta}}$, where the parameters $\boldsymbol{\theta}$ span a set of values Θ. In particular, the prior is represented by $f_{\underline{\boldsymbol{\theta}}}$ and the posterior (4) becomes $f_{\bar{\boldsymbol{\theta}}}$,

$$\bar{\boldsymbol{\theta}} \equiv \underset{\boldsymbol{\theta} \in \mathcal{V}}{\operatorname{argmin}} \, \mathcal{E}(\boldsymbol{\theta} \| \underline{\boldsymbol{\theta}}). \tag{6}$$

where, with minor abuse of notation $\boldsymbol{\theta} \in \mathcal{V} \Leftrightarrow \{f_{\boldsymbol{\theta}} \in \mathcal{V}, \boldsymbol{\theta} \in \Theta\}$ and $\mathcal{E}(\boldsymbol{\theta} \| \underline{\boldsymbol{\theta}}) \Leftrightarrow \mathcal{E}(f_{\boldsymbol{\theta}} \| f_{\underline{\boldsymbol{\theta}}})$.

A special case of the parametric approach is the normal assumption

$$f_{\boldsymbol{\mu},\boldsymbol{\sigma}^2}(\boldsymbol{x}) \equiv (2\pi)^{-\frac{\bar{n}}{2}} |\boldsymbol{\sigma}^2|^{-\frac{1}{2}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'(\boldsymbol{\sigma}^2)^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}, \tag{7}$$

where $\boldsymbol{\mu}$ is an $\bar{n}$ vector of expectations and $\boldsymbol{\sigma}^2$ is an $\bar{n}$ x $\bar{n}$ symmetric and positive definite covariance matrix. Therefore, the parameters are $\boldsymbol{\theta} \equiv (\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$, where $\boldsymbol{\sigma}^2$ is constrained to be symmetric and positive definite, which we denote by $\boldsymbol{\sigma}^2 \succ 0$. Accordingly, under normality the parametric problem (6) becomes

$$(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\sigma}}^2) \equiv \underset{(\boldsymbol{\mu},\boldsymbol{\sigma}^2)\in\mathcal{V}}{\operatorname{argmin}}\, \mathcal{E}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2 \| \underline{\boldsymbol{\mu}}, \underline{\boldsymbol{\sigma}}^2), \tag{8}$$

where $(\boldsymbol{\mu}, \boldsymbol{\sigma}^2) \in \mathcal{V} \Leftrightarrow \{f_{\boldsymbol{\mu},\boldsymbol{\sigma}^2} \in \mathcal{V}, \boldsymbol{\sigma}^2 \succ 0\}$. The relative entropy between two normal distributions can be computed explicitly and reads

$$\begin{aligned} \mathcal{E}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2 \| \underline{\boldsymbol{\mu}}, \underline{\boldsymbol{\sigma}}^2) = \tfrac{1}{2}(&\operatorname{tr}(\boldsymbol{\sigma}^2(\underline{\boldsymbol{\sigma}}^2)^{-1}) - \ln|\boldsymbol{\sigma}^2(\underline{\boldsymbol{\sigma}}^2)^{-1}| \\ &+ (\boldsymbol{\mu} - \underline{\boldsymbol{\mu}})'(\underline{\boldsymbol{\sigma}}^2)^{-1}(\boldsymbol{\mu} - \underline{\boldsymbol{\mu}}) - \bar{n}). \end{aligned} \tag{9}$$

Suppose that the views are equality statements on expectations $\mathbb{E}\{\boldsymbol{aX}\} \equiv \boldsymbol{\xi}$ and covariances $\mathbb{C}v\{\boldsymbol{cX}\} \equiv \phi^2$, or

$$(\boldsymbol{\mu}, \boldsymbol{\sigma}^2) \in \mathcal{V} \quad \Longleftrightarrow \quad \boldsymbol{a\mu} \equiv \boldsymbol{\xi}, \quad \boldsymbol{c\sigma}^2\boldsymbol{c}' \equiv \phi^2, \tag{10}$$

where $\boldsymbol{a}$ is a full-rank $\bar{m}$ x $\bar{n}$ matrix that de nes the combinations in the views for the expectations and $\boldsymbol{\xi}$ is an $\bar{m}$ x 1 vector that quantifies the views on the expectations; and $\mathbf{c}$ is a full-rank $\bar{s}$ x $\bar{n}$ matrix that defines the combinations in the views for the covariances and ϕ^2 is an $\bar{s}$ x $\bar{s}$ positive definite matrix that quantifies the views on the covariances. Then the normal Entropy Pooling problem (8) can be solved analytically and the solution of (8)-(10) reads

$$\bar{\boldsymbol{\mu}} = \underline{\boldsymbol{\mu}} + \underline{\boldsymbol{\sigma}}^2\boldsymbol{a}'(\boldsymbol{a}\underline{\boldsymbol{\sigma}}^2\boldsymbol{a}')^{-1}(\boldsymbol{\xi} - \boldsymbol{a}\underline{\boldsymbol{\mu}}) \tag{11}$$

$$\bar{\boldsymbol{\sigma}}^2 = \underline{\boldsymbol{\sigma}}^2 + \underline{\boldsymbol{\sigma}}^2\boldsymbol{c}'[(\boldsymbol{c}\underline{\boldsymbol{\sigma}}^2\boldsymbol{c}')^{-1}\phi^2(\boldsymbol{c}\underline{\boldsymbol{\sigma}}^2\boldsymbol{c}')^{-1} - (\boldsymbol{c}\underline{\boldsymbol{\sigma}}^2\boldsymbol{c}')^{-1}]\boldsymbol{c}\underline{\boldsymbol{\sigma}}^2. \tag{12}$$

It is immediate to check that the posterior $(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\sigma}}^2)$ satisfies the views (10).

The analytical solution (11)-(12) gives the optimal multivariate normal density that satisfies the views (10). A straightforward application of (11)-(12) is in the context of mean-variance portfolio optimisation, when the portfolio manager has views on some portfolios and their correlations, and/or on the correlations of other portfolios. Another feature of the analytical solution is that the portfolio manager can easily compute the effect on the final allocation of small changes in the views, and tweak the views accordingly. This gives interesting insights into the sensitivity of the portfolio with respect to the views applied.

The limitations of the analytical solution is that the views (10), although quite flexible, are still fairly restrictive. Non-linear views and inequality or ranking views are not addressedby (11)-(12). We proceed to discuss how to process these views in the next section.

3. Factor Entropy Pooling

Here we derive results that allow for the implementation of Entropy Pooling in its parametric form (8), with fully flexible views $\mathcal{V}$ beyond mean and covariance (10). In this case, the solution must be computed numerically. To this purpose, we impose that the covariances are of "factor" type

$$\boldsymbol{\sigma}^2 \equiv \boldsymbol{bb}' + Diag(\boldsymbol{d} \circ \boldsymbol{d}), \tag{13}$$

where $\boldsymbol{b}$ is an $\bar{n}$ x $\bar{k}$ matrix ($\bar{k} \ll \bar{n}$), $\boldsymbol{d}$ is an $\bar{n}$ x 1 vector; the operator *Diag*(***v***) embeds the $\bar{n}$x1 vector ***v*** into the principal diagonal of a square matrix which is zero anywhere else (hence *Diag*(***v***) is a matrix); and $\circ$ is the Hadamard product, i.e. the entry-by-entry multiplication of vectors.

The structure (13) is consistent with a linear factor model assumption $\boldsymbol{X} = \boldsymbol{b}\boldsymbol{Z} + \boldsymbol{U}$, where the factors covariance is the low-dimensional $\bar{k}$ x $\bar{k}$ identity matrix ($\boldsymbol{\sigma}^2_{\boldsymbol{Z}} = \boldsymbol{i}_{\bar{k}\times\bar{k}}$), the residuals have idiosyncratic diagonal covariance ($\boldsymbol{\sigma}^2_{\boldsymbol{U}} = Diag(\boldsymbol{d} \circ \boldsymbol{d})$), and the factors are systematic, in that they are uncorrelated with the residuals ($\boldsymbol{\sigma}^2_{\boldsymbol{Z},\boldsymbol{U}} = \boldsymbol{0}_{\bar{k}\times\bar{n}}$). This last remark explains the name of the present approach as Factor Entropy Pooling.

With the factor parametrisation (13), the Entropy Pooling problem (8) becomes the following Factor Entropy Pooling optimisation

$$(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{b}}, \bar{\boldsymbol{d}}) \equiv \underset{(\boldsymbol{\mu},\boldsymbol{b},\boldsymbol{d})\in\mathcal{V}}{\operatorname{argmin}} \; \mathcal{E}(\boldsymbol{\mu}, \boldsymbol{b}\boldsymbol{b}' + Diag(\boldsymbol{d} \circ \boldsymbol{d}) \| \underline{\boldsymbol{\mu}}, \underline{\boldsymbol{\sigma}}^2). \tag{14}$$

As we show in Appendix A.7, the optimisation target (8) in general is not a convex function of the entries ($\boldsymbol{\mu}$, ***b***, ***d***). However, we can enhance the computational efficiency of the optimisation by feeding the analytical expression of the gradient and the Hessian of the optimisation target (14) in the optimisation algorithm.

More precisely, using the notation discussed in Appendix A.2, the gradient of the optimisation target (14) reads

$$\nabla_{\boldsymbol{\mu}}\mathcal{E} = (\underline{\boldsymbol{\sigma}}^2)^{-1}(\boldsymbol{\mu} - \underline{\boldsymbol{\mu}}) \tag{15}$$

$$\nabla_{\boldsymbol{d}}\mathcal{E} = diag((\underline{\boldsymbol{\sigma}}^2)^{-1} - (\boldsymbol{\sigma}^2)^{-1}) \circ \boldsymbol{d} \tag{16}$$

$$\nabla_{\boldsymbol{b}}\mathcal{E} = vec(((\underline{\boldsymbol{\sigma}}^2)^{-1} - (\boldsymbol{\sigma}^2)^{-1})\boldsymbol{b}), \tag{17}$$

where *diag*(***m***) is the column vector formed from the elements of the main diagonal of the matrix ***m*** (hence *diag*(***m***) is a vector) and *vec*(***x***) is the column vector obtained by stacking columns of ***x***, see the proof in Appendix A.5. In Appendix A.6 we also provide the explicit formula for the Hessian of the optimisation target (14).

Notice that the high-dimensional inverses $(\boldsymbol{\sigma}^2)^{-1}$ that appear in the gradient and in the Hessian are easily obtained analytically in terms of low-cost, low-dimensional inverses. Indeed, from the binomial inverse theorem we obtain

$$(\boldsymbol{\sigma}^2)^{-1} = \boldsymbol{c} - \boldsymbol{c}\boldsymbol{b}(\boldsymbol{b}'\boldsymbol{c}\boldsymbol{b} + \boldsymbol{i}_{\bar{k}\times\bar{k}})^{-1}\boldsymbol{b}'\boldsymbol{c}, \tag{18}$$

where $\boldsymbol{c} \equiv Diag(\boldsymbol{1}_{\bar{n}\times 1}./(\boldsymbol{d} \circ \boldsymbol{d}))$ is a diagonal matrix ($\boldsymbol{1}_{\bar{n}\text{x}1}$ is a $\bar{n}$**x1** vector ones and ./ denotes the term-by-term division of two vectors).

To ensure that the views ($\boldsymbol{\mu}$, ***b***, ***d***) $\in \mathcal{V}$ in the Factor Entropy Pooling minimisation (14) are satisfied, we use one of two approaches.

The first case occurs when the views can be expressed directly as constraints on the parameters

$$(\boldsymbol{\mu}, \boldsymbol{b}, \boldsymbol{d}) \in \mathcal{V} \iff v(\boldsymbol{\mu}, \boldsymbol{b}, \boldsymbol{d}) \leq \boldsymbol{0} \tag{19}$$

for a suitable vector-valued function v. The simplest example are the equality views on means and covariances (10). As a less trivial occurs in quantitative portfolio construction, when the views are constraints on the Sharpe ratios, see (30) below. When the views can be expressed directly as constraints on the parameters as in (19) we can compute the gradient and the Hessian of the constraints vector *v* and thus further increase the speed of the Factor Entropy Pooling optimisation (14), refer again to the quantitative portfolio construction example below, formulas (38) and thereafter.

The second case occurs when the views cannot be expressed directly as constraints on the parameters. In this situation we rely on Monte Carlo simulations. More precisely, we generate once and for all a set $\{z_j\}_{j=1}^{\bar{j}}$ of joint uncorrelated independent standard normal draws $z_j \equiv (z_{1,j}, \ldots, z_{\bar{n},j})'$; then for any choice of the parameters ($\boldsymbol{\mu}$, $\boldsymbol{b}$, $\boldsymbol{d}$) we transform the scenarios z_j in joint scenarios $\boldsymbol{x}_j$ from the distribution $f_{\boldsymbol{\mu}, \boldsymbol{b}, \boldsymbol{d}}$ as follows

$$\boldsymbol{x}_j^{(\boldsymbol{\mu},\boldsymbol{b},\boldsymbol{d})} \equiv \boldsymbol{\mu} + \boldsymbol{\sigma} \boldsymbol{z}_j, \tag{20}$$

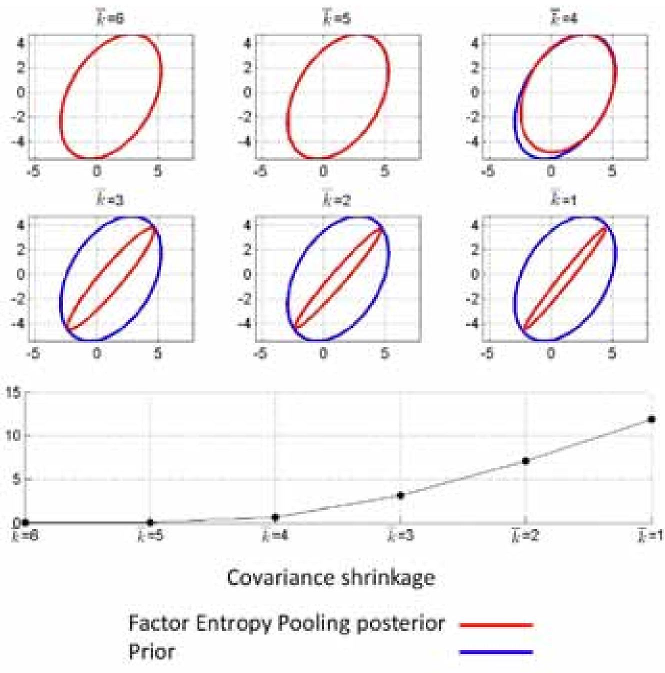
where $\boldsymbol{\sigma}$ is an $\bar{n}$ x $\bar{n}$ symmetric matrix such that $\boldsymbol{\sigma}\boldsymbol{\sigma}' = \boldsymbol{\sigma}^2 \equiv \boldsymbol{b}\boldsymbol{b}' + Diag(\boldsymbol{d} \circ \boldsymbol{d})$. Then we express the views as constraints on the Monte Carlo distribution

$$(\boldsymbol{\mu}, \boldsymbol{b}, \boldsymbol{d}) \in \mathcal{V} \iff \{\boldsymbol{x}_j^{(\boldsymbol{\mu},\boldsymbol{b},\boldsymbol{d})}\}_{j=1}^{\bar{j}} \in \mathcal{V}. \tag{21}$$

To increase the accuracy of the affine transformation (20) we match the mean and covariance of the simulations to the true parameters, as in [Meucci, 2009]. The affine transformation is fast even for large dimensions $\bar{n}$ and large number of scenarios $\bar{j}$ and thus it allows us to accurately and efficiently represent the views (21).

The Factor Entropy Pooling approach (14) presents a number of appealing features. From a statistical perspective, the factor parametrisation (13) is fully determined by a relatively small number $\bar{n}(\bar{k}$ + 2) of parameters $\boldsymbol{\theta}$ = (μ, b, d), instead of the large number $\bar{n}$ ($\bar{n}$ + 3)/2 of parameters in the full-blown specification $\boldsymbol{\theta}$ = (μ, σ^2). This parsimonious structure with limited parameters is an instance of shrinkage estimation, see e.g. [Meucci, 2005] for a review. As such, Factor Entropy Pooling provides statistically efficient estimates in large dimensional markets.

Figure 2: Factor Entropy Pooling as shrinkage estimation.

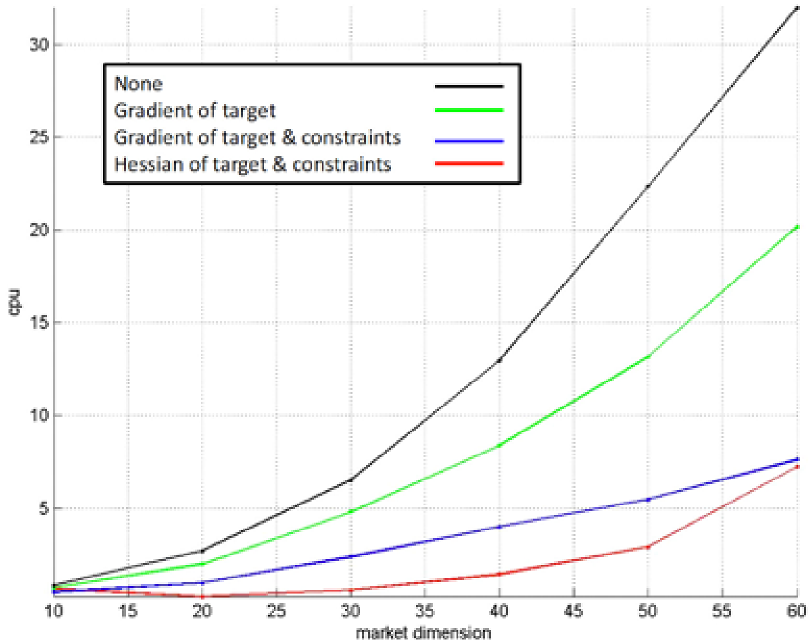


Example 1 We start with a randomly generated $\bar{n}$ = 6 dimensional market, we compute the Factor Entropy Pooling (14) without any view, progressively reducing the number of factors from $\bar{k}$ = $\bar{n}$ = 6 (no shrinkage) to $\bar{k}$ = 1 (maximum shrinkage). In Figure 2 we display the result. The top panel represents geometrically the expectations and the covariances via ellipsoids, as in [Meucci, 2005]: as the number of factors $\bar{k}$ decreases, Factor Entropy Pooling shrinks away from the starting no-shrinkage market. The bottom plot displays the relative entropy (amount of shrinkage) of Factor Entropy Pooling with respect to prior. For more details, please refer to the code available at symmys.com/node/160.

Furthermore, the Factor Entropy Pooling parametrisation allows us to efficiently handle both the optimisation target and the views. Indeed, the parsimonious parametrisation $\boldsymbol{\theta} \equiv (\boldsymbol{\mu}, \boldsymbol{b}, \boldsymbol{d})$ is unconstrained, as the parameters can freely range in the space $\Theta \equiv \mathbb{R}^{\bar{n}} \times \mathbb{R}^{\bar{n}\times\bar{k}} \times \mathbb{R}^{\bar{n}}$. Instead, in the full specification $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$ the matrix $\boldsymbol{\sigma}^2$ is constrained to be symmetric and positive definite.

Finally, the analytical expression of the gradient (15)-(17) and the Hessian, and the analytical inversion of the covariance (18) makes the convergence of off-the-shelf algorithms much faster.

Figure 3: Computational efficiency of the Factor Entropy Pooling optimisation: base case (black); adding use of analytical gradient of target (green); further adding use of analytical gradient of constraints (blue); further adding use of analytical Hessian of both target and constraints (red).



Example 2 We show in Figure 3 how feeding the analytical derivatives can significantly speed up the optimisation algorithm. For more details, please refer to the code available at symmys.com/node/160.

4. Implied Expected Returns

In this section we use the Factor Entropy Pooling to determine the "implied returns", namely the distribution consistent with an optimal target portfolio, which lies at the foundation of the portfolio construction approach in [Black and Litterman, 1990].

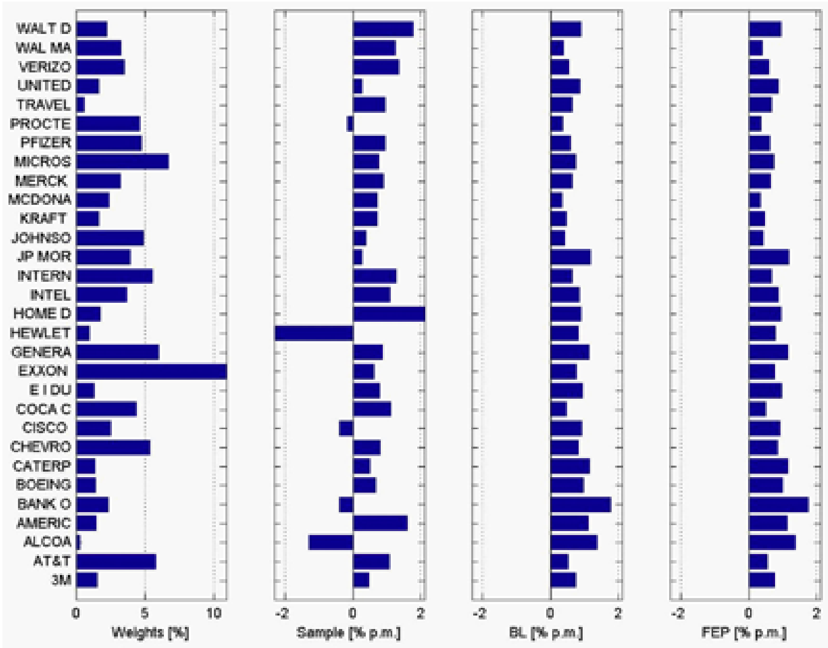
Consider a market of $\bar{n}$ assets. Under some assumptions on the market distribution and the preferences of the investors, the Capital Asset Pricing Model purports that the equilibrium market capitalisation portfolio, as represented by the $\bar{n}$ x 1 vector of weights w_{eq}, is linked to the $\bar{n}$ x 1 vector $\boldsymbol{\mu}$ of returns expectations and the $\bar{n}$ x $\bar{n}$ matrix $\boldsymbol{\sigma}^2$ of returns covariances by the following identity

$$\boldsymbol{\mu} - \gamma\boldsymbol{\sigma}^2 \boldsymbol{w}_{eq} \equiv \mathbf{0}_{\bar{n}\times 1}, \tag{22}$$

where $\gamma > 0$ is a risk aversion parameter.

In practical applications w_{eq} is not necessarily the CAPM equilibrium portfolio, but rather a target optimal portfolio that the portfolio manager would use in the absence of additional views on the market, such as a maximum-diversification/risk-parity portfolio.

Figure 4: Implied expected returns: The left-plot reports the weights of the Dow Jones portfolio. The other bar plots report the annualised expected returns of the various equities computed as sample means (second plot), using Black-Litterman (third plot) and using Factor Entropy Pooling (last plot).



For portfolio construction purposes, the equilibrium constraints (22) guarantees that a mean-variance optimisation yields the portfolio w_{eq}. However, the equilibrium constraint (22) is not satisfied empirically by standard estimates of the expectations $\hat{\boldsymbol{\mu}}$ and the covariances $\hat{\boldsymbol{\sigma}}^2$, such as, say, historical mean and historical covariance.

To enforce the constraint (22), [Black and Litterman, 1990] propose a two-step approach. In the first step, we fit a covariance matrix $(\boldsymbol{\sigma}^2)^{BL} \equiv \hat{\boldsymbol{\sigma}}^2$ to empirical observations by means of standard techniques such as exponential smoothing or maximum likelihood; for fairness, we enhance this estimate with a factor structure as in (13): In the second step, we compute the so-called "implied expected returns", namely the expectations that satisfy the equilibrium constraint (22)

$$(\boldsymbol{\sigma}^2)^{BL} \equiv \hat{\boldsymbol{\sigma}}^2, \quad \boldsymbol{\mu}^{BL} \equiv \gamma \hat{\boldsymbol{\sigma}}^2 \boldsymbol{w}_{eq}. \tag{23}$$

Although the parameters (23) are consistent with the equilibrium constraint (22), they present two problems: no estimation error is assumed on the covariances, and the equilibrium means can depart substantially from the data.

To partly address this issue, [Levy and Roll, 2010], propose to fit a correlation matrix ĉ to empirical observations and then ensure that the equilibrium constraint (22) is satisfied by modifying both the expectations and the variances. More precisely, defining $\boldsymbol{\sigma}_{vec} \equiv \sqrt{diag(\boldsymbol{\sigma}^2)}$, the authors introduce a distance $\mathcal{D}$ between the estimates of the expectations and the standard deviations $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\sigma}}_{vec})$ and the yet to be defined parameters $(\boldsymbol{\mu}, \boldsymbol{\sigma}_{vec})$ as follows

$$\mathcal{D}(\boldsymbol{\mu}, \boldsymbol{\sigma}_{vec} \| \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\sigma}}_{vec}) \equiv (\alpha \|(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})./\hat{\boldsymbol{\sigma}}_{vec}\|^2 + (1-\alpha)\| (\boldsymbol{\sigma}_{vec} - \hat{\boldsymbol{\sigma}}_{vec})./\hat{\boldsymbol{\sigma}}_{vec}\|^2)^{\frac{1}{2}}, \tag{24}$$

where we recall that ./ denotes the entry-by-entry division, and the authors set $\alpha = 0.75$.

Then the authors compute the parameters $(\boldsymbol{\mu}, \boldsymbol{\sigma}_{vec})$ that minimise the distance with respect to the estimated parameters

$$(\boldsymbol{\mu}^{LR}, \boldsymbol{\sigma}^{LR}_{vec}) \equiv \underset{\boldsymbol{\mu}, \boldsymbol{\sigma}^2 \in \mathcal{V}}{\operatorname{argmin}} \mathcal{D}(\boldsymbol{\mu}, \boldsymbol{\sigma}_{vec} \| \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\sigma}}_{vec}). \tag{25}$$

Finally, the authors set $(\boldsymbol{\sigma}^2)^{LR} \equiv Diag(\boldsymbol{\sigma}^{LR}_{vec})\,\hat{\boldsymbol{c}}\,Diag(\boldsymbol{\sigma}^{LR}_{vec})^2$. The parameters $(\boldsymbol{\mu}^{LR}, (\boldsymbol{\sigma}^2)^{LR})$ are consistent with the equilibrium condition (22). Furthermore, they give rise to better trading strategies, see [Ni *et al.*, 2011]. However, they still present one problem: no estimation error is assumed on the correlation, and thus more estimation error is loaded on the means.

To improve further the estimation of the equilibrium distribution we can use our Factor Entropy Pooling framework. Accordingly, we replace the Euclidean distance minimisation (25) with the relative entropy minimisation (14), which we report here

$$(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{b}}, \bar{\boldsymbol{d}}) \equiv \underset{(\boldsymbol{\mu},\boldsymbol{b},\boldsymbol{d})\in\mathcal{V}}{\text{argmin}}\ \mathcal{E}(\boldsymbol{\mu}, \boldsymbol{b}\boldsymbol{b}' + Diag(\boldsymbol{d}\circ\boldsymbol{d})\|\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\sigma}}^2), \tag{26}$$

where the equilibrium constraint (22) now becomes the following view

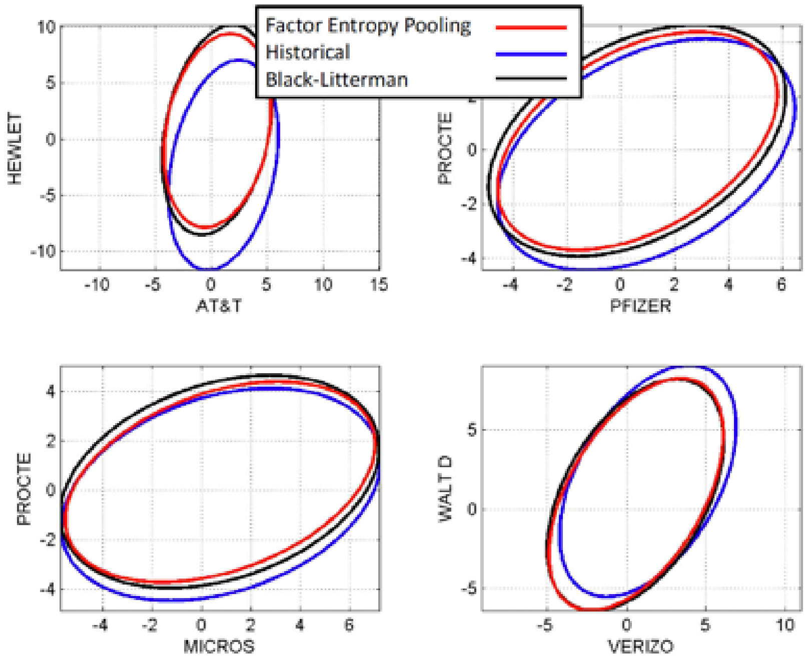
$$\mathcal{V}: \qquad \boldsymbol{\mu} - \gamma(\boldsymbol{b}\boldsymbol{b}' + Diag(\boldsymbol{d}\circ\boldsymbol{d}))\boldsymbol{w}_{eq} = \boldsymbol{0}_{\bar{n}\times 1}. \tag{27}$$

There are many combinations (μ, b, d) that satisfy those views, but only one is the closest to the data, as represented by the historical mean and the historical covariance. The Factor Entropy Pooling posterior (26) determines that solution. Then we obtain the Factor Entropy Pooling generalised equilibrium parameters

$$\bar{\boldsymbol{\mu}}, \quad \bar{\boldsymbol{\sigma}}^2 \equiv \bar{\boldsymbol{b}}\bar{\boldsymbol{b}}' + Diag(\bar{\boldsymbol{d}}\circ\bar{\boldsymbol{d}}). \tag{28}$$

The Factor Entropy Pooling equilibrium estimates (28) improve on the previous approaches in three directions. First, Factor Entropy Pooling replaces the somewhat arbitrary Euclidean distance between historical and equilibrium estimates with relative entropy, a statistically sound measure of discrepancy between distributions. Second, Factor Entropy Pooling simultaneously adjusts not only expectations and variances, but also correlations. Third, the parsimonious "low-rank-diagonal" specification (13) improves the statistical efficiency of the estimates. As a result, the equilibrium the Factor Entropy Pooling parameters $(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\sigma}}^2)$ are potentially less noisy.

Figure 5: Historical means and covariances (blue) for various pairs of stocks versus respective implied expected returns and covariances: Black-Litterman (black) and Factor Entropy Pooling (red).



Example 3 To illustrate the Factor Entropy Pooling equilibrium (28) in practice, we consider a market of $\bar{n}$ = 30 equities in the Dow Jones Index. For those equities, we consider weekly prices from January 2002 to June 2012. We fit compute the historical mean $\hat{\boldsymbol{\mu}}$ and the historical covariance $\hat{\boldsymbol{\sigma}}^2$ of the weekly returns. The market capitalisation weights w_{eq} is taken as of June

2 - For simplicity, we drop an additional parameter that enforces beta-neutrality. That parameter, along with other constraints mentioned by the authors, can easily be encompassed in our approach.

27th, 2012. In Figure 4 we report the weights w_{eq}, the historical means $\hat{\boldsymbol{\mu}}$, the implied expected returns a-la Black-Litterman $\boldsymbol{\mu}^{BL}$, as in (23), and the Factor Entropy Pooling equilibrium expected returns $\bar{\boldsymbol{\mu}}$, as in (28), computed with $\bar{k}$ = 3 hidden factors. As expected, the Factor Entropy Pooling parameters are more in line with the historical parameters than the Black-Litterman parameters:

$$\mathcal{E}(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\sigma}}^2 \| \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\sigma}}^2) = 1.83, \qquad \mathcal{E}(\boldsymbol{\mu}^{BL}, (\boldsymbol{\sigma}^2)^{BL} \| \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\sigma}}^2) = 2.41 \tag{29}$$

To illustrate the effect of Factor Entropy Pooling on correlations, in Figure 5, we display the location-dispersion ellipsoids that represent geometrically expectations and covariances, see [Meucci, 2005]. We display the ellipsoids for the historical distribution ($\hat{\boldsymbol{\mu}}$, $\hat{\boldsymbol{\sigma}}^2$), the Black-Litterman equilibrium ($\boldsymbol{\mu}^{BL}$, $\boldsymbol{\sigma}^2)^{BL}$) and the Factor Entropy Pooling equilibrium($\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\sigma}}^2$) of a few stock pairs. In Black-Litterman ellipsoid, the shape and orientation (covariance) are the same as in the historical ellipsoid, whereas the centre (expectation) is shifted. On the other hand, with Factor Entropy Pooling the centre, the dispersion and orientation of the ellipsoid are all modified. However, such modification are minimal by construction and thus the Factor Entropy Pooling parameters are more in line with the historical parameters than the Black-Litterman parameters, as highlighted in (29). This also illustrates that a constant correlation structure as in [Levy and Roll, 2010] is restrictive. Please refer to the code available at symmys.com/node/160 for more details.

5. Ranking views

In this section we use Factor Entropy Pooling to build enhanced systematic strategies, optimally processing ranking (inequality) trading signals.

The most standard approach to this problem, popularised by, among others, [De Bondt and Thaler, 1985], [Grinold and Kahn, 1999], [Park, 2010], [Wang and Kochard, 2011], [Moskowitz et al., 2012], [Asness et al., 2013a], [Asness et al., 2013b], [Menchero et al., 2013], proceeds by backtesting signals, as follows.

Step 1. At each generic time *t*, we focus on an observable characteristic of a set of $\bar{n}$ assets, which is deemed to have predictive power, say for instance, for stocks, a momentum/reversal indicator, or a value indicator such as the price/earnings ratio. Then we sort the $\bar{n}$ assets according to the value of the given characteristic. In our example, the stock n = 1 has the lowest momentum, the stock n = 2 has the second-lowest momentum, and so on, until the stock n = n has the highest momentum. The rationale of this step is that, if the signal is truly predictive, a lower ranking should give rise to a lower Sharpe ratio

$$\frac{\mu_{n,t}}{\sigma_{n,t}} \leq \frac{\mu_{n+1,t}}{\sigma_{n+1,t}} - q, \quad n = 1, \ldots, \bar{n} - 1, \tag{30}$$

where $\mu_{n,t} \equiv \mathbb{E}\{R_{n,t\to t+1} | \boldsymbol{i}_t\}$ and $\sigma_{n,t} \equiv \mathbb{S}d\{R_{n,t\to t+1} | \boldsymbol{i}_t\}$ denote respectively the expected value and the standard deviation of the next-period return, conditioned on the multidimensional information $\boldsymbol{i}_t$, available at time *t*; and $q \geq 0$ is a buffer that induces stronger inequalities.

Example 4 To illustrate our approach, we backtest a standard reversal strategy in the equity market. More precisely, in Step 1 above we construct the signal by a stylised version of the above references, as follows. For each stock *n*, at the current time *t*, we define as "momentum" the quotient of a short term momentum and a long term standard deviation estimated by exponentially weighted moving average:

$$mom^{\lambda,\gamma}_{n,t} \equiv \frac{\sum_{s\geq 0} e^{-\lambda s} r_{n,t-s}}{\sum_{s\geq 0} e^{-\lambda s}} / \sqrt{\frac{\sum_{s\geq 0} e^{-\gamma s} r^2_{n,t-s}}{\sum_{s\geq 0} e^{-\gamma s}}}. \tag{31}$$

In the above expression, typical values for the short-term decay coefficient λ correspond to a half-life of the order of a few days to a few weeks and typical values for the long-term decay coefficient γ correspond to a half-life of the order of a few weeks to a few months. Then, we reorder the stocks in such a way that $-mom_{1,t}^{\lambda} \leq \cdots \leq -mom_{\bar{n},t}^{\lambda}$, where the minus sign is set to implement a "reversal" strategy (plus sign for «momentum» strategy). The new ordering of stocks n = 1,..., $\bar{n}$ implies the signal (30).

Step 2. We estimate the next-period standard deviations $\hat{\boldsymbol{\sigma}}_{n,t}$ of the assets returns, say using exponentially weighted moving average or GARCH as in [Moskowitz et al., 2012], if it applies.

Step 3. We estimate the next-period correlations of the assets $\hat{c}_{m,n,t}$ with standard techniques. Jointly with the standard deviations $\hat{\boldsymbol{\sigma}}_{n,t}$, the correlations yield the estimated covariance matrix. More precisely, defining $\hat{\boldsymbol{\sigma}}_{vec,t} \equiv (\hat{\sigma}_{1,t}, \ldots, \hat{\sigma}_{\bar{n},t})$ and organising the correlations $\hat{c}_{m,n,t}$ in a $\bar{n}$ x $\bar{n}$ matrix $\hat{\boldsymbol{c}}_t$, we obtain the covariance matrix as follows

$$\hat{\boldsymbol{\sigma}}_t^2 \equiv Diag(\hat{\boldsymbol{\sigma}}_{vec,t})\, \hat{\boldsymbol{c}}_t\, Diag(\hat{\boldsymbol{\sigma}}_{vec,t}) \tag{32}$$

Step 4. We update the estimate of the expected returns of the assets, assuming that they are proportional to their relative ranking and to the volatility

$$\tilde{\mu}_{n,t} \equiv \eta \hat{\sigma}_{n,t}(n - \frac{\bar{n}+1}{2}), \quad n = 1, \ldots, \bar{n}, \tag{33}$$

where the constant η is the information content of the characteristic we are using, such as momentum.

Step 5. We construct an optimal portfolio, based on the covariances $\hat{\boldsymbol{\sigma}}_t^2$ and the expected returns $\tilde{\boldsymbol{\mu}}_t$. To construct the portfolio, we compute the maximum-expected-return long-short portfolio with constant target volatility, and we impose constraints on each position to arrive at a well balanced portfolio which is not too concentrated in a single position. To do so, we simplify [Lobo et al., 2007], replacing the $\bar{n}$ decision variables, namely the long-short weights ω, with four sets of positive variables (ω^+, ω^-, $\delta\omega^+$, $\delta\omega^-$), each of dimension $\bar{n}$, as follows: $\omega^+ \equiv \max(\omega, \mathbf{0})$ represents the positive part of the weights and $\omega^- \equiv \max(-\omega, \mathbf{0})$ its negative part; $\delta\omega^+ \equiv \max(\omega - \omega_{t-1}, \mathbf{0})$ represents the positive part of the transactions and $\delta\omega^- \equiv \max(\omega_{t-1} - \omega, \mathbf{0})$ its negative part, where ω_{t-1} is the legacy portfolio from the previous period. Then the weights read $\omega = \omega^+ - \omega^-$ and the absolute value of the transactions read $|\omega - \omega_{t-1}| = \delta\omega^+ + \delta\omega^-$.

Let us denote $\boldsymbol{t}$ the vector of the transaction costs and by σ_* the upper limit on risk $\boldsymbol{w}'\hat{\boldsymbol{\sigma}}_t^2\boldsymbol{w} \leq \sigma_*^2$. Then the portfolio optimisation can be expressed as the maximisation of a linear target

$$(\boldsymbol{w}_t^{\pm}, \delta\boldsymbol{w}_t^{\pm}) \equiv \underset{(\boldsymbol{w}^{\pm}, \delta\boldsymbol{w}^{\pm}) \in C}{\operatorname{argmax}} \, (\underbrace{\tilde{\boldsymbol{\mu}}_t'(\boldsymbol{w}^+ - \boldsymbol{w}^-)}_{\text{exp. return}} - \underbrace{\boldsymbol{t}'(\delta\boldsymbol{w}^+ + \delta\boldsymbol{w}^-)}_{\text{trans. cost}}), \tag{34}$$

under second-order conic constraints

$$C : \begin{cases} \boldsymbol{w}^{\pm}, \delta\boldsymbol{w}^{\pm} \geq \mathbf{0} & \text{positive parts} \\ \mathbf{1}'\boldsymbol{w}^+ \equiv \mathbf{1}'\boldsymbol{w}^- & \text{long} = \text{short} \\ \|\hat{\boldsymbol{\sigma}}_t \boldsymbol{w}\| \leq \sigma_* & \text{target risk} \\ \max|\boldsymbol{w}| \leq \bar{w} & \text{concentration} \end{cases} \tag{35}$$

where $\|\boldsymbol{v}\| \equiv \sqrt{\boldsymbol{v}'\boldsymbol{v}}$ is the standard Euclidean norm, and $\hat{\boldsymbol{\sigma}}_t$ is the Riccati root of $\hat{\boldsymbol{\sigma}}_t^2$, see [Meucci, 2009]. Then the optimal portfolio weights then read $\boldsymbol{w}_t \equiv \boldsymbol{w}_t^+ - \boldsymbol{w}_t^-$.

Example 5 To illustrate the standard backtesting approach outlined in Steps 1-5 above, we consider the same market as in Example 3, namely $\bar{n}$ = 30 equities in the Dow Jones Index, with data from January 2002 to June 2012. The backtest starts in January 2006 and portfolios are constructed every Wednesday for a total of 338 rebalancing dates. We estimate the historical means $\hat{\boldsymbol{\mu}}_t$ rolling on one year of data, and similarly the historical correlations and standard deviations, building the one-year rolling historical covariance matrix $\hat{\boldsymbol{\sigma}}_t^2$. Next, we replace the historical means $\hat{\boldsymbol{\mu}}_t$ with the signal means $\widetilde{\boldsymbol{\mu}}_t$ computed via the proportional assumption (33). Then, we build the backtest (34), where we set the transaction costs t as 5 basis points of the market value, and where we set the volatility target in (35) such that the dollar volatility is bounded at 100,000$.

Figure 1 displays the cumulative P&L ensuing from this standard backtest. For more details, refer to the code available at symmys.com/node/160. We stress that in general Factor Entropy Pooling does not guarantee to achieve a larger (risk adjusted) P&L, but rather it is a flexible generalisation of standard approaches to portfolio construction, which allows for a wider spectrum of results.

The most sensitive part in the above process is the proportional assumption (33) which quantifies the ranking signal (30). The proportional assumption presents two problems. First, it makes a much bolder statement on the expected return than the actual signal implies. In other words, the proportional assumption (33) corresponds to ex-ante Sharpe ratios $\tilde{\mu}_n/\hat{\sigma}_n = \eta(n - \frac{\bar{n}+1}{2})$ which never change through time. Such Sharpe ratios satisfy the signal (30).

The second problem is that the proportional assumption (33) does not change the volatilities in (32), whereas the trading signal inequalities (30) also involves the volatilities. [Almgren and Chriss, 2006] provide an alternative approach to process inequality views. The authors set the vector of expected returns as the "centroid", i.e. the average among all possible expected returns consistent with the ranking (30). However, the centroid presents the same problems as the standard approach. First, it does not depend on the observed empirical data: two completely different sets of securities with the same relative rankings give rise to the same expected returns. Second, the centroid approach does not alter the volatilities. Factor Entropy Pooling addresses both issues, as follows.

Step 4. The ranking signal that generates portfolios from sorts (30) is clearly a view in the constraint format (19), where the constraint function reads

$$\mathcal{V}: \qquad v_n(\boldsymbol{\mu}_t, \boldsymbol{b}_t, \boldsymbol{d}_t) \equiv \frac{\mu_{n,t}}{\sigma_{n,t}} - \frac{\mu_{n+1,t}}{\sigma_{n+1,t}} + q \leq 0, \quad n = 1, \ldots, \bar{n} - 1, \tag{36}$$

where $\sigma_{n,t} \equiv ([\boldsymbol{b}_t \boldsymbol{b}_t' + Diag(\boldsymbol{d}_t \circ \boldsymbol{d}_t)]_{n,n})^{1/2}$.

Hence, we just impose the constraints (36) in the Factor Entropy Pooling optimisation (14) process. Among all the distributions that satisfy the signal inequalities, Factor Entropy Pooling chooses the one which is the closest to the data, as represented by the estimated covariances $\hat{\boldsymbol{\sigma}}_t^2$ and expected returns $\hat{\boldsymbol{\mu}}_t$

$$(\bar{\boldsymbol{\mu}}_t, \bar{\boldsymbol{b}}_t, \bar{\boldsymbol{d}}_t) \equiv \underset{(\boldsymbol{\mu}, \boldsymbol{b}, \boldsymbol{d}) \in \mathcal{V}}{\operatorname{argmin}} \ \mathcal{E}(\boldsymbol{\mu}, \boldsymbol{b}\boldsymbol{b}' + Diag(\boldsymbol{d} \circ \boldsymbol{d}) \| \hat{\boldsymbol{\mu}}_t, \hat{\boldsymbol{\sigma}}_t^2). \tag{37}$$

Notice that this step can be further improved by replacing $\hat{\boldsymbol{\mu}}_t$ and $\hat{\boldsymbol{\sigma}}_t^2$ with the generalised equilibrium parameters (28).

To speed up further the optimisation, we compute in Appendix A.8 the gradient of the views (36), which reads

$$\nabla_{\boldsymbol{\mu}} v = Diag(1./\boldsymbol{\sigma}_{vec})\boldsymbol{a}' \tag{38}$$

$$\nabla_{\boldsymbol{d}} v = -Diag((\boldsymbol{\mu} \circ \boldsymbol{d})./\boldsymbol{\sigma}_{vec}^3)\boldsymbol{a}' \tag{39}$$

$$\nabla_{\boldsymbol{b}} v = -(\boldsymbol{b}' \otimes \boldsymbol{i}_{\bar{n} \times \bar{n}}) Diag(vec(\boldsymbol{i}_{\bar{n} \times \bar{n}}))(\boldsymbol{a}' \otimes (\boldsymbol{\mu}./\boldsymbol{\sigma}_{vec}^3)), \tag{40}$$

where ./ is the entry-by-entry division between vectors, $\boldsymbol{\sigma}_{vec} \equiv \sqrt{diag(\boldsymbol{\sigma}^2)}$, ⊗ is the tensor product, and **a** is an $(\bar{n} - 1) \times \bar{n}$ defined as follows

$$a_{m,n} \equiv 1_{m=n} - 1_{m=n-1} \tag{41}$$

In Appendix A.9 we also provide the Hessian of the views (36).
Then the Factor Entropy Pooling covariances are reconstructed from the Factor Entropy Pooling parameters as follows

$$\bar{\boldsymbol{\sigma}}_t^2 \equiv \bar{\boldsymbol{b}}_t \bar{\boldsymbol{b}}_t' + Diag(\bar{\boldsymbol{d}}_t \circ \bar{\boldsymbol{d}}_t). \tag{42}$$

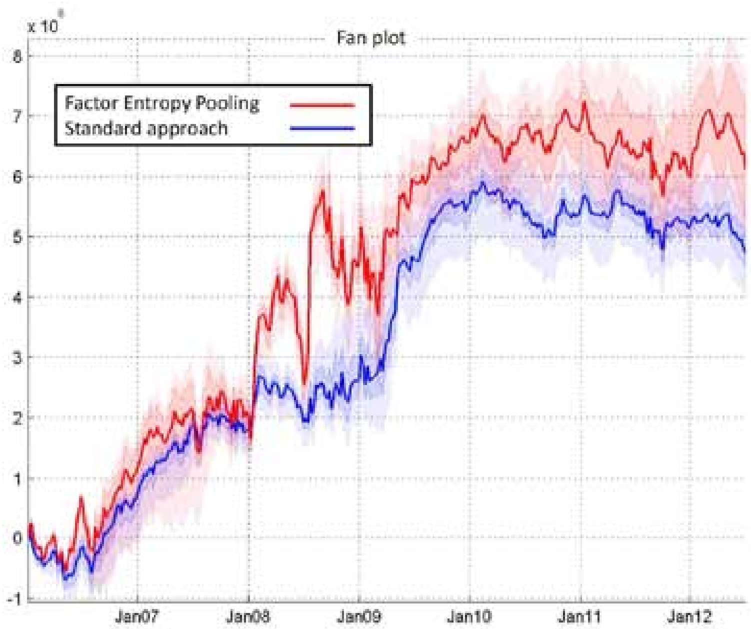
We can then proceed with step 5 above, constructing the optimal portfolio, based on the Factor Entropy Pooling expected returns $\bar{\boldsymbol{\mu}}_t$ and the Factor Entropy Pooling covariances $\hat{\boldsymbol{\sigma}}_t^2$.

Unlike the ex-ante Sharpe ratios $\tilde{\mu}_n/\hat{\sigma}_n = \eta(n - \frac{\bar{n}+1}{2})$ ensuing from the common approach (33), or the ex-ante Sharpe ratios in the centroid approach, the ex-ante Sharpe ratios $\bar{\mu}_n/\bar{\sigma}_n$ stemming from the Factor Entropy Pooling posterior (37)-(42) satisfy the ranking views (30) and at the same time change with the empirical data ($\hat{\boldsymbol{\mu}}$, $\hat{\boldsymbol{\sigma}}^2$).the parametric approach discussed above must match the analytical results obtained in (11)-(12).

Example 6 Continuing with the same framework as Example 5, we use $\bar{k}$ = 1 hidden factor in the Factor Entropy Pooling minimisation (37), in order to provide maximum shrinkage and maximum backtesting speed. Furthermore, we set the inequality buffer in (36) as $q = 1=(\bar{n} - 1)$. Then we set the decay parameters (λ,γ) in the signal construction (31), so that the half-lives are 8 and 52 weeks respectively. In Figure 1 we display the result.

For fairness, we performed the backtest with di¤erent values for the decay parameters (λ,γ): we set λ to span a half-life from 2 to 14 weeks, with a step size of one week, and we set γ to span a half-life from 40 to 60 weeks, with a step size of one week, for a total of 169 con gurations. The number of hidden factors $\bar{k}$ and the inequality buffer q provide additional parameters over which to tweak the Sharpe ratio in the backtest. In Figure 6 we report the percentiles of the ensuing backtest P&L. We plot both the outcomes of the Factor Entropy Pooling approach (discussed in this example) and of the standard approach (discussed in Example 5). For more details, refer to the code available at symmys.com/node/160.

Figure 6: Cumulative P&L generated by the reversal strategy backtest for various parametrisations. The plot reports the median (solid line), the 50% percentile range (dim shading) and the 90% percentile range (dimmer shading).



6. Conclusions

We have introduced the Factor Entropy Pooling technique, an efficient algorithm for the parametric implementation of Entropy Pooling. First, we used Factor Entropy Pooling to calibrate the implied returns, which can then be used for portfolio construction in the absence of market views. We illustrated its usefulness with two applications. Second, we used Factor Entropy Pooling to build and backtest a systematic strategy based on ranking trade signals.

An additional area of application of Factor Entropy Pooling other than portfolio construction is heavy stress-testing, whereby we subject the market to disruptive potential scenarios and observe their effect on the portfolio losses. However, heavy stress-testing does not require to optimise a portfolio based on the Entropy Pooling posterior, and thus computational speed is typically not as relevant. We explore a more computationally intensive evolution of Factor Entropy Pooling particularly suitable for heavy stress-testing in the companion paper [Ardia and Meucci, 2013].

A Appendix

In this appendix we discuss technical results that can be skipped at first reading.

A.1 Ranking via Factor Entropy Pooling

We consider the ranking views (30) together with a stronger specification which guarantees that the Sharpe ratios are bounded in [−1; 1]:

$$\mathcal{V}: \quad \begin{cases} \dfrac{\mu_n}{\sigma_n} \leq \dfrac{\mu_{n+1}}{\sigma_{n+1}} - q, \quad n = 1, \ldots, \bar{n}-1, \quad q \geq 0 \\ \qquad \dfrac{\mu_1}{\sigma_1} = -1, \quad \dfrac{\mu_{\bar{n}}}{\sigma_{\bar{n}}} = 1 \end{cases} . \tag{43}$$

First we notice that the conditions in (43) easily implies that q must belong to a bounded range. Indeed from (43) we have

$$\begin{aligned} (\bar{n}-1)q &= \textstyle\sum_{n=1}^{\bar{n}-1} q \\ &\leq \textstyle\sum_{n=1}^{\bar{n}-1} \dfrac{\mu_{n+1}}{\sigma_{n+1}} - \dfrac{\mu_n}{\sigma_n} \\ &= \dfrac{\mu_{\bar{n}}}{\sigma_{\bar{n}}} - \dfrac{\mu_1}{\sigma_1} \\ &= 1 - (-1) = 2. \end{aligned} \tag{44}$$

and so $q \in [0, \frac{2}{\bar{n}-1}]$. Also, setting $q = \frac{2}{\bar{n}-1}$, if there exist a vector of expected returns μ and a matrix of covariances σ^2 which satisfies (43), then

$$\frac{\mu_n}{\sigma_n} = \frac{\mu_{n+1}}{\sigma_{n+1}} - \frac{2}{\bar{n}-1}, \quad n = 1, \ldots, \bar{n}-1, \tag{45}$$

otherwise we would have

$$2 = \textstyle\sum_{n=1}^{\bar{n}-1} \dfrac{\mu_{n+1}}{\sigma_{n+1}} - \dfrac{\mu_n}{\sigma_n} > (\bar{n}-1)\dfrac{2}{\bar{n}-1} = 2, \tag{46}$$

which is impossible. Then from (45) we obtain

$$\begin{aligned} \frac{\mu_{n+1}}{\sigma_{n+1}} &= \frac{\mu_n}{\sigma_n} + \frac{2}{\bar{n}-1} = \frac{\mu_{n-1}}{\sigma_{n-1}} + 2\frac{2}{\bar{n}-1} \\ &= \ldots = \frac{\mu_1}{\sigma_1} + n\frac{2}{\bar{n}-1} = -1 + n\frac{2}{\bar{n}-1} \\ &= \frac{2}{\bar{n}-1}(n - \frac{\bar{n}-1}{2}) \end{aligned}$$

or equivalently

$$\frac{\mu_n}{\sigma_n} \equiv \frac{2}{\bar{n}-1}(n - \frac{\bar{n}+1}{2}), \tag{47}$$

which are the *ex-ante* Sharpe ratios (33) for $\eta \equiv \frac{2}{\bar{n}-1}$

A.2 Notational conventions

Let $g(\boldsymbol{x}) = (g_1(\boldsymbol{x}),\ldots, g_{\bar{m}}(\boldsymbol{x}))'$ be a multivariate function of a $\bar{n}$ x $\bar{k}$ matrix $\boldsymbol{x}$.

We arrange the gradient $\nabla_{\boldsymbol{x}} g(\boldsymbol{x})$ into an $\bar{n}\bar{k}$ x $\bar{m}$ block matrix as follows

$$\nabla_{\boldsymbol{x}} g \equiv (\nabla_{\boldsymbol{x}} g_1 | \ldots | \nabla_{\boldsymbol{x}} g_{\bar{m}}) \tag{48}$$

where $\nabla_{\boldsymbol{x}} g_m$ denotes an $\bar{n}\bar{k}$ x1 vector defined as follows

$$\nabla_{\boldsymbol{x}} g_m \equiv \begin{pmatrix} \nabla_{\boldsymbol{x}_{\cdot,1}} g_m \\ \vdots \\ \nabla_{\boldsymbol{x}_{\cdot,\bar{k}}} g_m \end{pmatrix}, \tag{49}$$

and where $\nabla_{\boldsymbol{x}_{\cdot,k}} g_m$ is an $\bar{n}$ x 1 vector with entries as follows

$$[\nabla_{\boldsymbol{x}_{\cdot,k}} g_m]_n \equiv \frac{\partial g_m}{\partial x_{n,k}}. \tag{50}$$

Using notation (49) we can write the variation due to $\boldsymbol{x}$ in compact form as follows

$$dg_m(\boldsymbol{x}) \equiv \textstyle\sum_{n,k} \frac{\partial g_m(\boldsymbol{x})}{\partial x_{n,k}} dx_{n,k} = (\nabla_{\boldsymbol{x}} g_m(\boldsymbol{x}))' vec(d\boldsymbol{x}). \tag{51}$$

Indeed

$$\begin{aligned} \textstyle\sum_{n,k} \frac{\partial g_m}{\partial x_{n,k}} dx_{n,k} &= \textstyle\sum_k (\nabla_{\boldsymbol{x}_{\cdot,k}} g_m)' d\boldsymbol{x}_{\cdot,k} \\ &= (\nabla_{\boldsymbol{x}} g_m(\boldsymbol{x}))' vec(d\boldsymbol{x}). \end{aligned} \tag{52}$$

Let $g(\boldsymbol{x},\boldsymbol{y}) = (g_1(\boldsymbol{x},\boldsymbol{y}),\ldots, g_{\bar{m}}(\boldsymbol{x},\boldsymbol{y}))'$ be a multivariate function of an $\bar{n}$ x $\bar{k}$ matrix $\boldsymbol{x}$ and a $\bar{r}$ x $\bar{s}$ matrix $\boldsymbol{y}$.

We arrange the Hessian $\nabla^2_{\boldsymbol{x},\boldsymbol{y}} g$ into an $\bar{n}\bar{k} \times \bar{r}\bar{s}\bar{m}$ block matrix as follows

$$\nabla^2_{\boldsymbol{x},\boldsymbol{y}} g \equiv (\nabla^2_{\boldsymbol{x},\boldsymbol{y}} g_1 | \ldots | \nabla^2_{\boldsymbol{x},\boldsymbol{y}} g_{\bar{m}}), \tag{53}$$

where $\nabla^2_{\boldsymbol{x},\boldsymbol{y}} g_m$ denotes an $\bar{n}\bar{k}$ x $\bar{r}\bar{s}$ block matrix as follows

$$\nabla^2_{\boldsymbol{x},\boldsymbol{y}} g_m \equiv \begin{pmatrix} \nabla^2_{\boldsymbol{x}_{\cdot,1},\boldsymbol{y}_{\cdot,1}} g_m & \nabla^2_{\boldsymbol{x}_{\cdot,1},\boldsymbol{y}_{\cdot,2}} g_m & \cdots & \nabla^2_{\boldsymbol{x}_{\cdot,1},\boldsymbol{y}_{\cdot,\bar{s}}} g_m \\ \vdots & \vdots & & \vdots \\ \nabla^2_{\boldsymbol{x}_{\cdot,\bar{k}},\boldsymbol{y}_{\cdot,1}} g_m & \nabla^2_{\boldsymbol{x}_{\cdot,\bar{k}},\boldsymbol{y}_{\cdot,2}} g_m & \cdots & \nabla^2_{\boldsymbol{x}_{\cdot,\bar{k}},\boldsymbol{y}_{\cdot,\bar{s}}} g_m \end{pmatrix}, \tag{54}$$

and where $\nabla^2_{\boldsymbol{x}_{\cdot,k},\boldsymbol{y}_{\cdot,s}} g_m$ is an $\bar{n}$ x $\bar{r}$ matrix with entries as follows

$$[\nabla^2_{\boldsymbol{x}_{\cdot,k},\boldsymbol{y}_{\cdot,s}} g_m]_{n,r} \equiv \frac{\partial g_m}{\partial x_{n,k} \partial y_{r,s}}. \tag{55}$$

Using notation (54) we have

$$\begin{aligned} &\textstyle\sum_{n,k,r,s} \frac{\partial^2 g_m(\boldsymbol{x},\boldsymbol{y})}{\partial x_{n,k} \partial y_{r,s}} dx_{n,k} dy_{r,s} \\ &= \textstyle\sum_{k,s} (d\boldsymbol{x}_{\cdot,k})' \times (\nabla^2_{\boldsymbol{x}_{\cdot,k},\boldsymbol{y}_{\cdot,s}} g_m(\boldsymbol{x},\boldsymbol{y})) \times d\boldsymbol{y}_{\cdot,s} \\ &= vec(d\boldsymbol{x})' \times \nabla^2_{\boldsymbol{x},\boldsymbol{y}} g_m(\boldsymbol{x},\boldsymbol{y}) \times vec(d\boldsymbol{y}). \end{aligned} \tag{56}$$

A.3 Gradient of relative entropy

Let us define the function

$$f(\mu, s \equiv \xi(\mu, s \mid \underline{\mu}, \underline{s}), \tag{57}$$

where μ is an $\bar{n}$-dimensional vector, $\boldsymbol{s}$ is an $\bar{n}$ x $\bar{n}$ symmetric invertible matrix, and $\xi(\mu, s \mid \underline{\mu}, \underline{s})$ is the relative entropy defined in (9).

We can write the first order differential of (57) as follows

$$\begin{aligned} df(\boldsymbol{\mu}, \boldsymbol{s}) &= \tfrac{1}{2}(d\boldsymbol{\mu}'\underline{\boldsymbol{s}}^{-1}(\boldsymbol{\mu} - \underline{\boldsymbol{\mu}}) + (\boldsymbol{\mu} - \underline{\boldsymbol{\mu}})'\underline{\boldsymbol{s}}^{-1}d\boldsymbol{\mu} \\ &+ \operatorname{tr}(\underline{\boldsymbol{s}}^{-1}d\boldsymbol{s}) - \operatorname{tr}(\boldsymbol{s}^{-1}d\boldsymbol{s})) \\ &= (\boldsymbol{\mu} - \underline{\boldsymbol{\mu}})'\underline{\boldsymbol{s}}^{-1}d\,\boldsymbol{\mu} + \tfrac{1}{2}\operatorname{tr}((\underline{\boldsymbol{s}}^{-1} - \boldsymbol{s}^{-1})d\boldsymbol{s}) \\ &= (\boldsymbol{\mu} - \underline{\boldsymbol{\mu}})'\underline{\boldsymbol{s}}^{-1}d\,\boldsymbol{\mu} + \tfrac{1}{2}vec(\underline{\boldsymbol{s}}^{-1} - \boldsymbol{s}^{-1})'vec(d\boldsymbol{s}). \end{aligned} \tag{58}$$

On the other hand, the general expression of the differential reads

$$\begin{aligned} df(\boldsymbol{\mu}, \boldsymbol{s}) &= \textstyle\sum_k \frac{\partial f(\boldsymbol{\mu}, \boldsymbol{s})}{\partial \mu_k} d\mu_k + \sum_{n,m} \frac{\partial f(\boldsymbol{\mu}, \boldsymbol{s})}{\partial s_{n,m}} ds_{n,m} \\ &\overset{(51)}{=} (\nabla_{\boldsymbol{\mu}} f)' vec(d\boldsymbol{\mu}) + (\nabla_{\boldsymbol{s}} f)' vec(d\boldsymbol{s}) \\ &= (\nabla_{\boldsymbol{\mu}} f)' d\boldsymbol{\mu} + (\nabla_{\boldsymbol{s}} f)' vec(d\boldsymbol{s}). \end{aligned} \tag{59}$$

Hence, comparing (58) with (59), we obtain

$$\nabla_{\boldsymbol{\mu}} f = \underline{\boldsymbol{s}}^{-1}(\boldsymbol{\mu} - \underline{\boldsymbol{\mu}}), \tag{60}$$

$$\nabla_{\boldsymbol{s}} f = \tfrac{1}{2} vec(\underline{\boldsymbol{s}}^{-1} - \boldsymbol{s}^{-1}). \tag{61}$$

A.4 Hessian of relative entropy

Let us consider the same setup in Appendix A.3.
The second order differential of (57) reads

$$\begin{aligned} d^2 f(\boldsymbol{\mu}, \boldsymbol{s}) &= d(df(\boldsymbol{\mu}, \boldsymbol{s})) \\ &= d((\boldsymbol{\mu} - \underline{\boldsymbol{\mu}})'\underline{\boldsymbol{s}}^{-1}d\boldsymbol{\mu}) + \tfrac{1}{2} d\operatorname{tr}((\underline{\boldsymbol{s}}^{-1} - \boldsymbol{s}^{-1})d\boldsymbol{s}) \\ &= d\boldsymbol{\mu}' \times \underline{\boldsymbol{s}}^{-1} \times d\,\boldsymbol{\mu} + \tfrac{1}{2}\operatorname{tr}((\boldsymbol{s}^{-1}(d\boldsymbol{s})\boldsymbol{s}^{-1})d\boldsymbol{s}) \\ &= d\boldsymbol{\mu}' \times \underline{\boldsymbol{s}}^{-1} \times d\,\boldsymbol{\mu} + \tfrac{1}{2}\operatorname{tr}(d\boldsymbol{s}(\boldsymbol{s}^{-1}(d\boldsymbol{s})\boldsymbol{s}^{-1})) \\ &= d\boldsymbol{\mu}' \times \underline{\boldsymbol{s}}^{-1} \times d\,\boldsymbol{\mu} + \tfrac{1}{2}\operatorname{tr}((d\boldsymbol{s})'(\boldsymbol{s}^{-1}(d\boldsymbol{s})\boldsymbol{s}^{-1})) \\ &= d\boldsymbol{\mu}' \times \underline{\boldsymbol{s}}^{-1} \times d\,\boldsymbol{\mu} + \tfrac{1}{2} vec(d\boldsymbol{s})' vec(\boldsymbol{s}^{-1}(d\boldsymbol{s})\boldsymbol{s}^{-1}) \\ &= d\boldsymbol{\mu}' \times \underline{\boldsymbol{s}}^{-1} \times d\boldsymbol{\mu} + \tfrac{1}{2} vec(d\boldsymbol{s})' \times (\boldsymbol{s}^{-1} \otimes \boldsymbol{s}^{-1}) \times vec(d\boldsymbol{s}). \end{aligned} \tag{62}$$

On the other hand, differentiating the general expression of the differential (59), we have

$$\begin{aligned} d^2 f(\boldsymbol{\mu}, \boldsymbol{s}) &= d(df(\boldsymbol{\mu}, \boldsymbol{s})) \\ &= d((\nabla_{\boldsymbol{\mu}} f)' d\boldsymbol{\mu}) + d\operatorname{tr}((\nabla_{\boldsymbol{s}} f)' d\boldsymbol{s})) \\ &= d\textstyle\sum_k [\nabla_{\boldsymbol{\mu}} f]_k d[\boldsymbol{\mu}]_k + d\sum_{n,m} [\nabla_{\boldsymbol{s}} f]_{n,m} d[\boldsymbol{s}]_{n,m} \\ &= \textstyle\sum_k (\sum_h \frac{\partial^2 f(\boldsymbol{\mu}, \boldsymbol{s})}{\partial \mu_h \partial \mu_k} d\mu_h d\mu_k + \sum_{l,r} \frac{\partial^2 f(\boldsymbol{\mu}, \boldsymbol{s})}{\partial s_{l,r} \partial \mu_k} ds_{l,r} d\mu_k) \\ &+ \textstyle\sum_{n,m} (\sum_h \frac{\partial^2 f(\boldsymbol{\mu}, \boldsymbol{s})}{\partial \mu_h \partial s_{n,m}} d\mu_h ds_{n,m} + \sum_{l,r} \frac{\partial^2 f(\boldsymbol{\mu}, \boldsymbol{s})}{\partial s_{l,r} \partial s_{n,m}} ds_{l,r} ds_{n,m}) \\ &\overset{(56)}{=} d\boldsymbol{\mu}' \times \nabla^2_{\boldsymbol{\mu},\boldsymbol{\mu}} f \times d\boldsymbol{\mu} + \textstyle\sum_r (d\boldsymbol{s}_{\cdot,r})' \times \nabla^2_{\boldsymbol{s}_{\cdot,r},\boldsymbol{\mu}} f \times d\boldsymbol{\mu} \\ &+ \textstyle\sum_m d\boldsymbol{\mu}' \times \nabla^2_{\boldsymbol{\mu},\boldsymbol{s}_{\cdot,m}} f \times d\boldsymbol{s}_{\cdot,m} \\ &+ \textstyle\sum_{r,m} (d\boldsymbol{s}_{\cdot,r})' \times \nabla^2_{\boldsymbol{s}_{\cdot,r},\boldsymbol{s}_{\cdot,m}} f \times d\boldsymbol{s}_{\cdot,m} \\ &\overset{(56)}{=} d\boldsymbol{\mu}' \times \nabla^2_{\boldsymbol{\mu},\boldsymbol{\mu}} f \times d\boldsymbol{\mu} + vec(d\boldsymbol{s})' \times \nabla^2_{\boldsymbol{s},\boldsymbol{\mu}} f \times d\boldsymbol{\mu} \\ &+ d\boldsymbol{\mu}' \times \nabla^2_{\boldsymbol{\mu},\boldsymbol{s}} f \times vec(d\boldsymbol{s}) + vec(d\boldsymbol{s})' \times \nabla^2_{\boldsymbol{s},\boldsymbol{s}} f \times vec(d\boldsymbol{s}). \end{aligned} \tag{63}$$

Hence, comparing (62) with (63) we obtain

$$\nabla^2_{\boldsymbol{\mu},\boldsymbol{\mu}} f = \underline{\boldsymbol{s}}^{-1}, \tag{64}$$

$$\nabla^2_{\boldsymbol{s},\boldsymbol{s}} f = \tfrac{1}{2}(\boldsymbol{s}^{-1} \otimes \boldsymbol{s}^{-1}), \tag{65}$$

$$\nabla^2_{\boldsymbol{s},\boldsymbol{\mu}} f = (\nabla^2_{\boldsymbol{\mu},\boldsymbol{s}} f)' = \mathbf{0}_{\bar{n}^2\times\bar{n}}. \tag{66}$$

A.5 Gradient of optimisation target

We consider the function

$$e\,(\boldsymbol{\mu}, \boldsymbol{b}, \boldsymbol{d}) \equiv \mathcal{E}(\boldsymbol{\mu}, \boldsymbol{b}\boldsymbol{b}' + Diag(\boldsymbol{d}\circ\boldsymbol{d}) \| \underline{\boldsymbol{\mu}}, \underline{\boldsymbol{s}}), \tag{67}$$

where $\boldsymbol{\mu}$ is an $\bar{n}$-dimensional vector, $\boldsymbol{b}$ is a $\bar{n}$ x $\bar{k}$ matrix, $\boldsymbol{d}$ is an $\bar{n}$-dimensional vector and $\mathcal{E}(\boldsymbol{\mu}, \boldsymbol{b}\boldsymbol{b}' + Diag(\boldsymbol{d}\circ\boldsymbol{d}) \| \underline{\boldsymbol{\mu}}, \underline{\boldsymbol{s}})$ is the relative entropy defined in (9).

Let us define

$$\boldsymbol{s} \equiv \boldsymbol{b}\boldsymbol{b}' + Diag(\boldsymbol{d}\circ\boldsymbol{d}). \tag{68}$$

Then we can express

$$s_{i,j} = \textstyle\sum_l b_{i,l} b_{j,l} + \sum_k d_k^2 1_{i=k} 1_{j=k}, \tag{69}$$

where 1_x is 1 if x is true and 0 otherwise.

Thus we obtain

$$\frac{\partial s_{i,j}}{\partial d_n} = 2 d_n 1_{i=n} 1_{j=n}, \tag{70}$$

and

$$\begin{aligned} \frac{\partial s_{i,j}}{\partial b_{n,k}} &= \textstyle\sum_l \left(\frac{\partial b_{i,l}}{\partial b_{n,k}} b_{j,l} + \frac{\partial b_{j,l}}{\partial b_{n,k}} b_{i,l}\right) \\ &= 1_{i=n} b_{j,k} + 1_{j=n} b_{i,k}. \end{aligned} \tag{71}$$

We can arrange (70) and (71) in a compact form as follows

$$\frac{\partial vec(\boldsymbol{s})}{\partial d_n} \equiv \begin{pmatrix} \frac{\partial \boldsymbol{s}_{\cdot,1}}{\partial d_n} \\ \vdots \\ \frac{\partial \boldsymbol{s}_{\cdot,n}}{\partial d_n} \\ \vdots \\ \frac{\partial \boldsymbol{s}_{\cdot,\bar{n}}}{\partial d_n} \end{pmatrix} = \begin{pmatrix} 0_{\bar{n}\times 1} \\ \vdots \\ 2 d_n \boldsymbol{\delta}^{(n)}_{\bar{n}\times 1} \\ \vdots \\ 0_{\bar{n}\times 1} \end{pmatrix}, \tag{72}$$

and

$$\frac{\partial vec(\boldsymbol{s})}{\partial b_{n,k}} \equiv \begin{pmatrix} \frac{\partial \boldsymbol{s}_{\cdot,1}}{\partial b_{n,k}} \\ \vdots \\ \frac{\partial \boldsymbol{s}_{\cdot,n}}{\partial b_{n,k}} \\ \vdots \\ \frac{\partial \boldsymbol{s}_{\cdot,\bar{n}}}{\partial b_{n,k}} \end{pmatrix} = \begin{pmatrix} b_{1,k} \boldsymbol{\delta}^{(n)}_{\bar{n}\times 1} \\ \vdots \\ b_{n,k} \boldsymbol{\delta}^{(n)}_{\bar{n}\times 1} \\ \vdots \\ b_{\bar{n},k} \boldsymbol{\delta}^{(n)}_{\bar{n}\times 1} \end{pmatrix} + \begin{pmatrix} 0_{\bar{n}\times 1} \\ \vdots \\ \boldsymbol{b}_{\cdot,k} \\ \vdots \\ 0_{\bar{n}\times 1} \end{pmatrix}, \tag{73}$$

where $\boldsymbol{\delta}^{(n)}_{\bar{n}\times 1}$ is an $\bar{n}$ x 1 vector with 1 in the n-th entry and 0 otherwise.

From Appendix A.3 we easily get

$$\nabla_{\boldsymbol{\mu}} e \overset{(60)}{=} \underline{\boldsymbol{s}}^{-1}(\boldsymbol{\mu} - \underline{\boldsymbol{\mu}}). \tag{74}$$

Using the chain rule, the derivatives with respect to ***d*** read

$$\frac{\partial e}{\partial d_n} = \sum_{i,j} \frac{\partial e}{\partial s_{i,j}} \frac{\partial s_{i,j}}{\partial d_n} \overset{(70)}{=} \sum_{i,j} \frac{\partial e}{\partial s_{i,j}} (2d_n 1_{i=n} 1_{j=n})$$
$$= 2\frac{\partial e}{\partial s_{n,n}} d_n \overset{(61)}{=} [(\underline{\boldsymbol{s}}^{-1} - \boldsymbol{s}^{-1})]_{n,n} [\boldsymbol{d}]_n. \tag{75}$$

Hence

$$\nabla_{\boldsymbol{d}} h = diag(\underline{\boldsymbol{s}}^{-1} - \boldsymbol{s}^{-1}) \circ \boldsymbol{d}. \tag{76}$$

Using the chain rule, the derivatives with respect to ***b*** read

$$\frac{\partial e}{\partial b_{n,k}} = \sum_{i,j} \frac{\partial e}{\partial s_{i,j}} \frac{\partial s_{i,j}}{\partial b_{n,k}} \overset{(71)}{=} \sum_{i,j} \frac{\partial e}{\partial s_{i,j}} (1_{i,n} b_{j,k} + 1_{j,n} b_{i,k})$$
$$= \tfrac{1}{2}(\sum_j \frac{\partial e}{\partial s_{n,j}} b_{j,k} + \sum_i \frac{\partial e}{\partial s_{i,n}} b_{i,k})$$
$$= \tfrac{1}{2}(\sum_j \frac{\partial e}{\partial s_{n,j}} b_{j,k} + \sum_i \frac{\partial e}{\partial s_{n,i}} b_{i,k})$$
$$\overset{(61)}{=} \sum_j [(\underline{\boldsymbol{s}}^{-1} - \boldsymbol{s}^{-1})]_{n,j} b_{j,k}$$
$$= [(\underline{\boldsymbol{s}}^{-1} - \boldsymbol{s}^{-1})\boldsymbol{b}]_{n,k} = [(\underline{\boldsymbol{s}}^{-1} - \boldsymbol{s}^{-1})\boldsymbol{b}_{\cdot,k}]_n. \tag{77}$$

Then from (50) we get

$$\nabla_{\boldsymbol{b}_{\cdot,k}} e = (\underline{\boldsymbol{s}}^{-1} - \boldsymbol{s}^{-1})\boldsymbol{b}_{\cdot,k}$$

and hence

$$\nabla_{\boldsymbol{b}} e = vec((\underline{\boldsymbol{s}}^{-1} - \boldsymbol{s}^{-1})\boldsymbol{b}). \tag{78}$$

A.6 Hessian of optimisation target

Let us consider the same setup in Appendix A.5

From Appendix A.4 we have

$$\nabla^2_{\boldsymbol{\mu},\boldsymbol{\mu}} e \overset{(64)}{=} \underline{\boldsymbol{s}}^{-1}. \tag{79}$$

Furthermore, since (74) does not depend from ***d*** and ***b*** the second mixed derivatives read

$$\nabla^2_{\boldsymbol{d},\boldsymbol{\mu}} e = \boldsymbol{0}_{\bar{n}\times\bar{n}}, \tag{80}$$
$$\nabla^2_{\boldsymbol{d},\boldsymbol{\mu}} e = \boldsymbol{0}_{\bar{n}\bar{k}\times\bar{n}}. \tag{81}$$

The derivative of (75) with respect to d_k reads

$$\frac{\partial^2 e}{\partial d_k \partial d_n} = (\frac{\partial vec(\boldsymbol{s})}{\partial d_k})' \times \nabla^2_{\boldsymbol{s},\boldsymbol{s}} e \times \frac{\partial vec(\boldsymbol{s})}{\partial d_n} + vec(\nabla_{\boldsymbol{s}} e)' \frac{\partial^2 vec(\boldsymbol{s})}{\partial d_k \partial d_n}$$
$$\overset{(72)}{=} 4d_k(\boldsymbol{\delta}^{(k)}_{\bar{n}\times 1})'(\nabla^2_{\boldsymbol{s}_{\cdot,k},\boldsymbol{s}_{\cdot,n}} e) d_n \boldsymbol{\delta}^{(n)}_{\bar{n}\times 1} + (\nabla_{\boldsymbol{s}_{\cdot,n}} e)' \frac{\partial}{\partial d_k}(2d_n \boldsymbol{\delta}^{(n)}_{\bar{n}\times 1})$$
$$\overset{(65)\text{-}(61)}{=} 2(\boldsymbol{\delta}^{(k)}_{\bar{n}\times 1})'([\boldsymbol{s}^{-1}]_{k,n}\boldsymbol{s}^{-1})\boldsymbol{\delta}^{(n)}_{\bar{n}\times 1} d_k d_n + ([\underline{\boldsymbol{s}}^{-1} - \boldsymbol{s}^{-1}]_{\cdot,n})' \boldsymbol{\delta}^{(n)}_{\bar{n}\times 1} 1_{k=n}$$
$$= 2[\boldsymbol{s}^{-1}]_{k,n}[\boldsymbol{s}^{-1}]_{k,n} d_k d_n + [\underline{\boldsymbol{s}}^{-1} - \boldsymbol{s}^{-1}]_{n,n} 1_{k=n},$$
$$= 2d_k[\boldsymbol{s}^{-1}]_{k,n}[\boldsymbol{s}^{-1}]_{k,n} d_n + 1_{k=n}[\underline{\boldsymbol{s}}^{-1} - \boldsymbol{s}^{-1}]_{k,n}, \tag{82}$$

from which we deduce

$$\nabla^2_{\boldsymbol{d},\boldsymbol{d}} e = 2(Diag(\boldsymbol{d})\boldsymbol{s}^{-1}) \circ (\boldsymbol{s}^{-1} Diag(\boldsymbol{d})) + \boldsymbol{i}_{\bar{n}\times\bar{n}} \circ (\underline{\boldsymbol{s}}^{-1} - \boldsymbol{s}^{-1}). \tag{83}$$

The derivative of (75) with respect to $b_{l,k}$ reads

$$\begin{aligned}
&\frac{\partial^2 e}{\partial b_{l,k}\partial d_n} = (\frac{\partial vec(\boldsymbol{s})}{\partial b_{l,k}})' \times \nabla^2_{\boldsymbol{s},\boldsymbol{s}} e \times \frac{\partial vec(\boldsymbol{s})}{\partial d_n} + vec(\nabla_{\boldsymbol{s}} e)' \frac{\partial^2 vec(\boldsymbol{s})}{\partial b_{l,k}\partial d_n} \\
&\overset{(73)-(72)}{=} 2\textstyle\sum_i (b_{i,k}\boldsymbol{\delta}^{(l)}_{\bar{n}\times 1} + 1_{i=l}\boldsymbol{b}_{\cdot,k})'\nabla^2_{\boldsymbol{s}_{\cdot,i},\boldsymbol{s}_{\cdot,n}} e(d_n\boldsymbol{\delta}^{(n)}_{\bar{n}\times 1}) \\
&+ 2(\nabla_{\boldsymbol{s}_{\cdot,n}} e)' \frac{\partial}{\partial b_{l,k}}(d_n\boldsymbol{\delta}^{(n)}_{\bar{n}\times 1}) \\
&= 2\textstyle\sum_i b_{i,k}(\boldsymbol{\delta}^{(l)}_{\bar{n}\times 1})'[\nabla^2_{\boldsymbol{s}_{\cdot,i},\boldsymbol{s}_{\cdot,n}} e] d_n \boldsymbol{\delta}^{(n)}_{\bar{n}\times 1} + 2(\boldsymbol{b}_{\cdot,k})'[\nabla^2_{\boldsymbol{s}_{\cdot,l},\boldsymbol{s}_{\cdot,n}} e] d_n\boldsymbol{\delta}^{(n)}_{\bar{n}\times 1} \\
&\overset{(65)}{=} \textstyle\sum_i b_{i,k} d_n (\boldsymbol{\delta}^{(l)}_{\bar{n}\times 1})'([\boldsymbol{s}^{-1}]_{i,n}\boldsymbol{s}^{-1})\boldsymbol{\delta}^{(n)}_{\bar{n}\times 1} + d_n(\boldsymbol{b}_{\cdot,k})'([\boldsymbol{s}^{-1}]_{l,n}\boldsymbol{s}^{-1})\boldsymbol{\delta}^{(n)}_{\bar{n}\times 1} \\
&= \textstyle\sum_i [\boldsymbol{b}']_{k,i}[\boldsymbol{s}^{-1}]_{i,n} d_n [\boldsymbol{s}^{-1}]_{l,n} + [\boldsymbol{s}^{-1}]_{l,n} d_n (\boldsymbol{b}_{\cdot,k})'[\boldsymbol{s}^{-1}]_{\cdot,n} \\
&= 2[\boldsymbol{b}'\boldsymbol{s}^{-1}]_{k,n}[\boldsymbol{s}^{-1}]_{l,n} d_n.
\end{aligned} \tag{84}$$

Using the following result

$$\textstyle\sum_{i,j} [\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}(\boldsymbol{\delta}^{(j)}_{\bar{n}\times 1})']_{r,r}[\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}(\boldsymbol{\delta}^{(j)}_{\bar{n}\times 1})']_{n,n} = 1_{r=n}, \tag{85}$$

we rewrite (84) as follows

$$\begin{aligned}
\frac{\partial^2 e}{\partial b_{l,k}\partial d_n} &= 2\textstyle\sum_r [\boldsymbol{b}'\boldsymbol{s}^{-1}]_{k,r} 1_{r=n} [\boldsymbol{s}^{-1} Diag(\boldsymbol{d})]_{l,n} \\
&= 2\textstyle\sum_r [\boldsymbol{b}'\boldsymbol{s}^{-1}]_{k,r} \sum_{i,j} [\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}(\boldsymbol{\delta}^{(j)}_{\bar{n}\times 1})']_{r,r} \\
&\times [\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}(\boldsymbol{\delta}^{(j)}_{\bar{n}\times 1})']_{n,n} [\boldsymbol{s}^{-1} Diag(\boldsymbol{d})]_{l,n} \\
&= 2\textstyle\sum_{i,j} (\sum_r [\boldsymbol{b}'\boldsymbol{s}^{-1}]_{k,r} [\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}(\boldsymbol{\delta}^{(j)}_{\bar{n}\times 1})']_{r,r}) \\
&\times [\boldsymbol{s}^{-1} Diag(\boldsymbol{d})]_{l,n} [\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}(\boldsymbol{\delta}^{(j)}_{\bar{n}\times 1})']_{n,n} \\
&= 2\textstyle\sum_{i,j} [\boldsymbol{b}'\boldsymbol{s}^{-1} diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}(\boldsymbol{\delta}^{(j)}_{\bar{n}\times 1})')]_k \\
&\times [\boldsymbol{s}^{-1} Diag(\boldsymbol{d}) Diag(diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}(\boldsymbol{\delta}^{(j)}_{\bar{n}\times 1})'))]_{l,n},
\end{aligned} \tag{86}$$

and since

$$diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}(\boldsymbol{\delta}^{(j)}_{\bar{n}\times 1})') = 1_{i=j}\boldsymbol{\delta}^{(j)}_{\bar{n}\times 1}, \tag{87}$$

then (86) becomes

$$\frac{\partial^2 e}{\partial b_{l,k}\partial d_n} = 2\textstyle\sum_i [\boldsymbol{b}'\boldsymbol{s}^{-1}\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}]_k [\boldsymbol{s}^{-1} Diag(\boldsymbol{d}) Diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1})]_{l,n}. \tag{88}$$

Hence

$$\begin{aligned}
\nabla^2_{\boldsymbol{b},\boldsymbol{d}} e &= 2\textstyle\sum_i (\boldsymbol{b}'\boldsymbol{s}^{-1}\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}) \otimes (\boldsymbol{s}^{-1} Diag(\boldsymbol{d}) Diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1})) \\
&= 2\textstyle\sum_i (\boldsymbol{b}'\boldsymbol{s}^{-1} \otimes \boldsymbol{s}^{-1} Diag(\boldsymbol{d}))(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1} \otimes Diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1})) \\
&= 2(\boldsymbol{b}'\boldsymbol{s}^{-1} \otimes \boldsymbol{s}^{-1} Diag(\boldsymbol{d}))\boldsymbol{h}_{\bar{n}^2\times\bar{n}},
\end{aligned} \tag{89}$$

where $\boldsymbol{h}_{\bar{n}^2\times\bar{n}}$ is an $\bar{n}^{-2}$ x $\bar{n}$ matrix defined as follows

$$\boldsymbol{h}_{\bar{n}^2\times\bar{n}} \equiv \textstyle\sum_i [\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1} \otimes Diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1})]. \tag{90}$$

The derivative of (75) with respect to $b_{l,r}$ reads

$$
\begin{aligned}
\frac{\partial^2 e}{\partial b_{l,r}\partial b_{n,k}} &= (\frac{\partial vec(\boldsymbol{s})}{\partial b_{l,k}})' \times \nabla^2_{\boldsymbol{s},\boldsymbol{s}} e \times \frac{\partial vec(\boldsymbol{s})}{\partial b_{n,k}} + vec(\nabla_{\boldsymbol{s}} e)' \frac{\partial^2 vec(\boldsymbol{s})}{\partial b_{l,r}\partial b_{n,k}} \\
&\overset{(73)}{=} \textstyle\sum_{i,j} b_{i,r} (\boldsymbol{\delta}^{(l)}_{\bar{n}\times 1})' (\nabla^2_{\boldsymbol{s}_{\cdot,i},\boldsymbol{s}_{\cdot,j}} e) b_{j,k} \boldsymbol{\delta}^{(n)}_{\bar{n}\times 1} + b_{i,r} \boldsymbol{e}'_l (\nabla^2_{\boldsymbol{s}_{\cdot,i},\boldsymbol{s}_{\cdot,j}} e) 1_{j=n} \boldsymbol{b}_{\cdot,k} \\
&+ 1_{i=l} (\boldsymbol{b}_{\cdot,r})' (\nabla^2_{\boldsymbol{s}_{\cdot,i},\boldsymbol{s}_{\cdot,j}} e) b_{j,k} \boldsymbol{\delta}^{(n)}_{\bar{n}\times 1} + 1_{i=l} (\boldsymbol{b}_{\cdot,r})' \nabla^2_{\boldsymbol{s}_{\cdot,i},\boldsymbol{s}_{\cdot,j}} 1_{j=n} \boldsymbol{b}_{\cdot,k} \\
&+ 1_{r=k} ((\nabla_{\boldsymbol{s}_{\cdot,l}} e)' \boldsymbol{\delta}^{(n)}_{\bar{n}\times 1} + (\nabla_{\boldsymbol{s}_{\cdot,n}} e)' \boldsymbol{\delta}^{(l)}_{\bar{n}\times 1}) \\
&\overset{(65)\text{-}(61)}{=} \tfrac{1}{2} (\textstyle\sum_{i,j} b_{i,r} b_{j,k} (\boldsymbol{\delta}^{(l)}_{\bar{n}\times 1})' ([\boldsymbol{s}^{-1}]_{i,j} \boldsymbol{s}^{-1}) \boldsymbol{\delta}^{(n)}_{\bar{n}\times 1} \\
&+ \textstyle\sum_{i,j} b_{i,r} [\boldsymbol{s}^{-1}]_{i,n} 1_{j=n} (\boldsymbol{\delta}^{(l)}_{\bar{n}\times 1})' \boldsymbol{s}^{-1} \boldsymbol{b}_{\cdot,k} \\
&+ (\boldsymbol{b}_{\cdot,r})' \boldsymbol{s}^{-1} \boldsymbol{\delta}^{(n)}_{\bar{n}\times 1} \textstyle\sum_{i,j} 1_{i=l} [\boldsymbol{s}^{-1}]_{l,j} b_{j,k} + (\boldsymbol{b}_{\cdot,r})' ([\boldsymbol{s}^{-1}]_{l,n} \boldsymbol{s}^{-1}) \boldsymbol{b}_{\cdot,k} \\
&+ 1_{r=k} (((\underline{\boldsymbol{s}}^{-1} - \boldsymbol{s}^{-1})_{\cdot,l})' \boldsymbol{\delta}^{(n)}_{\bar{n}\times 1} + ((\underline{\boldsymbol{s}}^{-1} - \boldsymbol{s}^{-1})_{\cdot,n})' \boldsymbol{\delta}^{(l)}_{\bar{n}\times 1}) \\
&= \tfrac{1}{2} (\textstyle\sum_{i,j} b_{i,r} b_{j,k} [\boldsymbol{s}^{-1}]_{i,j} [\boldsymbol{s}^{-1}]_{l,n} + (\boldsymbol{b}_{\cdot,r})' \boldsymbol{s}^{-1} \boldsymbol{\delta}^{(n)}_{\bar{n}\times 1} (\boldsymbol{\delta}^{(l)}_{\bar{n}\times 1})' \boldsymbol{s}^{-1} \boldsymbol{b}_{\cdot,k} \\
&+ (\boldsymbol{b}_{\cdot,r})' \boldsymbol{s}^{-1} \boldsymbol{\delta}^{(n)}_{\bar{n}\times 1} (\boldsymbol{\delta}^{(l)}_{\bar{n}\times 1})' \boldsymbol{s}^{-1} \boldsymbol{b}_{\cdot,k} + [\boldsymbol{s}^{-1}]_{l,n} (\boldsymbol{b}_{\cdot,r})' \boldsymbol{s}^{-1} \boldsymbol{b}_{\cdot,k} \\
&+ 1_{r=k} ([\underline{\boldsymbol{s}}^{-1} - \boldsymbol{s}^{-1}]_{n,l} + [\underline{\boldsymbol{s}}^{-1} - \boldsymbol{s}^{-1}]_{l,n})) \\
&= [\boldsymbol{b}' \boldsymbol{s}^{-1} \boldsymbol{b}]_{r,k} [\boldsymbol{s}^{-1}]_{l,n} + [\boldsymbol{s}^{-1} \boldsymbol{b}]_{l,k} [\boldsymbol{b}' \boldsymbol{s}^{-1}]_{r,n} + 1_{r=k} [\underline{\boldsymbol{s}}^{-1} - \boldsymbol{s}^{-1}]_{l,n}.
\end{aligned} \tag{91}
$$

Using the following identity

$$
[\boldsymbol{s}^{-1}\boldsymbol{b}]_{l,k} [\boldsymbol{b}'\boldsymbol{s}^{-1}]_{r,n} = \textstyle\sum_{i,j} [\boldsymbol{\delta}^{(i)}_{\bar{\boldsymbol{k}}\times \boldsymbol{1}} (\boldsymbol{\delta}^{(j)}_{\bar{\boldsymbol{n}}\times \boldsymbol{1}})' \boldsymbol{s}^{-1} \boldsymbol{b}]_{r,k} [\boldsymbol{\delta}^{(j)}_{\bar{\boldsymbol{n}}\times \boldsymbol{1}} (\boldsymbol{\delta}^{(i)}_{\bar{\boldsymbol{k}}\times \boldsymbol{1}})' \boldsymbol{b}' \boldsymbol{s}^{-1}]_{l,n}, \tag{92}
$$

we obtain

$$
\begin{aligned}
\nabla^2_{\boldsymbol{b},\boldsymbol{b}} e &= (\boldsymbol{b}'\boldsymbol{s}^{-1}\boldsymbol{b}) \otimes \boldsymbol{s}^{-1} + \textstyle\sum_{i,j} (\boldsymbol{\delta}^{(i)}_{\bar{\boldsymbol{k}}\times \boldsymbol{1}} (\boldsymbol{\delta}^{(j)}_{\bar{\boldsymbol{n}}\times \boldsymbol{1}})' \boldsymbol{s}^{-1} \boldsymbol{b}) \otimes (\boldsymbol{\delta}^{(j)}_{\bar{\boldsymbol{n}}\times \boldsymbol{1}} (\boldsymbol{\delta}^{(i)}_{\bar{\boldsymbol{k}}\times \boldsymbol{1}})' \boldsymbol{b}' \boldsymbol{s}^{-1}) \\
&+ \boldsymbol{i}_{\bar{k}\times\bar{k}} \otimes (\underline{\boldsymbol{s}}^{-1} - \boldsymbol{s}^{-1}) \\
&= (\boldsymbol{b}'\boldsymbol{s}^{-1}\boldsymbol{b}) \otimes \boldsymbol{s}^{-1} + \textstyle\sum_{i,j} (\boldsymbol{\delta}^{(i)}_{\bar{\boldsymbol{k}}\times \boldsymbol{1}} (\boldsymbol{\delta}^{(j)}_{\bar{\boldsymbol{n}}\times \boldsymbol{1}})' \otimes \boldsymbol{\delta}^{(j)}_{\bar{\boldsymbol{n}}\times \boldsymbol{1}} (\boldsymbol{\delta}^{(i)}_{\bar{\boldsymbol{k}}\times \boldsymbol{1}})') (\boldsymbol{s}^{-1}\boldsymbol{b} \otimes \boldsymbol{b}'\boldsymbol{s}^{-1}) \\
&+ \boldsymbol{i}_{\bar{k}\times\bar{k}} \otimes (\underline{\boldsymbol{s}}^{-1} - \boldsymbol{s}^{-1}) \\
&= (\boldsymbol{b}'\boldsymbol{s}^{-1}\boldsymbol{b}) \otimes \boldsymbol{s}^{-1} + \boldsymbol{k}_{\bar{k}\bar{n}\times\bar{k}\bar{n}} (\boldsymbol{s}^{-1}\boldsymbol{b} \otimes (\boldsymbol{s}^{-1}\boldsymbol{b})') + \boldsymbol{i}_{\bar{k}\times\bar{k}} \otimes (\underline{\boldsymbol{s}}^{-1} - \boldsymbol{s}^{-1}),
\end{aligned} \tag{93}
$$

where $\boldsymbol{k}_{\bar{k}\bar{n}\times\bar{k}\bar{n}}$ is a $\bar{k}\bar{n} \times \bar{k}\bar{n}$ matrix defined as follows

$$
\boldsymbol{k}_{\bar{k}\bar{n}\times\bar{k}\bar{n}} \equiv \textstyle\sum_{i,j} [\boldsymbol{\delta}^{(i)}_{\bar{\boldsymbol{k}}\times \boldsymbol{1}} (\boldsymbol{\delta}^{(j)}_{\bar{\boldsymbol{n}}\times \boldsymbol{1}})'] \otimes [\boldsymbol{\delta}^{(j)}_{\bar{\boldsymbol{n}}\times \boldsymbol{1}} (\boldsymbol{\delta}^{(i)}_{\bar{\boldsymbol{k}}\times \boldsymbol{1}})']. \tag{94}
$$

Remark: $\boldsymbol{k}_{\bar{k}\bar{n}\times\bar{k}\bar{n}}$ is the *commutation matrix*, i.e. the $\bar{k}\bar{n}$ x $\bar{k}\bar{n}$ matrix such that

$$
\boldsymbol{k}_{\bar{k}\bar{n}\times\bar{k}\bar{n}} vec(\boldsymbol{x}) = vec(\boldsymbol{x}'), \tag{95}
$$

for any $\bar{k}$ x $\bar{n}$ matrix $\boldsymbol{x}$. Also, it is possible to rewrite (94) as

$$
\boldsymbol{k}_{\bar{k}\bar{n}\times\bar{k}\bar{n}} \equiv \textstyle\sum_{i} [\boldsymbol{\delta}^{(i)}_{\bar{\boldsymbol{k}}\times \boldsymbol{1}} \otimes \boldsymbol{i}_{\bar{n}\times\bar{n}} \otimes (\boldsymbol{\delta}^{(i)}_{\bar{\boldsymbol{k}}\times \boldsymbol{1}})']. \tag{96}
$$

For more details see e.g. in [Magnus and Neudecker, 1979].

To summarise, using the compact notation of Appendix A.2, the Hessian of the optimisation target

$$\nabla^2_{\boldsymbol{\mu},\boldsymbol{\mu}} e = (\underline{\boldsymbol{\sigma}}^2)^{-1} \tag{97}$$

$$\nabla^2_{\boldsymbol{\mu},\boldsymbol{d}} e = \mathbf{0}_{\bar{n}\times\bar{n}} \tag{98}$$

$$\nabla^2_{\boldsymbol{\mu},\boldsymbol{b}} e = \mathbf{0}_{\bar{n}\times\bar{n}\bar{k}} \tag{99}$$

$$\begin{aligned}\nabla^2_{\boldsymbol{d},\boldsymbol{d}} e &= 2(Diag(\boldsymbol{d})(\boldsymbol{\sigma}^2)^{-1}) \circ ((\boldsymbol{\sigma}^2)^{-1} Diag(\boldsymbol{d})) \\ &\quad + \boldsymbol{i}_{\bar{n}\times\bar{n}} \circ ((\underline{\boldsymbol{\sigma}}^2)^{-1} - (\boldsymbol{\sigma}^2)^{-1})\end{aligned} \tag{100}$$

$$\nabla^2_{\boldsymbol{b},\boldsymbol{d}} e = 2(\boldsymbol{b}'(\boldsymbol{\sigma}^2)^{-1} \otimes (\boldsymbol{\sigma}^2)^{-1} Diag(\boldsymbol{d}))\boldsymbol{h}_{\bar{n}^2\times\bar{n}} \tag{101}$$

$$\begin{aligned}\nabla^2_{\boldsymbol{b},\boldsymbol{b}} e &= (\boldsymbol{b}'(\boldsymbol{\sigma}^2)^{-1}\boldsymbol{b}) \otimes (\boldsymbol{\sigma}^2)^{-1} + \boldsymbol{k}_{\bar{k}\bar{n}\times\bar{k}\bar{n}}((\boldsymbol{\sigma}^2)^{-1}\boldsymbol{b} \otimes ((\boldsymbol{\sigma}^2)^{-1}\boldsymbol{b})') \\ &\quad + \boldsymbol{i}_{\bar{k}\times\bar{k}} \otimes ((\underline{\boldsymbol{\sigma}}^2)^{-1} - (\boldsymbol{\sigma}^2)^{-1}),\end{aligned} \tag{102}$$

where $\boldsymbol{h}_{\bar{n}^2\times\bar{n}}$ is the $\bar{n}^2$ x $\bar{n}$ matrix (90) and $\boldsymbol{k}_{\bar{k}\bar{n}\times\bar{k}\bar{n}}$ is the $\bar{k}\bar{n} \times \bar{k}\bar{n}$ commutation matrix (94)-(96).

A.7 Convexity of optimisation target

Let us recall that a differentiable function

$$f(\boldsymbol{x}) = f(x_1, x_2, \ldots, x_{\bar{n}}), \tag{103}$$

is convex if and only if

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})'(\boldsymbol{y} - \boldsymbol{x}), \tag{104}$$

for any $\boldsymbol{x}$, $\boldsymbol{y}$. Thus, if f is a convex function we have

$$f(\boldsymbol{x} + \boldsymbol{z}) - f(\boldsymbol{x}) \geq 0, \tag{105}$$

for any $\boldsymbol{x}$ and vector $\boldsymbol{z}$ which is orthogonal to $\nabla f(\boldsymbol{x})$.

Thus, to verify that the entropy (8) is not a convex function of the entries ($\boldsymbol{\mu}$, $\boldsymbol{b}$, $\boldsymbol{d}$) we perform a numerical testing as follows. We generate randomly market parameters $\theta^{(j)} \equiv \boldsymbol{\mu}^{(j)}$, $\boldsymbol{b}^{(j)}$, $\boldsymbol{d}^{(j)}$) in dimension $\bar{n}$ = 7 and $\bar{k}$ = 1 for a large number of scenarios $\bar{j}$ = 1000. Then, for each market scenario j, we compute a basis $\{\boldsymbol{z}_1^{(j)}, \boldsymbol{z}_2^{(j)}, \ldots, \boldsymbol{z}_{\bar{n}(2+\bar{k})-1}^{(j)}\}$ of the tangent space of e at $\theta^{(j)}$.

Finally we consider the $\bar{n}(2 + \bar{k}) - 1$ x 1 vectors

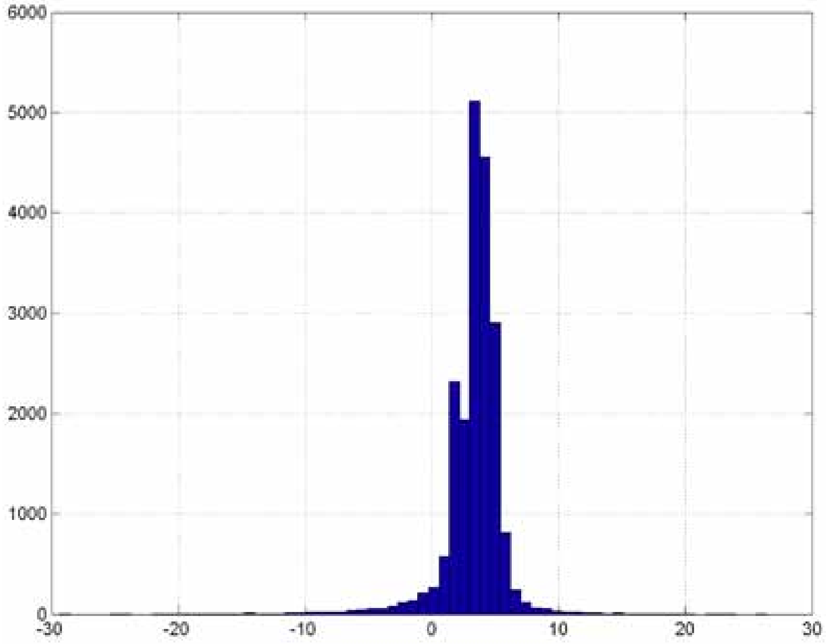
$$\boldsymbol{u}^{(j)} \equiv \begin{pmatrix} e(\boldsymbol{\theta}^{(j)}+\boldsymbol{z}_1^{(j)})-e(\boldsymbol{\theta}^{(j)}) \\ e(\boldsymbol{\theta}^{(j)}+\boldsymbol{z}_2^{(j)})-e(\boldsymbol{\theta}^{(j)}) \\ \vdots \\ e(\boldsymbol{\theta}^{(j)}+\boldsymbol{z}_{\bar{n}(2+\bar{k})-1}^{(j)})-e(\boldsymbol{\theta}^{(j)}) \end{pmatrix}, \tag{106}$$

for any $j = 1,\ldots, \bar{j}$ which we arrange in a compact vector $\boldsymbol{u}$ as follows

$$\boldsymbol{u} \equiv \begin{pmatrix} \boldsymbol{u}_1 \\ \vdots \\ \boldsymbol{u}_{\bar{j}} \end{pmatrix}. \tag{107}$$

As we can see in Figure 7 we realise that there are components that do not verify (105) by plotting the histogram of (107).

Figure 7: Convexity test.



A.8 Gradient of signal-to-noise ranking view

We consider the multivariate function

$$g(\boldsymbol{\mu}, \boldsymbol{b}, \boldsymbol{d}) = (g_1(\boldsymbol{\mu}, \boldsymbol{b}, \boldsymbol{d}), \ldots, g_{\bar{m}}(\boldsymbol{\mu}, \boldsymbol{b}, \boldsymbol{d}))' \equiv \boldsymbol{a}(\boldsymbol{\mu}./\sqrt{diag(\boldsymbol{s})}) - \boldsymbol{q}, \tag{108}$$

where μ is an $\bar{n}$-dimensional vector, ***b*** is a $\bar{n}$ x $\bar{k}$ matrix, ***d*** is an $\bar{n}$-dimensional vector, ***a*** is an $\bar{m}$ x $\bar{n}$ matrix and s is an $\bar{n}$ x $\bar{n}$ matrix defined as follows

$$\boldsymbol{s} \equiv \boldsymbol{b}\boldsymbol{b}' + Diag(\boldsymbol{d} \circ \boldsymbol{d}). \tag{109}$$

From (108) we have

$$g_m = \sum_i a_{m,i} \frac{\mu_i}{\sqrt{s_{i,i}}} - q_m, \tag{110}$$

and then

$$\begin{aligned} \frac{\partial g_m}{\partial s_{l,r}} &= \frac{\partial}{\partial s_{l,r}} (\sum_i a_{m,i} \frac{\mu_i}{\sqrt{s_{i,i}}} - q_m) \\ &= \sum_i a_{m,i} \mu_i \frac{\partial}{\partial s_{l,r}} (\frac{1}{\sqrt{s_{i,i}}}) \\ &= -a_{m,l} \frac{\mu_l}{2(s_{l,l})^{\frac{3}{2}}} 1_{l=r}. \end{aligned} \tag{111}$$

The gradient of g with respect to μ reads

$$\begin{aligned} \frac{\partial g_m}{\partial \mu_n} &= \sum_i a_{m,i} \frac{\partial}{\partial \mu_n} (\frac{\mu_i}{\sqrt{s_{i,i}}}) \\ &= \sum_i a_{m,i} \frac{1_{i=n}}{\sqrt{s_{i,i}}} = a_{m,n} \frac{1}{\sqrt{s_{n,n}}} \\ &= [Diag(diag(1./\sqrt{\boldsymbol{s}}))]_{n,n} [\boldsymbol{a}']_{n,m} \\ &= [Diag(diag(1./\sqrt{\boldsymbol{s}}))\boldsymbol{a}']_{n,m}. \end{aligned} \tag{112}$$

Hence

$$\nabla_{\boldsymbol{\mu}} g = Diag(diag(1./\sqrt{\boldsymbol{s}}))\boldsymbol{a}'. \tag{113}$$

Using the chain rule the gradient of g with respect to $\boldsymbol{d}$ reads

$$\begin{aligned}
\frac{\partial g_m}{\partial d_n} &= \sum_{l,r} \frac{\partial g_m}{\partial s_{l,r}} \frac{\partial s_{l,r}}{\partial d_n} \\
&\overset{(70)\text{-}(111)}{=} -\sum_{l,r} a_{m,l} \frac{\mu_l}{(s_{l,l})^{\frac{3}{2}}} d_n 1_{l=r} 1_{r=n} \\
&= -a_{m,n} \frac{\mu_n d_n}{(s_{n,n})^{\frac{3}{2}}} = -[Diag(\boldsymbol{\alpha})]_{n,n} [\boldsymbol{a}']_{n,m} \\
&= -[Diag(\boldsymbol{\alpha})\boldsymbol{a}']_{n,m},
\end{aligned} \tag{114}$$

where α is an $\bar{n}$ x 1 vector defined as follows

$$\begin{aligned}
\boldsymbol{\alpha} &\equiv \left(\frac{\mu_1 d_1}{(s_{1,1})^{\frac{3}{2}}}, \ldots, \frac{\mu_{\bar{n}} d_{\bar{n}}}{(s_{\bar{n},\bar{n}})^{\frac{3}{2}}}\right)' \\
&= (\boldsymbol{\mu} \circ \boldsymbol{d})./diag(\boldsymbol{s})^{\frac{3}{2}}.
\end{aligned} \tag{115}$$

Thus

$$\nabla_{\boldsymbol{d}} g = -Diag((\boldsymbol{\mu} \circ \boldsymbol{d})./diag(\boldsymbol{s})^{\frac{3}{2}})\boldsymbol{a}'. \tag{116}$$

Using the chain rule the gradient of g with respect to $\boldsymbol{b}$ reads

$$\begin{aligned}
\frac{\partial g_m}{\partial b_{n,k}} &= \sum_{l,r} \frac{\partial g_m}{\partial s_{l,r}} \frac{\partial s_{l,r}}{\partial b_{n,k}} \\
&\overset{(71)\text{-}(111)}{=} -\sum_{l,r} a_{m,l} \frac{\mu_l}{2(s_{l,l})^{\frac{3}{2}}} 1_{l=r} (1_{l=n} b_{r,k} + 1_{r=n} b_{l,k}) \\
&= -a_{m,n} \frac{\mu_n}{(s_{n,n})^{\frac{3}{2}}} b_{n,k}.
\end{aligned} \tag{117}$$

Then from (85), (117) becomes

$$\begin{aligned}
\frac{\partial g_m}{\partial b_{n,k}} &= -\sum_r b_{r,k} 1_{r=n} a_{m,n} \frac{\mu_n}{(s_{n,n})^{\frac{3}{2}}} \\
&\overset{(85)}{=} -\sum_{i,j} \sum_r [\boldsymbol{b}']_{k,r} [\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1} (\boldsymbol{\delta}^{(j)}_{\bar{n}\times 1})']_{r,r} [\boldsymbol{a}']_{r,m} [\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1} (\boldsymbol{\delta}^{(j)}_{\bar{n}\times 1})']_{n,n} \frac{\mu_n}{(s_{n,n})^{\frac{3}{2}}} \\
&= -\sum_{i,j} [\boldsymbol{b}' Diag(diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1} (\boldsymbol{\delta}^{(j)}_{\bar{n}\times 1})'))\boldsymbol{a}']_{k,m} [Diag(diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1} (\boldsymbol{\delta}^{(j)}_{\bar{n}\times 1})'))\boldsymbol{\beta}]_n \\
&\overset{(87)}{=} -\sum_i [\boldsymbol{b}' Diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1})\boldsymbol{a}']_{k,m} [Diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1})\boldsymbol{\beta}]_n,
\end{aligned} \tag{118}$$

where β is an $\bar{n}$ x 1 vector defined as follows

$$\begin{aligned}
\boldsymbol{\beta} &\equiv \left(\frac{\mu_1}{(s_{1,1})^{\frac{3}{2}}}, \ldots, \frac{\mu_{\bar{n}}}{(s_{\bar{n},\bar{n}})^{\frac{3}{2}}}\right)' \\
&= \boldsymbol{\mu}./diag(\boldsymbol{s})^{\frac{3}{2}}.
\end{aligned} \tag{119}$$

Then

$$\begin{aligned}
\nabla_{\boldsymbol{b}} g &= -\sum_i (\boldsymbol{b}' Diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1})\boldsymbol{a}') \otimes (Diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1})\boldsymbol{\beta}) \\
&= -\sum_i (\boldsymbol{b}' \otimes \boldsymbol{i}_{\bar{n}\times\bar{n}})(Diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}) \otimes Diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}))(\boldsymbol{a}' \otimes \boldsymbol{\beta}) \\
&= -(\boldsymbol{b}' \otimes \boldsymbol{i}_{\bar{n}\times\bar{n}})\boldsymbol{h}_{\bar{n}^2\times\bar{n}^2}(\boldsymbol{a}' \otimes (\boldsymbol{\mu}./diag(\boldsymbol{s})^{\frac{3}{2}})),
\end{aligned} \tag{120}$$

where $\boldsymbol{h}_{\bar{n}^2\times\bar{n}^2}$ is an $\bar{n}^2$ x $\bar{n}^2$ matrix defined as follows

$$\begin{aligned}\boldsymbol{h}_{\bar{n}^2\times\bar{n}^2} &\equiv \sum_i Diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}) \otimes Diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1})\\ &= Diag(\sum_i \boldsymbol{\delta}^{(i)}_{\bar{n}\times 1} \otimes \boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}) = Diag(\sum_i vec(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1})'))\\ &= Diag(vec(\sum_i \boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1})')) = Diag(vec(\boldsymbol{i}_{\bar{n}\times\bar{n}})).\end{aligned} \tag{121}$$

A.9 Hessian of signal-to-noise ranking view

Let us consider the same setup in Appendix A.8.

Let us start with the derivative of (112) with respect to μ

$$\frac{\partial^2 g_m}{\partial\mu_k\partial\mu_n} = \frac{\partial}{\partial\mu_k}(a_{m,n}\frac{1}{\sqrt{s_{n,n}}}) = 0, \tag{122}$$

and hence

$$\nabla^2_{\boldsymbol{\mu},\boldsymbol{\mu}} g = \mathbf{0}_{\bar{n}\times\bar{n}\bar{m}}. \tag{123}$$

The derivative of (112) with respect to $\boldsymbol{d}$ reads

$$\begin{aligned}\frac{\partial^2 g_m}{\partial d_n\partial\mu_t} &= \frac{\partial}{\partial d_n}(a_{m,t}\frac{1}{\sqrt{s_{t,t}}}) = \sum_{l,r}\frac{\partial}{\partial s_{l,r}}(\frac{a_{m,t}}{\sqrt{s_{t,t}}})\frac{\partial s_{l,r}}{\partial d_n}\\ &\overset{(70)}{=} -\frac{a_{m,t}}{(s_{t,t})^{\frac{3}{2}}}d_n 1_{t=n} = -d_t[\boldsymbol{a}']_{t,m}\frac{1_{t=n}}{(s_{t,t})^{\frac{3}{2}}} = -\sum_r d_r[\boldsymbol{a}']_{r,m}1_{r=t}\frac{1_{t=n}}{(s_{t,t})^{\frac{3}{2}}}\\ &\overset{(85)}{=} -\sum_{i,j}\sum_r d_r[\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}(\boldsymbol{\delta}^{(j)}_{\bar{n}\times 1})']_{r,r}[\boldsymbol{a}']_{r,m}\frac{[\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}(\boldsymbol{\delta}^{(j)}_{\bar{n}\times 1})']_{t,t}1_{t=n}}{(s_{t,t})^{\frac{3}{2}}}\\ &= -\sum_{i,j}\sum_r[\boldsymbol{d}'Diag(diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}(\boldsymbol{\delta}^{(j)}_{\bar{n}\times 1})'))]_r[\boldsymbol{a}']_{r,m}\\ &\times[Diag(diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}(\boldsymbol{\delta}^{(j)}_{\bar{n}\times 1})'))Diag(diag(1./\boldsymbol{s}))^{\frac{3}{2}}]_{n,t}\\ &\overset{(87)}{=} -\sum_i[\boldsymbol{d}'Diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1})\boldsymbol{a}']_m[Diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1})Diag(diag(1./\boldsymbol{s}))^{\frac{3}{2}}]_{n,t}.\end{aligned} \tag{124}$$

Thus

$$\begin{aligned}\nabla^2_{\boldsymbol{d},\boldsymbol{\mu}} g &= -\sum_i(\boldsymbol{d}'Diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1})\boldsymbol{a}')\otimes(Diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1})Diag(diag(1./\boldsymbol{s}))^{\frac{3}{2}})\\ &= -\sum_i(\boldsymbol{d}'\otimes\boldsymbol{i}_{\bar{n}\times\bar{n}})(Diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1})\otimes Diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}))(\boldsymbol{a}'\otimes Diag(diag(1./\boldsymbol{s}))^{\frac{3}{2}})\\ &\overset{(121)}{=} -(\boldsymbol{d}'\otimes\boldsymbol{i}_{\bar{n}\times\bar{n}})\boldsymbol{h}_{\bar{n}^2\times\bar{n}^2}(\boldsymbol{a}'\otimes Diag(diag(1./\boldsymbol{s}))^{\frac{3}{2}}).\end{aligned} \tag{125}$$

The derivative of (112) with respect to $\boldsymbol{b}$ reads

$$\begin{aligned}\frac{\partial^2 g_m}{\partial b_{n,k}\partial\mu_t} &= \frac{\partial}{\partial b_{n,k}}(a_{m,t}\frac{1}{\sqrt{s_{t,t}}}) = \sum_{l,r}\frac{\partial}{\partial s_{l,r}}(\frac{a_{m,t}}{\sqrt{s_{t,t}}})\frac{\partial s_{l,r}}{\partial b_{n,k}}\\ &\overset{(71)}{=} -\frac{a_{m,t}}{(s_{t,t})^{\frac{3}{2}}}b_{t,k}1_{t=n} = -\sum_r[\boldsymbol{b}']_{k,r}1_{r=t}[\boldsymbol{a}']_{r,m}\frac{1_{t=n}}{(s_{t,t})^{\frac{3}{2}}}\\ &\overset{(85)}{=} -\sum_{i,j}\sum_r[\boldsymbol{b}']_{k,r}[\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}(\boldsymbol{\delta}^{(j)}_{\bar{n}\times 1})']_{r,r}[\boldsymbol{a}']_{r,m}\frac{[\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}(\boldsymbol{\delta}^{(j)}_{\bar{n}\times 1})']_{t,t}1_{t=n}}{(s_{t,t})^{\frac{3}{2}}}\\ &= -\sum_{i,j}\sum_r[\boldsymbol{b}'Diag(diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}(\boldsymbol{\delta}^{(j)}_{\bar{n}\times 1})'))]_{k,r}[\boldsymbol{a}']_{r,m}\\ &\times[Diag(diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}(\boldsymbol{\delta}^{(j)}_{\bar{n}\times 1})'))Diag(diag(1./\boldsymbol{s}))^{\frac{3}{2}}]_{n,t}\\ &\overset{(87)}{=} -\sum_i[\boldsymbol{b}'Diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1})\boldsymbol{a}']_{k,m}[Diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1})Diag(diag(1./\boldsymbol{s}))^{\frac{3}{2}}]_{n,t}.\end{aligned} \tag{126}$$

Thus

$$\nabla^2_{\boldsymbol{b},\boldsymbol{\mu}} g = -\textstyle\sum_i (\boldsymbol{b}' Diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1})\boldsymbol{a}') \otimes (Diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}) Diag(diag(1./\boldsymbol{s}))^{\frac{3}{2}})$$
$$= -\textstyle\sum_i (\boldsymbol{b}' \otimes \boldsymbol{i}_{\bar{n}\times\bar{n}})(Diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}) \otimes Diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}))(\boldsymbol{a}' \otimes Diag(diag(1./\boldsymbol{s}))^{\frac{3}{2}})$$
$$\overset{(121)}{=} -(\boldsymbol{b}' \otimes \boldsymbol{i}_{\bar{n}\times\bar{n}})\boldsymbol{h}_{\bar{n}^2\times\bar{n}^2}(\boldsymbol{a}' \otimes Diag(diag(1./\boldsymbol{s}))^{\frac{3}{2}}). \quad (127)$$

The derivative of (114) with respect to ***b*** reads

$$\frac{\partial^2 g_m}{\partial b_{n,k}\partial d_t} = -\frac{\partial}{\partial b_{n,k}}(a_{m,t}\frac{\mu_t d_t}{(s_{t,t})^{\frac{3}{2}}}) = -a_{m,t}\mu_t d_t \frac{\partial}{\partial b_{n,k}}((s_{t,t})^{-\frac{3}{2}})$$
$$= -a_{m,t}\mu_t d_t \textstyle\sum_{l,r} \frac{\partial}{\partial s_{l,r}}((s_{t,t})^{-\frac{3}{2}})\frac{\partial s_{l,r}}{\partial b_{n,k}}$$
$$\overset{(71)}{=} 3a_{m,t}\mu_t d_t (s_{t,t})^{-\frac{5}{2}} 1_{t=n} b_{t,k} = 3\textstyle\sum_r [\boldsymbol{b}']_{k,r} 1_{r=t}[\boldsymbol{a}']_{r,m}\frac{\mu_t d_t}{(s_{t,t})^{\frac{5}{2}}} 1_{t=n}$$
$$\overset{(85)}{=} 3\textstyle\sum_{i,j}\sum_r [\boldsymbol{b}']_{k,r}[\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}(\boldsymbol{\delta}^{(j)}_{\bar{n}\times 1})']_{r,r}[\boldsymbol{a}']_{r,m}\frac{\mu_t d_t[\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}(\boldsymbol{\delta}^{(j)}_{\bar{n}\times 1})']_{t,t}}{(s_{t,t})^{\frac{5}{2}}} 1_{t=n}$$
$$= 3\textstyle\sum_{i,j}\sum_r [\boldsymbol{b}' Diag(diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}(\boldsymbol{\delta}^{(j)}_{\bar{n}\times 1})'))]_{k,r}[\boldsymbol{a}']_{r,m}$$
$$\times [Diag(diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}(\boldsymbol{\delta}^{(j)}_{\bar{n}\times 1})'))Diag(\boldsymbol{\alpha})]_{n,t}$$
$$\overset{(87)}{=} 3\textstyle\sum_i\sum_r [\boldsymbol{b}' Diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1})]_{k,r}[\boldsymbol{a}']_{r,m}[Diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1})Diag(\boldsymbol{\alpha})]_{n,t} \quad (128)$$

where α is an $\bar{n}$ x 1 vector defined as follows

$$\boldsymbol{\alpha} \equiv (\frac{\mu_1 d_1}{(s_{1,1})^{\frac{5}{2}}}, \ldots, \frac{\mu_{\bar{n}} d_{\bar{n}}}{(s_{\bar{n},\bar{n}})^{\frac{5}{2}}})'$$
$$= (\boldsymbol{\mu} \circ \boldsymbol{d})./diag(\boldsymbol{s})^{\frac{5}{2}}. \quad (129)$$

Then

$$\nabla^2_{\boldsymbol{b},\boldsymbol{d}} g = 3\textstyle\sum_i (\boldsymbol{b}' Diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1})\boldsymbol{a}') \otimes (Diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}) Diag((\boldsymbol{\mu}\circ\boldsymbol{d})./diag(\boldsymbol{s})^{\frac{5}{2}}))$$
$$= 3\textstyle\sum_i (\boldsymbol{b}' \otimes \boldsymbol{i}_{\bar{n}\times\bar{n}})(Diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}) \otimes Diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}))$$
$$\times (\boldsymbol{a}' \otimes Diag((\boldsymbol{\mu}\circ\boldsymbol{d})./diag(\boldsymbol{s})^{\frac{5}{2}}))$$
$$\overset{(121)}{=} 3(\boldsymbol{b}' \otimes \boldsymbol{i}_{\bar{n}\times\bar{n}})\boldsymbol{h}_{\bar{n}^2\times\bar{n}^2}(\boldsymbol{a}' \otimes Diag((\boldsymbol{\mu}\circ\boldsymbol{d})./diag(\boldsymbol{s})^{\frac{5}{2}})). \quad (130)$$

The derivative of (114) with respect to ***d*** reads

$$\frac{\partial^2 g_m}{\partial d_n \partial d_t} = -\frac{\partial}{\partial d_n}(a_{m,t}\frac{\mu_t d_t}{(s_{t,t})^{\frac{3}{2}}})$$
$$= -a_{m,t}\mu_t(\frac{1_{n=t}}{(s_{t,t})^{\frac{3}{2}}} + d_t\textstyle\sum_{l,r}\frac{\partial}{\partial s_{l,r}}((s_{t,t})^{-\frac{3}{2}})\frac{\partial s_{l,r}}{\partial d_n})$$
$$\overset{(70)}{=} -a_{m,t}\mu_t(\frac{1_{n=t}}{(s_{t,t})^{\frac{3}{2}}} - 3d_t(s_{t,t})^{-\frac{5}{2}} d_n 1_{n=t})$$
$$= a_{m,t}\mu_t(3\frac{d_t^2}{(s_{t,t})^{\frac{5}{2}}} - \frac{1}{(s_{t,t})^{\frac{3}{2}}})1_{t=n}$$
$$= \textstyle\sum_r 1_{r=t}[\boldsymbol{a}']_{r,m}\mu_t(3\frac{d_t^2}{(s_{t,t})^{\frac{5}{2}}} - \frac{1}{(s_{t,t})^{\frac{3}{2}}})1_{t=n}$$

$$
\begin{aligned}
&\overset{(85)}{=} \textstyle\sum_{i,j}\sum_r [\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}(\boldsymbol{\delta}^{(j)}_{\bar{n}\times 1})']_{r,r}[\boldsymbol{a}']_{r,m}[\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}(\boldsymbol{\delta}^{(j)}_{\bar{n}\times 1})']_{t,t}\mu_t \\
&\times (3\frac{d_t^2}{(s_{t,t})^{\frac{5}{2}}} - \frac{1}{(s_{t,t})^{\frac{3}{2}}})1_{t=n} \\
&= \textstyle\sum_{i,j}\sum_r [diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}(\boldsymbol{\delta}^{(j)}_{\bar{n}\times 1})')']_r[\boldsymbol{a}']_{r,m} \\
&\times [Diag(diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}(\boldsymbol{\delta}^{(j)}_{\bar{n}\times 1})'))]_{n,n}[Diag(\boldsymbol{\beta})]_{t,t} \\
&\overset{(87)}{=} \textstyle\sum_i [(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1})'\boldsymbol{a}']_m[Diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1})Diag(\boldsymbol{\beta})]_{n,t},
\end{aligned} \tag{131}
$$

where $\boldsymbol{\beta}$ is an $\bar{n}$ x 1 vector defined as follows

$$
\begin{aligned}
\boldsymbol{\beta} &\equiv (\mu_1(3\frac{d_1^2}{(s_{1,1})^{\frac{5}{2}}} - \frac{1}{(s_{1,1})^{\frac{3}{2}}}), \ldots, \mu_{\bar{n}}(3\frac{d_{\bar{n}}^2}{(s_{\bar{n},\bar{n}})^{\frac{5}{2}}} - \frac{1}{(s_{\bar{n},\bar{n}})^{\frac{3}{2}}}))' \\
&= \boldsymbol{\mu} \circ (3(\boldsymbol{d} \circ \boldsymbol{d})./diag(\boldsymbol{s})^{\frac{5}{2}} - 1./diag(\boldsymbol{s})^{\frac{3}{2}}).
\end{aligned} \tag{132}
$$

Hence

$$
\begin{aligned}
\nabla^2_{\boldsymbol{d},\boldsymbol{d}} g &= \textstyle\sum_i ((\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1})'\boldsymbol{a}') \otimes (Diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}) \\
&\times Diag(\boldsymbol{\mu} \circ (3(\boldsymbol{d} \circ \boldsymbol{d})./diag(\boldsymbol{s})^{\frac{5}{2}} - ./diag(\boldsymbol{s})^{\frac{3}{2}})) \\
&= \boldsymbol{h}_{\bar{n}\times\bar{n}^2}(\boldsymbol{a}' \otimes Diag(\boldsymbol{\mu} \circ (3(\boldsymbol{d} \circ \boldsymbol{d})./diag(\boldsymbol{s})^{\frac{5}{2}} - 1./diag(\boldsymbol{s})^{\frac{3}{2}}))),
\end{aligned} \tag{133}
$$

where $\boldsymbol{h}_{\bar{n}^2\times\bar{n}^2}$ is an $\bar{n}$ x $\bar{n}^2$ matrix defined as follows

$$
\boldsymbol{h}_{\bar{n}\times\bar{n}^2} \equiv \textstyle\sum_i [(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1})' \otimes Diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1})]. \tag{134}
$$

The derivative of (118) with respect to $\boldsymbol{b}$ reads

$$
\begin{aligned}
\frac{\partial^2 g_m}{\partial b_{s,t}\partial b_{n,k}} &= -\frac{\partial}{\partial b_{s,t}}(a_{m,n}\frac{\mu_n}{(s_{n,n})^{\frac{3}{2}}}b_{n,k}) = -a_{m,n}\mu_n\frac{\partial}{\partial b_{s,t}}(\frac{b_{n,k}}{(s_{n,n})^{\frac{3}{2}}}) \\
&= -a_{m,n}\mu_n(\frac{1_{s=n}1_{t=k}}{(s_{n,n})^{\frac{3}{2}}} + b_{n,k}\textstyle\sum_{l,r}\frac{\partial}{\partial s_{l,r}}((s_{n,n})^{-\frac{3}{2}})\frac{\partial s_{l,r}}{\partial b_{n,k}} \\
&\overset{(71)}{=} -a_{m,n}\mu_n(\frac{1_{s=n}1_{t=k}}{(s_{n,n})^{\frac{3}{2}}} - 3\frac{b_{n,k}}{(s_{n,n})^{\frac{5}{2}}}1_{n=s}b_{n,t}) \\
&= 3\textstyle\sum_r 1_{r=n}[\boldsymbol{a}']_{r,m}\mu_n(3\frac{b_{n,k}b_{n,t}}{(s_{n,n})^{\frac{5}{2}}}1_{n=s} - \frac{1_{t=k}}{(s_{n,n})^{\frac{3}{2}}})1_{s=n} \\
&\overset{(85)}{=} \textstyle\sum_{i,j}\sum_r [\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}(\boldsymbol{\delta}^{(j)}_{\bar{n}\times 1})']_{r,r}[\boldsymbol{a}']_{r,m}[\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}(\boldsymbol{\delta}^{(j)}_{\bar{n}\times 1})']_{n,n} \\
&\times \mu_n(3\frac{b_{n,k}b_{n,t}}{(s_{n,n})^{\frac{5}{2}}}1_{n=s} - \frac{1_{t=k}}{(s_{n,n})^{\frac{3}{2}}})1_{s=n} \\
&= \textstyle\sum_{i,j}[diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}(\boldsymbol{\delta}^{(j)}_{\bar{n}\times 1})')'\boldsymbol{a}']_m\mu_n[\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}(\boldsymbol{\delta}^{(j)}_{\bar{n}\times 1})']_{n,n} \\
&\times (3\frac{b_{n,k}b_{n,t}}{(s_{n,n})^{\frac{5}{2}}}1_{n=s} - \frac{1_{t=k}}{(s_{n,n})^{\frac{3}{2}}})1_{s=n} \\
&= \textstyle\sum_i [(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1})'\boldsymbol{a}']_m\mu_n[\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1})']_{n,n}(3\frac{b_{n,k}b_{n,t}}{(s_{n,n})^{\frac{5}{2}}}1_{n=s} - \frac{1_{t=k}}{(s_{n,n})^{\frac{3}{2}}})1_{s=n}.
\end{aligned} \tag{135}
$$

We rewrite

$$
\phi \equiv \mu_n[\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1})']_{n,n}\frac{b_{n,k}b_{n,t}}{(s_{n,n})^{\frac{5}{2}}}1_{s=n} \tag{136}
$$

as follows

$$
\begin{aligned}
\phi &\equiv \textstyle\sum_r [\boldsymbol{b}']_{t,r} b_{r,k} 1_{r=n} \frac{\mu_n [\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1})']_{n,n}}{(s_{n,n})^{\frac{5}{2}}} 1_{s=n} \\
&\overset{(85)}{=} \textstyle\sum_{u,v}\sum_r [\boldsymbol{b}']_{t,r}[\boldsymbol{\delta}^{(u)}_{\bar{n}\times 1}(\boldsymbol{\delta}^{(v)}_{\bar{n}\times 1})']_{r,r} b_{r,k} \\
&\times \frac{\mu_n [\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1})']_{n,n}[\boldsymbol{\delta}^{(u)}_{\bar{n}\times 1}(\boldsymbol{\delta}^{(v)}_{\bar{n}\times 1})']_{n,n}}{(s_{n,n})^{\frac{5}{2}}} 1_{s=n} \\
&= \textstyle\sum_{u,v} [\boldsymbol{b}' Diag(diag(\boldsymbol{\delta}^{(u)}_{\bar{n}\times 1}(\boldsymbol{\delta}^{(v)}_{\bar{n}\times 1})'))\boldsymbol{b}]_{t,k} \\
&\times [Diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}) Diag(diag(\boldsymbol{\delta}^{(u)}_{\bar{n}\times 1}(\boldsymbol{\delta}^{(v)}_{\bar{n}\times 1})')) Diag(\boldsymbol{\gamma}_1)]_{s,n} \\
&\overset{(87)}{=} \textstyle\sum_u [\boldsymbol{b}' Diag(\boldsymbol{\delta}^{(u)}_{\bar{n}\times 1})\boldsymbol{b}]_{t,k}[Diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}) Diag(\boldsymbol{\delta}^{(u)}_{\bar{n}\times 1}) Diag(\boldsymbol{\gamma}_1)]_{s,n},
\end{aligned}
\tag{137}
$$

where γ_1 is an $\bar{n}$ x 1 vector defined as follows

$$
\begin{aligned}
\boldsymbol{\gamma}_1 &\equiv (\frac{\mu_1}{(s_{1,1})^{\frac{5}{2}}}, \dots, \frac{\mu_{\bar{n}}}{(s_{\bar{n},\bar{n}})^{\frac{5}{2}}})' \\
&= \boldsymbol{\mu}./diag(\boldsymbol{s})^{\frac{5}{2}}.
\end{aligned}
\tag{138}
$$

Replacing (137) in (135) we obtain

$$
\begin{aligned}
\frac{\partial^2 g_m}{\partial b_{s,t}\partial b_{n,k}} &= \textstyle\sum_{i,u} [(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1})'\boldsymbol{a}']_m (3[\boldsymbol{b}' Diag(\boldsymbol{\delta}^{(u)}_{\bar{n}\times 1})\boldsymbol{b}]_{t,k} \\
&\times [Diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1})) Diag(\boldsymbol{\delta}^{(u)}_{\bar{n}\times 1}) Diag(\boldsymbol{\gamma}_1)]_{s,n} \\
&- [\boldsymbol{i}_{\bar{k}\times\bar{k}}]_{t,k}[Diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}) Diag(\boldsymbol{\gamma}_2)]_{s,n}),
\end{aligned}
\tag{139}
$$

where γ_2 is an $\bar{n}$ x 1 vector defined as follows

$$
\begin{aligned}
\boldsymbol{\gamma}_2 &\equiv (\frac{\mu_1}{(s_{1,1})^{\frac{3}{2}}}, \dots, \frac{\mu_{\bar{n}}}{(s_{\bar{n},\bar{n}})^{\frac{3}{2}}})' \\
&= \boldsymbol{\mu}./diag(\boldsymbol{s})^{\frac{3}{2}}.
\end{aligned}
\tag{140}
$$

Hence

$$
\begin{aligned}
\nabla^2_{\boldsymbol{b},\boldsymbol{b}} g &= \textstyle\sum_{i,u}((\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1})'\boldsymbol{a}') \otimes (3(\boldsymbol{b}' Diag(\boldsymbol{\delta}^{(u)}_{\bar{n}\times 1})\boldsymbol{b}) \otimes (Diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}) \\
&\times Diag(\boldsymbol{\delta}^{(u)}_{\bar{n}\times 1}) Diag(\boldsymbol{\gamma}_1)) - \boldsymbol{i}_{\bar{k}\times\bar{k}} \otimes (Diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1}) Diag(\boldsymbol{\gamma}_2))) \\
&= 3\textstyle\sum_i ((\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1})'\boldsymbol{a}') \otimes ((\boldsymbol{b}' \otimes Diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1})) \\
&\times Diag(\textstyle\sum_u \boldsymbol{\delta}^{(u)}_{\bar{n}\times 1} \otimes \boldsymbol{\delta}^{(u)}_{\bar{n}\times 1})(\boldsymbol{b} \otimes Diag(\boldsymbol{\gamma}_1))) \\
&- \textstyle\sum_i ((\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1})'\boldsymbol{a}') \otimes [(\boldsymbol{i}_{\bar{k}\times\bar{k}} \otimes Diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1})(\boldsymbol{i}_{\bar{k}\times\bar{k}} \otimes Diag(\boldsymbol{\gamma}_2)))] \\
&\overset{(121)}{=} 3\textstyle\sum_i [(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1})' \otimes \boldsymbol{i}_{\bar{k}\times\bar{k}} \otimes Diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1})] \\
&\times \boldsymbol{a}' \otimes [(\boldsymbol{b}' \otimes \boldsymbol{i}_{\bar{n}\times\bar{n}})\boldsymbol{h}_{\bar{n}^2\times\bar{n}^2}(\boldsymbol{b} \otimes Diag(\boldsymbol{\gamma}_1))] \\
&- \textstyle\sum_i [(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1})' \otimes \boldsymbol{i}_{\bar{k}\times\bar{k}} \otimes Diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1})]\boldsymbol{a}' \otimes [\boldsymbol{i}_{\bar{k}\times\bar{k}} \otimes Diag(\boldsymbol{\gamma}_2))] \\
&\overset{(138)\text{-}(140)}{=} \boldsymbol{k}_{\bar{k}\bar{n}\times\bar{k}\bar{n}^2}\{\boldsymbol{a}' \otimes [3(\boldsymbol{b}' \otimes \boldsymbol{i}_{\bar{n}\times\bar{n}})\boldsymbol{h}_{\bar{n}^2\times\bar{n}^2} \\
&\times (\boldsymbol{b} \otimes Diag(\boldsymbol{\mu}./diag(\boldsymbol{s})^{\frac{5}{2}})) - (\boldsymbol{i}_{\bar{k}\times\bar{k}} \otimes Diag(\boldsymbol{\mu}./diag(\boldsymbol{s})^{\frac{3}{2}}))]\},
\end{aligned}
\tag{141}
$$

where $\boldsymbol{k}_{\bar{k}\bar{n}\times\bar{k}\bar{n}^2}$ is an $\bar{k}\bar{n}\times\bar{k}\bar{n}^2$ matrix defined as follows

$$\boldsymbol{k}_{\bar{k}\bar{n}\times\bar{k}\bar{n}^2} \equiv \textstyle\sum_i [(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1})' \otimes \boldsymbol{i}_{\bar{k}\times\bar{k}} \otimes Diag(\boldsymbol{\delta}^{(i)}_{\bar{n}\times 1})]. \tag{142}$$

Hence, to summarise, using the compact notation of Appendix A.2, the Hessian of the views (36) reads

$$\nabla^2_{\boldsymbol{\mu},\boldsymbol{\mu}} v = \boldsymbol{0}_{\bar{n}\times\bar{n}\bar{m}} \tag{143}$$

$$\nabla^2_{\boldsymbol{d},\boldsymbol{\mu}} v = -(\boldsymbol{d}' \otimes \boldsymbol{i}_{\bar{n}\times\bar{n}})\boldsymbol{h}_{\bar{n}^2\times\bar{n}^2}(\boldsymbol{a}' \otimes Diag(\boldsymbol{\sigma}^3_{vec})) \tag{144}$$

$$\nabla^2_{\boldsymbol{b},\boldsymbol{\mu}} v = -(\boldsymbol{b}' \otimes \boldsymbol{i}_{\bar{n}\times\bar{n}})\boldsymbol{h}_{\bar{n}^2\times\bar{n}^2}(\boldsymbol{a}' \otimes Diag(\boldsymbol{\sigma}^3_{vec})) \tag{145}$$

$$\nabla^2_{\boldsymbol{b},\boldsymbol{d}} v = 3(\boldsymbol{b}' \otimes \boldsymbol{i}_{\bar{n}\times\bar{n}})\boldsymbol{h}_{\bar{n}^2\times\bar{n}^2}(\boldsymbol{a}' \otimes Diag((\boldsymbol{\mu}\circ\boldsymbol{d})./\boldsymbol{\sigma}^5_{vec})) \tag{146}$$

$$\nabla^2_{\boldsymbol{d},\boldsymbol{d}} v = \boldsymbol{h}_{\bar{n}\times\bar{n}^2}(\boldsymbol{a}' \otimes Diag(\boldsymbol{\mu}\circ(3(\boldsymbol{d}\circ\boldsymbol{d})./\boldsymbol{\sigma}^5_{vec} - 1./\boldsymbol{\sigma}^3_{vec}))) \tag{147}$$

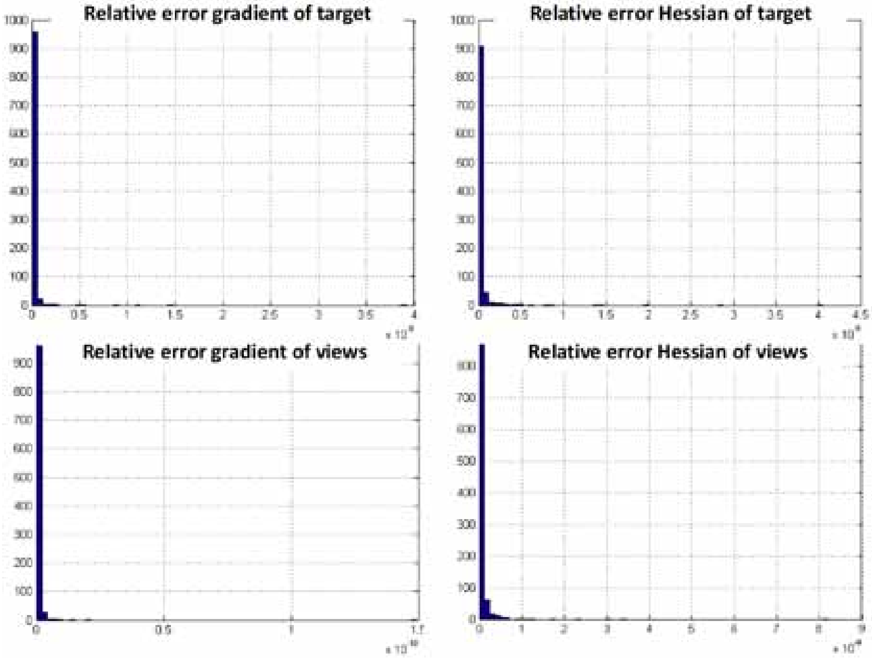
$$\nabla^2_{\boldsymbol{b},\boldsymbol{b}} v = \boldsymbol{k}_{\bar{k}\bar{n}\times\bar{k}\bar{n}^2}\{\boldsymbol{a}' \otimes [3(\boldsymbol{b}' \otimes \boldsymbol{i}_{\bar{n}\times\bar{n}})\boldsymbol{h}_{\bar{n}^2\times\bar{n}^2}(\boldsymbol{b} \otimes Diag(\boldsymbol{\mu}./\boldsymbol{\sigma}^5_{vec})) - (\boldsymbol{i}_{\bar{k}\times\bar{k}} \otimes Diag(\boldsymbol{\mu}./\boldsymbol{\sigma}^3_{vec}))]\}, \tag{148}$$

where $\boldsymbol{h}_{\bar{n}\times\bar{n}^2}$ is defined in (134) and $\bar{k}\bar{n}\times\bar{k}\bar{n}^2$ is defined in (142).

A.10 Formulas verification

To verify that the formulas in Appendix A.5-A.6-A.8-A.9 are correct, we perform extensive numerical testing. More precisely, we generate randomly market parameters $\theta^{(j)} \equiv \boldsymbol{\mu}^{(j)}, \boldsymbol{b}^{(j)}, \boldsymbol{d}^{(j)}$ in dimension $\bar{n}$ = 8 and $\bar{k}$ = 3 for a large number of scenarios $\bar{j}$ = 1000.

Figure 8: Relative error of analytical formulas with respect to nite di¤erence approximation.



Then, for each market scenario j, we compute the analytical gradients $\nabla_{\theta^{(j)}}$ in (15)-(17) and in (38)-(40). Next, we compute the numerical counterparts of the gradients $\hat{\nabla}_{\boldsymbol{\theta}^{(j)}}$, using the finite difference approximation routines from [D'Errico, 2007].

Then, we compute the relative error in each market scenario

Finally, in Figure 8 we visualise the extent of the relative error by plotting the histogram of

$$h^{(j)} \equiv \|\nabla_{\boldsymbol{\theta}^{(j)}} - \hat{\nabla}_{\boldsymbol{\theta}^{(j)}}\| / \|\nabla_{\boldsymbol{\theta}^{(j)}}\|. \tag{149}$$

We perform the same evaluation for the analytical Hessians $\nabla_{\theta^{(j)}}$ in (97)-(102) and in (143)-(148), computing the relative errors

$$h^{(j)} \equiv \|\nabla^2_{\boldsymbol{\theta}^{(j)}} - \hat{\nabla}^2_{\boldsymbol{\theta}^{(j)}}\|_F / \|\nabla^2_{\boldsymbol{\theta}^{(j)}}\|_F, \tag{150}$$

where $\| . \|_F$ denotes the Frobenius norm. In Figure 8 we also visualise the extent of the relative error by plotting the histogram of (150). As we can appreciate in these gures, the error, which is due to the finite-difference approximations, is consistently negligible.

References

• [Almgren and Chriss, 2006] Almgren, R. and Chriss, N. (2006). Optimal portfolios from ordering information. *Journal of Risk*, 9:1—47.

• [Ardia and Meucci, 2013] Ardia, D. and Meucci, A. (2013). Heavy stress-testing in non-normal markets via entropy pooling. http://symmys.com/node/160. Working Paper.

• [Asness et al., 2013a] Asness, C., Frazzini, A., and Pedersen, L. (2013a). Quality minus junk. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2312432. Working Paper.

• [Asness et al., 2013b] Asness, C. S., Moskowitz, T. J., and Pedersen, L. H. (2013b). Value and momentum everywhere. *The Journal of Finance*, 68(3):929—985.

• [Bailey et al., 2013] Bailey, D. H., Borwein, J. M., Lopez de Prado, M., and Zhu, Q. J. (2013). Pseudo-mathematics and financial charlatanism: The effects of backtest overfitting on out-of-sample performance. http://ssrn.com/abstract=2308659. Working Paper.

• [Black and Litterman, 1990] Black, F. and Litterman, R. (1990). Asset allocation: combining investor views with market equilibrium. Goldman Sachs Fixed Income Research.

• [De Bondt and Thaler, 1985] De Bondt, W. and Thaler, R. (1985). Does the stock market overreact. *The Journal of Finance*, 40(3):793â¼A ¸S805.

• [D'Errico, 2007] D'Errico, J. (2007). Adaptive robust numerical differentiation. http://www.mathworks.com/matlabcentral/fileexchange/13490.

• [Grinold and Kahn, 1999] Grinold, R. C. and Kahn, R. (1999). Active Portfolio Management. A Quantitative Approach for Producing Superior Returns and Controlling Risk. McGraw-Hill, 2nd edition.

• [Levy and Roll, 2010] Levy, M. and Roll, R. (2010). The market portfolio may be Mean/Variance efficient after all. *The Review of Financial Studies*, 23(6):2464—2491.

• [Lobo et al., 2007] Lobo, M., Fazel, M., and Boyd, B. (2007). Portfolio optimisation with linear and xed transaction costs. *Annals of Operations Research*, 152(1):341—365.

• [Magnus and Neudecker, 1979] Magnus, J. R. and Neudecker, H. (1979). The commutation matrix: Some properties and applications. *Annals of Statistics*, 7:381—394.

• [Menchero et al., 2013] Menchero, J., Morozov, A., and Pasqua, A. (2013). Predicting risk at short horizons. http://www.lqg.org.uk/wp-content/uploads/2013/02/ Predicting-Risk-at-Short-Horizons-Model-Insight.pdf.

• [Meucci, 2005] Meucci, A. (2005). Risk and asset allocation. http://symmys.com. Springer Finance.

• [Meucci, 2008] Meucci, A. (2008). Fully Flexible Views: Theory and practice. http://symmys.com/node/158. Risk, 21(10), 97-102.

• [Meucci, 2009] Meucci, A. (2009). Simulations with exact means and covariances. http://symmys.com/node/162. Risk, 22(7), 89-91.

• [Meucci, 2010] Meucci, A. (2010). Historical scenarios with Fully Flexible Probabilities. http://symmys.com/node/150. Working Paper.

• [Meucci, 2012] Meucci, A. (2012). Effective number of scenarios with Fully Flexible Probabilities. http://symmys.com/node/362. GARP Risk Professional, 45-46.

• [Meucci, 2013] Meucci, A. (2013). Estimation and stress-testing via time- and market- conditional flexible probabilities. http://symmys.com/node/600. Working Paper.

• [Meucci et al., 2011] Meucci, A., Ardia, D., and Keel, S. (2011). Fully Flexible Extreme Views. http://symmys.com/node/159. *Journal of Risk*, 14(2), 39-49.

• [Moskowitz et al., 2012] Moskowitz, T., Ooi, Y. H., and Pedersen, L. H. (2012). Time series momentum. *Journal of Financial Economics*, 104(228-250):2.

• [Ni et al., 2011] Ni, X., Malevergne, Y., Sornette, D., and Woehrmann, P. (2011). Robust reverse engineering of cross sectional returns and improved portfolio allocation performance using the CAPM. *The Journal of Portfolio Management*, 37:76—85.

• [Park, 2010] Park, S. C. (2010). The moving average ratio and momentum. *Financial Review*, 45(2):415—447.

• [Wang and Kochard, 2011] Wang, P. and Kochard, L. (2011). Using a z-score approach to combine value and momentum in tactical asset allocation. http://papers.ssrn.com/ sol3/papers.cfm?abstract_id=1726443. Working Paper.-

Founded in 1906, EDHEC Business School offers management education at undergraduate, graduate, post-graduate and executive levels. Holding the AACSB, AMBA and EQUIS accreditations and regularly ranked among Europe's leading institutions, EDHEC Business School delivers degree courses to over 6,000 students from the world over and trains 5,500 professionals yearly through executive courses and research events. The School's 'Research for Business' policy focuses on issues that correspond to genuine industry and community expectations.

Established in 2001, EDHEC-Risk Institute has become the premier academic centre for industry-relevant financial research. In partnership with large financial institutions, its team of ninety permanent professors, engineers, and support staff, and forty-eight research associates and affiliate professors, implements six research programmes and sixteen research chairs and strategic research projects focusing on asset allocation and risk management. EDHEC-Risk Institute also has highly significant executive education activities for professionals. It has an original PhD in Finance programme which has an executive track for high level professionals. Complementing the core faculty, this unique PhD in Finance programme has highly prestigious affiliate faculty from universities such as Princeton, Wharton, Oxford, Chicago and CalTech.

In 2012, EDHEC-Risk Institute signed two strategic partnership agreements with the Operations Research and Financial Engineering department of Princeton University to set up a joint research programme in the area of risk and investment management, and with Yale School of Management to set up joint certified executive training courses in North America and Europe in the area of investment management.

For more information, please contact:
Carolyn Essid on +33 493 187 824
or by e-mail to: carolyn.essid@edhec-risk.com

EDHEC-Risk Institute
393 promenade des Anglais
BP 3116 - 06202 Nice Cedex 3
France
Tel: +33 (0)4 93 18 78 24

EDHEC Risk Institute—North America
One Boston Place, 201 Washington Street
Suite 2608/2640
Boston, MA 02108
Tel: +1 857 239 8891

EDHEC Risk Institute—Europe
10 Fleet Place, Ludgate
London EC4M 7RB
United Kingdom
Tel: +44 207 871 6740

EDHEC Risk Institute—France
16-18 rue du 4 septembre
75002 Paris
France
Tel: +33 (0)1 53 32 76 30

EDHEC Risk Institute—Asia
1 George Street
#07-02
Singapore 049145
Tel: +65 6438 0030

www.edhec-risk.com