



Centre Interuniversitaire sur le Risque,
les Politiques Économiques et l'Emploi

Cahier de recherche/Working Paper **13-11**

Fully Flexible Views in Multivariate Normal Markets

Attilio Meucci
David Ardia
Simon Keel

Mai/May 2013

Meucci : SYMMYS

attilio_meucci@symmys.com

Ardia : Université Laval, Faculté des sciences de l'administration, Département de finance, assurance et immobilier;
et CIRPÉE

david.ardia@fsa.ulaval.ca

Keel : Aeris Capital AG

simon.keel@aeris-capital.com

Abstract:

The Entropy Pooling approach in Meucci (2008) is a versatile, general framework to process market views in portfolio construction and generalized stress-tests in risk management. Here we present an efficient algorithm to implement Entropy Pooling with fully general views in multivariate normal markets.

Then we discuss two applications. First, we use normal Entropy Pooling to estimate a market distribution consistent with the CAPM equilibrium, which improves on the “implied returns” a-la-Black and Litterman (1990) and can be used as the starting point for portfolio construction. Second, we use normal Entropy Pooling to process ranking signals for alpha-generation.

Keywords: Portfolio construction, tactical allocation, Entropy Pooling, Kullback-Leibler, Black-Litterman, equilibrium prior, portfolios from sorts, ranking, alpha, signals, factor models, risk management

JEL Classification: C1, G11

1 Introduction

The combination of subjective views on the market within a prior risk model to compute an optimal allocation that incorporates the views 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 generalized Bayesian approach "entropy pooling" (EP) is a general, flexible framework to process views and generalized stress-tests.

The general theoretical framework for EP was laid out in Meucci (2008). EP combines an arbitrary market model, which is referred to as the "prior" and fully general views or stress-tests on the underlying market. The output is a distribution, referred to as the "posterior", which incorporates all the inputs and which can be used for risk management and portfolio optimization. In EP, 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 minimizes the entropy relative to the prior, which is the natural measure of discrepancy between two distributions.

The EP framework can be implemented in two ways: non-parametrically, by representing the prior and the posterior in terms of a scenarios-probabilities pair; and parametrically, by making parametric assumptions on the prior and the posterior.

The nonparametric implementation of EP was studied in the original article Meucci (2008) and further extended to handle views on tail risk in Meucci, Ardia, and Keel (2011). The non-parametric approach is very flexible, but subject to the curse of dimensionality. Therefore, it is most effectively applied in risk management contexts, see Meucci (2010).

The parametric implementation of EP was studied in the original article Meucci (2008) under the normal assumption with views set as equalities on expectations and covariances. Under these special types of views, the posterior can be computed analytically.

In this article we study the parametric implementation of EP in normal markets, but with more general views. To compute the posterior, we propose an efficient numerical approach. First, we impose structure on the correlations, thereby increasing the statistical efficiency of our estimates. Second, we compute analytically the gradient of our objective function, namely the relative entropy with the prior, thereby speeding the convergence of our numerical approach. Third, we feed our inputs into an interior point optimizer.

Our framework, namely normal EP with general views, allows us to address a variety of problems. In particular, we discuss two applications.

The first application is the estimation of a market distribution consistent with equilibrium. Such equilibrium estimate is the starting point of portfolio construction in Black and Litterman (1990). In the original paper a covariance of returns is estimated and the "implied" expected returns are then computed accordingly. However, these expected returns are substantially different from their historical counterparts. This issue has been addressed by Levy and Roll (2010)

by simultaneously computing implied expected returns and implied volatilities. Normal EP with general views computes at the same time implied expected returns, implied volatilities and implied correlations, thereby obtaining a closer distribution to the empirical observations.

The second application of the normal EP framework is the extraction of ranking signals for alpha-generation. In the standard approach, discussed e.g. in Grinold and Kahn (1999), the expected return of all the securities in a given market is set proportional to the z-score of a given predictive signal. However, assuming the expected return proportional to the z-score imposes spurious information in the optimization process. Almgren and Chriss (2006) first addressed this issue, obtaining expected returns that do not overly spurious information. However, their solution does not take empirical data into account. EP effectively estimates ranking-consistent expected returns that do not impose spurious information and at the same time starts from the empirical observations. The case study in Meucci (2008) addresses the low-dimensional case non-parametrically. The present normal EP framework overcomes the curse of dimensionality in large markets.

We emphasize that the multivariate normal specification for the risk drivers that lies at the foundation of our approach is by no means restricted to model normal returns. By applying non-linear pricing functions to the drivers, the normal specification is suitable to model for instance highly skewed markets of options, see Meucci (2009).

The remainder of this article is organized as follows. In Section 2 we review the original general EP framework. In Section 3 we discuss the normal EP implementation. In Section 4 we show how to use normal EP to specify the equilibrium prior. In Section 5 we illustrate how to use normal EP to process ranking signals for portfolio construction.

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

2 Review of Entropy Pooling

In this section we draw from Meucci (2008), please refer to that publication for more details.

EP proceeds in three main steps. The first step of EP is the estimation of a "prior" distribution for a set of N risk drivers $X \equiv (X_1, \dots, X_N)$ in the market, as represented by its pdf, which we denote by \underline{f}

$$X \equiv (X_1, \dots, X_N) \sim \underline{f}. \quad (1)$$

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

The second step of EP is expressing the views or stress-tests \mathcal{V} . These are statements on expectations, correlations, tail risk conditions, etc. that we want the yet to be defined posterior distribution to satisfy. Therefore, views and stress-tests \mathcal{V} are constraints on the posterior. We denote that a generic

distribution f satisfies these constraints as follows

$$f \in \mathcal{V}. \quad (2)$$

The third step of EP is the computation of the posterior distribution \bar{f} for the risk drivers, which incorporates the views 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}

$$\mathcal{E}(f|\underline{f}) \equiv \int f(x) \ln \frac{f(x)}{\underline{f}(x)} dx. \quad (3)$$

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 the prior, satisfies the views \mathcal{V} . Therefore, we define the posterior as follows

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

The posterior distribution \bar{f} is then used as input to an optimizer 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.

EP can be implemented in two ways: non-parametric and parametric.

In the non-parametric approach the prior \underline{f} is represented in terms of a large number S of joint scenarios for the risk drivers and the associated probabilities $\{x_{1,s}, \dots, x_{N,s}; \underline{p}_s\}_{s=1, \dots, S}$. Then the posterior (4) is represented by the same scenarios with a new set of probabilities $\{x_{1,s}, \dots, x_{N,s}; \bar{p}_s\}_{s=1, \dots, S}$ defined as follows

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

where with minor abuse we let $\mathcal{E}(p|\underline{p}) \equiv \sum_{s=1}^S p_s \ln(p_s/\underline{p}_s)$ denote the discrete counterpart of the relative entropy (3). As it turns out, for several types of views the optimization (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. $f \equiv f_\theta$, where the parameters θ span a set of values Θ . In particular, the prior is represented by f_θ and the posterior (4) becomes

$$f_{\bar{\theta}} \equiv \underset{\substack{\theta \in \Theta \\ f_\theta \in \mathcal{V}}}{\operatorname{argmin}} \mathcal{E}(f_\theta|f_\theta). \quad (6)$$

A special case of the parametric approach is the normal assumption

$$f_{\mu, \sigma^2}(x) = (2\pi)^{-\frac{N}{2}} |\sigma^2|^{-\frac{1}{2}} e^{-\frac{1}{2}(x-\mu)' \sigma^{-2}(x-\mu)}, \quad (7)$$

where μ is a $N \times 1$ vector of expectations and σ^2 is a $N \times N$ covariance matrix. Therefore, the parameters are $\theta \equiv (\mu, \sigma^2)$, where σ^2 is constrained to be

symmetric and positive definite, which we denote by $\sigma^2 \succ 0$. Accordingly, under normality the parametric problem (6) becomes

$$(\bar{\mu}, \bar{\sigma}^2) \equiv \underset{\substack{\sigma^2 \succ 0 \\ \mu, \sigma^2 \in \mathcal{V}}}{\operatorname{argmin}} \mathcal{E}(\mu, \sigma^2 | \underline{\mu}, \underline{\sigma}^2), \quad (8)$$

where the relative entropy between two normal distributions can be computed explicitly and reads

$$2\mathcal{E}(\mu, \sigma^2 | \underline{\mu}, \underline{\sigma}^2) = \operatorname{tr}(\sigma^2 \underline{\sigma}^2^{-1}) - \ln |\sigma^2 \underline{\sigma}^2^{-1}| + (\mu - \underline{\mu})' \underline{\sigma}^2^{-1} (\mu - \underline{\mu}) - N. \quad (9)$$

The normal EP problem (8) can be solved analytically when the views are equality statements on expectations $E\{aX\} \equiv \xi$ and covariances $\operatorname{Cov}\{bX\} \equiv \phi$, or

$$\mathcal{V}: \quad a\mu \equiv \xi, \quad b\sigma^2 b' \equiv \phi. \quad (10)$$

Then the solution of (8)-(10) reads

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

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

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

3 Normal EP with fully flexible views

Here we address the parametric implementation of EP under normality (8), with fully flexible views \mathcal{V} beyond (10). In this case, the solution must be computed numerically.

To this purpose, we impose that the covariances be of "low rank + diagonal" type

$$\sigma^2 \equiv bb' + \delta^2, \quad (13)$$

where b is a $N \times K$ matrix, δ is a $N \times N$ diagonal matrix, and, in large markets, $K \ll N$. With the parametrization (13), the EP problem (8) becomes

$$(\bar{\mu}, \bar{b}, \bar{\delta}) \equiv \underset{\mu, b, \delta \in \mathcal{V}}{\operatorname{argmin}} \mathcal{E}(\mu, bb' + \delta^2 | \underline{\mu}, \underline{\sigma}^2). \quad (14)$$

The normal EP optimization (14) presents palatable statistical and numerical features.

From a statistical perspective, the "low rank + diagonal" parametrization of the normal distribution is fully determined by a relatively small number $N(K + 2)$ of parameters $\theta \equiv (\mu, b, \delta)$, instead of the large number $N(N + 3)/2$ of parameters in the full-blown specification $\theta \equiv (\mu, \sigma^2)$. The structure imposed on the correlations by the limited number of parameters substantially enhances the statistical efficiency of the estimates in large dimensional markets.

From a numerical point of view, the parsimonious parametrization $\theta \equiv (\mu, b, \delta)$ is unconstrained, as the parameters can freely range in the space $\Theta \equiv$

$\mathbb{R}^N \times \mathbb{R}^{N \times K} \times \mathbb{R}^N$. Instead, in the full specification $\theta \equiv (\mu, \sigma^2)$ the matrix σ^2 is constrained to be symmetric and positive definite. Furthermore, we can enhance the computational efficiency of the optimization by feeding the analytical expression of the gradient in the interior point algorithms. Indeed, in the technical appendix, we obtain the following results

$$\frac{\partial}{\partial \mu} \mathcal{E}(\mu, bb' + \delta^2 | \underline{\mu}, \underline{\sigma}^2) = (\mu - \underline{\mu})' \underline{\sigma}^2{}^{-1} \quad (15)$$

$$\frac{\partial}{\partial \delta} \mathcal{E}(\mu, bb' + \delta^2 | \underline{\mu}, \underline{\sigma}^2) = 2 \text{diag}(\underline{\sigma}^2{}^{-1} - \sigma^2{}^{-1}) \delta \quad (16)$$

$$\frac{\partial}{\partial b_{n,k}} \mathcal{E}(\mu, bb' + \delta^2 | \underline{\mu}, \underline{\sigma}^2) = 2 \sum_{m=1}^N (\underline{\sigma}^2{}^{-1} - \sigma^2{}^{-1})_{n,m} b_{m,k}, \quad (17)$$

where all the high-dimensional inverses $\sigma^2{}^{-1}$ and $\underline{\sigma}^2{}^{-1}$ are easily obtained analytically in terms of a low-cost, low-dimensional inverses as follows

$$\sigma^2{}^{-1} = \delta^{-2} - \delta^{-2} b (b' \delta^{-2} b + \mathbf{i}_K)^{-1} b' \delta^{-2}, \quad (18)$$

where \mathbf{i}_K is $K \times K$ identity matrix.

The parsimonious number of entries $N(K+2)$ in the parameters (μ, b, δ) , the fact that the parameters span the unconstrained set $\Theta \equiv \mathbb{R}^N \times \mathbb{R}^{N \times K} \times \mathbb{R}^N$, the analytical expression of the gradient (15)-(17) and the analytical inversion (18) makes it possible to converge efficiently to the optimal solution $(\bar{\mu}, \bar{b}, \bar{\delta})$ in the optimization (14) by means of off-the-shelf interior point algorithms. We proceed to show this in two case studies.

4 Case study: the equilibrium prior

In our first case study we use the normal EP framework with general views to determine the equilibrium distribution that lies at the foundation of the portfolio construction approach in Black and Litterman (1990).

Under some assumptions on the market distribution and the preferences of the investors, the capital asset pricing model purports that the market capitalization portfolio w_{MC} is linked to the distribution of the market returns by the identity

$$E\{R\} - \gamma Cov\{R\} w_{MC} \equiv 0, \quad (19)$$

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

For portfolio construction purposes, the condition (19) guarantees that, in the absence of specific views, a mean-variance optimization yields the portfolio w_{MC} . However, the condition (19) is not satisfied empirically by sample estimates of the expectations and the covariances.

To enforce (19), Black and Litterman (1990) propose a two-step approach. First, a covariance matrix $Cov\{R\}$ is fitted to empirical observations by means of standard techniques such as exponential smoothing or maximum likelihood; then the equilibrium constraint (19) is solved for the expectations $E\{R\}$. This

generates the so-called "implied expected returns". The implied expected returns are consistent with equilibrium, but can differ substantially from the true expected values, because no attempt is made to fit jointly covariances and expectations to the observed data.

To address this issue, Levy and Roll (2010), propose to fit a correlation matrix $\hat{\beta}$ to empirical observations and then ensure that the equilibrium constraint (19) is satisfied by modifying both the expectations and the variances. More precisely, the authors introduce a distance \mathcal{D} between the sample estimates of the expectations and the standard deviations $(\underline{\mu}, \underline{\sigma})$ and the yet to be defined parameters (μ, σ) as follows

$$\mathcal{D}(\mu, \sigma; \underline{\mu}, \underline{\sigma}) \equiv \left(\alpha \left\| \frac{\mu - \underline{\mu}}{\underline{\sigma}} \right\|^2 + (1 - \alpha) \left\| \frac{\sigma - \underline{\sigma}}{\underline{\sigma}} \right\|^2 \right)^{\frac{1}{2}}, \quad (20)$$

where $\|\cdot\|$ denotes the standard Euclidean norm, the division is meant entry-by-entry, and the authors set $\alpha \equiv 0.75$. Then the authors minimize \mathcal{D} with respect to (μ, σ)

$$(\bar{\mu}, \bar{\sigma}) \equiv \underset{\mu, \sigma \in \mathcal{V}}{\operatorname{argmin}} \mathcal{D}(\mu, \sigma; \underline{\mu}, \underline{\sigma}), \quad (21)$$

where the equilibrium constraint (19) now reads¹

$$\mathcal{V}: \quad \mu - \gamma \operatorname{diag}(\sigma) \hat{\beta} \operatorname{diag}(\sigma) w_{MC} \equiv 0. \quad (22)$$

Unlike in Black and Litterman (1990), the parameters $(\bar{\mu}, \bar{\sigma})$ are statistically indistinguishable from the sample estimates $(\underline{\mu}, \underline{\sigma})$ and thus equally acceptable from an estimation perspective. Furthermore, they give rise to better trading strategies, see Ni, Malevergne, Sornette, and Woehrmann (2011).

To further improve the estimation of the equilibrium distribution we can use our normal EP framework. Accordingly, we replace the Euclidean distance minimization (21) with the relative entropy minimization (14), which we report here

$$(\bar{\mu}, \bar{b}, \bar{\delta}) \equiv \underset{\mu, b, \delta \in \mathcal{V}}{\operatorname{argmin}} \mathcal{E}(\mu, bb' + \delta^2 | \underline{\mu}, \underline{bb}' + \underline{\delta}^2), \quad (23)$$

where the equilibrium constraint (19) now becomes the following view

$$\mathcal{V}: \quad \mu - \gamma (bb' + \delta^2) w_{MC} \equiv 0. \quad (24)$$

The EP equilibrium estimates $\bar{\mu}$ and $\bar{\sigma}^2 \equiv \bar{bb}' + \bar{\delta}^2$ improve on the previous approaches in three directions. First, EP replaces the somewhat arbitrary Euclidean distance between sample and equilibrium estimates with relative entropy, a statistically sound measure of discrepancy between distributions. Second, EP 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 parameters $(\bar{\mu}, \bar{b}, \bar{\delta})$ become less noisy and thus perform better out of sample.

¹For 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.

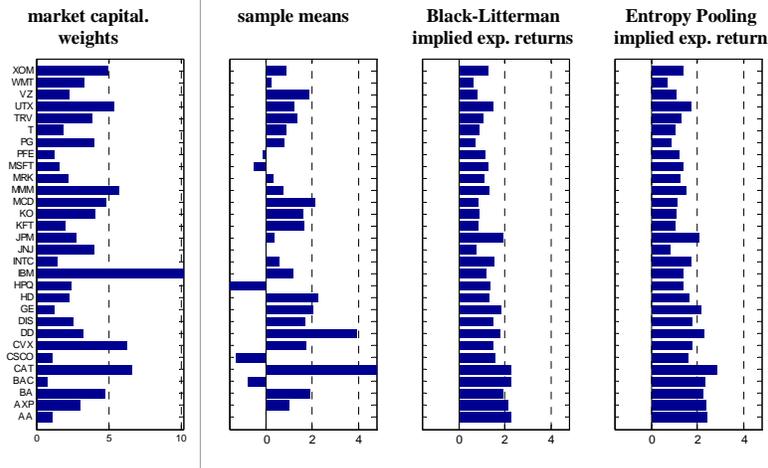


Figure 1: Equilibrium expected returns: Black-Litterman vs. Entropy Pooling

To illustrate the normal EP approach (23)-(24) in practice, we consider a market of $N \equiv 30$ equities in the Dow Jones Index. For those equities, we consider the time series of the most recent two year of daily returns. Then we fit the empirical prior parameter $\underline{\mu}$ as the sample mean, and $\underline{\sigma}^2 \equiv \underline{bb}' + \underline{\delta}^2$ by factor analysis decomposition of the sample covariance with $K \equiv 3$ hidden factors. Then we determine the equilibrium parameters $(\bar{\mu}, \bar{b}, \bar{\delta})$. Please refer to the code available at www.symmys.com/node/160 for more details.

In Figure 1 we report the capitalization weights w_{MC} , the sample means $\underline{\mu}$, the implied expected returns a-la Black-Litterman $\mu_{BL} \equiv \gamma \underline{\sigma}^2 w_{MC}$, and the EP equilibrium expected returns $\bar{\mu}$. As expected, the EP parameters $\bar{\mu}$ are more in line with the sample parameters $\underline{\mu}$ than the Black-Litterman parameters μ_{BL} .

To illustrate the effect of EP on correlations, in Figure 2, we display the location-dispersion ellipsoids that represent geometrically expectations and covariances, see Meucci (2005). We display the ellipsoids for the sample distribution $(\underline{\mu}, \underline{\sigma}^2)$, the Black-Litterman equilibrium $(\mu_{BL}, \underline{\sigma}^2)$ and the EP equilibrium $(\bar{\mu}, \bar{\sigma}^2)$ of three stocks.

Notice how the center (expectation), dispersion (variance) and orientation (correlation) of the ellipsoids are modified by the EP approach. Not only the locations are impacted by the views on the tangent portfolio, but the overall covariance structure is modified as well. Again, as expected, the EP parameters are more in line with the sample parameters than the Black-Litterman parameters. This illustrates that a constant correlation structure as in Levy and Roll (2010) is restrictive.

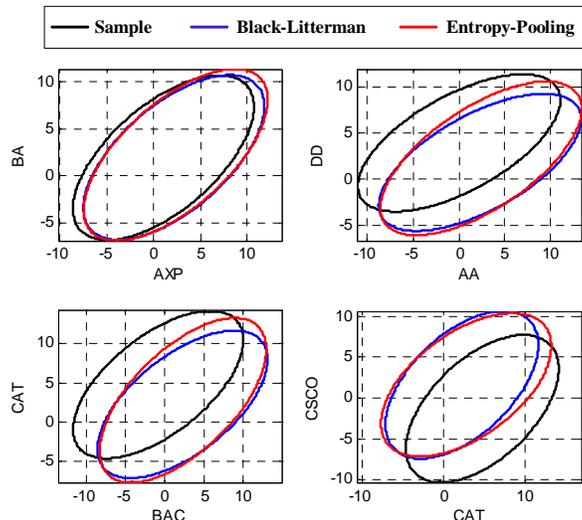


Figure 2: Equilibrium expected returns: Black-Litterman vs. Entropy Pooling

5 Case study: portfolios from sorts

The second case study addresses portfolio construction with predictive ranking signals. The most standard approach to this problem, popularized by Grinold and Kahn (1999), proceeds as follows.

First, an observable characteristic of a set of N securities, say for instance the price/earnings ratio for stocks, is believed to have predictive power. Then, the N securities are sorted according to the value of the given characteristic. In our example, the stock $n = 1$ has the lowest price/earnings, the stock $n = 2$ has the second-lowest price/earnings, and so on, until the stock $n = N$ has the highest price/earnings ratio. Next, an assumption is made that the expected returns of the securities are proportional to their relative ranking

$$E\{R_n\} \equiv \hat{\sigma}_n (n - N/2), \quad n = 1, \dots, N, \quad (25)$$

where $\hat{\sigma}_n$ is an estimate of the standard deviation of the n -th return; other specifications, where the expectation is set as proportional to the current historical z-score, are also common, but we do not explore them here for simplicity. Next, the correlation of the securities returns is estimated with standard techniques. Finally, an optimal portfolio is constructed by mean-variance optimization.

The most sensitive step in the above process is setting the expectations proportional to the ranking as in (25). The rationale behind this choice is that, if the signal is truly predictive, a lower ranking should give rise to a lower return

$$E\{R_n\} \leq E\{R_{n+1}\}, \quad n = 1, \dots, N - 1. \quad (26)$$

However, the proportional assumption (25) is much more invasive than the rationale (26).

To address this issue, Almgren and Chriss (2006) set the vector of expected returns as the "centroid", i.e. the average among all possible expected returns consistent with the ranking (26). However, the centroid 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.

EP allows us to process only the information available from the ranking signal (26) without imposing the spurious proportional structure (25), but at the same time keeping into full account the empirical data. To do so, we start by fitting a prior estimate for the expectations $\underline{\mu}$ and the covariances $\underline{\sigma}^2 \equiv \underline{bb}' + \underline{\delta}^2$ to the data available. Then we compute the posterior parameters by minimizing the relative entropy between the prior and the posterior as in (14), which we report here

$$(\bar{\mu}, \bar{b}, \bar{\delta}) \equiv \underset{\mu, b, \delta \in \mathcal{V}}{\operatorname{argmin}} \mathcal{E}(\mu, bb' + \delta^2 | \underline{\mu}, \underline{bb}' + \underline{\delta}^2). \quad (27)$$

In this case, the views are the ranking inequalities (26), which read as follows on the parameters

$$\mathcal{V}: \quad \mu_n \leq \mu_{n+1}, \quad n = 1, \dots, N - 1. \quad (28)$$

The posterior distribution defined by $(\bar{\mu}, \bar{b}, \bar{\delta})$ can then be used to optimize the portfolio.

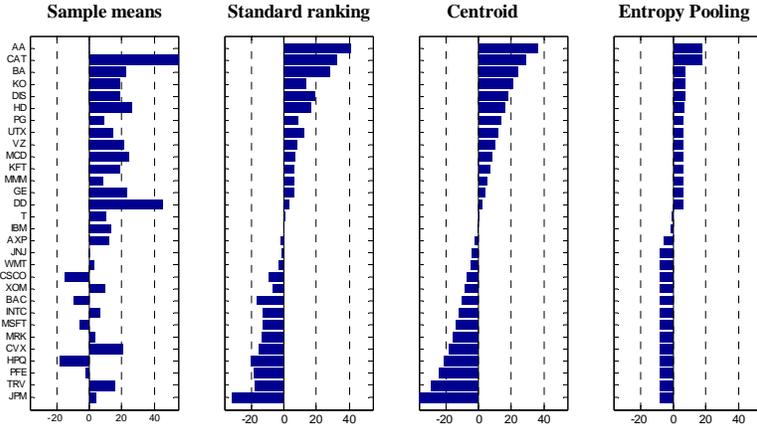


Figure 3: Expected returns for portfolio from sorts

To illustrate how EP applies to building portfolios from sorts, we consider again a market of $N \equiv 30$ equities in the Dow Jones Index. The prior is constructed from the available sample as in the previous case study. In Figure

3 we report the sample mean returns $\bar{\mu}$; the expected returns stemming from the standard approach (25); the rescaled centroid; and the EP-ranked expected returns (27)-(28), with the additional long-short constraint $\sum_{n=1}^N \mu_n \equiv 0$ to better compare the results with the other cases.

As expected, the EP results respect the desired ranking, but at the same time remain as close as possible to the empirical observations summarized by the sample means. Interestingly, though not unexpectedly, the EP expected returns appear to coincide with the centroid when the information from the empirical data is disregarded. For more details, please refer to the code available at www.symmys.com/node/160.

References

- Almgren, R., and N. Chriss, 2006, Optimal portfolios from ordering information, *Journal of Risk* 9, 1–47.
- Black, F., and R. Litterman, 1990, Asset allocation: combining investor views with market equilibrium, *Goldman Sachs Fixed Income Research*.
- Grinold, R. C., and R. Kahn, 1999, *Active Portfolio Management. A Quantitative Approach for Producing Superior Returns and Controlling Risk* (McGraw-Hill) 2nd edn.
- Levy, M., and R. Roll, 2010, The market portfolio may be Mean/Variance efficient after all, *The Review of Financial Studies* 23, 2464–2491.
- Meucci, A., 2005, *Risk and Asset Allocation* (Springer) Available at <http://symmys.com>.
- , 2008, Fully flexible views: Theory and practice, *Risk* 21, 97–102 Article and code available at <http://symmys.com/node/158>.
- , 2009, Enhancing the Black-Litterman and related approaches: Views and stress-test on risk factors, *Journal of Asset Management* 10, 89–96 Article and code available at <http://symmys.com/node/155>.
- , 2010, Historical scenarios with fully flexible probabilities, *GARP Risk Professional - "The Quant Classroom by Attilio Meucci"* December, 40–43 Article and code available at <http://symmys.com/node/150>.
- , D. Ardia, and S. Keel, 2011, Fully flexible extreme views, *Journal of Risk (to appear)* Article and code available at <http://symmys.com/node/159>.
- Ni, X., Y. Malevergne, D. Sornette, and P. Woehrmann, 2011, Robust reverse engineering of cross sectional returns and improved portfolio allocation performance using the CAPM, *Swiss Finance Institute Research Paper No. 11-03*.

A Gradient of relative entropy between normal distributions

Consider the relative entropy between two multivariate normal distributions

$$2\mathcal{E}(\mu, \sigma^2 | \underline{\mu}, \underline{\sigma}^2) = \text{tr}(\sigma^2 \underline{\sigma}^{-1}) - \ln |\sigma^2 \underline{\sigma}^{-1}| + (\mu - \underline{\mu})' \underline{\sigma}^{-1} (\mu - \underline{\mu}) - N. \quad (29)$$

Recall the following differentials with respect to the posterior

$$d \ln |\sigma^2 \underline{\sigma}^{-1}| = \text{tr}(\sigma^{-1} d\sigma^2) \quad (30)$$

$$d \text{tr}(\sigma^2 \underline{\sigma}^{-1}) = \text{tr}(\underline{\sigma}^{-1} d\sigma^2) \quad (31)$$

$$d\left((\mu - \underline{\mu})' \underline{\sigma}^{-1} (\mu - \underline{\mu})\right) = 2(\mu - \underline{\mu})' \underline{\sigma}^{-1} d\mu. \quad (32)$$

Then

$$d\mathcal{E} = \frac{1}{2} \text{tr}((\underline{\sigma}^{-1} - \sigma^{-1}) d\sigma^2) + (\mu - \underline{\mu})' \underline{\sigma}^{-1} d\mu. \quad (33)$$

We can impose

$$\sigma^2 \equiv bb' + \delta^2. \quad (34)$$

Then the inversion is fast

$$\sigma^{-1} = \delta^{-2} - \delta^{-2} b (b' \delta^{-2} b + \mathbf{i}_K)^{-1} b' \delta^{-2} \quad (35)$$

and the differential reads

$$d\sigma^2 = (db)b' + b(db') + 2\delta d\delta. \quad (36)$$

Therefore, we can minimize the relative entropy under linear constraints on σ^2 and μ with steepest ascent.

To compute the gradient, we write the first term ("f.t.") in (33) more explicitly as

$$\begin{aligned} \text{f.t.} &= \text{tr}((\underline{\sigma}^{-1} - \sigma^{-1}) d\sigma^2) = \sum_{n,m} (\underline{\sigma}^{-1} - \sigma^{-1})_{n,m} d\sigma_{n,m}^2 \\ &= \sum_{n,m} (\underline{\sigma}^{-1} - \sigma^{-1})_{n,m} (\sum_k b_{m,k} db_{n,k} + \sum_k b_{n,k} db_{m,k} + 2\Delta_{n,m} \delta_m d\delta_m) \\ &= \sum_{u,m,k} (\underline{\sigma}^{-1} - \sigma^{-1})_{u,m} b_{m,k} db_{u,k} \\ &\quad + \sum_{n,u,k} (\underline{\sigma}^{-1} - \sigma^{-1})_{n,u} b_{n,k} db_{u,k} \\ &\quad + 2 \sum_n (\underline{\sigma}^{-1} - \sigma^{-1})_{n,n} \delta_n d\delta_n \\ &= 2 \sum_{n,u,k} (\underline{\sigma}^{-1} - \sigma^{-1})_{u,n} b_{n,k} db_{u,k} + 2 \sum_n (\underline{\sigma}^{-1} - \sigma^{-1})_{n,n} \delta_n d\delta_n \\ &= \sum_{u,k} q_{u,k} db_{u,k} + \sum_n p_n d\delta_n, \end{aligned} \quad (37)$$

where

$$q_{n,k} \equiv 2 \sum_m (\underline{\sigma}^{-1} - \sigma^{-1})_{n,m} b_{m,k} \quad (38)$$

$$p_n \equiv 2(\underline{\sigma}^{-1} - \sigma^{-1})_{n,n} \delta_n. \quad (39)$$