Portfolio construction and systematic trading with factor entropy pooling

Construction of large portfolios consistent with investors' views and stress test scenarios is a challenging task, considering the volume of information to be processed. Attilio Meucci, David Ardia and Marcello Colasante introduce a technique that significantly reduces the effort needed and can account for more flexible views compared with existing methods

Processing trading signals or 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 known as entropy pooling, which is laid out in full generality in Meucci (2008), is a flexible framework for processing views and embedding generalised stress tests. The inputs to entropy pooling are an arbitrary market distribution called the prior distribution and generalised views on the market, such as views held by the portfolio manager on spread, tail distribution, volatility and correlation, or stress tests. The output is a distribution called the posterior distribution that is of minimum deviation from the prior, but at the same time satisfies the generalised market views, often unlike the prior. The posterior is then used to generate mean-variance optimal portfolios that reflect the market views or the stressed scenarios.

The parametric implementation of entropy pooling was studied in the original article under more restrictive special views – namely, equalities on expectations and covariances – where the posterior could be computed analytically. Here, this is extended to fully general views, such as inequality statements, non-linear views and rankings based on market observables, which are more flexible and allow construction of systematic trading strategies. The limitation is that the posterior cannot be computed analytically, making implementation a big challenge, especially for large markets. We introduce an efficient numerical method called factor entropy pooling (FEP), which reduces the dimension of the asset correlation structure using a factor model methodology and selects coordinates such that the optimisation target becomes unconstrained. FEP can be used in large-dimensional problems typical of portfolio construction. In particular, two applications are discussed.

The first application of FEP is the estimation of the implied returns – backed out from market weights on assets and historical covariances – consistent with a target optimal portfolio, such as one with maximum diversification/risk parity, or a capital asset pricing model-like equilibrium. Implied returns were first proposed in Black and Litterman (1990) as the starting point for the construction of a sensible mean-variance portfolio. The implied returns based on FEP improve on the Black-Litterman approach by being closer to the original market data.

The second application of FEP is construction of quantitative trading strategies based on ranking signals for alpha generation, or constructing the so-called portfolios from sorts, where the alphas (returns in excess of those predicted) of the assets are assumed to be an increasing function (ranking) of certain characteristics of the assets.

In the standard approach (discussed, for example, 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 on the optimisation process. Indeed, all that we should assume is that the stronger the signal, the higher the alpha, without imposing a strict linear relationship between signal and alpha. Almgren and Chriss (2006) were the first to address this issue, but their solution does not take empirical data into account. FEP effectively estimates ranking-consistent expected returns that do not impose spurious information and at the same time starts from empirical observations.

Notation.

$\boldsymbol{m} = \mathrm{DgMatr}(\boldsymbol{v})$	$\bar{n} \times \bar{n}$ matrix of zeros, except principal diagonal, which is $\bar{n} \times 1$ vector $\boldsymbol{v}$
$\boldsymbol{v} = \mathrm{DgVect}(\boldsymbol{m})$	$\bar{n} \times 1$ vector equals principal diagonal of $\bar{n} \times \bar{n}$ matrix $\boldsymbol{m}$
$\boldsymbol{\sigma}^2$	$\bar{n} \times \bar{n}$ symmetric, positive (semi-)definite matrix
$\boldsymbol{\sigma}_{\mathrm{vec}} \equiv \sqrt{\mathrm{DgVect}(\boldsymbol{\sigma}^2)}$	$\bar{n} \times 1$ vector of square roots of principal diagonal of $\boldsymbol{\sigma}^2$
$\boldsymbol{\sigma}$	$\bar{n} \times \bar{n}$ symmetric matrix $\boldsymbol{\sigma}$ such that $\boldsymbol{\sigma}\boldsymbol{\sigma}' = \boldsymbol{\sigma}^2$
$n = 1, \ldots, \bar{n}$	indexes of market entries ($\bar{n}$ is the market dimension)
$k = 1, \ldots, \bar{k}$	factor indexes ($\bar{k}$ is the number of factors)
$m = 1, \ldots, \bar{m}$	view/constraint indexes ($\bar{m}$ is the number of views/constraints)
$j = 1, \ldots, \bar{j}$	scenario indexes ($\bar{j}$ is the number of scenarios)

Review of entropy pooling

This section draws from Meucci (2008). Entropy pooling (EP) proceeds via three main steps. The first step 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, which we denote by $f_{\underline{\boldsymbol{\theta}}}$, where $\underline{\boldsymbol{\theta}}$ is a set of parameters that fully determines the prior:

$$\boldsymbol{X} \sim f_{\underline{\boldsymbol{\theta}}} \tag{1}$$

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

The second step of EP is expressing the views or embedding stress tests $\mathcal{V}$. These are statements on expectations, correlations, tail risk conditions, etc, that possibly contradict the prior, but should be included in risk management or allocation. For instance, the prior could represent a regular regime in the markets, and the views/stress test could be a regime in which 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 of the market. We denote that a parametric distribution $f_{\boldsymbol{\theta}}$ satisfies these constraints as follows:

$$\boldsymbol{\theta} \in \mathcal{V} \tag{2}$$

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

$$\mathcal{E}(\boldsymbol{\theta} \| \underline{\boldsymbol{\theta}}) \equiv \int f_{\boldsymbol{\theta}}(\boldsymbol{x}) \ln\left(\frac{f_{\boldsymbol{\theta}}(\boldsymbol{x})}{f_{\underline{\boldsymbol{\theta}}}(\boldsymbol{x})}\right) \mathrm{d}\boldsymbol{x} \tag{3}$$

We then define the posterior $f_{\bar{\boldsymbol{\theta}}}$ as the distribution that is most similar to the prior $f_{\underline{\boldsymbol{\theta}}}$ but at the same time, and often unlike the prior, satisfies the views $\mathcal{V}$. Therefore, we define the posterior $f_{\bar{\boldsymbol{\theta}}}$ as the distribution determined by the following set of parameters:

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

The posterior $f_{\bar{\boldsymbol{\theta}}}$ is then used as an 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.

A special case of the parametric approach is the normal assumption $f_{\boldsymbol{\mu},\boldsymbol{\sigma}^2}$, where $\boldsymbol{\mu}$ denotes the mean and $\boldsymbol{\sigma}^2$ the covariance matrix of the risk drivers. The relative entropy (3) then reads explicitly:

$$\begin{aligned}\mathcal{E}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2 \| \underline{\boldsymbol{\mu}}, \underline{\boldsymbol{\sigma}}^2) = \tfrac{1}{2}(&\operatorname{tr}(\boldsymbol{\sigma}^2(\underline{\boldsymbol{\sigma}}^2)^{-1}) - \ln|\boldsymbol{\sigma}^2(\underline{\boldsymbol{\sigma}}^2)^{-1}| \\ &+ (\boldsymbol{\mu} - \underline{\boldsymbol{\mu}})'(\underline{\boldsymbol{\sigma}}^2)^{-1}(\boldsymbol{\mu} - \underline{\boldsymbol{\mu}}) - \bar{n})\end{aligned} \tag{5}$$

where $\operatorname{tr}(\boldsymbol{x})$ is the trace: that is, the sum of the entries of the main diagonal of $\boldsymbol{x}$. A special type of view is an equality statement on expectations and covariances of linear combinations of the risk drivers, such as the spreads of returns of different asset classes. In this case, the normal EP problem (4) can be solved analytically (see Meucci 2008). However, the views are quite restrictive. Non-linear views and inequality or ranking views, which are vital for portfolio construction, are not addressed by the analytical solution. We proceed to discuss how to process such views in the next section.

Factor entropy pooling

Here we derive results that allow for the implementation of EP in its parametric form (4), with fully flexible views $\mathcal{V}$ beyond mean and covariance. In this case, the solution must be computed numerically. For this purpose, we impose the stipulation that the covariances are of factor type:

$$\boldsymbol{\sigma}^2 \equiv \boldsymbol{b}\boldsymbol{b}' + \operatorname{DgMatr}(\boldsymbol{d} \circ \boldsymbol{d}) \tag{6}$$

where $\boldsymbol{b}$ is an $\bar{n} \times \bar{k}$ matrix ($\bar{k} \ll \bar{n}$), $\boldsymbol{d}$ is an $\bar{n} \times 1$ vector, the operator $\operatorname{DgMatr}(\boldsymbol{v})$ embeds the $\bar{n} \times 1$ vector $\boldsymbol{v}$ into the principal diagonal of a square matrix that is zero anywhere else (hence $\operatorname{DgMatr}(\boldsymbol{v})$ is a matrix), and $\circ$ is the Hadamard product: that is, the entry-by-entry multiplication of vectors.

The structure (6) is consistent with a linear factor model assumption $\boldsymbol{X} = \boldsymbol{b}\boldsymbol{Z} + \boldsymbol{U}$, where the covariance of the factors is the low-dimensional $\bar{k} \times \bar{k}$ identity matrix ($\boldsymbol{\sigma}^2_{\boldsymbol{Z}} = \boldsymbol{i}_{\bar{k}\times\bar{k}}$), the residuals have idiosyncratic diagonal covariance ($\boldsymbol{\sigma}^2_{\boldsymbol{U}} = \operatorname{DgMatr}(\boldsymbol{d} \circ \boldsymbol{d})$), 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 why the name of the present approach is factor entropy pooling.

With the factor parametrisation (6), the EP problem (4) becomes the following FEP optimisation:

$$(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{b}}, \bar{\boldsymbol{d}}) \equiv \underset{(\boldsymbol{\mu},\boldsymbol{b},\boldsymbol{d}) \in \mathcal{V}}{\operatorname{argmin}}\, \mathcal{E}(\boldsymbol{\mu}, \boldsymbol{b}\boldsymbol{b}' + \operatorname{DgMatr}(\boldsymbol{d} \circ \boldsymbol{d}) \| \underline{\boldsymbol{\mu}}, \underline{\boldsymbol{\sigma}}^2) \tag{7}$$

where the optimisation target $\mathcal{E}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2 \| \underline{\boldsymbol{\mu}}, \underline{\boldsymbol{\sigma}}^2)$ is provided explicitly in (5).

The optimisation target (5) is not, in general, a convex function of the entries $(\boldsymbol{\mu}, \boldsymbol{b}, \boldsymbol{d})$. 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 (5) into the optimisation algorithm.

Moreover, the high-dimensional inverses that appear in the gradient and in the Hessian are easily obtained analytically in terms of low-cost, low-dimensional inverses. For details, refer to Meucci, Ardia and Colasante (2011).

To ensure that the views $(\boldsymbol{\mu}, \boldsymbol{b}, \boldsymbol{d}) \in \mathcal{V}$ in the FEP minimisation (7) are satisfied, we use one of two approaches.

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

$$(\boldsymbol{\mu}, \boldsymbol{b}, \boldsymbol{d}) \in \mathcal{V} \iff v(\boldsymbol{\mu}, \boldsymbol{b}, \boldsymbol{d}) \leqslant \boldsymbol{0} \tag{8}$$

for a suitable vector-valued function v. The simplest example is when we have equality views on means and covariances: a scenario highlighted at the end of the previous section. Less trivial examples occur in quantitative portfolio construction, when the views are constraints on the Sharpe ratios (see (18)). When the views can be expressed directly as constraints on the parameters, as in (8), we can compute the gradient and the Hessian of the constraints vector v and thus further increase the speed of the FEP optimisation (7); again see the quantitative portfolio construction example below.

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. We generate draws $\{\boldsymbol{x}^{(j)}_{\boldsymbol{\mu},\boldsymbol{b},\boldsymbol{d}}\}_{j=1}^{\bar{j}}$ from a multivariate normal distribution with expectation $\boldsymbol{\mu}$ and covariance $\boldsymbol{b}\boldsymbol{b}' + \operatorname{DgMatr}(\boldsymbol{d} \circ \boldsymbol{d})$. We then express the views as constraints on the Monte Carlo distribution:

$$(\boldsymbol{\mu}, \boldsymbol{b}, \boldsymbol{d}) \in \mathcal{V} \iff \{\boldsymbol{x}^{(j)}_{\boldsymbol{\mu},\boldsymbol{b},\boldsymbol{d}}\}_{j=1}^{\bar{j}} \in \mathcal{V} \tag{9}$$

Accurate and efficient simulation of draws can be achieved by generating once and for all a set of uncorrelated normal draws and then, given a set of parameters $(\boldsymbol{\mu}, \boldsymbol{b}, \boldsymbol{d})$, mapping them onto $\{\boldsymbol{x}^{(j)}_{\boldsymbol{\mu},\boldsymbol{b},\boldsymbol{d}}\}_{j=1}^{\bar{j}}$ by simple linear transformation.

The FEP approach (7) has a number of appealing features. From a statistical perspective, the factor parametrisation (6) is fully determined by a relatively small number $\bar{n}(\bar{k} + 2)$ of parameters $\boldsymbol{\theta} \equiv (\boldsymbol{\mu}, \boldsymbol{b}, \boldsymbol{d})$, instead of the large number $\bar{n}(\bar{n}+3)/2$ of parameters in the full-blown specification $\boldsymbol{\theta} \equiv (\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$. This parsimonious structure with limited parameters is an instance of shrinkage estimation (see, for example, Meucci (2005, page 200) for a review). As such, FEP provides statistically efficient estimates in large-dimensional markets.

Furthermore, the FEP parametrisation allows us to handle both the optimisation target and the views efficiently. 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 $\boldsymbol{\theta} \equiv (\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$, the matrix $\boldsymbol{\sigma}^2$ is constrained to be symmetric and positive definite.

Finally, the analytical expressions of the gradient and the Hessian along with the analytical inversions make the convergence of off-the-shelf algorithms much faster.

Implied expected returns

In this section we use the FEP 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 $\bar{n} \times 1$ vector of weights $\boldsymbol{w}_{\text{eq}}$, is linked to the $\bar{n} \times 1$ vector $\boldsymbol{\mu}$ of returns expectations and the $\bar{n} \times \bar{n}$ matrix $\boldsymbol{\sigma}^2$ of returns covariances by the following identity:

$$\boldsymbol{\mu} - \gamma\boldsymbol{\sigma}^2\boldsymbol{w}_{\text{eq}} \equiv \mathbf{0}_{\bar{n}\times 1} \tag{10}$$

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

In practical applications, $\boldsymbol{w}_{\text{eq}}$ is not necessarily the capital asset pricing model 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.

For portfolio construction purposes, the equilibrium constraint (10) guarantees that a mean-variance optimisation yields the portfolio $\boldsymbol{w}_{\text{eq}}$. However, the equilibrium constraint (10) 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 (10), Black and Litterman (1990) propose a two-step approach. In the first step, we fit a covariance matrix $(\boldsymbol{\sigma}^2)^{\text{BL}} \equiv \hat{\boldsymbol{\sigma}}^2$ 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 (6). In the second step, we compute the so-called implied expected returns, namely the expectations that satisfy the equilibrium constraint (10):

$$(\boldsymbol{\sigma}^2)^{\text{BL}} \equiv \hat{\boldsymbol{\sigma}}^2, \quad \boldsymbol{\mu}^{\text{BL}} \equiv \gamma\hat{\boldsymbol{\sigma}}^2\boldsymbol{w}_{\text{eq}} \tag{11}$$

Although the parameters (11) are consistent with the equilibrium constraint (10), 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 fitting a correlation matrix $\hat{\boldsymbol{c}}^2$ to empirical observations and then ensuring the equilibrium constraint (10) is satisfied by modifying both the expectations and the variances. More precisely, defining $\boldsymbol{\sigma}_{\text{vec}} \equiv \sqrt{\text{DgVect}(\boldsymbol{\sigma}^2)}$, the authors introduce a distance $\mathcal{D}$ between the estimates of the expectations and the standard deviations $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\sigma}}_{\text{vec}})$ and the yet to be defined parameters $(\boldsymbol{\mu}, \boldsymbol{\sigma}_{\text{vec}})$ as follows:

$$\begin{aligned}&\mathcal{D}(\boldsymbol{\mu}, \boldsymbol{\sigma}_{\text{vec}} \| \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\sigma}}_{\text{vec}})\\ &\quad \equiv (\alpha\|(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})./\hat{\boldsymbol{\sigma}}_{\text{vec}}\|^2 + (1-\alpha)\|(\boldsymbol{\sigma}_{\text{vec}} - \hat{\boldsymbol{\sigma}}_{\text{vec}})./\hat{\boldsymbol{\sigma}}_{\text{vec}}\|^2)^{1/2}\end{aligned} \tag{12}$$

where the authors set $\alpha \equiv 0.75$ and where ./ denotes entry-by-entry division. The authors then compute the parameters $(\boldsymbol{\mu}, \boldsymbol{\sigma}_{\text{vec}})$ that minimise the distance with respect to the estimated parameters:

$$(\boldsymbol{\mu}^{\text{LR}}, \boldsymbol{\sigma}_{\text{vec}}^{\text{LR}}) \equiv \underset{\boldsymbol{\mu},\boldsymbol{\sigma}^2 \in \mathcal{V}}{\text{argmin}}\ \mathcal{D}(\boldsymbol{\mu}, \boldsymbol{\sigma}_{\text{vec}} \| \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\sigma}}_{\text{vec}}) \tag{13}$$

Finally, the authors set:

$$(\boldsymbol{\sigma}^2)^{\text{LR}} \equiv \text{DgMatr}(\boldsymbol{\sigma}_{\text{vec}}^{\text{LR}})\hat{\boldsymbol{c}}^2\,\text{DgMatr}(\boldsymbol{\sigma}_{\text{vec}}^{\text{LR}})$$

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. The parameters $(\boldsymbol{\mu}^{\text{LR}}, (\boldsymbol{\sigma}^2)^{\text{LR}})$ are consistent with the equilibrium condition (10). 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 onto the means.

To further improve the estimation of the equilibrium distribution we can use our FEP framework. Accordingly, we replace the Euclidean distance minimisation (13) with the relative entropy minimisation (7), which we report here:

$$(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{b}}, \bar{\boldsymbol{d}}) \equiv \underset{(\boldsymbol{\mu},\boldsymbol{b},\boldsymbol{d})\in\mathcal{V}}{\text{argmin}}\ \mathcal{E}(\boldsymbol{\mu}, \boldsymbol{b}\boldsymbol{b}' + \text{DgMatr}(\boldsymbol{d} \circ \boldsymbol{d}) \| \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\sigma}}^2) \tag{14}$$

where the equilibrium constraint (10) now becomes the following view:

$$\mathcal{V}\colon \boldsymbol{\mu} - \gamma(\boldsymbol{b}\boldsymbol{b}' + \text{DgMatr}(\boldsymbol{d} \circ \boldsymbol{d}))\boldsymbol{w}_{\text{eq}} = \mathbf{0}_{\bar{n}\times 1} \tag{15}$$

There are many combinations $(\boldsymbol{\mu}, \boldsymbol{b}, \boldsymbol{d})$ that satisfy those views, but only one is the closest to the data, as represented by the historical mean and the historical covariance. The FEP posterior (14) determines that solution. We then obtain the FEP generalised equilibrium parameters:

$$\bar{\boldsymbol{\mu}}, \quad \bar{\boldsymbol{\sigma}}^2 \equiv \bar{\boldsymbol{b}}\bar{\boldsymbol{b}}' + \text{DgMatr}(\bar{\boldsymbol{d}} \circ \bar{\boldsymbol{d}}) \tag{16}$$

The FEP equilibrium estimates in (16) improve on the previous approaches in three ways. First, FEP replaces the somewhat arbitrary Euclidean distance between historical and equilibrium estimates with relative entropy, a statistically sound measure of discrepancy between distributions. Second, FEP simultaneously adjusts not only expectations and variances but also correlations. Third, the parsimonious 'low-rank-diagonal' specification (6) improves the statistical efficiency of the estimates. As a result, the equilibrium FEP parameters $(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\sigma}}^2)$ are potentially less noisy.

To illustrate the FEP equilibrium (16) in practice, we consider, for example, 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 compute the historical mean $\hat{\boldsymbol{\mu}}$ and the historical covariance $\hat{\boldsymbol{\sigma}}^2$ of the weekly returns. The market capitalisation weights $\boldsymbol{w}_{\text{eq}}$ are taken as of June 27, 2012. As expected, the FEP parameters are more in line with the historical parameters than the Black-Litterman parameters:

$$\mathcal{E}(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\sigma}}^2 \| \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\sigma}}^2) = 1.83, \quad \mathcal{E}(\boldsymbol{\mu}^{\text{BL}}, (\boldsymbol{\sigma}^2)^{\text{BL}} \| \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\sigma}}^2) = 2.41 \qquad (17)$$

In the extended version of this paper (Meucci, Ardia and Colasante 2011) we illustrate the effect of FEP on correlations and we compare the expectations and covariances of the historical distribution $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\sigma}}^2)$, the Black-Litterman equilibrium $(\boldsymbol{\mu}^{\text{BL}}, (\boldsymbol{\sigma}^2)^{\text{BL}})$ and the FEP equilibrium $(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\sigma}}^2)$ via the location–dispersion ellipsoids (see Meucci 2005, page 54) of a few stock pairs. In a 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 FEP, the centre, the dispersion and the orientation of the ellipsoid are all modified. However, such modifications are minimal by construction and thus the FEP parameters are more in line with the historical parameters than the Black-Litterman parameters, as highlighted in (17). This also illustrates that a constant correlation structure as in Levy and Roll (2010) is restrictive.

Ranking views

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

The most standard approach to this problem, popularised by Grinold and Kahn (1999) among others, proceeds by back-testing signals, as follows.

■ **Step 1.** At each generic time t we focus on an observable characteristic of a set of $\bar{n}$ assets that is deemed to have predictive power: a momentum/reversal indicator for stocks, for example, or a value indicator such as the price/earnings ratio. We then 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 $n = \bar{n}$ that 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:

$$\frac{\mu_{n,t}}{\sigma_{n,t}} \leqslant \frac{\mu_{n+1,t}}{\sigma_{n+1,t}} - q, \quad n = 1, \ldots, \bar{n} - 1 \qquad (18)$$

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, conditional on the multidimensional information $\boldsymbol{i}_t$ available at time t, and where $q \geqslant 0$ is a buffer that induces stronger inequalities.

■ **Step 2.** We estimate the next-period standard deviations $\hat{\sigma}_{n,t}$ of the assets returns: using exponentially weighted moving averages, say.

■ **Step 3.** We estimate the next-period correlations of the assets $\hat{c}_{m,n,t}$ with standard techniques. Jointly with the standard deviations $\hat{\sigma}_{n,t}$, the correlations yield the estimated covariance matrix. More precisely, defining $\hat{\boldsymbol{\sigma}}_{\text{vec},t} \equiv (\hat{\sigma}_{1,t}, \ldots, \hat{\sigma}_{\bar{n},t})$ and organising the correlations $\hat{c}_{m,n,t}$ into an $\bar{n} \times \bar{n}$ matrix $\hat{\boldsymbol{c}}_t^2$, we obtain the covariance matrix as follows:

$$\hat{\boldsymbol{\sigma}}_t^2 \equiv \text{DgMatr}(\hat{\boldsymbol{\sigma}}_{\text{vec},t}) \hat{\boldsymbol{c}}_t^2 \, \text{DgMatr}(\hat{\boldsymbol{\sigma}}_{\text{vec},t}) \qquad (19)$$

■ **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:

$$\tilde{\mu}_{n,t} \equiv \eta \hat{\sigma}_{n,t}(n - (\bar{n}+1)/2), \quad n = 1, \ldots, \bar{n} \qquad (20)$$

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 that is not too concentrated in a single position. To do so, we simplify Lobo, Fazel and Boyd (2007), replacing the $\bar{n}$ decision variables, namely the long-short weights $\boldsymbol{w}$, with four sets of positive variables $(\boldsymbol{w}^+, \boldsymbol{w}^-, \delta\boldsymbol{w}^+, \delta\boldsymbol{w}^-)$, each of dimension $\bar{n}$, as follows: $\boldsymbol{w}^+ \equiv \max(\boldsymbol{w}, \boldsymbol{0})$ represents the positive part of the weights and $\boldsymbol{w}^- \equiv \max(-\boldsymbol{w}, \boldsymbol{0})$ its negative part; $\delta\boldsymbol{w}^+ \equiv \max(\boldsymbol{w} - \boldsymbol{w}_{t-1}, \boldsymbol{0})$ represents the positive part of the transactions and $\delta\boldsymbol{w}^- \equiv \max(\boldsymbol{w}_{t-1} - \boldsymbol{w}, \boldsymbol{0})$ its negative part, where $\boldsymbol{w}_{t-1}$ is the legacy portfolio from the previous period. The weights then read $\boldsymbol{w} = \boldsymbol{w}^+ - \boldsymbol{w}^-$ and the absolute value of the transactions reads $|\boldsymbol{w} - \boldsymbol{w}_{t-1}| = \delta\boldsymbol{w}^+ + \delta\boldsymbol{w}^-$.

If we denote by $\boldsymbol{t}$ the vector of the transaction costs, then the portfolio optimisation can be expressed as the maximisation of a linear target:

$$(\boldsymbol{w}_t^{\pm}, \delta\boldsymbol{w}_t^{\pm}) \equiv \underset{(\boldsymbol{w}^{\pm}, \delta\boldsymbol{w}^{\pm}) \in \mathcal{C}}{\operatorname{argmax}} \; (\underbrace{\tilde{\boldsymbol{\mu}}_t'(\boldsymbol{w}^+ - \boldsymbol{w}^-)}_{\text{expected return}} - \underbrace{\boldsymbol{t}'(\delta\boldsymbol{w}^+ + \delta\boldsymbol{w}^-)}_{\text{transaction cost}}) \qquad (21)$$

under second-order conic constraints $\mathcal{C}$ that include the upper limit on risk $\|\hat{\boldsymbol{\sigma}}_t \boldsymbol{w}\| \leqslant \sigma_*$, 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). The optimal portfolio weights then read $\boldsymbol{w}_t \equiv \boldsymbol{w}_t^+ - \boldsymbol{w}_t^-$.

The most sensitive part in the above process is the proportional assumption (20) that quantifies the ranking signal (18).

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 (20) corresponds to *ex ante* Sharpe ratios $\tilde{\mu}_n / \hat{\sigma}_n = \eta(n - (\bar{n}+1)/2)$, which never change through time. Such Sharpe ratios satisfy the signal (18).

The second problem is that the proportional assumption (20) does not change the volatilities in (19), whereas the trading signal inequalities (18) also involve 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': that is, the average among all possible expected returns consistent with the ranking (18). 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. FEP addresses both issues, as follows.

■ **Step 4′.** The ranking signal that generates portfolios from sorts (18) is clearly a view in the constraint format (8), where the constraint function reads:

$$\mathcal{V}: v_n(\boldsymbol{\mu}_t, \boldsymbol{b}_t, \boldsymbol{d}_t) \equiv \frac{\mu_{n,t}}{\sigma_{n,t}} - \frac{\mu_{n+1,t}}{\sigma_{n+1,t}} + q \leqslant 0, \quad n = 1, \ldots, \bar{n} - 1 \tag{22}$$

where:

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

Hence, we just impose the constraints (22) in the FEP optimisation (7) process. Among all the distributions that satisfy the signal inequalities, FEP chooses the one that is closest to the data, as represented by the estimated covariances $\hat{\boldsymbol{\sigma}}_t^2$ and expected returns $\hat{\boldsymbol{\mu}}_t$:

$$(\bar{\boldsymbol{\mu}}_t, \bar{\boldsymbol{b}}_t, \bar{\boldsymbol{d}}_t) \equiv \underset{(\boldsymbol{\mu}, \boldsymbol{b}, \boldsymbol{d}) \in \mathcal{V}}{\text{argmin}} \ \mathcal{E}(\boldsymbol{\mu}, \boldsymbol{b}\boldsymbol{b}' + \text{DgMatr}(\boldsymbol{d} \circ \boldsymbol{d}) \| \hat{\boldsymbol{\mu}}_t, \hat{\boldsymbol{\sigma}}_t^2) \tag{23}$$

Note that this step can be further improved by replacing $\hat{\boldsymbol{\mu}}_t$ and $\hat{\boldsymbol{\sigma}}_t^2$ with the generalised equilibrium parameters (16).

To further speed up the optimisation, we also provide the gradient and the Hessian of the views (22). (For the explicit formulas, refer to Meucci, Ardia and Colasante (2011).) The FEP covariances are then reconstructed from the FEP parameters as follows:

$$\bar{\boldsymbol{\sigma}}_t^2 \equiv \bar{\boldsymbol{b}}_t \bar{\boldsymbol{b}}_t' + \text{DgMatr}(\bar{\boldsymbol{d}}_t \circ \bar{\boldsymbol{d}}_t) \tag{24}$$

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

Unlike the *ex ante* Sharpe ratios $\tilde{\mu}_n/\hat{\sigma}_n = \eta(n-(\bar{n}+1)/2)$ ensuing from the common approach (20), or the *ex ante* Sharpe ratios in the centroid approach, the *ex ante* Sharpe ratios $\bar{\mu}_n/\bar{\sigma}_n$ stemming from the FEP posterior (23)–(24) satisfy the ranking views (18) and at the same time change with the empirical data $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\sigma}}^2)$.

To illustrate with an example the standard back-testing approach outlined in steps 1–5 above, as well as the modified step 4′ obtained via FEP, we back test a standard reversal strategy in the equity market. More precisely, in step 1 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 averages:

$$\text{mom}_{n,t}^{\lambda,\gamma} \equiv \frac{\sum_{s\geqslant 0} e^{-\lambda s} r_{n,t-s}}{\sum_{s\geqslant 0} e^{-\lambda s}} \Bigg/ \sqrt{\frac{\sum_{s\geqslant 0} e^{-\gamma s} r_{n,t-s}^2}{\sum_{s\geqslant 0} e^{-\gamma s}}} \tag{25}$$

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 $-\text{mom}_{1,t}^{\lambda} \leqslant \cdots \leqslant -\text{mom}_{\bar{n},t}^{\lambda}$, where the minus sign is set to implement a 'reversal' strategy (a plus sign for a 'momentum' strategy). The new ordering of stocks $n = 1, \ldots, \bar{n}$ implies the signal (18).

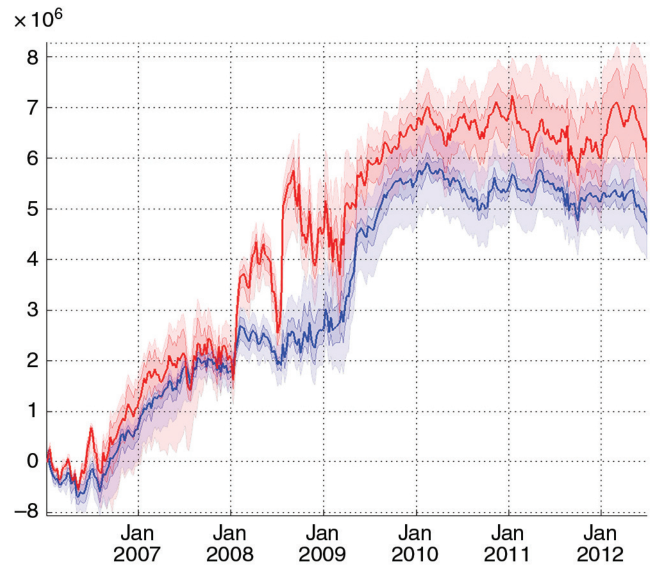
We then consider the same market as in our previous example, namely $\bar{n} = 30$ equities in the Dow Jones index, with data from January 2002 to June 2012. The back test starts in January 2006 and portfolios are constructed every Wednesday for a total of 338 rebalancing dates. We estimate the historical 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 $\tilde{\boldsymbol{\mu}}_t$ computed via the proportional assumption (20). We then build the back test (21), where we set the transaction costs $\boldsymbol{t}$ as 5 basis points of the market value, and where we set the volatility target such that the dollar volatility is bounded at US$100,000.

Next, we replace step 4 above with step 4′, based on the FEP framework. We use $\bar{k} = 1$ hidden factor in the FEP minimisation (23), in order to provide maximum shrinkage and maximum back-testing speed. Furthermore, we set the inequality buffer in (22) as $q = 1/(\bar{n} - 1)$. We then set the decay parameters (λ, γ) in the signal construction (25), so that the half-lives are 8 and 52 weeks, respectively.

For fairness we performed the back test with different 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 configurations. The number of hidden factors $\bar{k}$ and the inequality buffer q provide additional parameters over which to tweak the Sharpe ratio in the back test. In figure 1 we report the percentiles of the ensuing back-test P&L. We plot the outcomes of both the FEP approach (discussed in this example) and the standard approach (discussed in the previous example).

For more details, refer to the code available in Meucci, Ardia and Colasante (2011).

1 Cumulative P&L generated by the reversal strategy back test for various parametrisations



Note: the plot reports the median (solid line), the 50% percentile range (dim shading) and the 90% percentile range (dimmer shading). Red lines and shading denote factor entropy pooling while blue lines and shading denote the standard approach

Conclusion

Entropy pooling is a method used to estimate distributions of market observables consistent with historical data and market views. This paper introduces an efficient numerical algorithm called factor entropy pooling that improves the parametric implementation of entropy pooling by allowing more flexible views on the market compared with existing methods. First, it was used to calibrate implied returns, originally part of the Black-Litterman framework, which can then be used for portfolio construction in the absence of market views. Second, it was used to build and back test a systematic strategy based on ranking trade signals.

An additional area of application of factor entropy pooling beyond portfolio construction is heavy stress testing, where the market is subject to disruptive potential scenarios and their effect on the portfolio losses is observed. However, heavy stress testing does not require portfolio optimisation based on the entropy pooling posterior, so the computational speed is not relevant. A more computationally intensive evolution of factor entropy pooling that is particularly suitable for heavy stress testing is discussed in the companion paper Ardia and Meucci (2013). **R**

Attilio Meucci is the chief risk officer at KKR and founder of Symmys. He is based in New York. David Ardia is an assistant professor in the department of finance, insurance and real estate at Laval University in Québec. Marcello Colasante is a researcher at Symmys. **The authors are grateful to Simon Keel for his support and feedback throughout and to Lennart Hoogerheide for his constructive comments.** **Email: attilio.meucci@symmys.com, david.ardia@fsa.ulaval.ca, marcello.colasante@symmys.com**

REFERENCES

Almgren R and N Chriss, 2006
Optimal portfolios from ordering information
Journal of Risk 9, pages 1–47

Ardia D and A Meucci, 2013
Stress-testing in non-normal markets via entropy pooling (http://symmys.com/node/719)
Working Paper

Black F and R Litterman, 1990
Asset allocation: combining investor views with market equilibrium
Goldman Sachs Fixed Income Research

Grinold RC and R Kahn, 1999
Active Portfolio Management: A Quantitative Approach for Producing Superior Returns and Controlling Risk, 2nd edition
McGraw-Hill

Levy M and R Roll, 2010
The market portfolio may be mean/variance efficient after all
Review of Financial Studies 23(6), pages 2464–2491

Lobo M, M Fazel and B Boyd, 2007
Portfolio optimization with linear and fixed transaction costs
Annals of Operations Research 152(1), pages 341–365

Meucci A, 2005
Risk and asset allocation
Springer Finance
Available at http://symmys.com/attilio-meucci/book

Meucci A, 2008
Fully flexible views: theory and practice (http://symmys.com/node/158)
Risk October, pages 97–102

Meucci A, 2009
Simulations with exact means and covariances (http://symmys.com/node/162)
Risk July, pages 89–91

Meucci A, D Ardia and M Colasante, 2011
Portfolio construction and systematic trading with factor entropy pooling
http://symmys.com/node/160

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
Journal of Portfolio Management 37, pages 76–85

GUIDELINES FOR THE SUBMISSION OF TECHNICAL ARTICLES

Risk welcomes the submission of technical articles on topics relevant to our readership. Core areas include market and credit risk measurement and management, the pricing and hedging of derivatives and/or structured securities, and the theoretical modelling and empirical observation of markets and portfolios. This list is not exhaustive.

The most important publication criteria are originality, exclusivity and relevance – we attempt to strike a balance between these. Given that *Risk* technical articles are shorter than those in dedicated academic journals, clarity of exposition is another yardstick for publication. Once received by the technical editor and his team, submissions are logged and checked against these criteria. Articles that fail to meet the criteria are rejected at this stage.

Articles are then sent without author details to one or more anonymous referees for peer review. Our referees are drawn from the research groups, risk management departments and trading desks of major financial institutions, in addition to academia. Many have already published articles in *Risk*. Authors should allow four to eight weeks for the refereeing process. Depending on the feedback from referees, the author may attempt to revise the manuscript. Based on this process, the technical editor makes a decision to reject or accept the submitted article. His decision is final.

Submissions should be sent to the technical team, preferably by email addressed to technical@incisivemedia.com. LaTeX is the preferred format for submissions (your LaTeX file should be accompanied by a PDF of the output from it), although Word documents are also acceptable. It is helpful for graphs and figures to be submitted as separate Excel, postscript or EPS files. Vector graphics are strongly preferred, but rasterised images with a minimum resolution of 300dpi are sometimes acceptable.

The maximum recommended length for articles is 3,500 words, with some allowance for charts and/or formulas. We expect all articles to contain references to previous literature. We reserve the right to cut accepted articles to satisfy production considerations.