Parametric Stress–Testing in Non-Normal Markets via Copula–Marginal Entropy Pooling[☆]

David Ardia[a,b], Attilio Meucci[c]

[a] *Département de finance, assurance et immobilier, Université Laval, Québec City (Québec), Canada*
[b] *Centre interuniversitaire sur le risque, les politiques économiques et l'emploi, Québec, Canada*
[c] *SYMMYS, New-York, USA*

This revision: 7 February 2015
Latest revision and code: `http://symmys.com/node/719`

Abstract

A novel approach for stress–testing (portfolios of) financial assets is presented. The technique extends the parametric Entropy Pooling approach to skewed and thick-tailed markets. The technique rests on a copula–marginal decomposition for the entropy together with several approximation schemes which renders the numerical computations feasible for real-life problems. An illustration with a portfolio of European options is presented.

Keywords: Entropy Pooling, Kullback-Leibler, copula–marginal, stress–test, risk management

1. Introduction

The combination of trading signals, or more in general 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.

Standard approaches to process views include the celebrated model by Black and Litterman (1990). The most popular methodology to perform stress–testing is the linear regression-based *RiskManager* model (RiskMetrics, 2013). However, this model only outputs one specific stressed market scenario, instead of stressed market distribution. Furthermore,

[☆]Work in progress; comments are welcome. We thank Guido Bolliger, Kris Boudt, Marcello Colasante, Lennart Hoogerheide and Richard Luger for their comments. David Ardia gratefully acknowledges the financial support from IFM2. Any remaining errors or shortcomings are the authors' responsibility.

Email addresses: `david.ardia@fsa.ulaval.ca` (David Ardia), `attilio.meucci@symmys.com` (Attilio Meucci)

the regression based methodology implicitly assumes a normal or elliptically distributed market.

The generalized Bayesian approach *Entropy Pooling* (hereafter EP) is a flexible framework to process views and generalized stress–tests. The theoretical framework for EP was laid out in full generality 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 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 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 a 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 scenarios with *flexible probabilities*; and parametrically, by making parametric assumptions on the prior and the posterior distributions.

The non-parametric flexible probabilities is useful for *soft* stress–testing, i.e., to assess the effect on the portfolio of i) non-extreme market regimes (e.g., high/low volatility, high/low inflation, etc.) or ii) qualitative causal statements (e.g., "rates will rise only if the market soars"). The non-parametric implementation has been studied in the original article, and developed in Meucci et al. (2011a) and Meucci (2010c, 2013). See also Marriott (2009), Dey and Juneja (2012) and Rebonato and Denev (2014) for instance.

The parametric normal implementation has been studied in the original paper under views on means and covariances, and in Meucci et al. (2014) with flexible views, such as ranking trading signals. The parametric normal implementation of EP is useful for constructing portfolios in quantitative systematic strategies. We emphasize that the parametric normal implementation is general enough to model exactly highly non-linear instruments such as options, as long as the risk drivers, such as the implied volatility surfaces, are assumed normally distributed.

A summary of the various approaches available for EP is provided in Table 1.

Type of Entropy Pooling	Pros.	Cons.	Applications
Non-parametric (flexible probabilities) → Meucci (2008)	Flexibility	Breaks in large dimensions	Soft stress–testing
Parametric normal (analytical) → Meucci (2008)	High Speed	Normal market & basic views only	Portfolio construction
Parametric normal (FEP) → Meucci et al. (2014)	High speed	Normal market	Portfolio construction
Parametric generic (CMEP) → This paper	Flexibility	Low speed	Hard stress–testing

Table 1: The various Entropy Pooling approaches.

In this article, we address the issue of distributional stress–testing for non-normal markets. To achieve our goal, we rely on a parametric implementation of the Entropy Pooling approach, which we name Copula–Marginal Entropy Pooling (CMEP). Our approach rests on the copula–marginal decomposition of the joint distribution by Sklar's. This allows us to model skew and fat tails via marginals such as the skewed Student-t distribution. We show how this decomposition can be used to decompose the relative entropy itself, which allows us to reduce the computational burden with various approximations. We illustrate its usefulness with the stress–test of a portfolio of European options.

The remainder of this article is organized as follows. In Section 2 we review the various approaches for EP. In Section 3 we describe the CMEP approach. In Section 4 we evaluate the performance of the CMEP implementation. In Section 5 we illustrate how CMEP can be used. Section 6 concludes.

2. Review of Entropy Pooling

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 $\boldsymbol{X} \equiv (X_1, \ldots, X_N)'$ in the market, as represented by its probability density function (pdf), which we denote by $\underline{f}$: $\boldsymbol{X} \sim \underline{f}$. The risk drivers are any set of random variables that fully determine the securities P&L, such as log-returns, interest rates, implied volatility surfaces, 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 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 $f \in \mathcal{V}$.

Since the views possibly contradict the prior, the prior distribution does not necessary 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 EP 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 (or Kullback-Leibler divergence), denoted $\mathcal{E}$, a measure of the similarity of a distribution f with respect to a reference distribution, in our case the prior $\underline{f}$:

$$\mathcal{E}(f \,|\, \underline{f}) \equiv \int f(\mathbf{x}) \ln \frac{f(\mathbf{x})}{\underline{f}(\mathbf{x})} d\mathbf{x} \,. \tag{1}$$

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:

$$\bar{f} \equiv \underset{f \in \mathcal{V}}{\operatorname{argmin}} \, \mathcal{E}(f \,|\, \underline{f}) \,. \tag{2}$$

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

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

$$\bar{\mathbf{p}} \equiv \underset{\mathbf{p} \in \mathcal{V}}{\operatorname{argmin}} \, \mathcal{E}(\mathbf{p} \,|\, \underline{\mathbf{p}}) \tag{3}$$

where with minor abuse of notation we let $\mathcal{E}(\mathbf{p} \,\|\, \underline{\mathbf{p}}) \equiv \sum_{j=1}^{J} p^{(j)} \ln(p^{(j)} / \underline{p}^{(j)})$ denote the discrete counterpart of the relative entropy (1). For several types of views the optimization (3) can be transformed in an instance of linear programming with a low number of variables, and thus it can be efficiently solved numerically (see Meucci, 2008). The non-parametric approach is very flexible but breaks in large dimensions due to the curse of dimensionality.

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

$$f_{\bar{\boldsymbol{\psi}}} \equiv \underset{\substack{f_{\boldsymbol{\psi}} \in \mathcal{V} \\ \boldsymbol{\psi} \in \Psi}}{\operatorname{argmin}} \mathcal{E}(f_{\boldsymbol{\psi}} \mid f_{\underline{\boldsymbol{\psi}}}) \,. \tag{4}$$

A special case of the parametric approach rests on the normal distribution:

$$f_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\mathbf{x}) \equiv (2\pi)^{-\frac{N}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left(-\tfrac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right) \,, \tag{5}$$

where $\boldsymbol{\mu}$ is an $N \times 1$ vector of expectations and $\boldsymbol{\Sigma}$ is an $N \times N$ symmetric and positive definite covariance matrix. Therefore, the parameters are $\boldsymbol{\psi} \equiv (\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is constrained to be symmetric and positive definite, which we denote by $\boldsymbol{\Sigma} \succ 0$. Accordingly, under normality the parametric problem (4) becomes:

$$f_{\bar{\boldsymbol{\mu}},\bar{\boldsymbol{\Sigma}}} \equiv \underset{\substack{f_{\boldsymbol{\mu},\boldsymbol{\Sigma}} \in \mathcal{V} \\ \boldsymbol{\Sigma} \succ 0}}{\operatorname{argmin}} \mathcal{E}(f_{\boldsymbol{\mu},\boldsymbol{\Sigma}} \mid f_{\underline{\boldsymbol{\mu}},\underline{\boldsymbol{\Sigma}}}) \,. \tag{6}$$

The relative entropy between two normal distributions can be computed explicitly and reads:

$$\mathcal{E}(f_{\boldsymbol{\mu},\boldsymbol{\Sigma}} \mid f_{\underline{\boldsymbol{\mu}},\underline{\boldsymbol{\Sigma}}}) = \tfrac{1}{2}\left(\operatorname{trace}(\boldsymbol{\Sigma}\underline{\boldsymbol{\Sigma}}^{-1}) - \ln|\boldsymbol{\Sigma}\underline{\boldsymbol{\Sigma}}^{-1}| + (\boldsymbol{\mu}-\underline{\boldsymbol{\mu}})'\underline{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{\mu}-\underline{\boldsymbol{\mu}}) - N\right) \,; \tag{7}$$

see Meucci (2008) for a proof. In the case of views on equality statements on expectations and covariances, the normal EP problem can be solved analytically; see Meucci (2008, formulas 21-22). Equality statements are useful in the context of mean–variance portfolio optimization, 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. Equality views on expectations and covariances are however still fairly restrictive. When views are more complex, such as non-linear views and inequality or ranking views, we can resort to the Factor Entropy

Pooling (FEP) technique as described in Meucci et al. (2014). FEP is an efficient algorithm for the normal implementation of EP with generic views. FEP is very suitable to build and backtest a systematic strategy based on ranking trade signals. An additional area of application of FEP other than portfolio construction is stress–testing, whereby we subject the market to disruptive potential scenarios and observe their effect on the portfolio losses. However, stress–testing does not require to optimize a portfolio based on the EP posterior, and thus computational speed is typically not as relevant.

In this paper, we address the solution of the parametric problem (4) under very general assumptions for the market distributions f and the views $\mathcal{V}$. When f is generic, the relative entropy between the prior $\underline{f}$ and f requires the evaluation of the multivariate integral (1). One option to compute this integral is Monte Carlo. However, Monte Carlo is computationally costly. Instead, we resort to an efficient copula–marginal decomposition, as we proceed to discuss.

3. Copula–Marginal Entropy Pooling

Let us denote by c the probability density function (pdf) of the copula of f, by f_n the pdf of the generic n-th marginal and by F_n the cumulative density function (cdf) of the generic n-th marginal. We use $\underline{c}$, $\underline{f}$, $\underline{f}_n$ and $\underline{F}_n$ to denote the same functions but for the prior. Using Sklar's theorem (Sklar, 1959), the density function f can be expressed as:

$$f(\mathbf{x}) = c(F_1(x_1), \ldots, F_N(x_N)) \prod_{n=1}^{N} f_n(x_n)\,. \tag{8}$$

When (8) is used for f and the prior $\underline{f}$, we can write the relative entropy (1) as the sum of the entropy from the marginals, the entropy from the copula and a cross-term:

$$\mathcal{E}(f\,|\,\underline{f}) = \sum_{n=1}^{N} \mathcal{E}(f_n\,|\,\underline{f}_n) + \mathcal{E}(c\,|\,\underline{c}) + \mathcal{E}_{cr}\,; \tag{9}$$

see Appendix A. The cross-term reads:

$$\mathcal{E}_{cr} \equiv \mathbb{E}\left\{ \ln \frac{\underline{c}(U_1, \ldots, U_N)}{\underline{c}(\underline{F}_1(F_1^{-1}(U_1)), \ldots, \underline{F}_N(F_N^{-1}(U_N)))} \right\}, \tag{10}$$

where $U_n \equiv F_n(X_n)$, with $(X_1, \ldots, X_N)' \sim f$. We can therefore tackle the evaluation of (1) by three sub-evaluations.

3.1. The marginals

We parametrize the generic univariate distribution f_n that we intend to use in terms of the expectation μ_n, the standard deviation σ_n, and a set of additional parameters $\boldsymbol{\theta}_n$, i.e., $f_n \equiv f_{\mu_n,\sigma_n,\boldsymbol{\theta}_n}$. If the moments are not defined, we resort to location and scatter parameters, such as the median and the interquartile range respectively.

Then the pure marginal terms in (9) can be computed by quadratures techniques. In Appendix B, we propose a fast shifted/scaled trapezoidal rule for the numerical integration of the univariate entropies. Note that in the case where (some) marginals are normal, the one-dimensional variant of (7) can be used to compute the relative entropy.

For our illustrations, we rely on the skewed Student-t distribution by Hansen (1994). The pdf of the skewed Student-t distribution depends on four parameters, namely the expectation μ, the standard deviation $\sigma > 0$, a tail parameter $\nu > 2$, and a skewness parameter $\lambda \in (-1, 1)$. Therefore, with the skewed Student-t distribution we have $\boldsymbol{\theta} \equiv (\nu, \lambda)'$ and the pdf of a given marginal writes:

$$f_{\mu,\sigma,\boldsymbol{\theta}}(x) \equiv \frac{\kappa\gamma}{\sigma}\left[1 + \frac{\kappa}{(\nu-2)\,(1 \pm \lambda)}\left(\frac{x-\mu}{\sigma} + \eta\right)^2\right]^{-(\nu+1)/2}, \tag{11}$$

where:

$$\gamma \equiv \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi(\nu-2)}\Gamma(\frac{\nu}{2})}, \quad \kappa \equiv 4\,\frac{\nu-2}{\nu-1}\lambda\gamma, \quad \eta \equiv \sqrt{1 + 3\lambda^2 - \kappa^2}, \tag{12}$$

and where in (11) the symbol $\pm$ is $+$ when $\frac{x-\mu}{\sigma} \geq -\gamma\eta$ and is $-$ otherwise.

3.2. The copula and the cross-term

We restrict our attention to normal copulas, noting that the present discussion can be extended to more general elliptical copulas that include the Student-t specification.

Consider a normal copula, which is fully determined by a $N \times N$ correlation matrix $\mathbf{C}$, or equivalently, by its Cholesky factorization $\mathbf{A}$, i.e., a $N \times N$ lower-triangular matrix $\mathbf{A}$ such that $\mathbf{C} \equiv \mathbf{A}\mathbf{A}'$. In this case, using (7) the second term in (9) reads:

$$\mathcal{E}(c\,|\,\underline{c}) = \ln|\underline{\mathbf{A}}| - \ln|\mathbf{A}| + \tfrac{1}{2}\mathrm{trace}(\mathbf{A}\mathbf{A}'\underline{\mathbf{A}}^{-1\prime}\underline{\mathbf{A}}^{-1}) - \tfrac{N}{2}\,; \tag{13}$$

see Meucci (2008). Furthermore, under the normal copula assumption (13), the cross-term in (9) reads:

$$\mathcal{E}_{cr} = -\tfrac{1}{2}\mathrm{trace}(\mathbf{A}\mathbf{A}'\underline{\mathbf{A}}^{-1\prime}\underline{\mathbf{A}}^{-1}) + \tfrac{1}{2}\mathrm{trace}((\underline{\mathbf{A}}^{-1\prime}\underline{\mathbf{A}}^{-1} - \mathbf{I}_N)\mathbb{E}\{\boldsymbol{Y}\boldsymbol{Y}'\}) + \tfrac{N}{2}\,, \tag{14}$$

with $\mathbf{I}_N$ the identity matrix of order N and $\boldsymbol{Y} \equiv (Y_1, \ldots, Y_N)'$ where:

$$Y_n \equiv (\Phi^{-1} \circ \underline{F}_n \circ F_n^{-1} \circ \Phi)[\mathbf{A}\boldsymbol{Z}]_n\,, \tag{15}$$

where Φ denotes the standard normal cdf, $\boldsymbol{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_N)$ and $[\mathbf{v}]_n$ denotes the n-th component of the $N \times 1$ vector $\mathbf{v}$; see Appendix C.

When adding (13) to (10), some terms cancel, such that the copula term and the cross-term can be evaluated as:

$$\mathcal{E}(c\,|\,\underline{c}) + \mathcal{E}_{cr} = \ln|\underline{\mathbf{A}}| - \ln|\mathbf{A}| + \tfrac{1}{2}\mathrm{trace}((\underline{\mathbf{A}}^{-1\prime}\underline{\mathbf{A}}^{-1} - \mathbf{I}_N)\mathbb{E}\{\boldsymbol{Y}\boldsymbol{Y}'\})\,. \tag{16}$$

The expectation in (16) can be estimated by Monte Carlo integration, but it is computationally costly. We propose to replace the expectation by a normal approximation:

$$\mathbb{E}\{\boldsymbol{Y}\boldsymbol{Y}'\} \approx \left(\frac{\boldsymbol{\mu} - \underline{\boldsymbol{\mu}}}{\underline{\boldsymbol{\sigma}}}\right)\left(\frac{\boldsymbol{\mu} - \underline{\boldsymbol{\mu}}}{\underline{\boldsymbol{\sigma}}}\right)' + \mathrm{diag}\left(\frac{\boldsymbol{\sigma}}{\underline{\boldsymbol{\sigma}}}\right)\mathbf{A}\mathbf{A}'\mathrm{diag}\left(\frac{\boldsymbol{\sigma}}{\underline{\boldsymbol{\sigma}}}\right)\,, \tag{17}$$

where $\boldsymbol{\mu} \equiv (\mu_1, \ldots, \mu_N)'$, $\boldsymbol{\sigma} \equiv (\sigma_1, \ldots, \sigma_N)'$, $\mathrm{diag}(\mathbf{v})$ is a $N \times N$ diagonal matrix containing the entries of the $N \times 1$ vector $\mathbf{v}$, and where the division is meant term-by-term; see Appendix D. Relying on the normal approximation (17) instead of Monte Carlo integration significantly speedups the estimation. As we show in Section 4, the normal approximation is good for a large range of parameters ν and λ.

3.3. Dimension reduction

Let us recall that, so far, we have specified the distribution of the market in terms of $4N$ (marginals) plus $N(N-1)/2$ (copula) parameters. As a result, performing any kind of optimal stress–testing in our market calls for an optimization with a prohibitively large number of variables. To address this dimensionality issue, we need to perform dimension reduction, with particular focus on the $N(N-1)/2$ parameters of the copula. We assume as in Meucci et al. (2014) the following N_F dimensional factor-based parametrization for the copula $\mathbf{A}$ and the univariate dispersions $\boldsymbol{\sigma}$:

$$\boldsymbol{\Sigma} \equiv \operatorname{diag}(\boldsymbol{\sigma})\mathbf{A}\mathbf{A}'\operatorname{diag}(\boldsymbol{\sigma}) \equiv \mathbf{B}\mathbf{B}' + \operatorname{diag}(\boldsymbol{\delta}^2)\,, \tag{18}$$

where $\mathbf{B}$ is a $N \times N_F$ matrix and $\boldsymbol{\delta}$ is a $N \times 1$ vector. Thus $\mathbf{A}\mathbf{A}'$ and $\boldsymbol{\sigma}$ are jointly parametrized by only $N(N_F+1)$ terms instead of $N(N+1)/2$. Typically $N_F \ll N$, and setting $N_F \equiv 1$ satisfies most practical applications. This parsimonious structure with limited parameters is an instance of shrinkage estimation (see Meucci, 2005, for a review). Moreover, the parsimonious parametrization $(\mathbf{B}, \boldsymbol{\delta})$ in (18) is unconstrained, as the parameters can freely range in the space $\mathbb{R}^{N\times N_F} \times \mathbb{R}^N$. Instead, in the specification $(\mathbf{A}, \boldsymbol{\sigma})$ in (18), the matrix $\boldsymbol{\Sigma}$ is constrained to be symmetric and positive definite.

3.4. The optimization

Using the factor-based parametrization (18), our distributions are parametrized in terms of $N(N_F+2) + \#\{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N\}$ parameters collected into $\boldsymbol{\psi}$, where $\#\{\boldsymbol{\theta}\}$ is the number of parameters in $\boldsymbol{\theta}$ (e.g., $\#\{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N\} = 2N$ if we use N skewed Student-t distributions as marginals):

$$(\mathbf{A}, \boldsymbol{\sigma}, \boldsymbol{\mu}, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N) \iff (\mathbf{B}, \boldsymbol{\delta}, \boldsymbol{\mu}, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N) \equiv \boldsymbol{\psi}\,. \tag{19}$$

Using our copula–marginal decomposition of the relative entropy (9) and the various approximations, we thus have:

$$\mathcal{E}(f_{\boldsymbol{\psi}} \,|\, f_{\underline{\boldsymbol{\psi}}}) \approx \hat{\mathcal{E}}(\boldsymbol{\psi}) \equiv \hat{\mathcal{E}}_{c+cr}(\mathbf{B}, \boldsymbol{\delta}, \boldsymbol{\mu}) + \sum_{n=1}^{N} \hat{\mathcal{E}}_n(\mu_n, \sigma_n, \boldsymbol{\theta}_n)\,, \tag{20}$$

where $\hat{\mathcal{E}}_{c+cr}$ is (16) with the normal approximation (17) for the expectation, as a function of the parameters $(\mathbf{B}, \boldsymbol{\delta}, \boldsymbol{\mu})$, and $\hat{\mathcal{E}}_n$ is the quadrature approximation of the n-th univariate marginal using the trapezoidal rule, as a function of the parameters $(\mu_n, \sigma_n, \boldsymbol{\theta}_n)$.

As in (4), the posterior distribution $\bar{\boldsymbol{\psi}}$ is obtained as follows:

$$\bar{\boldsymbol{\psi}} \equiv \underset{\substack{\boldsymbol{\psi}\in\mathcal{V} \\ \boldsymbol{\psi}\in\Psi}}{\operatorname{argmin}}\, \hat{\mathcal{E}}(\boldsymbol{\psi})\,, \tag{21}$$

where $\boldsymbol{\psi} \in \mathcal{V}$ indicates that $f_{\boldsymbol{\psi}}$ fulfills the views, and $\boldsymbol{\psi} \in \Psi$ that the parameters in $\boldsymbol{\psi}$ satisfy some domain constraints. In our case, the factor-based parametrization ensures a positive definite covariance matrix. Moreover, we can easily reparametrize parameters λ and ν such that the optimization becomes unconstrained.

3.5. The views

To ensure that the views $\boldsymbol{\psi} \in \mathcal{V}$ in the minimization (21) are satisfied, we use one of two approaches.

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

$$\boldsymbol{\psi} \in \mathcal{V} \iff v(\boldsymbol{\psi}) \leq \mathbf{0}\,, \tag{22}$$

for a suitable vector-valued function $v(\cdot)$. The simplest example are the equality views on means and covariances.

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 $\{\mathbf{z}^{(j)}\}_{j=1}^{J}$ of J uncorrelated normal draws $\mathbf{z}^{(j)} \equiv (z_1^{(j)}, \ldots, z_N^{(j)})'$. Then for any choice of the parameters $\boldsymbol{\psi}$ we map the scenarios $\mathbf{z}^{(j)}$ onto the joint scenarios $\mathbf{x}^{(j)}$ from the distribution $f_{\boldsymbol{\psi}}$, as follows:

$$\mathbf{x}_{\boldsymbol{\psi}}^{(j)} \equiv \boldsymbol{\mu} + \operatorname{diag}(\boldsymbol{\sigma})\mathbf{A}\mathbf{z}^{(j)}\,, \tag{23}$$

where the parametrization in $\boldsymbol{\psi}$ follows from (19). We then express the views as constraints on the Monte Carlo distribution:

$$\boldsymbol{\psi} \in \mathcal{V} \iff \{\mathbf{x}_{\boldsymbol{\psi}}^{(j)}\}_{j=1}^{J} \in \mathcal{V}\,. \tag{24}$$

We rely on a precomputed mesh approach in order to speedup the simulation procedure for checking the views; see Appendix E.

4. Numerical aspects

In this section we validate the implementation of the CMEP method and assess the precision of the normal approximation of the cross-term.

4.1. Validation of CMEP

We illustrate how the CMEP performs when a view on a tail (i.e., the Value-at-Risk) is imposed. This example is purely illustrative and aims at showing how the CMEP performs. To that purpose, we consider a bivariate example $N = 2$, with $N_F = 1$ where we take the prior as a multivariate normal distribution with zero mean and diagonal covariance matrix with unit variances. We set the view on the fifth quantile of the first marginal (i.e., a view on the Value-at-Risk at the 95% risk level). We consider three levels for the view. The first one is set to -1.64, the real quantile of a normal distribution; the second to -2.00, and the third to -2.50. We set the number of draws to $J = 10^4$ to check the views by simulation. Plots for the posteriors are displayed in Figure 1; contour plots are obtained via a bivariate kernel smoothing estimator based on the posterior draws. We clearly see the effect of a larger Value-at-Risk (in absolute value). This skews the first marginal, leading to higher relative entropy values.

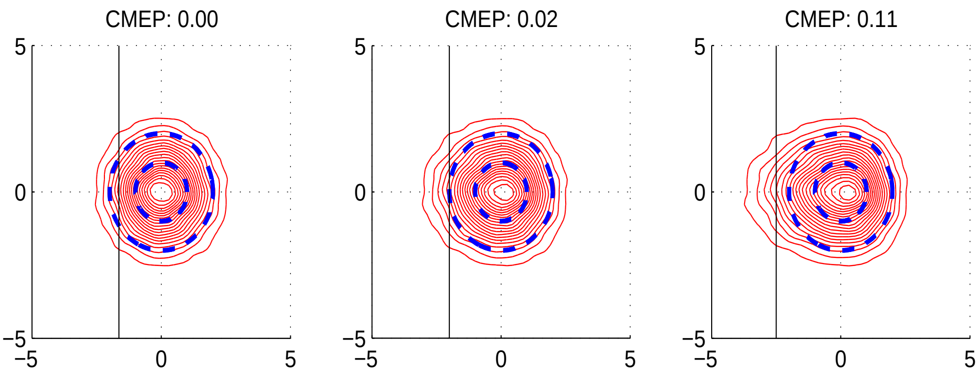

Figure 1: Contour plot of the CMEP posterior distribution for various views on Value-at-Risk at the 95% risk level (in solid red lines, obtained by bivariate kernel smoothing). The dashed blue lines report the contour of the normal prior, and the vertical black lines report the view on the Value-at-Risk (VaR95). Left plot: VaR95 = -1.64, middle plot: VaR95 = -2.00, right plot: VaR95 = -2.50. The numerical value reported on the top of the plot is the relative entropy of the posterior with respect to the normal prior.

4.2. Accuracy of the cross-term approximation

We measure the accuracy of the cross-term normal approximation (17) in the non-normal case. To do so, we compute the key term in (16):

$$\kappa \equiv \text{trace}(\mathbb{E}\{\boldsymbol{Y}\boldsymbol{Y}'\}) , \tag{25}$$

by Monte Carlo integration for a very large number of draws ($J = 10^5$), and compare it with the value obtained with the normal approximation for $\mathbb{E}\{\boldsymbol{Y}\boldsymbol{Y}'\}$ in (17). The Monte Carlo integration value is considered as the *exact* value. The relative error is then defined as:

$$\text{relative error} \equiv \frac{\kappa - \hat{\kappa}}{(|\kappa| + |\hat{\kappa}|)/2}\,, \tag{26}$$

where κ is the estimation obtained by Monte Carlo integration and $\hat{\kappa}$ is the estimation obtained via the normal approximation.

We consider a standard normal prior, and various posteriors where we modify either the tails via the degrees-of-freedom parameter ν, or the skewness via the asymmetry parameter λ (same value applied to all marginals). We perform this for dimensions $N = 5$ and $N = 25$ (and set $N_F = 1$ in each case). For both the prior and the posterior, the correlation is set to 0.1. For each setup, we replicate one hundred times the relative error and compute the 90% asymptotic normal confidence bands of the average based on the numerical standard errors.

Figure 2 displays the results of the sensitivity analysis for various ν (top) and λ (bottom); we only report positive values for λ as negative values mirror the plot. For ν, we see that for both dimensions, the approximation is very good, even for very low values. The worst relative error is -2% for $\nu = 4$. For $\nu > 9$, the relative error is not significantly different from zero. In the case of λ, as the asymmetry increases, we see that the normal approximation overestimates the true value of the cross-term. The worst relative error is obtained for $\lambda = 0.99$, with a value of -1.5%, which corresponds to an extreme degree of asymmetry. For $\lambda < 0.25$, the relative error is not significantly different from zero.

The computational cost for calculating the approximation is almost zero. On the contrary, the Monte Carlo requires time. On a Intel(R) Core(TM) i5-2557M CPI 1.7GHz machine, the average time is 0.8 seconds for $N = 2$ and 4 seconds for $N = 25$ (it is respectively 0.8 seconds and 4 seconds when $J = 10^4$). This time seems negligible, but it aggregates in the optimization process, as for each point tried by the optimizer the relative entropy, and therefore the cross-term, has to be computed. This renders the Monte Carlo approach almost infeasible for practical applications.

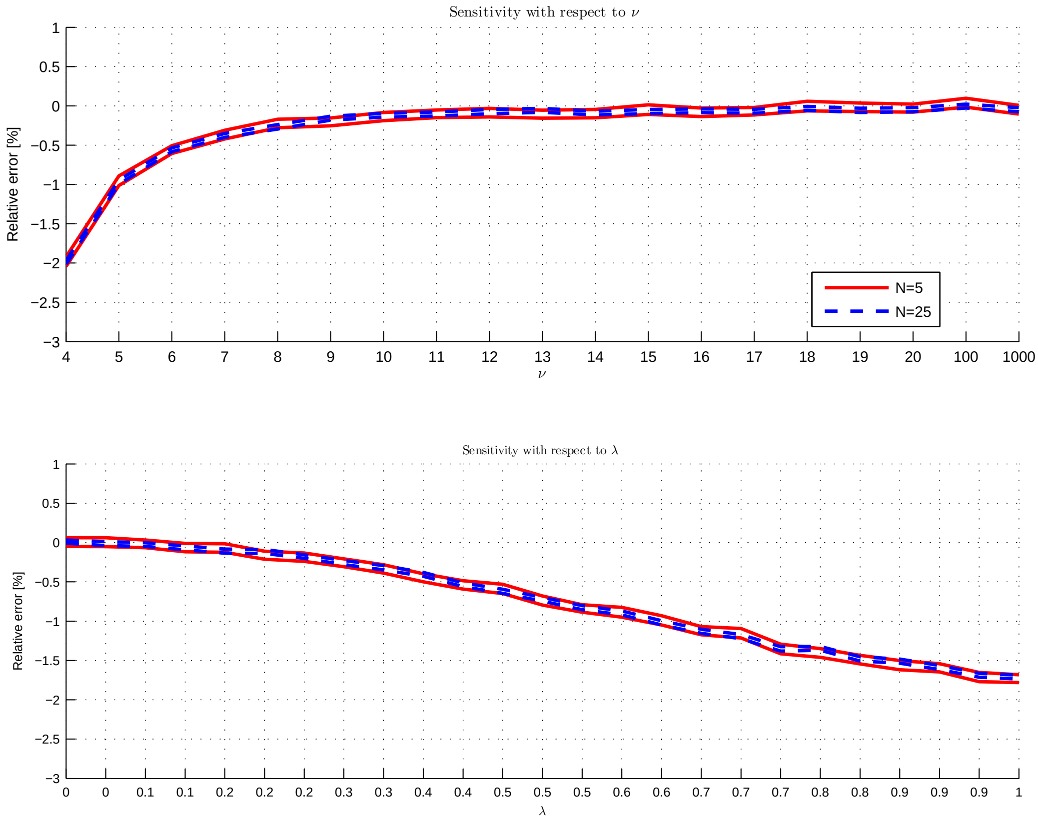

Figure 2: Relative error (26) for various values of ν for tail thickness (top plot) and λ for skewness (bottom plot). The lines report the 90% asymptotic normal confidence bands of the average based on the numerical standard errors (obtained from one hundred sets of Monte Carlo replications, $J = 10^5$ each). Solid red lines are for $N = 5$ and dashed blue lines for $N = 25$.

As an additional sensitivity check, we generate random skewed Student-t posteriors with random positive correlation structures in dimension $N = 25$. For each of the one thousand posteriors, we compute the relative error between the *exact* posterior and the posterior obtained by the Normal approximation. The obtained average (over the Monte Carlo replications) of the relative error is -1.8%. The 5%, 50% and 95% quantiles are -3.5%, -1.8% and -0.31%, respectively.

5. Empirical illustration

We illustrate a practical application of the CMEP approach in quantitative risk management. We consider a portfolio of European call and put options on the S&P 500 index.

Fully documented code is available at `http://symmys.com/node/719`.

Let us define the moneyness by $m \equiv K/S$, where K is the strike price of the option and S is the underlying value. Similarly to Malz (1997), we parametrize the volatility surface at time t by means of a quadratic interpolation of the smile for two times to maturity $\tau_{\min}$ and $\tau_{\max}$ of 30 and 60 days respectively:

$$\begin{aligned}\sigma_{t,\tau_{\min}}(m) &\equiv \beta_{t,1} + \beta_{t,2}(\beta_{t,3} - m)^2 \\ \sigma_{t,\tau_{\max}}(m) &\equiv \beta_{t,4} + \beta_{t,5}(\beta_{t,6} - m)^2 \,.\end{aligned} \tag{27}$$

where $\beta_{t,\cdot}$ are the smile parameters at time t. For tenors $\tau_{\min} \leq \tau \leq \tau_{\max}$ we interpolate the surface linearly between the above values. The parameters $\boldsymbol{\beta}_t \equiv (\beta_{t,1}, \ldots, \beta_{t,6})'$, together with the price S_t of the underlying S&P 500 index fully determine, via Black–Scholes pricing formula, the distribution of all the options on the index at the generic time t; see Meucci (2008). Thus, in our case we have a market with $N = 7$ risk drivers.

To calibrate the prior, we use daily prices of the index together with daily values of implied volatilities for moneyness $m \in \{0.75, 1, 1.25\}$ and tenor $\tau \in \{30, 60\}$ days. We consider a period ranging from 28 March, 2011, to 13 March, 2012, for a total of $T = 252$ observations. Historical risk factors are constructed from the differences in prices and parameters of the fitted quadratic smiles over time, $\boldsymbol{X}_t \equiv (\Delta S_t, \Delta\boldsymbol{\beta}_t')'$. For estimation of the prior, we consider a multivariate normal distribution, with exponential smoothing estimates $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ with half-life of two months. Note that the normal assumption for the prior is taken as an illustration. Alternatively, we could fit the copula skewed Student-t distribution as the prior. For the portfolio of options, we consider a book of three short straddles (at-the-money) with maturities of 30, 60 and 120 days. Straddles are simple combinations of one call and one put option with the same strike and maturity, and are mainly used to trade the implied volatility (instead of taking a bet on a direction of the underlying).

We use the CMEP approach to stress–test the portfolio under a scenario where the downside risk (i.e., the Expected Shortfall at the 95% risk level) of the S&P 500 index is expected to increase by 10%:

$$\mathcal{V} \,:\, \mathrm{ES}_{0.95}(S_{T+1}) \equiv \mathbb{E}\{S_{T+1} \,|\, S_{T+1} < \mathrm{VaR}_{0.95}(S_{T+1})\} \leq 0.90 \cdot \mathrm{ES}_{0.95}(S_T) \,, \tag{28}$$

where $\mathrm{ES}_{0.95}(S_T)$ is the current downside risk of the underlying, and where $\mathrm{VaR}_{0.95}(S_{T+1})$ denotes the next-step-ahead Value-at-Risk at the 95% risk level, i.e., the fifth percentile of the distribution of S_{T+1}. As it is not possible to frame such view as a constraint of the parameters, we rely on simulations. We set the number of draws to $J = 10^4$. We set $N_F = 3$ in the factor reduction of the copula.

In the top panel of Figure 3 we display the portfolio's P&L following from full-repricing with $J = 10^4$ Monte Carlo scenarios under the normal prior distribution. The P&L is defined as the difference between the simulated portfolio values at $T + 1$ and its current value. In the bottom panel, we display the P&L distribution following from full-repricing under the CMEP stress–test distribution. For each plot we also report the Value-at-Risk and the Expected Shortfall at the 95% risk level. We can notice how the stress–test distribution spreads the P&L distribution to the left, with higher risk measures (in absolute value), than the prior.

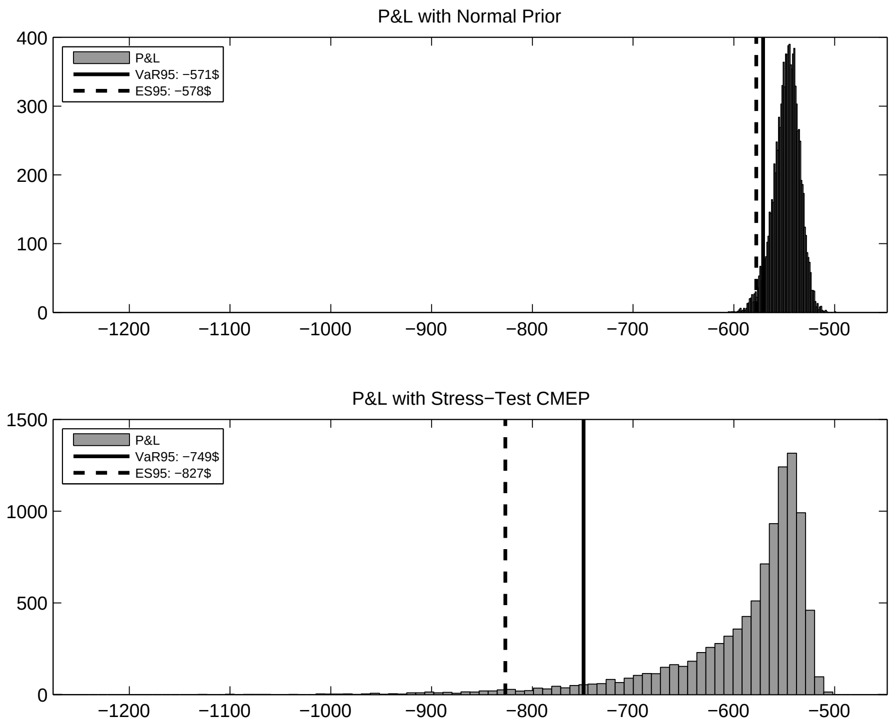

Figure 3: Normal prior (top plot) and CMEP posterior stress–test distribution (bottom plot). Vertical lines report the Value-at-Risk (solid line) and the Expected Shortfall (dashed line) risk measures at the 95% risk level. The number of Monte Carlo scenarios is set to $J = 10^4$.

In Figure 4 we display the distribution of the implied volatility smiles for a 30-day horizon obtained with the normal prior (left) and the stress–test CMEP posterior (right). The figure reports the median value (blue line), the 50% area (green region) and the 90% area (red region) of the quadratic smiles obtained from the parameters of the two distributions. We clearly notice an upward parallel shift of the smile and slight increase in its slope for the stress–test distribution, compared with the normal prior. In the CMEP case, we also notice a larger dispersion of smiles. This example illustrates how a view on the market's drop can be translated into scenarios for the shape of the volatility smile.

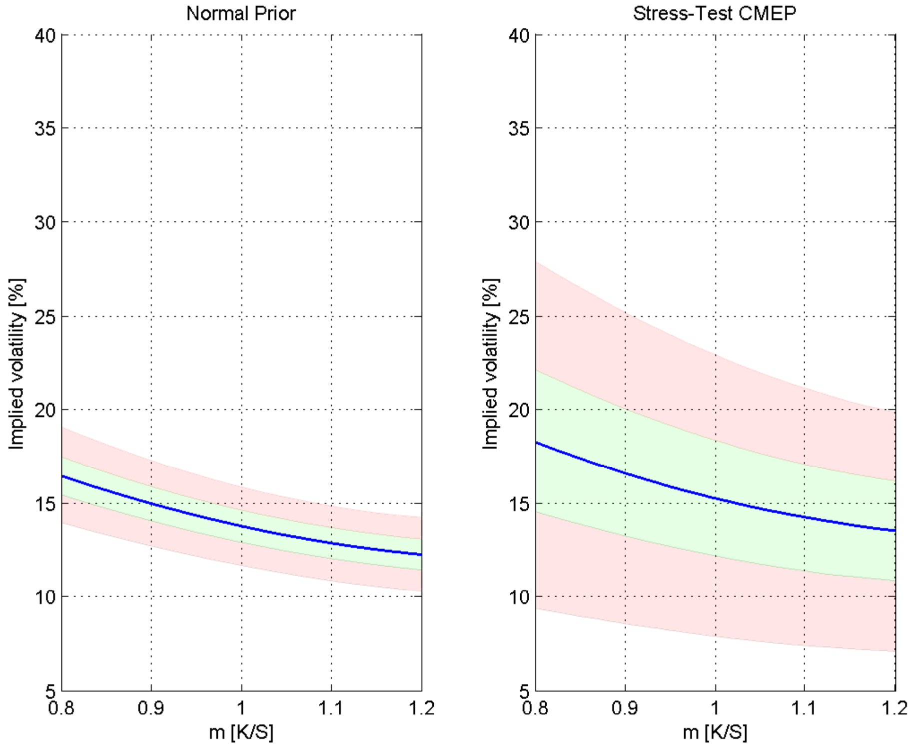

Figure 4: Distribution of smiles for normal prior (left) and the stress–test CMEP (right). The figure reports the median value (blue line), the 50% area (green region) and the 90% area (red region) of the quadratic smiles obtained with $J = 10^4$ draws from the two distributions.

In the left part of Figure 5 we display the pdf of the underlying S&P for the normal prior (dashed blue line) and the stress–test CMEP (solid red line). We notice the strong left-skew of the stress–test distribution. Also, the variance increases. In the right part of Figure 5 we display the location-dispersion ellipsoids that represent graphically expectations and covariances for some risk drivers. In particular, we notice how the the stress–test view modifies the correlation between the risk drivers.

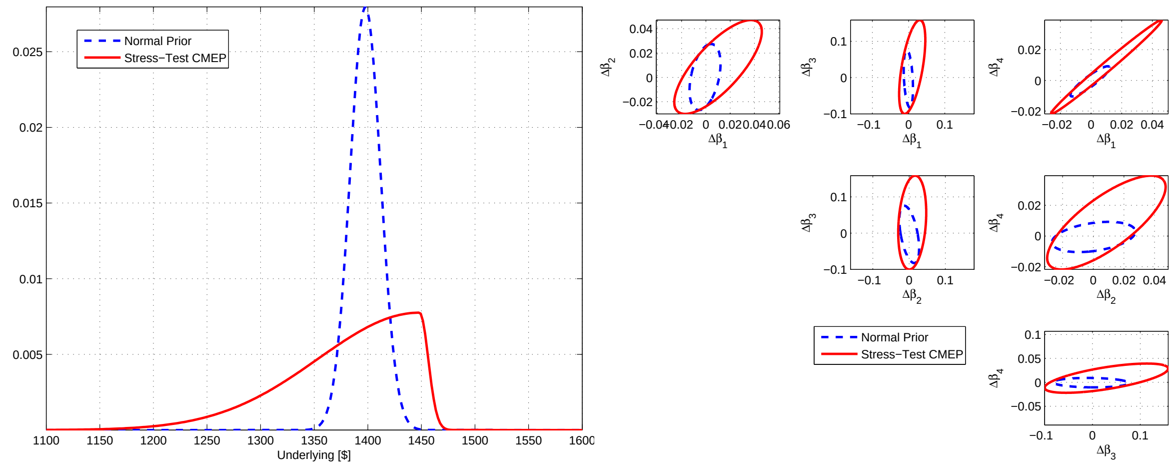

Figure 5: Left plot: Probability density function of the underlying S&P for the normal prior (dashed blue line) and the stress–test CMEP (solid red line). Right plot: Location-dispersion ellipsoids of some risk factors $\Delta\beta_i$ for the normal prior (dashed blue lines) and stress–test CMEP (solid red lines). Risk factors are first differences of parameters of the fitted quadratic smiles over time. Ellipsoids represent graphically expectations and covariances.

6. Conclusion

The Entropy Pooling approach developed by Meucci (2008) provides a flexible way to process views and generalized stress–tests. 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.

In this article, we extended the list of EP techniques and proposed a method to deal with skewed and thick-tailed markets. The methodology rests on a copula–marginal decomposition of the entropy and various approximations which speedup the estimation. We validated the novel approach with views on the Value-at-Risk and provided an empirical illustration with a portfolio of European options.

Potential next steps include i) the use of Student-t copula instead of normal copula and ii) the use of alternative parametric shapes for the marginals, in particular the g&h distribution.

References

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

Dey, S. and Juneja, S., 2012. Incorporating fat tails in financial models using entropic divergence measures.

Hansen, B. E., 1994. Autoregressive conditional density estimation. International Economic Review 35 (3), 705730.

Malz, A., 1997. Option-implied probability distributions and currency excess returns. Federal Reserve Bank of New York - Staff Reports.

Marriott, A., 2009. Giving credibility to judgment. Actuary, October, 34–35.

Meucci, A., 2005. Risk and Asset Allocation. Springer. Available at `http://symmys.com`.

Meucci, A., 2008. Fully Flexible Views: Theory and practice. Risk 21 (10), 97–102. Article and code available at `http://symmys.com/node/158`.

Meucci, A., 2010c. Historical scenarios with Fully Flexible Probabilities. GARP Risk Professional December, 40–43. Article and code available at `http://symmys.com/node/150`.

Meucci, A., 2013. Estimation and stress–testing via time- and market-conditional flexible probabilities. Article and code available at `http://symmys.com/node/600`.

Meucci, A., Ardia, D., Keel, S., 2011a. Fully Flexible Extreme Views. Journal of Risk 14 (2), 39–49. Article and code available at `http://symmys.com/node/159`.

Meucci, A., Ardia, D., Colasante, M., 2014. Quantitative Portfolio Construction and Systematic Trading Strategies Using Factor Entropy Pooling. Risk 27, 56–61. Article and code available at `http://symmys.com/node/160`.

Rebonato, R. and Denev, A., 2014. Portfolio management under stress: A Bayesian-net approach to coherent asset allocation. Chapter 10. Cambridge University Press.

RiskMetrics, 2013. RiskManager: A multi-asset class, scalable SaaS framework for entreprise-wide risk management. `http://www.msci.com/products/risk_management_analytics/riskmanager`

Sklar, A., 1959. Fonctions de répartition à n dimensions et leurs marges. Institut de Statistique de l'Université de Paris.

Appendix A. The copula–marginal decomposition of relative entropy

To prove (9) we use the copula–marginal decomposition of the pdf f:

$$f(\mathbf{x}) = c(F_1(x_1), \ldots, F_N(x_N)) \prod_{n=1}^{N} f_n(x_n)\,, \tag{A.1}$$

where c is the pdf of the copula and f_n (F_n) the pdf (cdf) of the n-th marginal. We use the same decomposition for the prior $\underline{f}$, where the underline symbols refers to prior pdf/cdf.

Then:

$$
\begin{aligned}
\mathcal{E}(f \,|\, \underline{f}) &\equiv \int f(\mathbf{x}) \ln \frac{f(\mathbf{x})}{\underline{f}(\mathbf{x})} d\mathbf{x} \\
&= \mathbb{E}\left\{ \ln \frac{f(\boldsymbol{X})}{\underline{f}(\boldsymbol{X})} \right\} \\
&= \sum_{n=1}^{N} \mathbb{E}\left\{ \ln \frac{f_n(X_n)}{\underline{f}_n(X_n)} \right\} + \mathbb{E}\left\{ \ln \frac{c(F_1(X_1), \dots, F_N(X_N))}{\underline{c}(\underline{F}_1(X_1), \dots, \underline{F}_N(X_N))} \right\} ,
\end{aligned} \tag{A.2}
$$

where $\boldsymbol{X} \equiv (X_1, \dots, X_N)' \sim f$. Defining the grades $U_n \equiv F_n(X_n)$, we have:

$$
\begin{aligned}
\mathcal{E}(f \,|\, \underline{f}) = \sum_{n=1}^{N} \mathbb{E}\left\{ \ln \frac{f_n(X_n)}{\underline{f}_n(X_n)} \right\} \\
+ \mathbb{E}\left\{ \ln \frac{c(U_1, \dots, U_N)}{\underline{c}(U_1, \dots, U_N)} \right\} \\
+ \mathbb{E}\left\{ \ln \frac{\underline{c}(U_1, \dots, U_N)}{\underline{c}(\underline{F}_1(F_1^{-1}(U_1)), \dots, \underline{F}_N(F_N^{-1}(U_N)))} \right\} ,
\end{aligned} \tag{A.3}
$$

which is (9), i.e., $\sum_{n=1}^{N} \mathcal{E}(f_n \,|\, \underline{f}_n) + \mathcal{E}(c \,|\, \underline{c}) + \mathcal{E}_{cr}$.

Appendix B. Fast adaptive trapezoidal scheme

Fast evaluation of the relative entropy of the univariate marginal:

$$
\mathcal{E}(f \,|\, \underline{f}) \equiv \int f(x) \ln \frac{f(x)}{\underline{f}(x)} dx \,, \tag{B.1}
$$

can be achieved by a shifted/scaled (adaptive) trapezoidal rule.

Consider a grid $\{\underline{x}_k\}$ determined by the points:

$$
\underline{x}_k \equiv \underline{x}_{\min} + \tfrac{1}{2}(2k-1)\Delta \,, \tag{B.2}
$$

for $k = 1, \dots, Q$ where $\Delta \equiv (\underline{x}_{\min} - \underline{x}_{\max})/Q$ is a constant width, and where $\underline{x}_{\min}$ and $\underline{x}_{\max}$ are typically chosen to span the domain of the prior. In our application, $Q = 250$ and $\underline{x}_{\min}$ and $\underline{x}_{\max}$ are determined as minimum and maximum values over a large range of draws from the prior. Then, the relative entropy can be approximated by the trapezoidal rule as:

$$\mathcal{E}(f \,|\, \underline{f}) \approx \sum_{k=1}^{Q} f(\underline{x}_k) \ln \frac{f(\underline{x}_k)}{\underline{f}(\underline{x}_k)} \Delta \,. \tag{B.3}$$

We can render the boundaries of the grid adaptive to span the domain of the prior $\underline{f}$ and f, thus rendering the estimation more stable. Simply define:

$$\begin{aligned} x_{\min} &\equiv \min\{\underline{x}_{\min}, \mu + \tfrac{\sigma}{\underline{\sigma}}(\underline{x}_{\min} - \underline{\mu})\} \\ x_{\max} &\equiv \max\{\underline{x}_{\max}, \mu + \tfrac{\sigma}{\underline{\sigma}}(\underline{x}_{\max} - \underline{\mu})\} \,, \end{aligned} \tag{B.4}$$

and compute the trapezoidal rule over the equally-spaced grid which spans $[x_{\min}, x_{\max}]$.

Appendix C. Reformulation of the cross-term

Consider the super-normalized grades $\boldsymbol{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_N)$ of the normal copula. We write the cross-term in (10) as:

$$\mathbb{E}\left\{ \ln \frac{\underline{c}(U_1, \ldots, U_N)}{\underline{c}(\underline{F}_1(F_1^{-1}(U_1)), \ldots, \underline{F}_N(F_N^{-1}(U_N)))} \right\} = \mathbb{E}\left\{ \ln \frac{\underline{c}(U_1, \ldots, U_N)}{\underline{c}(V_1, \ldots, V_N)} \right\} , \tag{C.1}$$

where $U_n \equiv \Phi([\mathbf{A}\boldsymbol{Z}]_n)$, $V_n \equiv \Phi(Y_n)$, $Y_n \equiv (\Phi^{-1} \circ \underline{F}_n \circ F_n^{-1} \circ \Phi)[\mathbf{A}\boldsymbol{Z}]_n$, Φ is the standard normal cdf and $[\mathbf{v}]_n$ denotes the n-th component of the $N \times 1$ vector $\mathbf{v}$. Let us denote by $f_{\mathbf{A}}^{\mathcal{N}}$ the pdf of a multivariate normal distribution with expectations zero, unit variances, and correlation matrix $\mathbf{A}\mathbf{A}'$, and by $f^{\mathcal{N}}$ the pdf of a univariate standard normal distribution. Since:

$$c(u_1, \ldots, u_N) = \frac{f_{\mathbf{A}}^{\mathcal{N}}(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_N))}{f^{\mathcal{N}}(\Phi^{-1}(u_1)) \cdots f^{\mathcal{N}}(\Phi^{-1}(u_N))} \,, \tag{C.2}$$

then:

$$\frac{\underline{c}(U_1, \ldots, U_N)}{\underline{c}(V_1, \ldots, V_N)} = \frac{f_{\underline{\mathbf{A}}}^{\mathcal{N}}(\mathbf{A}\boldsymbol{Z})}{f^{\mathcal{N}}([\mathbf{A}\boldsymbol{Z}]_1) \cdots f^{\mathcal{N}}([\mathbf{A}\boldsymbol{Z}]_N)} \frac{f^{\mathcal{N}}(Y_1) \cdots f^{\mathcal{N}}(Y_N)}{f_{\underline{\mathbf{A}}}^{\mathcal{N}}(\boldsymbol{Y})} \,, \tag{C.3}$$

and therefore:

$$\ln \frac{\underline{c}(U_1, \ldots, U_N)}{\underline{c}(V_1, \ldots, V_N)} = \ln f_{\underline{\mathbf{A}}}^{\mathcal{N}}(\mathbf{A}\boldsymbol{Z}) - \ln f_{\underline{\mathbf{A}}}^{\mathcal{N}}(\boldsymbol{Y}) + \ln f_{\mathbf{I}_N}^{\mathcal{N}}(\boldsymbol{Y}) - \ln f_{\mathbf{I}_N}^{\mathcal{N}}(\mathbf{A}\boldsymbol{Z}) \,, \tag{C.4}$$

where $\mathbf{I}_N$ the identity matrix of order N. Using:

$$\ln f^{\mathcal{N}}_{\mathbf{A}}(\mathbf{x}) = -\tfrac{N}{2}\ln(2\pi) - \ln|\mathbf{A}| - \tfrac{1}{2}\left\|\mathbf{A}^{-1}\mathbf{x}\right\|^2 , \tag{C.5}$$

and $\|\mathbf{v}\|^2 \equiv \mathbf{v}'\mathbf{v}$, we obtain:

$$\ln \frac{\underline{c}(U_1,\ldots,U_N)}{\underline{c}(V_1,\ldots,V_N)} = -\tfrac{1}{2}\|\underline{\mathbf{A}}^{-1}\mathbf{A}\boldsymbol{Z}\|^2 + \tfrac{1}{2}\|\underline{\mathbf{A}}^{-1}\boldsymbol{Y}\|^2 - \tfrac{1}{2}\|\boldsymbol{Y}\|^2 + \tfrac{1}{2}\|\mathbf{A}\boldsymbol{Z}\|^2 . \tag{C.6}$$

Then:

$$\mathbb{E}\left\{\ln \frac{\underline{c}(U_1,\ldots,U_N)}{\underline{c}(V_1,\ldots,V_N)}\right\} = \tfrac{1}{2}\left(\mathbb{E}\{\|\mathbf{A}\boldsymbol{Z}\|^2\} - \mathbb{E}\{\|\underline{\mathbf{A}}^{-1}\mathbf{A}\boldsymbol{Z}\|^2\} + \mathbb{E}\{\|\underline{\mathbf{A}}^{-1}\boldsymbol{Y}\|^2\} - \mathbb{E}\{\|\boldsymbol{Y}\|^2\}\right) . \tag{C.7}$$

For a general square matrix $\mathbf{B}$ and random vector $\boldsymbol{X}$ we have:

$$\mathbb{E}\{\|\mathbf{B}\boldsymbol{X}\|^2\} = \text{trace}\left(\mathbb{E}\{\mathbf{B}\boldsymbol{X}\boldsymbol{X}'\mathbf{B}'\}\right) = \text{trace}\left(\mathbf{B}\mathbb{E}\{\boldsymbol{X}\boldsymbol{X}'\}\mathbf{B}'\right) = \text{trace}\left(\mathbf{B}'\mathbf{B}\mathbb{E}\{\boldsymbol{X}\boldsymbol{X}'\}\right) . \tag{C.8}$$

Noticing that $\mathbb{E}\{\boldsymbol{Z}\boldsymbol{Z}'\} = \mathbf{I}_N$, and that $\text{trace}(\mathbf{A}\mathbf{A}') = N$ because $\mathbf{A}\mathbf{A}'$ is a correlation matrix, we finally obtain:

$$\begin{aligned}
&\mathbb{E}\left\{\ln \frac{\underline{c}(U_1,\ldots,U_N)}{\underline{c}(V_1,\ldots,V_N)}\right\} \\
=&\tfrac{1}{2}\Big(N - \text{trace}(\mathbf{A}\mathbf{A}'\underline{\mathbf{A}}^{-1\prime}\underline{\mathbf{A}}^{-1}) + \text{trace}(\underline{\mathbf{A}}^{-1\prime}\underline{\mathbf{A}}^{-1}\mathbb{E}\{\boldsymbol{Y}\boldsymbol{Y}'\}) - \text{trace}(\mathbb{E}\{\boldsymbol{Y}\boldsymbol{Y}'\})\Big) \\
=&\tfrac{1}{2}\Big(N - \text{trace}(\mathbf{A}\mathbf{A}'\underline{\mathbf{A}}^{-1\prime}\underline{\mathbf{A}}^{-1}) + \text{trace}((\underline{\mathbf{A}}^{-1\prime}\underline{\mathbf{A}}^{-1} - \mathbf{I}_N)\mathbb{E}\{\boldsymbol{Y}\boldsymbol{Y}'\})\Big) .
\end{aligned} \tag{C.9}$$

Appendix D. Normal approximation of the cross-term

For the normal cdf $F_{\mu,\sigma}$ the following is true:

$$F_{\mu,\sigma}(x) = \Phi\left(\frac{x-\mu}{\sigma}\right) , \tag{D.1}$$

and similarly for the inverse cdf:

$$F^{-1}_{\mu,\sigma}(u) = \mu + \sigma\Phi^{-1}(u) . \tag{D.2}$$

We approximate the generic n-th marginals $\underline{F}_n$ and F_n^{-1} by normal marginals $F_{\underline{\mu}_n,\underline{\sigma}_n}$ and $F^{-1}_{\mu_n,\sigma_n}$, leading to the following approximation for the composition (15):

$$\begin{aligned}(\Phi^{-1}\circ \underline{F}_n \circ F_n^{-1}\circ\Phi)(x) &\approx \Phi^{-1}(F_{\underline{\mu}_n,\underline{\sigma}_n}(F^{-1}_{\mu_n,\sigma_n}(\Phi(x))))\\ &= \Phi^{-1}(F_{\underline{\mu}_n,\underline{\sigma}_n}(\mu_n+\sigma_n x))\\ &= \Phi^{-1}\left(\Phi\left(\frac{\mu_n+\sigma_n x-\underline{\mu}_n}{\underline{\sigma}_n}\right)\right)\\ &= \frac{\mu_n-\underline{\mu}_n}{\underline{\sigma}_n}+\frac{\sigma_n}{\underline{\sigma}_n}x\,.\end{aligned} \tag{D.3}$$

This implies that in approximation:

$$\boldsymbol{Y} \overset{d}{=} \frac{\boldsymbol{\mu}-\underline{\boldsymbol{\mu}}}{\underline{\boldsymbol{\sigma}}} + \mathrm{diag}\left(\frac{\boldsymbol{\sigma}}{\underline{\boldsymbol{\sigma}}}\right)\mathbf{A}\boldsymbol{Z}\,, \tag{D.4}$$

where $\boldsymbol{Y}\equiv(Y_1,\ldots,Y_N)'$, $\boldsymbol{\mu}\equiv(\mu_1,\ldots,\mu_N)'$, $\boldsymbol{\sigma}\equiv(\sigma_1,\ldots,\sigma_N)'$, $\boldsymbol{Z}\sim\mathcal{N}(\mathbf{0},\mathbf{I}_N)$, $\mathrm{diag}(\mathbf{v})$ is a $N\times N$ diagonal matrix containing the entries of the $N\times 1$ vector $\mathbf{v}$, and the division is meant term-by-term. The normal approximation of the cross-term is therefore:

$$\mathbb{E}\{\boldsymbol{Y}\boldsymbol{Y}'\} \approx \left(\frac{\boldsymbol{\mu}-\underline{\boldsymbol{\mu}}}{\underline{\boldsymbol{\sigma}}}\right)\left(\frac{\boldsymbol{\mu}-\underline{\boldsymbol{\mu}}}{\underline{\boldsymbol{\sigma}}}\right)' + \mathrm{diag}\left(\frac{\boldsymbol{\sigma}}{\underline{\boldsymbol{\sigma}}}\right)\mathbf{A}\mathbf{A}'\mathrm{diag}\left(\frac{\boldsymbol{\sigma}}{\underline{\boldsymbol{\sigma}}}\right)\,. \tag{D.5}$$

Appendix E. Precomputed mesh for fast simulation

We rely on a precomputed mesh to speedup the simulation procedure for checking the views. This avoids the computation of the cumulative normal distribution and quantile function of the skewed Student-t marginal at Monte Carlo draws.

First, we consider a grid $\{\underline{x}_k\}$ determined by the points:

$$\underline{x}_k \equiv \underline{x}_{\min} + \tfrac{1}{2}(2k-1)\Delta\,, \tag{E.1}$$

for $k=1,\ldots,Q$ where $\Delta\equiv(\underline{x}_{\max}-\underline{x}_{\max})/Q$ is a constant width, and where $\underline{x}_{\max}$ and $\underline{x}_{\max}$ are typically chosen to span the domain of the (normalized) prior. In our application, we use $Q=500$, $\underline{x}_{\min}\equiv -7.5$ and $\underline{x}_{\max}\equiv 7.5$.

Then, we define a set $\{\boldsymbol{\vartheta}_m\}\equiv\{\cup(\nu_{m_1},\lambda_{m_2})';m_1=1,\ldots,M_1\,,m_2=1,\ldots,M_2\}$, with ν_{m_1} and λ_{m_2} parameter values of the skewed Student-t distribution spanning values in their domain. For this set of $M\equiv M_1\cdot M_2$ tuples and the grid $\{\underline{x}_k\}$, we compute (once for all) and store the points:

$$y_{k,m} \equiv F^{-1}_{0,1,\boldsymbol{\vartheta}_m}(\Phi(\underline{x}_k)) \,, \tag{E.2}$$

for $m = 1, \ldots, M$, where Φ is the cumulative normal distribution and $F^{-1}_{0,1,\boldsymbol{\vartheta}_m}$ is the inverse cdf function corresponding to the standardized marginal of $\boldsymbol{\vartheta}_m$. Note that in the case of normal marginals, $y_{k,m} \equiv \underline{x}_k$.

With this pre-computed mesh, it is then fast to generate (approximate) draws from the joint pdf f for a given set of parameters $(\mathbf{B}, \boldsymbol{\delta}, \boldsymbol{\mu}, \boldsymbol{\theta}_i)$. First and once for all, a set of J draws from a standardized N-dimensional normal random variables $\{\mathbf{z}^{(j)}\}_{j=1}^{J}$ are generated. Then, the correlated normal random draws are obtained as $\tilde{\mathbf{z}} \equiv \mathbf{A}\mathbf{z}$. In order to generate draws from the marginal $(\mu_i, \sigma_i, \boldsymbol{\theta}_i)$, we determine the appropriate mesh index m, interpolate the mapping values of the draws $\tilde{z}_i^{(j)}$ in the grid (E.2), and finally scale and translate the interpolated values, $\tilde{y}_i^{(j)} \equiv \mu_i + \sigma_i \hat{y}_i^{(j)}$ where $\hat{y}_i^{(j)}$ denotes the interpolated value in the grid $\{y_{k,i}\}$.