Historical Scenarios with Fully Flexible Probabilities

Attilio Meucci[1]
www.symmys.com

published[2]: December 01 2010
this revision: May 16 2011
last version of article and code at `http://symmys.com/node/150`

Abstract

After reviewing the parametric and scenario-based approaches to risk management, we discuss a methodology to enhance the flexibility of the scenario-based approach. We change the probability of each scenario, and then we compute the ensuing p&l distribution and all relevant statistics such as VaR and volatility. The probabilities can be changed to reflect specific market conditions, advanced estimation techniques, or partial information, using the entropy-based Fully Flexible Views technique in Meucci (2008). The implementation of this approach is trivial, as no costly repricing is needed. Commented code is available at http://symmys.com/node/150.

JEL Classification: C1, G11

Keywords: exponential smoothing, kernel smoothing, parametric VaR, Monte Carlo simulations, stress-test, Fully Flexible Views, entropy pooling, multivariate histogram, empirical distribution, Dirac delta, Kullback-Leibler divergence

[1] The author is grateful to Garli Beibi

[2] This article appears as Meucci, A. (2010), *Personalized Risk Management: Historical Scenarios with Fully Flexible Probabilities*, The Quant Classroom by Attilio Meucci - GARP Risk Professional, December 2010, p. 47-51

Consider a market driven by N risk drivers $X \equiv \{X_n\}_{n=1,\ldots,N}$, i.e. a market where the p&l Π of each security is fully determined by X through a known pricing function π:

$$\Pi = \pi(X). \tag{1}$$

For instance, in equity options the risk drivers are the log-changes of the implied volatility smile, of the underlying, and of the interest rates; and the p&l depends on the risk drivers through the Black-Scholes-Merton formula. For corporate bonds the drivers are the interest rates changes of a reference curve such as swap or government and the spreads changes over that curve; and the pricing function is the usual discount formula. We refer to Meucci (2010) for more details.

To manage risk we must estimate the joint distribution of the risk drivers, as represented by the probability density function (pdf) f_X, and analyze its impact on the distribution of the prospective p&l (1) of a given portfolio. There exist two approaches to achieve this, namely parametric and scenario-based, but only the latter is reliable upon across all asset classes and investment horizons.

In Section 1 we review these two approaches and we introduce the generalized empirical distribution, which allows us to greatly enhance the scenario-based approach by associating non-equal probabilities to different scenarios. In Section 2 we discuss techniques to specify exogenously all the probabilities of the generalized empirical distribution. These techniques include rolling window, exponential smoothing and market conditioning by kernel smoothing. In Section 3 we rely on the Fully Flexible Views approach in Meucci (2008) to specify only some features of the probabilities of the generalized empirical distribution, and then fill in the missing information by relative entropy minimization. This allows us for instance to perform market conditioning without having to specify the shape of the smoothing kernel.

1 Parametric versus scenario-based risk management

The first approach to risk management is parametric: the distribution of the risk drivers is modeled by a parametric distribution f_X^θ, where θ is a set of parameters. The most classical example in this direction is the normal distribution, where θ represent the expected values and the covariances of the risk drivers. Other examples are elliptical distributions, skew-t distributions, parametric marginal-copula decompositions, etc. The parametric approach is computationally cheap. However, for a wide spectrum of instruments the p&l (1) is a non-linear function of the risk drivers. In such cases, the distribution of the p&l cannot be computed exactly and one must resort to approximations.

A second approach to risk management represents the distribution of the risk drivers f_X in terms of a $T \times N$ panel of joint scenarios $\{x_{t,n}\}$. Each N dimensional row

$$x_{t,\cdot} \equiv (x_{t,1}, \ldots, x_{t,N}) \tag{2}$$

is an independent joint scenario from the distribution f_X, and each scenario occurs with constant probability $p_t \equiv 1/T$, see Figure 1. If the risk drivers are "invariants", i.e. if they can be assumed independently and identically distributed across time, the scenarios can be historical realizations. Alternatively, the scenarios can be Monte Carlo simulations. In the sequel we focus on historical scenarios, and t is to be interpreted as time.

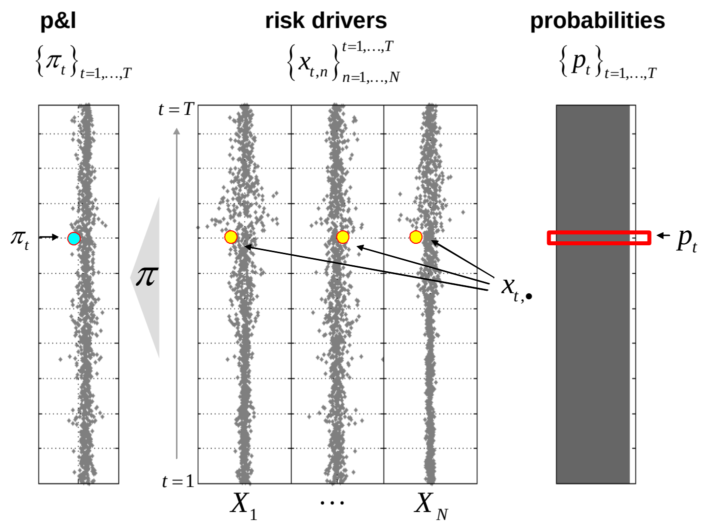

Figure 1: Distribution of p&l and risk-drivers as scenarios/probabilities

Using the scenario-based approach it is easy to generate scenarios for the portfolio p&l $\{\pi_t\}$ that represent the p&l distribution. Indeed, as in Figure 1, it suffices to apply the pricing function (1) to each scenario

$$\pi_t = \pi\left(x_{t,\cdot}\right), \quad t = 1, \ldots, T. \tag{3}$$

From these p&l scenarios it is then trivial to extract the relevant statistics, such as volatility and VaR, in terms of their sample counterparts.

Notice that the computational burden of pricing each scenario can be handled by parallel computing across both securities and scenarios. Therefore, even the computational challenge of the scenario-based approach can be overcome. As a result, the scenario-based approach has become the reference methodology for risk management in most financial institutions.

Assume that a panel of historical scenarios for the risk drivers $\{x_{t,n}\}$ is available, and that the respective p&l scenarios $\{\pi_t\}$ for the securities and the portfolio have been generated as in (3), possibly by a vendor during an overnight batch process. We can greatly enhance the flexibility of the scenario-based approach by suitably changing the probabilities p_t associated with the scenarios.

In other words, we represent the distribution of the risk drivers f_X by means of the generalized empirical distribution

$$f_X \equiv \sum_{t=1}^{T} p_t \delta^{(x_t)}. \tag{4}$$

In this expression $\delta^{(z)}$ denotes the Dirac delta centered in the generic point $z \in \mathbb{R}^N$, i.e. a spike of probability mass equal to one in z, refer to the appendix for more details. The standard empirical distribution is the special case of (4) where for all times t the probability is set as $p_t \equiv 1/T$.

Then, using the simple recipes in the appendix, we can recompute the p&l distribution with the new probabilities p_t and extract all the relevant statistics such as volatility and VaR. From an implementation perspective, this approach is trivial. Once the new probabilities have been defined, all the above computations can be performed in real time, without calling again the costly pricing functions (3).

2 Full probability specification

In this section we discuss a few approaches to specify exogenously all the probabilities $\{p_t\}$ in the generalized empirical distribution. In Section 3 we specify the probabilities $\{p_t\}$ based on partial information only.

2.1 Rolling window

The simplest possible customization for the probabilities $\{p_t\}$ of the historical scenarios $\{x_{t,n}\}$ is to make them constant over an arbitrary window of time and zero otherwise as in Figure 2:

$$p_t \propto \begin{cases} 1 & \text{if } \underline{t} \leq t \leq \overline{t} \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

The symbol $\propto$ in this expression denotes "proportional". Indeed, the probabilities must be re-normalized in such a way that they sum to one:

$$\sum_{t=1}^{T} p_t \equiv 1. \tag{6}$$

Typically, the window covers a fixed period τ from the latest available observation, i.e. in (5) one sets $\overline{t} \equiv T$ and $\underline{t} \equiv T - \tau$, and therefore this approach is named "rolling window".

To illustrate, we consider a long-short book of twenty close-to-expiry, highly non-linear options on the S&P 500, with strikes ranging from in-the-money to out-of-the-money. For such options, the risk drivers are the log-changes of the implied volatility, of the underlying and of the short term interest rate; and the p&l depends on the risk drivers as in (3) through the Black-Scholes-Merton formula.

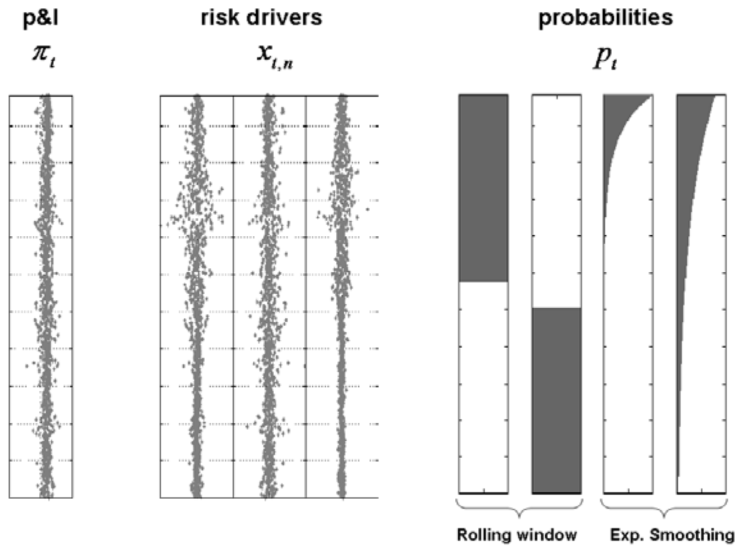

Figure 2: Probabilities emphasizing different time periods

In the top portion of Figure 3 we display the historical p&l corresponding to two different two-year periods. In Table (7) below we report select relevant statistics

Rolling window	Period '07-'08	Period '09-'10
Exp. value	381	385
St. dev.	240	214
VaR 99%	-287	-259
Cond. VaR 99%	-478	-303

(7)

We emphasize that in order to plot the histograms in Figure 3 or to compute the statistics in Table (7) we did not extract a sub-sample of the original panel of scenarios, but rather we changed the probabilities of all the original scenarios. For more details on the methodology please refer to the appendix. For more details on the computations please refer to the code available at http://symmys.com/node/150.

2.2 Exponential smoothing

Exponential smoothing is an estimation technique that accounts for the time-changing nature of volatility. In standard risk platforms exponential smoothing is applied as a four-step recipe. First, the expected values and the covariances of

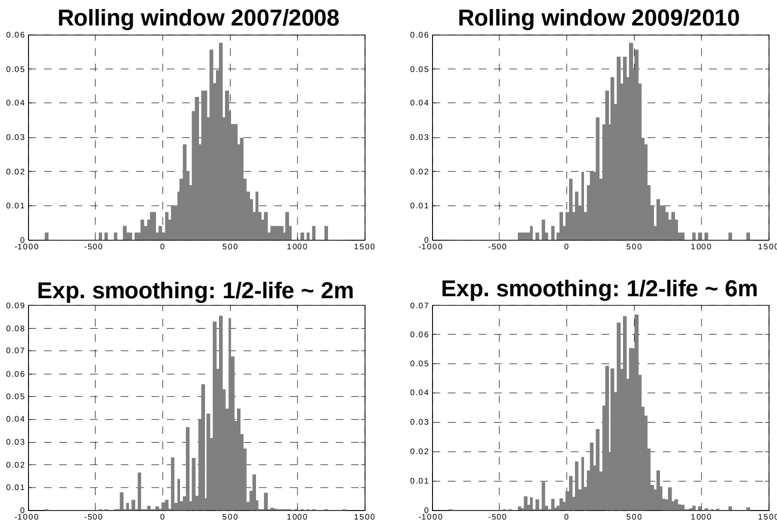

Figure 3: P&L distribution of option book under flexible stress-tests

the risk drivers X are estimated from the historical scenarios $\{x_{t,n}\}$ as follows

$$\mu_n^\lambda \equiv \gamma \sum_{t=1}^T e^{-\lambda(T-t)} x_{t,n}, \quad n = 1, \dots N. \tag{8}$$

$$\Sigma_{m,n}^\lambda \equiv \gamma \sum_{t=1}^T e^{-\lambda(T-t)} x_{t,m} x_{t,n} - \mu_m^\lambda \mu_n^\lambda, \quad 1 \leq m \leq n = 1, \dots N. \tag{9}$$

In this expression λ is a user-defined decay coefficient that determines the persistence of past observations, and $\gamma \equiv 1/\sum_{t=1}^T e^{-\lambda(T-t)}$. Second, a large number J of Monte Carlo scenarios for the risk drivers $\{\widetilde{x}_{j,n}\}_{n=1,\dots,N}^{j=1,\dots,J}$ are simulated, assuming in most cases a normal distribution $\mathrm{N}\left(\mu^\lambda, \Sigma^\lambda\right)$. Third, the Monte Carlo scenarios for the drivers are priced into p&l scenarios as in (3). Fourth, the p&l risk numbers are computed from the p&l scenarios.

We can estimate the expected values and covariances of the risk drivers as in (8)-(9), and generate the ensuing p&l distribution without making distributional assumptions and without generating Monte Carlo scenarios. To do so, we start with our panel $\{x_{t,n}\}$ of historical scenarios and we set the respective probabilities p_t proportional to the decay factor that appears in (8)-(9)

$$p_t \propto e^{-\lambda(T-t)}. \tag{10}$$

Then we determine the proportionality constant to ensure that the probabilities sum to one as in (6), see Figure 2. In words, more recent scenarios are considered more likely to occur again than older ones.

The empirical expectations and covariances of the risk drivers corresponding to the probabilities (10) then match exactly the smoothing recipe (8)-(9), as showed in the appendix. Furthermore, the relevant statistics for the portfolio p&l follow immediately without simulations and repricing. This methodology is used in BARRA (2010) to compute the exponentially smoothed VaR. However, it is easy to compute the whole p&l distribution, as well as any additional statistics, as discussed in the appendix.

To illustrate, we consider the above long-short options book. We set the decay coefficient in such a way that the half-life is two months. With daily observations, this means $\lambda \approx 0.0166$. Then we compare the results with a half-life decay of six months, or $\lambda \approx 0.0055$. In the bottom portion of Figure 3 we display the p&l distribution corresponding to the two different decay factors and in Table (11) below we report select relevant statistics

Exp. smoothing	Half-life $\approx$ 2 months	Half-life $\approx$ 6 months
Exp. value	394	389
St. dev.	183	203
VaR 99%	-259	-259
Cond. VaR 99%	-297	-307

(11)

We refer the reader to the appendix and to the code available at http://symmys.com/node/150 for more details.

2.3 Market conditioning

The rolling window methodology of Section 2.1 and the exponential smoothing approach of Section 2.2 can be interpreted as techniques to emphasize some scenarios over others according to when the scenarios occurred. It is also interesting to emphasize scenarios according to key macroeconomic conditions. To do so, let us select a relevant macro-economic indicator Y and a region of values $\mathcal{Y}$ that we intend to emphasize. Then we set the probabilities $\{p_t\}$ of the historical scenarios $\{x_{t,n}\}$ equal to a constant when the historical realization of the relevant indicator lies in the emphasis zone

$$p_t \propto \begin{cases} 1 & \text{if } y_t \in \mathcal{Y} \\ 0 & \text{otherwise,} \end{cases} \tag{12}$$

where, again, the probabilities must be normalized to sum to one as in (6). This digital, or "crisp", macro-economic conditioning is illustrated in the left portion of Figure 4.

To illustrate the crisp conditioning approach, we consider as macroeconomic indicator Y the ten-year inflation swap rate. We condition on a higher inflation environment, defined as $Y > 2.8\%$. This way the emphasis zone is $\mathcal{Y} \equiv (0.028, \infty)$. Then we apply the conditional distribution (12) to our long-short options book: In the top-left portion of Figure 5 we display the conditional

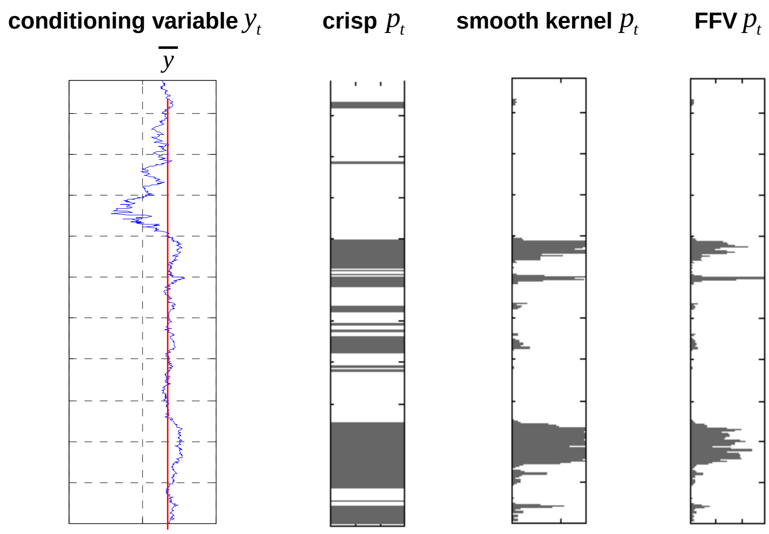

Figure 4: Probabilities emphasizing different market conditions

historical p&l and in the left column of Table (13) we report select relevant statistics

Conditioning	Crisp	Kernel	FFV
Exp. value	378	380	381
St. dev.	243	228	249
VaR 99%	-291	-287	-291
Cond. VaR 99%	-360	-397	-412

(13)

For more details we refer the reader to the appendix.

A more general approach to market conditioning replaces the crisp selection criterion (12) with smooth criteria. For instance, we can condition on a given target level $\overline{y}$ for the macro-economic indicator by using a normal kernel. In this framework, the historical realization of the indicator y_t is given a weight proportional to its distance from the target $\overline{y}$, as measured by a normal pdf around the target, refer to the center portion of Figure 4

$$p_t \propto e^{-(y_t-\overline{y})^2/2s^2}. \tag{14}$$

In this expression, the "volatility" s is knows as the "bandwidth" and controls the level of smoothing: a small s corresponds to a crisp selection criterion similar to (12), whereas a large s brings us back to the unconditional case where all the probabilities are approximately equal $p_t \approx 1/T$.

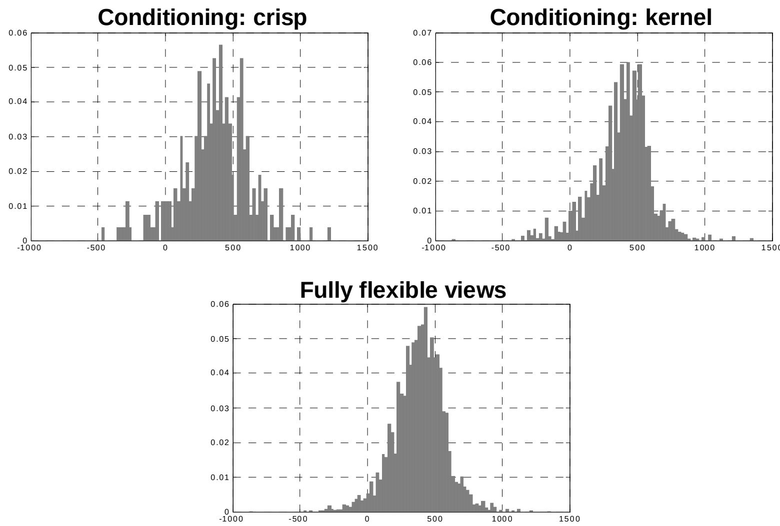

Figure 5: P&L distribution of option book under flexible stress-tests

To illustrate the smooth kernel conditioning approach, we consider again our options book and the macroeconomic indicator Y set as the ten-year inflation swap rate. We wish to stress-test our book in a relatively high inflationary environment. Therefore we set the target in (14) as $\overline{y} \approx 3\%$. Then we set the smoothing level as the unconditional sample standard deviation of the daily inflation rate changes, i.e. $s^2 \equiv \sum_{t=1}^{T} (\Delta y_t)^2 / T$.

In the top-right portion of Figure 5 we display the p&l distribution stemming from the probabilities (14) and in the center column of Table (13) we report select relevant statistics. We refer the reader to the appendix to see the computation of the summary statistics and to the code for more details.

We remark that it is immediate to generalize both the crisp conditioning (12) and the smooth kernel conditioning (14) to the case of several simultaneous conditioning indicators $Y \equiv \{Y_k\}_{k=1,\ldots,K}$.

3 Partial information: Fully Flexible Views

In the examples in Section 2, namely (5), (10), (12), (14), all the probabilities $\{p_t\}$ of the historical scenarios $\{x_{t,n}\}$ are completely determined by the chosen model.

A more general situation occurs when only partial information is available to determine the probability of each scenario. In these situations, we can use the

Fully Flexible Views approach in Meucci (2008) to fill in the missing information. Here we present two applications.

3.1 Partial information market conditioning

In the kernel smoothing approach we used the Gaussian kernel (14) to express the conditioning statement that the macroeconomic variable Y is expected to be close to the target $\overline{y}$ in the generalized empirical distribution (4), i.e.

$$\sum_{t=1}^{T} y_t p_t \equiv \overline{y}. \tag{15}$$

However, we could have used a different type of kernel, or, even with the same kernel, a different value for the bandwidth. Unless we feel very strong about the shape of the kernel and the size of the bandwidth, we should avoid imposing such strong constraints.

To choose the solution $\{p_t^*\}_{t=1,\ldots,T}$ that only reflects our desired statement (15) without imposing other constraints, we follow the Fully Flexible Views approach in Meucci (2008). Out of the many solutions $\{p_t\}$ that satisfy the macroeconomic conditioning statement (15), we select the closest to the historical distribution. To define the concept of the "closest" distribution p^* we draw on statistics, physics and information theory. More precisely, we define the "distance" of a distribution p from a reference "prior" distribution q, in our case the historical distribution with equal probabilities $q_t = 1/T$, by the Kullback-Leibler divergence, or relative entropy, $d\left(p,q\right) \equiv \sum_{t=1}^{T} p_t \ln\left(p_t/q_t\right)$. Therefore we define the solution as

$$\begin{aligned} p^* \quad &\equiv \quad \underset{p}{\operatorname{argmin}} \sum_{t=1}^{T} p_t \ln\left(T p_t\right) \\ & \quad\quad \text{such that } p \text{ satisfies } (15)\,. \end{aligned} \tag{16}$$

Qualitatively, the optimal probabilities p^* resemble the Gaussian kernel (14), see Figure 4.

To illustrate partial information market conditioning by means of Fully Flexible Views, we consider again as macroeconomic indicator Y the ten-year inflation swap rate. We set the target in (15) again as $\overline{y} \approx 3\%$. In the right column of Table (13) we report select relevant statistics stemming from the probabilities (16) as applied to our options book p&l scenarios. We refer the reader to the appendix to see the computation of the summary statistics and to the code at http://symmys.com/node/150 for more details.

3.2 Double-decay covariance

It is common practice to estimate the covariance of the risk drivers using a high decay (short half-life) $\overline{\lambda}$ for the volatilities and a low decay (long half-life) $\underline{\lambda}$

for the correlations, all while the expected values are set to zero, see De Santis, Litterman, Vesval, and Winkelmann (2003). In formulas this reads

$$\widehat{\mu}_n \equiv 0, \quad n = 1, \ldots N \tag{17}$$

$$\widehat{\Sigma}_{m,n} \equiv \sqrt{\Sigma^{\overline{\lambda}}_{m,m}\Sigma^{\overline{\lambda}}_{n,n}} \frac{\Sigma^{\underline{\lambda}}_{m,n}}{\sqrt{\Sigma^{\underline{\lambda}}_{m,m}\Sigma^{\underline{\lambda}}_{n,n}}}, \quad m, n = 1, \ldots N, \tag{18}$$

where we used the notation in (9).

This approach is a variation of the exponential smoothing (8)-(9). As in that case, to obtain the distribution of the p&l and all the relevant statistics, we could assume that the drivers are normally distributed $X \sim \mathrm{N}(\widehat{\mu}, \widehat{\Sigma})$, then generate a large number of Monte Carlo scenarios and price them into p&l scenarios as in (3).

If we try to avoid generating scenarios and assuming a specific distribution by tweaking the probabilities p_t of the historical scenarios as in (10), we face a conundrum. With two different decay rates $\overline{\lambda}$ and $\underline{\lambda}$ we obtain two different sets of probabilities $p_t \propto e^{\overline{\lambda}(T-t)}$ and $p_t \propto e^{\underline{\lambda}(T-t)}$. Is there a better p_t, which is consistent with both decay rates $\overline{\lambda}$ and $\underline{\lambda}$? To achieve consistency, p_t must be such that the sample moments of the empirical distribution match the estimators (17)-(18). In formulas, as we prove in the appendix, p_t must satisfy the following $N\,(N+3)\,/2$ equations

$$\widehat{\mu}_n = \sum_{t=1}^{T} p_t x_{t,n}, \quad n = 1, \ldots N \tag{19}$$

$$\widehat{\Sigma}_{m,n} = \sum_{t=1}^{T} p_t x_{t,m} x_{t,n} - \widehat{\mu}_m \widehat{\mu}_n, \quad 1 \leq m \leq n = 1, \ldots N. \tag{20}$$

In typical cases the number of scenarios T, and thus the number of probabilities p_t, exceeds the number $N\,(N+3)\,/2$ of equations: therefore the above equations do not provide enough information to fully determine p_t and thus there exist multiple solutions to this problem.

To choose the solution $p^* \equiv \{p^*_t\}_{t=1,\ldots,T}$, i.e. to fill in the missing information, we follow the Fully Flexible Views approach in Meucci (2008). Similarly to Section 3.1, out of the many solutions $\{p_t\}$ that satisfy the moments conditions (17)-(18) we select the closest to the prior, in this case the historical distribution. Therefore we define the solution as

$$\begin{aligned} p^* \equiv\ & \underset{p}{\operatorname{argmin}} \sum_{t=1}^{T} p_t \ln (T p_t) \\ & \text{such that } p \text{ satisfies } (19)\text{-}(20)\,. \end{aligned} \tag{21}$$

To illustrate the Fully Flexible Views approach in this context, we return to our long-short options book. First, we set $\overline{\lambda} \equiv 0.0166$ and $\underline{\lambda} \equiv 0.0055$ as model inputs. Next, we estimate the expectations and covariances of the risk drivers as in (17)-(18). Then we compute the fully flexible probabilities (21).

This way we have replaced any restrictive parametric assumption, such as normality, with the empirical distribution, without giving up the ability to in-corporate different decay factors $\overline{\lambda}$ and $\underline{\lambda}$ and the ability to generate all the statistics of the p&l distribution.

In the bottom portion of Figure 5 we display the p&l distribution and in Table (22) below we report select relevant statistics, computed as discussed in the appendix

Double-decay FFV	
Exp. value	381
St. dev.	189
VaR 99%	-141
Cond. VaR 99%	-237

(22)

We refer the reader to the code available at http://symmys.com/node/150 for more details.

References

BARRA, 2010, *BarraOne Analytics Guide* (MSCI Barra).

De Santis, G., R. Litterman, A. Vesval, and K. Winkelmann, 2003, Covariance matrix estimation, *"Modern investment management: an equilibrium approach", Wiley.*

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

———, 2008, Fully flexible views: Theory and practice, *Risk* 21, 97–102 Article and code available at http://symmys.com/node/158.

———, 2010, Factors on Demand - building a platform for portfolio managers risk managers and traders, *Risk* 23, 84–89 Article and code available at http://symmys.com/node/164.

A Appendix

Assume a market driven by a set of N of risk drivers $X \equiv (X_1, \ldots, X_N)'$. Assume as in (2) that we have a panel of T historical joint scenarios or Monte Carlo joint scenarios $x_{t,\cdot} \equiv (x_{t,1}, \ldots, x_{t,N})$, where $t = 1, \ldots, T$.

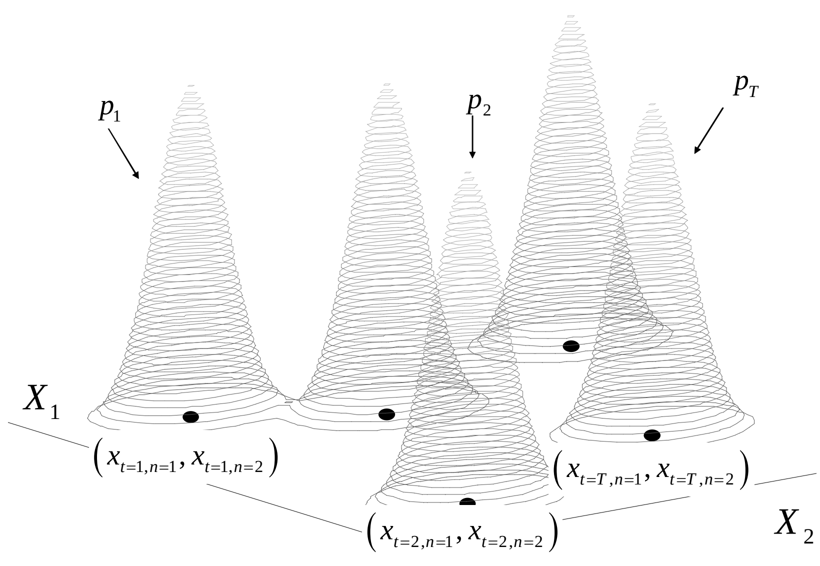

Figure 6: Flexible empirical distribution pdf, or multivariate histogram

In order to handle scenarios with non-even probabilities $\{p_t\}$, we model the distribution f_X of the risk drivers in terms of the "generalized empirical distribution"

$$f_X(x) \equiv \sum_{t=1}^{T} p_t \delta\left(x - x_{t,\cdot}\right). \tag{23}$$

In this expression δ is the "Dirac delta", a positive function which for any regular function g satisfies the following averaging property:

$$\int \delta(x - z)\, g(x)\, dx = g(z). \tag{24}$$

As a consequence, the Dirac$\delta(x - z)$ delta spikes to infinity at the point $x = z$, it tends to zero otherwise, and it integrates to one. As such, the Dirac delta corresponds to a unit probability mass, and the term $p_t\delta$ in (23) represents a probability mass p_t associated with the joint scenario $x_{t,\cdot}$. The flexible empirical distribution (23) can then be interpreted as a multivariate histogram with bins of infinitesimal width, see Figure 6. We refer to Meucci (2005) for more details on the Dirac delta in this context.

The generalized empirical distribution has a remarkable property, which is not shared by parametric representations f_X^{θ}: the distribution of an arbitrary function of the risk drivers $Y = y(X)$ is easily computed as follows

$$f_Y(y) \equiv \sum_{t=1}^{T} p_t \delta\left(y - y_t\right), \tag{25}$$

where $y_t \equiv y(x_{t,\cdot})$.is the arbitrary function y evaluated on the risk drivers scenarios $x_{t,\cdot}$ defined in (2). In words, the probability of Y spikes at each scenario y_t with a probability mass p_t. As a consequence, we can readily compute all the desired statistics of any variable that is fully determined by the risk drivers, such as the p&l (1), or the very risk drivers.

For instance, for the expected value of Y we obtain

$$\begin{aligned} \mu_Y &\equiv \mathrm{E}\{y(X)\} \equiv \int y f_Y(y)\, dy \\ &= \sum_{t=1}^{T} p_t \int y \delta(y - y_t)\, dy = \sum_{t=1}^{T} p_t y_t, \end{aligned} \tag{26}$$

where the last equality follows from (24).

A similar calculation for the covariances yields

$$\Sigma_{Y,Z} \equiv \mathrm{Cv}\{y(X), z(X)\} = \sum_{t=1}^{T} p_t y_t z_t - \mu_Y \mu_Z, \tag{27}$$

from which the correlations can also be computed. Skewness and kurtosis are computed in a similar way.

As an additional example, consider the value at risk, or quantile, with confidence c of the variable Y, which is defined as the number VaR_Y such that $\mathbb{P}\{Y \leq \mathrm{VaR}_Y\} = 1 - c$. When applied to the flexible empirical distribution (23) this definition yields

$$\mathrm{VaR}_Y = \max\{y_t\} \text{ such that } \sum_{y_t \leq \mathrm{VaR}_Y} p_t \leq 1 - c. \tag{28}$$

As a final example, consider the conditional value at risk with confidence c, or expected shortfall, which is defined as the expectation conditioned on outcomes below the value at risk, i.e. $\mathrm{VaR}_Y \equiv \mathbb{E}\{Y|Y \leq \mathrm{VaR}_Y\}$. When applied to the flexible empirical distribution (23) this definition yields

$$\mathrm{CVaR}_Y = \frac{\sum_t p_t y_t}{\sum_t p_t}, \text{ where } t \text{ such that } y_t \leq \mathrm{VaR}_Y \,. \tag{29}$$