

Department of Mathematics
University of Toronto, Room 205
1 Spadina Crescent, Toronto, ON M5S 3G3, Canada
Department of Mathematics
University of Toronto, Room 205
1 Spadina Crescent, Toronto, ON M5S 3G3, Canada

Non-gaussian Multivariate simulations in Mark-to-Future calculations

O. CROISSANT, GUSTAVO COMEZANA, MARCOS ESCOBAR, PABLO FERNÁNDEZ,
NICOLÁS HERNÁNDEZ AND LUIS ÁNGEL SECO¹

Abstract: Scenario generation techniques for gaussian markets are very well understood, and are based on Monte Carlo methodologies for multivariate normal variables determined by marginal means and covariance matrices. This paper presents an approach for the calibration and generation of nongaussian scenarios; it is first developed in the one-factor setting, and then extended multifactor situations. They are both based on extreme value theory.

1. Introduction

The problem of calibrating financial time series is at the heart of financial risk management. RiskMetrics, when it assumes linear portfolios of gaussian risk factors, needs estimates for covariance matrices. When RiskMetrics assumes nonlinear portfolios with GARCH underlyings, one needs to derive the parameters of the underlying marginal distributions, and even then one assumes that standard correlations are a suitable measure of dependence among the risk factors. In the Mark-to-Future framework, one is free from these assumptions but at the same time needs to have the ability to generate forward-looking scenarios that are compatible with historical observations.

The basic premise of this paper is that it is possible to split the calibration problem into two parts:

- Calibration of one-dimensional marginal distributions.
- Calculation of a dependence structure among the risk factors.

The dependence structure is discussed in section 3 below. Traditionally, the dependence structure is based on the correlation matrix. While this is totally appropriate for gaussian distributions, it does not adequately reflect the dependence structure for non-gaussian variables. In this paper, we shall use instead the normal rank correlation as a measure of dependence, which enjoys a number of advantages over the standard approach.

The fitting of one-dimensional distributions is seemingly simpler, although careful inspection of the problem leads to a number of possible fitting methodologies. These can basically be classified as parametric versus non-parametric. Parametric methods may give rise to vastly different distributions depending on the model adopted. Non-parametric methodologies tend to overfit the data, leading to unsatisfactory scenario simulations.

The purpose of this paper is to examine the validity of non-parametric multivariate predictions for risk management, when the dependence structure is given by the normal rank correlation. We will test the methodology in the case of natural gas forward prices. The ideas in this paper lead naturally to considering other one-dimensional fitting methodologies, as well as more general dependence structures, and also to generalizations to time-dependent (non-iid) distributions. We plan to tackle these in subsequent papers.

2. The Dependence Structure

In this section we wish to consider a general framework of dependence structures, which we will refer to as generalized correlations.

It is a well known fact that two gaussian variables are independent if and only if they are uncorrelated. In other words, a single number encapsulates their entire dependence structure. The same is not true for general random variables. In fact, the concept of correlation can be very deceptive outside the gaussian domain. We try to bridge this gap by introducing a more general correlation concept, as described below. Let us agree on some notation and fundamental concepts first: Here, and throughout this paper, we define

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp(-t^2/2) dt,$$

the cdf of the univariate normal.

Lemma 1 (Fisher Transformation). *If a random variable X has cumulative distribution function (cdf) P , then $P(X)$ is uniformly distributed in $[0,1]$. Similarly, $\phi^{-1}(P(X))$ is a gaussian random variable with mean 0 and variance 1.*

Next we discuss an extended notion of correlation. Consider two random variables X and Y , with cdf given by P and Q respectively, and two increasing functions, F and G . We define:

$$\sum_{F,G}(X, Y) = E\{F^{-1}P(X)G^{-1}Q(Y)\}.$$

The idea of this generalization is to calculate the correlation of a convenient transformation of X and Y , namely their remapping into a random variables with cumulative distributions given by F and G respectively.

This generalizes more familiar correlation concepts as follows:

- **Correlation**, given by:

$$Corr(X, Y) = \sum_{P_X, Q_Y}(X, Y).$$

- **Spearman Ratio**, given by:

$$E(P_X(X)P_Y(Y)) = \sum_{1,1}(X, Y),$$

where 1 denotes the cdf of the uniform distribution.

- **Normal Rank Correlation**, given by:

$$\mathcal{K}(X, Y) = \sum_{\phi, \phi}(X, Y).$$

Note that the normal rank correlation for gaussian variables coincides with the usual correlation.

The normal rank correlation is specially meaningful because of the following result:

Lemma 2. *Consider a random vector (X, Y) that maximizes the entropy among all random vectors with X and Y as marginals. Then X and Y are independent if and only if $\mathcal{K}(X, Y) = 0$.*

The advantages of this generalized concept of correlation include the fact that one can easily fit distributions in n factors to market data, represented simply by observed marginals with cdfs given by P_i and computed normal rank correlations V , as follows:

Let X be a gaussian vector with normal-0,1 marginals and variance/covariance matrix V . The random vector Y with components

$$Y_i = P_i^{-1}(\phi(X_i))$$

has normal rank correlation V , and marginal cdfs P_i and hence solves the problem.

3. One-factor models

The previous section managed to reduce the problem of calibration and generation of financial scenarios to a collection of decoupled, one-factor models. In this section we present some general ideas dealing with this problem. This will lead us to one specific method, which is the leitmotif of this paper: extreme value theory, Appendix A 8.

In what follows X will denote a real-valued random variable. We will specify the probability law associated to X by its cumulative distribution function (cdf), F . We will denote by the conventional moments of X , defined as

$$\mu_r = E(x - \mu)^r, \quad r = 1, 2, 3, \dots$$

Location. Intuitively a good measure of location should reflect the “center of mass” of the distribution. For symmetric distributions it is very clear how to define such a measure: the center of symmetry is the only logical choice. For more general distributions different measures have been proposed in the statistical literature, the best-known among being the mean of the random variable. The mean, usually denoted by μ , is defined to be the first moment of X .

Spread. A good measure of spread should reflect the degree dispersion of the values taken by X . Traditionally this has been measured by the standard deviation, $\sqrt{\mu_2}$.

Skewness. A good measure of skewness should account for the degree of asymmetry of a distribution. A widely used measure of skewness is based on the third moment, and is given by

$$s = \frac{\mu_3}{\mu_2^{3/2}}.$$

Kurtosis. A good measure of kurtosis should reflect certain features of the shape of a distribution, such as the presence of fat tails. Traditionally this has been measured by the ratio between the fourth moment and the square of the variance

$$k = \frac{\mu_4}{\mu_2^2}.$$

The sample analogues of these quantities are obtained by substituting the theoretical moments for their sample estimators. Traditionally these quantities have been used to reflect, in one way or another, typical deviations from normality like asymmetries or fatter tails. This line of thought has led some authors (see for example Hull and White (1998)) to extend the gaussian model to more general families of distributions by adding a number of additional parameters that can account for the empirical skewness and kurtosis observed in the sample.

This approach has some serious drawbacks. Estimators for the parameters of these distributions based on sample moments are as unlikely to be reliable as the sample moments on which they are based. This fact becomes more evident for long tails and/or very skewed distributions for which the empirical moments exhibit a high variability as well as an extreme sensitivity to outliers. In practice, an unrealistically large sample size would be required in order to obtain accurate estimators for both empirical moments and the parameters of the model in question.

In general, for data exhibiting a higher frequency of extreme observations than that expected for a normal model, more robust descriptive measures are needed. Recently, an alternative and appealing system of similar descriptors have been proposed by Hosking (1990). In analogy with the conventional moments, they are called L -moments. L -moments are linear statistics of the quantile function $Q(u) = F^{-1}(u)$ of the random variable X . They are defined as

$$\lambda_r = \int Q(u)P_{r-1}(u)du,$$

where $P_r(u)$ are orthogonal polynomials defined in the interval (0,1) of the form

$$(P_{r-1}(u)) = \sum p_{r,k}u^k,$$

where

$$p_{r,k} = \frac{(-1)^{r-k}(r+k)!}{(k!)^2(r-k)!}.$$

Like their traditional counterparts, L -moments can be interpreted as both intuitive and simple descriptors of the shape of a general distribution, as location, scale, asymmetry and kurtosis can be described in terms of the first four L -moments. However, they offer a number of advantages over conventional moments. They can be defined for a wider class of distributions, such as distributions that decay like power laws for which moments of higher order do not exist. Second, they completely characterize the probability law of the random variable, unlike conventional moments. Finally,

and perhaps most importantly of all, they can be accurately estimated by their empirical analogs even for distributions with fat tails. It has been shown in Monte Carlo studies (see for example, Hosking (1990) and Sankarasubramanian and Srinivasan (1999)) that typically they just require a moderate sample size to obtain a reasonable accuracy. This advantage in efficiency and many other desirable properties offered by L -moments have been discussed in detail by Hosking (1986; 1990). Based on the assumption that the sample L -moments are reasonably robust and efficient estimators for the true L -moments of the unknown distribution of the population, we can expect a model that matches the empirical L -moments to be a more accurate and robust approximation to the unknown distribution than other models based on a fitting to higher order sample moments. In an upcoming paper we discuss this problem in more detail and provide a solution to it.

The focus of this paper is to deal with the reconstruction of the marginal one-factor distributions by several methods. We review two of them: one based on the Hill estimator, another based on an explicit formula that invokes the L -moments to determine the model parameters.

4. Multivariate Model

The process of building a multivariate simulation up can be summarize into the following steps:

- Extraction of marginal distributions from historical samples.
- Computing of Normal Rank Correlation Matrix.
- Creation of a Gaussian simulation with “the correct” dependency structure.
- Applying inverse of Fischer transform on the marginals.

The pioneers of this ideas have been Hull and White, they applied it to mixtures of gaussian distributions. Our step ahead has been its extension to the non-parametric case.

5. Limitations of the model

The proposed model has a number of limitations, usually linked to the fact that the historical data available may not be stationary, or show trends, mean reversion, etc. In this section we document some of them, and propose a methodology to be studied in upcoming papers.

Our first exhibit about such limitations is a series of electricity price data.

The data presents a number of clear features, all of them rooted in the strong temporal structure of the price data, making the methodology presented in this paper inadequate:

- There is a clear correlation between a forward curve for a fixed term, and lagged versions of other terms.
- There is a clear seasonality in the series, as electricity prices tend to peak in the summer.

Note however, that the extreme tails observed in the data are not an obstacle, since the methodology presented here addresses precisely those issues. Compare, for instance, the histogram of the electricity forward price data and the one for forward gas prices: the gas price histogram clearly gives rise to a marginal structure that is harder to fit.

The methodology to be used in situations like this one is a generalization of the one presented here. Roughly speaking, we will apply filters to each of the marginals to achieve a series of residuals that are free from time-dependent effects (seasonality, autocorrelations, etc.) and are, in short, independent identically distributed observations. The resulting series of residuals will be dealt with the methods presented in this paper.

6. Traditional approaches to VaR

RiskMetrics-type VaR. The basic assumption is that the returns are normally distributed; the parameters (mean and volatility) are then estimated using the historical information available and the distribution obtained in this way is used to compute quantiles.

Problem: The financial data seem to have heavier tails than the normal distribution. If that is so the quantiles obtained with the normal approximation will be lower than they should.

Historical VaR. The basic assumption now is that the distribution of returns is unknown; one proceeds to compute quantiles using the empirical distribution obtained from the historical data of the portfolio.

Problem: Since we are trying to understand extreme losses, the empirical distribution will not give good information for high quantiles since there are, in general, only a few data points in the “extreme” range (or none at all!).

EVT provides a different approach to the problem, Appendix A 8. We may assume that the distribution F of the X 's is unknown but it satisfies the conditions for convergence of the maxima (i.e. the tail is like $x^{-1/k}$ for x large). We can then use the results we have reviewed to give a measure of the value at risk, in several ways:

- Get estimates of the tail of F and use it to compute quantiles that are taken as measure of value at risk.

- MaxVaR: instead of using the X , use new random variables Y , that correspond to maxima of the X in consecutive blocks of n days. Estimate the parameters of the limit distribution of the maxima, use that limit distribution to get the quantiles of the maxima and take that as measure of value at risk.
- In any of the two former cases instead of using quantiles as measure of value at risk, use the quantities given by $E[X|X > Q]$ and $E[Y|Y > Q]$ respectively, where Q is a given quantile of the distribution.

In all three cases there are procedural difficulties, such as deciding how large the n has to be in order that M_n has an approximate distribution given by $H_{k,\mu,\psi}$, or what should the initial threshold u be to start use values in excess of the threshold to approximate the tail of F .

7. Application to Crude Oil and Natural Gas portfolios

In this section we apply the multifactor methodology to the calculation of Value-at-risk numbers for a portfolios of forward price contracts for crude oil and natural gas.

In this situation, a gaussian fit produces very poor results: if one computes the frequency of outliers beyond the theoretical gaussian VaR numbers, one gets the following chart:

TABLE 1	
	% outliers
5% VAR	8%
1% VAR	4%

The conclusion is obvious: gaussian methodologies are slightly off for 95% levels, and yield unacceptable results for 99% VaR calculations. The methodology described in this paper obtains results slightly better than gaussian at 95% levels, and dramatically better at 99% levels, where they still maintain acceptable levels of accuracy in their VaR predictions.

The methodology to test the predictions of our results is an out-of-sample VaR comparison test, which uses the price series and a collection of futures portfolios. Roughly speaking, using subsets of the available history, we obtain out-of-sample predictions for the Value-at-Risk numbers for each of the portfolios under consideration, with confidence levels of 95% and 99%. Using the previously ignored history, we will then check if the actual losses that the portfolio incurred are above the VaR calculations or not. Finally, after doing this for a large collection of sub-histories, we check whether the loss outliers have the correct frequency: 5% and 1% respectively.

In greater detail: Consider futures portfolios Π_k ; constant through time. Let x denote each scenario for a futures curve, which is a vector of dimension equal to the number of available contracts at any point in time; in our case, this equals 12.

Our data set consists of a series of 800 daily observations of the forward curve, denoted x_i , $i = 1, \dots, 800$, respectively.

The test selects a number m , for instance 400, which will account for about one years worth of data. Starting at the oldest point in time for when data is available, we select an initial window of m values given by x_1, \dots, x_m , and we calibrate the multivariate distribution to this first window. Next, for the calibrated distribution, we generate a number of scenarios (1000, for example) which we then use to stress-test each portfolio. According to this, we compute the non-parametric 95% or 99% Value-at-Risk, $V_{k,m}$, of portfolio Π_k .

Using the scenario x_{k+1} which was available in our dataset, but so far ignored in our mark-to-future VaR calculation, we check whether

$$\Pi(x_{m+1}) > V_{m,k}$$

or not. This has the effect to test the observed outliers in the P&L distribution with the VaR calculations according to our model, for this particular window.

The window is then rolled one unit to the right, until our available data is exhausted, and the corresponding check is performed each time.

Finally, we compare how frequently the portfolio values exceed the projected value-at-risk number, with the theoretical one, that is, 5% (or 1%).

The results obtained were applied to 12 different futures portfolios, and are summarized in the table below. The numbers quoted include, for both the 95% and 99% VaR, the smaller outlier frequency, the largest frequency and the average frequency across the 12 different portfolios under consideration.

TABLE 2

	Smallest	Average	Largest
5% VAR	3.00%	6.05%	8.02%
1% VAR	0.25%	0.67%	1.50%

References

- [1] BOOTHE, P. AND GLASSMAN, D. (1987): “The Statistical Distribution of Exchange Rates: Economic Evidence and Economic Implications”. *Journal of International Economics* **22**, 297–319.
- [2] BORWEIN, J. W. AND LEWIS, A. S. (1991): “Convergence of best entropy estimates”. *SIAM Journal of Optimization* **1** (2), 191–205.
- [3] CORRADO, C. AND SU, T. (1997): “Implied Volatility Skews and Stock Index Skewness and Kurtosis implied by S&P 500 Index Option Prices”. *Journal of Derivatives* **4**, 8–19.
- [4] CROW, E. L. AND SIDDIQUI, M. M. (1967): “Robust estimation of location”. *Journal of the American Statistical Association* **62**, 353–389.
- [5] EMBRECHTS, P., KLÜPPELBERG, C., MIKOSCH, T.: *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag, 1997.
- [6] HULL, J. AND WHITE, A. (1998): “Value at Risk When Daily Changes in Market Variables are not Normally Distributed”. *Journal of Derivatives* **5**, no. 3, 9–19
- [7] KOTZ, S., MADARAJAH, S.: “Extreme Value Distributions”. Imperial College Press, 2000.
- [8] HOSKING, J. R. M. (1986): “The theory of probability weighted moments”. Research Report RC12210, IBM Research Division, Yorktown Heights, N.Y.
- [9] HOSKING, J. R. M. (1990): “L-moments: Analysis and estimation of distributions using linear combinations of order statistics”. *Journal of the Royal Statistics Society, Series B*, **52**, 105–24.

8. Appendix A: Extreme Value Theory

Extreme Value Theory (EVT) can be thought as the study of tails of distributions. It has been widely used in Engineering for studying extremal events (such as earthquakes, floods, etc.). More recently, it has become popular in the financial context. Our setting is as follows:

Consider a random variable X (as a model of, say, the daily returns of a portfolio), with cumulative distribution function (cdf) F . One could postulate a certain distribution function for the random variable (for example, a lognormal one) and fit

the parameters by means of historical data. If interested in extremal events, one may compute the quantiles and from there obtain risk measurements, such as VaR. However it is well known that returns do not, in general, follow a log-normal distribution, but exhibit fat tails: normality or lognormality lead to underestimation of tail events, and losses that exceed VaR bounds occur more often than predicted. What EVT proposes is to study the maximum of a random sample of n values of X , that is

$$M_n = \max(X_1, \dots, X_n)$$

where the X are independent and identically distributed (i.i.d) with common distribution function F . If the X_j represent the daily returns of a portfolio then $\text{Prob}(M_n > X)$ represents the chance of having at least a loss that exceeds x in a period of n days.

The advantage of considering the maximum value is that, no matter what F is

$$\lim_{n \rightarrow \infty} \text{Prob}(M_n \leq X) = H_{\xi, \mu, \psi}(X)$$

where

$$H_{\xi, \mu, \psi}(X) = \exp(-(1 + \xi(x - \mu) \div \psi)^{-1/\xi})$$

ξ, μ, ψ are the parameters of location, scale and shape, respectively, and $a = \max(a, 0)$. Here the case $\psi = 0$ has to be understood in the limiting sense, that is

$$H_{0, \mu, \psi}(X) = \exp(-\exp((x - \mu) \div \psi))$$

The normal and lognormal distributions correspond to the case $\psi = 0$, but for many financial series, ψ seems to be positive. This parameter is in some sense related to the size of the tail of the distribution:

$$1 - F(x) \cong x^{-1/\psi}$$

for x large.

This paper considers two methods for the estimation of the parameters ξ, μ, ψ (the biggest difficulty lying in the parameter ψ): maximum likelihood, and L-moments. This ψ also appears when one considers the excess distribution function

$$F_u = P(X - u \leq x | X > u) = (F(x + u) - F(u)) \div F(u)$$

and the mean excess function

$$\Pi(x_{m+1}) > V_{m,k}$$

If the behaviour of the maximum is like then looks like a the generalized Pareto distributions where

$$\Pi(x_{m+1}) > V_{m,k}$$

and

$$\Pi(x_{m+1}) > V_{m,k}$$

This is behind some of the methods that are used to estimate the tail of $F(x)$. In fact, since

$$\Pi(x_{m+1}) > V_{m,k}$$

one can use the empirical distribution of the sample to approximate $F(u)$ and a parametric method, for example maximum likelihood, to estimate through a GPD. We need therefore: A choice of the initial threshold u . Estimators and of and . One way to choose u is based upon the following observation: for a generalized Pareto distribution with , the mean excess function is given by

$$\Pi(x_{m+1}) > V_{m,k}$$

in particular, it is linear in u . We can now take the empirical mean excess function

$$\Pi(x_{m+1}) > V_{m,k}$$

and choose u in such a way that is approximately linear for . The estimators can be obtained, for example, by using the maximum likelihood method to adjust a to the data points in excess of u (the).

O. Croissant
@

Gustavo Comezaña
Sigmanalysis
Suite 340 - The Fields Institute
222 College St., Toronto, ON M5T 3J1, Canada
gustavo@sigmanalysis.com

Marcos Escobar
marcos@math.toronto.edu

Pablo Fernández
Departamento de Matemáticas
Universidad Autónoma de Madrid
Ciudad Universitaria de Cantoblanco, s/n. 28049-Madrid, Spain
pablo.fernandez@uam.es

Nicolás Hernández
nicolas@math.toronto.edu

Luis Ángel Seco
RiskLab
University of Toronto, Room 205
1 Spadina Crescent, Toronto, ON M5S 3G3, Canada
seco@math.toronto.edu