Fitting financial time series returns distributions: a mixture normality approach

Riccardo Bramante and Diego Zappa

Abstract

Value at Risk has emerged as a useful tool to risk management. A relevant driving force has been the diffusion considered as a benchmark of JP Morgan RiskMetrics™ methodology and the subsequent BIS adoption for all trading portfolios of financial institutions. In this paper we propose the use of mixture of truncated normal distributions in returns modelling. An optimization algorithm has been developed to obtain the best fit both in the univariate and in the bivariate case. The approach has shown evidence to fit the return distribution better than the RiskMetrics™ method completely maintaining local normality properties in the model.

In applied financial literature a relevant issue is the modelling of the empirical distribution of returns since many decision-making and asset pricing models depend on the assumptions related to the stochastic model underlying the data. Research on probability models in finance has given rise to several works (see e.g. Ruppert, 2011) but many questions are open:

a) Are the models supported by financial markets data?
b) How are the parameters in these models estimated?
c) Can the models be simplified?

Quoting the very famous fact by George Box “All models are false but some models are useful”, it can generally be agreed that complex models may be closer to reality but often involve many parameters and are not easy to be interpreted; on the other hand, too simple models may not capture important features of the data and can lead to serious bias.

Starting from these preliminaries, we made some considerations on how and whether the normal distribution might be exploited in financial modelling (Bramante, Zappa, 2012).

It is obvious to say that the normal distribution is too simple in this context, but it is far to be excluded in practice. In fact, many practitioners still make extensive use of normal distribution for returns even if different approaches are described in a very rich literature (McNeil et al., 2005 for an extensive review).

Extending encouraging results obtained in the univariate approach, by fitting returns of an in-
vestment conditional to a benchmark, we propose in this paper a procedure that appears a good compromise between theory and practice exploring how and when the global beta coefficient may be decomposed into local ones referred to specific regions of the domain giving raise to a different interpretation of dependence among returns.

I. Methodology

A. Fitting Distributions: notation and preliminary exploratory results

Before proceeding further, let us fix some ideas.

Let \((X, Y) \sim N_2(\mu, \Sigma)\) be a bivariate normal random variable. Let

\[
(F_n(x, y), F_{X,Y}(x, y; \theta))
\]

be, respectively, the empirical and the cumulative distribution function (ECDF and CDF) of \((X, Y)\). Let \((X, Y)_n\) be a sample drawn from \((X, Y)\). The bivariate QQ-plot of

\[
\{(x, y)_i, F_{X,Y}^{-1}(F_n(x, y)_i; \theta)\} \quad \text{for} \quad i = 1, 2, \ldots, n
\]

can be represented by plotting for some \(p\) the distance

\[
\| (x, y)_i - F_{X,Y}^{-1}(F_n(x, y)_i; \theta) \|_p \quad \text{for} \quad i = 1, 2, \ldots, n
\] (1)

If \((X, Y)_n\) comes from a distribution \(W\) different from \((X, Y)\) we expect that locus of points is locally different from the constant zero.

In an analogous manner, the analysis of the PP-plot over the normalized domain \([0,1]^2\) may be considered, i.e. the plot of

\[
\{ F_n((x, y)_i), F_{X,Y}((x, y)_i; \theta) \} \quad \text{for} \quad i = 1, 2, \ldots, n
\]

This choice avoids to select the parameter \(p\) in (1).

Parameter estimation is typically based on standard maximum likelihood (ML) or by robust estimation procedures, i.e. the median and the median absolute deviation from median (MAD) (Ruppert, 2011). In order to let the fitting process as flexible and maximally data dependent as possible, we have used the minimum distance estimation method (MDA) (Parr, 1985). It consists in
solving the general unconstrained problem

\[ \min_{\theta} d \left( \hat{F}_n(x, y), F_{XY}(x, y; \theta) \right) \quad X, Y \in \mathbb{R}^2 \]  

(2)

where \( d(\cdot) \) is an appropriate measure of discrepancy (or loss function). If \( F_{XY}(x, y; \theta) \) is the “true” distribution then the unconstrained estimator \( \hat{\theta} \) minimizing (2) has been shown to possess to be strongly consistent.

Depending on which \( d(\cdot) \) is used, further properties, e.g. robustness to extreme influence values, may be defined. Let \( A(x) \) and \( B_{\theta}(x) \) be continuous functions. Examples of \( d(\cdot) \) are

\[ d(A(x), B_{\theta}(x)) := \begin{cases} 
    KS: & \sup |A(y) - B_{\theta}(y)| \\
    MH: & E[|A(y) - B_{\theta}(xy)|] \\
    MQ: & E\left[|A(y) - B_{\theta}(y)|^2\right]
\end{cases} \]  

(3)

i.e. the Kolgomorov, Manhattan, Euclidean (Cramer – von Mises) distances, respectively. In particular the MQ loss function is the one adopted in this paper. To give more weights to extreme returns (2) has been modified as follows. Let \( z = [x_i y_i]' \)

\[ \min_{\theta} \sum_i d \left( \hat{F}_n(x_i, y_i), F_{X,Y}(x_i, y_i; \theta) \right) \| z_i \|^2 \quad X, Y \in \mathbb{R}^2 \]  

(4)

B. Fitting Bivariate Truncated Normal Distributions

Optimization in (4) can be applied even to conditional regions of the empirical returns distribution domain. If the solution doesn’t improve from the one obtained by using the whole domain, this should be an evidence that locally the process is not different from the unconditional one. By contrast if constraining (4) over \( \mathbb{R}_+^2 \subseteq \mathbb{R}^2 \) we get a solution better than the unconditional one, there is an evidence that the underlying process may be considered as a mixture of distributions (see also Kon, 1984). In this paper we have studied how to fit bivariate truncated Gaussian mixtures conditioning the domain with respect to return World Index. This is analogous to the standard exploratory investigation of dependence between market returns and stock returns in CAPM modelling.

Suppose that the location \( \bar{\mu} \) and variance-covariance matrix \( \bar{\Sigma} \) are the ones obtained by (4) (without containing \( \mathbb{R}^2 \)). Let \( x \) be a generic threshold for the marginal \( X \) such that
with $x_{tr_0} = -\infty$ and $x_{tr_k} = +\infty$. Let

$$f_Z(z; \bar{\mu}; \bar{\Sigma}; x_{tr_{i-1}}, x_{tr_i}) = \frac{\exp \left(-\frac{1}{2}(z-\bar{\mu})'\bar{\Sigma}(z-\bar{\mu}) \right)}{f(x_{tr_i}) \exp \left(-\frac{1}{2}(z-\bar{\mu})'\bar{\Sigma}(z-\bar{\mu}) dx \right) dy} \quad (6)$$

be the truncated normal pdf conditional to $\{(x \in x_{tr_{i-1}} \rightarrow x_{tr_i}) \cap (y \in \mathbb{R})\}$. Then

$$F_M(z; \bar{\mu}; \bar{\Sigma}) = \sum_{j=1}^{l_i} f_Z(z; \bar{\mu}; \bar{\Sigma}; x_{tr_{j-1}}, x_{tr_j})w_j$$

with $\sum_{j=1}^{l_k-1} w_j = 1 - w_k \quad (7)$

is the CDF in $z^1$. The matter is how to estimate the optimal weights and thresholds. Analogously to the problem stated in (4) we have implemented a procedure to search a solution to

$$\text{find } w_1, \ldots, w_{k-1} \text{ and } tr_1, \ldots, tr_{k-1} : \min_{w, k} \sum_i d \left( \hat{F}_n(x_i, y_i), F_M(x_i, y_i; \bar{\mu}, \bar{\Sigma}) \right) \|z_i\|_2^2 \quad (8)$$

i.e. we look for the smallest partition and the weights of the partition such that (8) is fulfilled. If $k = 1$ the problem in (8) reduces to (4) \footnote{From [7], $F_M(x, y; \bar{\mu}, \bar{\Sigma})$ may also be interpreted as a weighted sum of disjoint truncated distributions.}. Given the estimated thresholds and the corresponding domain partitions, it is easy to see that it is possible to evaluate local betas that represent relative risk to the market in each subset. The overall beta coefficient $\beta_G$ can be decomposed into a weighted average of all the local betas, where the weight is given be the fraction of each partition deviance, plus the beta parameter $\beta_M$ of the linear regression through the conditional means, where the weight is the deviance between them, that is

$$\beta_G = \sum_{j=1}^{k} \beta_j \frac{\text{Dev}_j}{\text{Dev}_G} + \beta_M \frac{\text{Dev}_M}{\text{Dev}_G} \quad (9)$$

\footnote{Using (1) on each element of the partition, we may estimate the best conditional location and scale parameter. From the risk manager point of view, only one reliable risk parameter estimate is needed in decision making. If we assume for each return series the existence of different estimates we may induce some difficulties in the decision process and the simplicity of our approach may get lost.}
II. Empirical Results

To give evidence of the aspects captured by the proposed model, we show results by using the returns of 300 Morgan Stanley Capital International (MSCI) indices, provided in the country and sector sub set, along with the World Index which is assumed to be the benchmark in the bivariate case. All the indices are denominated in US dollars and cover the period from January 1996 to November 2011.

The sample period was divided into pre specified consecutive intervals with a fixed length of 500 observations using a 20 trading session rolling window procedure: totally, approximately 20,000 data periods were tested.

In the univariate case, the optimization tool was used to automatically estimate – within the minimum distance approach modified by a weight function that emphasises the tail probabilities – mean and variance of the distribution. Results, in terms of degree of fit, are encouraging since a significant increment, 76% on average, was provided with no significant differences due to the index type. In the second step, the threshold optimization algorithm was employed to dynamically estimate the up and down thresholds and converge to the optimum solution of our minimization problem (see Figure 1 for an example regarding the World Index).

Figure 1: World Index Marginal Truncated Normal Distribution

Table 1 reports results on the relative gain in the discrepancy measure by using the MQ fit with respect to the ML estimators for both the normal and the Skew-t distribution. Figures in the table suggest that the introduction of a mixture of truncated normal distributions has significantly in-
creased the accuracy. With respect to the Skew-t, starting practically from the case with 3 thresholds (i.e. in the 92.42% of cases), the distribution captures at a satisfactory degree the most relevant aspects of daily returns empirical distribution, by regulating both kurtosis and fat tails.

Table 1: Threshold distribution (univariate case)

<table>
<thead>
<tr>
<th>Number of Threshold</th>
<th>% of Cases</th>
<th>Normal</th>
<th>Skew-t</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7.58</td>
<td>81.16</td>
<td>87.71</td>
</tr>
<tr>
<td>3</td>
<td>73.81</td>
<td>87.02</td>
<td>88.59</td>
</tr>
<tr>
<td>5</td>
<td>15.60</td>
<td>88.81</td>
<td>85.77</td>
</tr>
<tr>
<td>7</td>
<td>2.69</td>
<td>90.04</td>
<td>83.07</td>
</tr>
<tr>
<td>9</td>
<td>0.27</td>
<td>87.48</td>
<td>67.04</td>
</tr>
<tr>
<td>11</td>
<td>0.04</td>
<td>94.45</td>
<td>75.56</td>
</tr>
<tr>
<td>15</td>
<td>0.01</td>
<td>96.93</td>
<td>82.7%</td>
</tr>
</tbody>
</table>

For the bivariate case, the optimization tool was used in a similar way to estimate, in the first step, the vector of the unknown parameters where the criterion function measures the weighted squared distance between the empirical and the bivariate normal distribution and to select, in the second step, the optimal truncation thresholds conditional to the benchmark (see Figure 2 for an example regarding the MSCI Italy Index). In table 2 results regarding the relative gain in the discrepancy measure, for the two steps, are reported.

Table 2: Optimization results (bivariate case)

<table>
<thead>
<tr>
<th>Type of Index</th>
<th>First Opt. Step</th>
<th>Second Opt. Step</th>
</tr>
</thead>
<tbody>
<tr>
<td>Developed Markets Index</td>
<td>43,38</td>
<td>3,77</td>
</tr>
<tr>
<td>Emerging Markets Index</td>
<td>37,47</td>
<td>2,85</td>
</tr>
<tr>
<td>Emerging Sector Index</td>
<td>41,55</td>
<td>4,62</td>
</tr>
<tr>
<td>Europe Sector Index</td>
<td>45,28</td>
<td>4,00</td>
</tr>
<tr>
<td>World Sector Index</td>
<td>46,76</td>
<td>3,96</td>
</tr>
</tbody>
</table>

If compared to univariate outcomes, fitting results are less positive mainly in the truncation part of the model. Nevertheless, the analysis of where thresholds are set provides usual information on the behaviour of the index with respect to the benchmark. First, the partitioning algorithm returned in the 88% of cases at least three cut off points (table 3), where the approach with one threshold and a breakdown in the positive domain has been predominant in this context.
Moreover, along with a reduction in variance parameters similar to the one obtained in the univariate case, certain sign inversion in the bivariate normal parameters were reported in the first optimization step (table 4). Above all, reverse in sign in the correlation coefficient (12% on average) were obtained: generally this occurs when this parameter in not statistically significant.
Table 4: Bivariate Normal parameters

<table>
<thead>
<tr>
<th>Number of Threshold</th>
<th>Means</th>
<th>Correlation Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>14,13</td>
<td>7,16</td>
</tr>
<tr>
<td>2</td>
<td>15,93</td>
<td>22,38</td>
</tr>
<tr>
<td>3</td>
<td>25,10</td>
<td>12,95</td>
</tr>
<tr>
<td>4</td>
<td>32,69</td>
<td>12,50</td>
</tr>
<tr>
<td>5</td>
<td>42,37</td>
<td>3,39</td>
</tr>
<tr>
<td>6</td>
<td>25,00</td>
<td>-</td>
</tr>
</tbody>
</table>

As for relative risk analysis – typically measured by the classical beta coefficient – domain partitioning provides a set of disjoint conditional regions where the local relationship between the index and the benchmark can be slightly different with respect to the one on the domain as a whole. In particular, it is interesting to analyse reverse in sign in the index – benchmark relation (from a global positive beta to at least one negative local beta and the inverse case) which occurs on average in 61% of the total cases (table 5): this provides strong evidences that relative risk varies conditionally with the benchmark return regions which are bounded by the computed thresholds.

Table 5: Beta parameters sign inversion

<table>
<thead>
<tr>
<th>Number of Threshold</th>
<th>% negative</th>
<th>% positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>57,83</td>
<td>39,74</td>
</tr>
<tr>
<td>2</td>
<td>69,78</td>
<td>70,31</td>
</tr>
<tr>
<td>3</td>
<td>60,99</td>
<td>77,29</td>
</tr>
<tr>
<td>4</td>
<td>57,53</td>
<td>89,63</td>
</tr>
<tr>
<td>5</td>
<td>66,67</td>
<td>78,05</td>
</tr>
<tr>
<td>6</td>
<td>50,00</td>
<td>100,00</td>
</tr>
</tbody>
</table>
REFERENCES


Bramante R., Zappa D., 2012, Value at Risk Estimation in a Mixture Normality Framework, Submitted


