In defence of the Gaussian copula

It's been called the model that killed Wall Street, but *Jean-David Fermanian* argues that the structured credit market's Gaussian copula model has been unfairly maligned

From the beginning of the basket credit derivatives story, practitioners looked for a benchmark that would capture the dependence between several intricate risks, in particular between underlying default events and credit spread moves. Rapidly, the Gaussian copula model (GCM) of Li (2000) was adopted almost unanimously by major financial institutions.

The model provided a relatively easy way of pricing, quoting and hedging most credit derivatives (standard CDOs particularly), and was the backbone that supported the growth of the correlation trading business.

The success of GCM was partly due to its simplicity. As with any copula-based (static) model, the Gaussian copula model is consistent with the individual (spot) CDS curves of the names in the underlying basket.

Moreover, in its simplest one-factor version, a single correlation is equivalent to a price and summarises the primary risk. Thus, correlation traders were able to think and work in terms of correlation levels exactly as vanilla traders think/work in terms of the implied volatilities of calls/puts in the Black Scholes model.

Initially the framework was too crude to manage different maturities, but soon term structures of correlations were introduced.

Concerns are raised about the Gaussian copula model

Nonetheless, the GCM appeared iconoclastic in the fixed income world, especially compared with more sophisticated models, like HJM (Heath Jarrow Morton) or BGM (Brace Gatarek Musiela) in the interest rate derivative business. Also, the development of more exotic credit derivatives structures highlighted its limitations, and it was not long before academics such as Darrell Duffie, and leading authors such as Paul Wilmott and Nassim Taleb raised concerns. Their arguments were convincing and can be summarized as follows:

a. The GCM is the result of an ad-hoc specification, inspired by an actuarial point of view. Unlike the Black Scholes model, the GCM cannot be justified through replication/no-arbitrage arguments.
b. It does not include dynamics on the underlying financial assets, for instance individual credit spread curves or asset values, despite the interpretation of GCM as a strongly simplified structural model.
c. There is no way of having an idea of the "true" correlation parameter value, based on sound theoretical arguments. This induces delicate calibration issues, increases model risk and the uncertainty around available market quotes.
d. A single number (a correlation level) is not able to reflect the complexity of the dependence between dozens of assets or underlying risks. In particular, the Gaussian specification is unable to generate significant clusters of default events.
e. Some days during the credit crisis, no base correlations were able to fit the market prices.

Validation of the Gaussian copula

The Gaussian copula model (GCM) of Li (2000) assumes that the joint law of n default times T_k, $k = 1, \dots, n$, is specified through a Gaussian copula: for every date $t_1, \dots, t_n$, we have:

$$P(T_1 < t_1, \dots, T_n < t_n) = \Phi_\Sigma(\Phi^{-1}(P(T_1 < t_1)),\dots, \Phi^{-1}(P(T_n < t_n))$$

where Φ (resp. Φ_Σ) denotes the cumulative distribution function of a standard Gaussian random variable (resp. of an n-dimensional centred Gaussian vector, whose correlation matrix is given by Σ).

Consider Q_{kt} the risk-neutral survival probability of a given name k up to a fixed future date. It is read continuously in the market, through its corresponding CDS at time t. To validate the GCM framework, the historical dynamics followed by Q_{kt} have to be of the type:

$$dQ_{kt} = \varphi(\Phi^{-1}(Q_{kt}))\xi_t\sigma_k dW_{kt} + \mu_{kt}dt$$

where φ denotes the density of a standard Gaussian random variable, ξ_t is a deterministic function, σ_k is a firm-specific volatility-like term, μ_{kt} is an (arbitrary) random process and W_{kt} are n correlated Brownian motions.

Let us denote by ρ_{kl} these empirical correlations between the spread moves of the names k and l. Then, the break-even matrix, i.e. the correlation levels that must be used in a GCM pricing model, are given by:

$$\rho_{kl,Break\text{-}even} = 2\rho_{kl}\sigma_k\sigma_l \,/\, (\sigma_k^2 + \sigma_l^2)$$

These difficulties, along with the collapse of the structured credit derivatives market, led to GCM being called "the formula that killed Wall Street" (Salmon, 2009).

Counter arguments to the objections

However, there are arguments to counter these objections. For example Fermanian and Vigneron (2008, 2010) deduce the dynamics of individual credit spreads that are compatible with a pricing by replication of a CDO tranche under the GCM.

In this case, continuous delta-hedging strategies are fully efficient, and CDO prices become martingales, exactly as in the Black Scholes model. Such processes belong to the standard family of diffusions driven by Brownian motions. Even if their definition appears rather odd, they admit closed-form solutions and generate realistic spread curves. Thus, the GCM specification can be encompassed inside a broader fully dynamic framework, and points **a** and **b** above are much weakened.

Even more strikingly, it is possible to find the "right" correlation parameters of the Gaussian copula that should be used in a CDO pricer. These correlations depend on the realised correlations between individual credit spread moves of the names in the basket, combined with some credit spread volatility-like factors.

Contrary to popular opinion and usual practice (for example the use of equity return correlations as proxies for asset correlations for the GCM correlations themselves), spread volatilities matter, and in a non-trivial way in the GCM. In a certain sense, the previous point **c** is solved. The solution is given by the so-called "break-even correlation", a single number that should be put into a one-factor GCM pricer to hedge a CDO tranche with individual CDS. Such a parameter can be estimated empirically from historical series of CDS spreads. In theory, the implied (risk-neutral) value of the break-even correlation, as deduced from tranche quotes, should be close to this empirical break-even value, at least on average and exactly like historical and pricing (risk-neutral) volatilities should be in the Black Scholes framework.

Adding extra factors to the model

To respond to **d**, we stress the previous results are not restricted to the one-factor framework. By extending the GCM towards two, three or even more factors, we enrich significantly their perspective, at the price of increasing the number of parameters.

Therefore, we can show that a break-even correlation will be replaced by a break-even vector or even a break-even correlation matrix. We can then explain how to estimate them consistently with similar replication arguments and get closed-form formulas. It is true that practitioners rarely use the GCM with more than one factor, but this is only due to common practice and technical issues (calculation time) that do not seem insurmountable today.

Finally, **e** relies on the lack of quotes during the most hectic days in the market in 2007 and 2008. In our opinion, this cannot be interpreted as a weakness in the GCM specification, but rather by arbitrage opportunities in the market. These opportunities lasted a few days, because of the lack of liquidity just after Lehman's collapse, for instance. Indeed, when the GCM correlation levels go to one, the degree of dependence between the underlying default events is maximal "in absolute terms", and such a dependence (measured through copulas) cannot be higher mathematically. In other words, no alternative model would be able to generate more dependence in these extreme situations.

Restrictions and assumptions

To moderate this enthusiasm, we note that the results above have been obtained under some restricted conditions: European payoffs, only up-front payments (no running premia), a single time horizon, fixed recoveries and no sudden jump-to-default. The first two assumptions have been made purely for convenience, but the last three points are more demanding. For instance, "no jump-to-default" means that future default events will be anticipated by the market. In other words, we assume that correlation traders can hedge their books just before a default event. Historically, such an assumption does not seem unrealistic, despite some recent counter examples, for example Fannie Mae and Freddie Mac.

Most of these results have been proven only recently. It seems to us that this is because the GCM has been considered a crude approach, so that going into it in depth appeared a futile exercise for most academics. For them, it is more interesting to work on fully different classes of models. Indeed, the academic productivity on the credit side has been impressive for a decade. Some of the proposals were outstanding, but none of these approaches has replaced the GCM as a benchmark, and it is very likely the base-correlation paradigm will survive the current credit crisis.

Explaining this domination by "the first-mover" advantage is probably short-sighted.

References and notes for this piece may be found online.

Jean-David Fermanian is a professor at ENSAE (Paris), Centre for Research in Economics and Statistics.