Consequences of Picking the Wrong Model
Andrew Smith
andrewdsmith8@deloitte.co.uk

Failures in Pricing Models

Gaussian copula distribution models are an overly simplistic and inadequate means of valuing tranches of collateralized debt obligations (CDOs) and other structured products, warned a senior quant yesterday.

Speaking at the Quant Congress USA in New York, Joe Gregory, formerly global head of credit quantitative analytics at Barclays Capital in London, told delegates that the Gaussian copula: "fails quite dramatically when applied in practical terms to the credit market" and does not legislate for the possibility of dissonant or systemic defaults.

"Focusing particularly on super senior tranches, Gregory indicated that, under a Gaussian model, a super senior tranche of a CDO referencing 125 investment-grade assets should theoretically be able to withstand 45 individual default events before the super senior tranche experienced any loss of principal.

"But the result is that the holder of one of these tranches, with spreads having blown out massively over the past 12 months, would be able to receive as much as $50 billion plus for holding a CDO like this, so this basis, apparently has no risk. That is opposed to the three- or four-year events an investor would have received on this kind of structure in early 2007," said Gregory.

Gregory also pointed out that in the assumptions Gaussian models make on the maturities of losses on a super senior tranche, given that a Gaussian distribution draws a straight upward-sloping line between the expected loss on an equity tranche of a CDO and the maturity of the structure.
Failures of Value-at-Risk Models

Historic Equity Returns in €

Source: The Banker and Bank calculations.
(a) Weighted average risk weights of Deutsche Bank, HSBC, BNP Paribas, BNP Paribas, Citigroup, UBS, BAML, BONY, Commerzbank, ING, JPM, UBS, RBS, Santander, RBS, Standard Chartered, and other large global banks. Data are not available for the remaining 9-11 banks.

Will yesterday’s fit work tomorrow?

EGB2 distribution densities
Fitted using Method of Moments
To Overlapping annual log returns.

Percentiles: Substitution Method

Return period = 1/probability

-40.34% to October 2008
-40.47% to March 2003
-41.78% to November 2008
-43.29% to December 2008
-43.39% to February 2009
-43.67% fitted 1-in-200 event
What About Parameter Estimation Error?

Example “consistency test” based on logistic reference model

What About Parameter Estimation Error?

Four Possible Questions

What is the weakest stress test I could possibly justify, to use as an opening gambit in negotiations but leaving some wiggle room to strengthen if forced to do so?

What is the 0.5% percentile asset return using substitution (ie my best estimate assumptions and ignoring any possibility that the model or parameters may be mis-specified?)

How can I construct an interval that has at least a 99.5% chance of including the true 0.5%-ile in the presence of model & parameter uncertainty?

How can I construct a prediction interval that has at least a 99.5% probability of containing the next observation in the presence of model & parameter uncertainty?
Understanding a Range of Models

Swiss Re Catastrophe History

Source: Sigma reports
Fitted 2013 Distributions (GLM + MOM)

![Graph showing 2013 Loss distributions for Lognormal and GPD models.]

Question: What is Prob(loss > $500bn)?

- 99.5%-ile
- 90%-ile
- Median

$500bn = 1-in-5000$ $500bn = \text{impossible}$

A Range of Model Risk Tools

<table>
<thead>
<tr>
<th>Benchmarking Approach</th>
<th>Consistency Tests</th>
<th>Robustness Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description</strong></td>
<td>Comparing the outputs of different models calibrated to the same data.</td>
<td>Generating random data from a model, and feeding that data back into the calibration process to see if you recover the parameters you started with.</td>
</tr>
<tr>
<td><strong>What it tells you</strong></td>
<td>The range of different experts’ estimates given the data.</td>
<td>The likely accuracy of parameter estimates, both in terms of bias and variability.</td>
</tr>
<tr>
<td><strong>What it doesn’t tell you</strong></td>
<td>How much the results might be distorted by random fluctuations in the observed history.</td>
<td>What happens if the model specification is incorrect?</td>
</tr>
</tbody>
</table>

For more details, see [http://www.theactuary.com/features/2013/10/gi-prepare-for-the-worst/](http://www.theactuary.com/features/2013/10/gi-prepare-for-the-worst/)
Testing Prediction Intervals

Historic data
Market risks
Non-market risks
Reference model
Future Profits
Non-market risks
Market risks
Simulated Profits
0.5%-ile estimate
Parameter estimates
Capital Calculation
Verify Percentile

Lognormal fit is statistically robust

Prediction Frequency
Prob(next obs ≤ fitted 99.5%-ile)

Shape (99.5%-ile / median)

- True = GPD, fit = LN (robustness test)
- True = LN, fit = LN (consistency test)
- True = GPD, fit = GPD (consistency test)
- True = LN, fit = GPD (robustness test)
The study contains numerous other data sets, but there are concerns over accuracy (for example, negative lapse rates).
Assume Logistic Distribution for Increments

\[
F(x) = \frac{1}{1 + \exp\left(-\frac{x}{\beta}\right)}
\]

\[
f(x) = \frac{\exp\left(-\frac{x}{\beta}\right)}{\beta\left(1 + \exp\left(-\frac{x}{\beta}\right)\right)^2}
\]

\[
E(X) = 0
\]

\[
\text{Stdev}(X) = \frac{\pi\beta}{\sqrt{3}}
\]

Some Unrealistic Assumptions

<table>
<thead>
<tr>
<th>Assumption</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log[lapse rate / (1-lapse rate) ] performs a random walk</td>
<td>???</td>
</tr>
<tr>
<td>Increments have a logistic distribution</td>
<td>???</td>
</tr>
<tr>
<td>Sample standard deviation is a good way to measure dispersion of a logistic distribution.</td>
<td>???</td>
</tr>
<tr>
<td>We know the standard deviation of the increments</td>
<td>???</td>
</tr>
<tr>
<td>The same model applies to the future as to the past</td>
<td>???</td>
</tr>
</tbody>
</table>
**Prediction Test : Substitution Method**

Impact of calculating stress based on estimated stdev and not on the reference stdev.

**What is going on?**

This is sometimes called the "T" effect because, if the underlying distribution is normal, prediction intervals should use the Student T distribution instead.
The T effect Disappears for Large Samples

Alternative Models: Noise & Walk
Impact of Mis-specified Models

![Graph showing probability of exceedance for different models over observations](image)

**Unrealistic Assumptions Revisited**

<table>
<thead>
<tr>
<th>Assumption</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log(lapse rate / (1-lapse rate)) performs a random walk</td>
<td>Prediction interval is cautious if the lapse rates are independent.</td>
</tr>
<tr>
<td>Increments have a logistic distribution</td>
<td>Prediction interval is cautious if we assume normal distributions instead,</td>
</tr>
<tr>
<td>Sample standard deviation is a good way to measure dispersion of a logistic distribution.</td>
<td>The prediction test is evidence that the method works; how we derived the estimates is irrelevant.</td>
</tr>
<tr>
<td>We know the standard deviation of the increments</td>
<td>Use a larger multiple of estimated standard deviation</td>
</tr>
<tr>
<td>The same model applies to the future as to the past</td>
<td>You cannot get rid of all limitations and exclusions with clever statistics.</td>
</tr>
</tbody>
</table>
Implied Lapse Stresses

Key:  
- 99.5%-ile incorporating parameter error
- 99.5%-ile ignoring parameter error
- Latest lapse rate
- 0.5%-ile ignoring parameter error
- 0.5%-ile incorporating parameter error

Source: Dickson & Smith
Life Convention 2013

Serious about Model Risk: Who Wins?

Illustrative: for discussion only

- ROE bonus
- Insurer CXO
- UK supervisors / govt
- Ponzi schemes
- Insurer employees
- Job security/ satisfaction
- Management flexibility
- Insurer shareholders
- Franchise value
- Competition
- Consumers
- Benefit security
- Actuarial profession
- Equitable Life #2
- Taxpayers / FSCS
- Bailout cost

Strongly reject onerous requirements
Indifferent
Strongly support onerous requirements
Conclusions

• A regulatory requirement to validate a single “best fit” internal model may address model shopping more than it addresses model error.

• “Picking the right model” is not a practical solution to model error; inevitably there are many possible models capable of passing validation and we cannot know which (if any) is correct.

• A theoretical approach to model and parameter risk is to randomise data sets using reference models in order to test prediction intervals, but this is not yet market practice in financial firms. There is a limit to statistical methods which inevitably make some form of “future will be like the past” assumption.

• There is a need for reflexivity: ability or willingness by employees within an organisation to question its dominant beliefs, norms and expectations (Spicer & Alvesson)

• Commercial incentives, functional stupidity and personal integrity.