**Gaussian Slug** – Simple Nonlinearity Enhancement to the 1-Factor and Gaussian Copula Models in Finance, with Parametric Estimation and Goodness-of-Fit Tests on US and Thai Equity Data

22nd Australasian Finance and Banking Conference

16th – 18th December 2009, Sydney

Poomjai Nacaskul & Worawut Sabborriboon

Bank of Thailand
I. Introduction, Motivation, Conception, Formulation & Encapsulation
   – from ‘Gaussian’ to ‘Gaussian Slug’

II. Devising a Copula Test Methodology
   – Non-parametric \( \chi^2 \) Goodness-of-Fit

III. Experiments
   – Results from US & Thai Equity Data

IV. Q&A
I. Introduction, Motivation, Conception, Formulation & Encapsulation

...
Introduction

- **Multi(bi)variate Normal Distribution**
  - as a packaged foundation
    - Normal marginals \( \Rightarrow \) ‘mu’ & ‘sigma’ vectors
    - Linear (Pearson’s) Correlation \( \Rightarrow \) ‘Rho’ matrix
      - Hence our familiarity/comfort w/ positive semi-definiteness, quadratic form, Cholesky decomposition …
    - Decoupling \( \Rightarrow \text{Gaussian copula} \) & the marginals
  - of modern finance
    - *Modern Portfolio Theory (MPT)* \( \Rightarrow \) investment management
    - *Capital Asset Pricing Model (CAPM)* \( \Rightarrow \) financial econometrics
    - *Asymptotic Single Risk Factor (ASRF)* \( \Rightarrow \) bank capital adequacy
Motivation

- **Bi(multi)variate Normal Distribution**
  - as a linear conditional expectation function
    \[
    \begin{align*}
    f_{\Pi}(x,y) &= \exp \left( -\frac{1}{2\rho^2} \left( \frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x \sigma_y} \right) \right) \\
    & \quad \times \frac{1}{2\pi \sigma_x \sigma_y \sqrt{1-\rho^2}}
    \end{align*}
    \]

- **Gaussian copula**
  - as a foundation of credit risk modelling & credit derivatives pricing
    - Duffie & Singleton (1999), Li (2000)
      
      Recently vilified—*quite unjustly, methinks*—as “The Formula That Killed Wall Street”,
    - as a modelling limitation
      - Cannot capture extreme *tail*, *asymmetric*, or *nonlinear* dependency structure
(Quite refutable) stylised observation

- When a shock hits emerging econ. equity market
  - ‘Blue Chips’ quick to move (initially), but (further) moves muted

Think initially increasing, then decreasing, ‘marginal dependency’

- Hence a **Sigmoidal response/transfer function**

\[
B[Y|x] = \alpha + \beta x + \gamma f(x)
\]

\[
B[Y|x]_1(x) = \alpha + \beta x + \frac{\gamma}{1+e^{-\beta x}}
\]

\[
B[Y|x]_2(x) = \alpha + \beta x + \gamma \tanh(x), \quad \tanh(x) = \frac{e^{\beta x} - 1}{e^{\beta x} + 1}
\]

\[
B[Y|x]_3(x) = \text{arsinh}(\alpha + \beta x), \quad \text{arsinh}(x) = \ln(x + \sqrt{x^2 + 1})
\]

\[
B[Y|x]_4(x) = \sqrt[3]{\alpha + \beta x}
\]
Formulation

- (Simpler w/) our ‘change of variable’ approach

\[
    f_{II}(x, y) \mapsto f_{II}(x, k(y)), \quad \begin{align*}
        k_0(y) &= y^3 \\
        k_n(y) &= y^{2n-1}, \quad k \in \{0,1,2,\ldots\} \\
        k_m(y) &= y + \frac{y^3}{3!} + \frac{y^5}{5!} + \cdots = \sinh(y) \\
        k_0(y) &= \beta y + \beta y^3 + \beta y^5 + \cdots
    \end{align*}
\]

\[
    f_{II}(x, y^2) = \exp\left( -\frac{1}{2\rho^2} \left( \frac{(x - \mu_x)^2}{\sigma_x^2} + \frac{(y^2 - \mu_y)^2}{\sigma_y^2} - \frac{2\rho(x - \mu_x)(y^2 - \mu_y)}{\sigma_x \sigma_y} \right) \right) \frac{1}{2\pi \sigma_x \sigma_y \sqrt{1 - \rho^2}}
\]

- Normalise \( \textbf{Kummer's confluent hypergeometric function of the 1st kind} \)

\[
    \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{II}(x, y) \, dx \, dy = \frac{\Gamma\left(\frac{3}{2}\right)}{\sqrt{\pi} \sigma_y} \left( \frac{2}{\sigma_x} \right)^{3/2} H \left( \frac{3}{2}, \frac{1}{2}, \frac{\mu_y}{\sigma_y} \right),
\]

\[
    H(a, b, z) = 1 + \frac{a}{b} z + \frac{a(a+1) z^2}{b(b+1) 2!} + \frac{a(a+1)(a+2) z^3}{b(b+1)(b+2) 3!} + \cdots
\]
(Nearly there) but above \textit{p.d.f.} is not \textit{location-scale invariant}, hence:

- Step 1: reduce this 5-parameter p.d.f. to a 1-parameter\n
\textbf{standard(ised) Gaussian Slug distribution}

\[
g_{\text{st}}(x,y) = \exp \left( -\frac{(x^2 + y^2 - 2\rho y)}{2(1 - \rho^2)} \right) \frac{1}{2^{\nu/2} \Gamma(\nu/2) \pi^{-\nu/2}}
\]

- Step 2: reintroduce the conditional means & variance parameters for a 5-parameter \textbf{Gaussian Slug distribution}

\[
g_{\nu}(x,y) = g_{\text{st}} \left( \frac{x - \mu_x}{\sigma_x}, \frac{y - \mu_y}{\sigma_y} \right) \exp \left( -\frac{\left( \frac{x - \mu_x}{\sigma_x} \right)^3 + \left( \frac{y - \mu_y}{\sigma_y} \right)^4 - 2\rho \left( \frac{x - \mu_x}{\sigma_x} \right) \left( \frac{y - \mu_y}{\sigma_y} \right)^3}{2(1 - \rho^2)} \right) \frac{1}{2^{\nu/2} \Gamma(\nu/2) \pi^{-\nu/2}}
\]
Figure 1: Bivariate Normal vs. ‘Gaussian Slug’ p.d.f. – 3D Plots
Bivariate Normal vs. ‘Gaussian Slug’ p.d.f. – Contour Plots

Figure 2: Bivariate Normal vs. ‘Gaussian Slug’ p.d.f. – Contour Plots
The ‘Gaussian Slug’ Distribution

- Nonlinear Dependency Structure
  - where ‘rho’ now denotes the (Gaussian Slug) dependency parameter

\[
E[y|x] = y^* \frac{\partial \pi_{x,y}}{\partial y} = 0 \implies E[y|x] = \mu_r + \sigma_y \cdot \frac{(x - \mu_x)}{\sigma_x}
\]

- which can be estimated
  - i.e. via Ordinary Least Square (OLS) type fit

\[
\hat{\rho}_1(\rho) = \hat{\mu}_r + \hat{\sigma}_y \cdot \frac{\rho(x_1 - \hat{\mu}_x)}{\hat{\sigma}_x} \implies \rho_{\text{OLS}} = \min \left\{ \sum_{i=1}^{n} (\hat{y}_i(\rho) - y_i)^2 \right\}
\]

- esp. as Maximum Likelihood Function (MLE) requires solving

\[
\frac{\partial}{\partial \rho} \left( \sum_{i=1}^{n} \log(g_{x,y}(x_i, y_i)) \right) = \sum_{i=1}^{n} \frac{xy_i^3 + \rho - \rho^3 - \rho x_i^3 + \rho^2 x_i y_i^3 - \rho y_i^4}{1 - 2\rho^2 + \rho^4} = 0
\]
Recall how a Gaussian copula function is constructed

- as a bivariate standard normal c.d.f. of inverses of univariate normal c.d.f.

\[
C^{\text{Gaussian}}(u, v, \rho) = \Phi(\Phi^{-1}(u), \Phi^{-1}(v), \rho) \\
= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{(u^2 + v^2 - 2uv\rho)}{2(1-\rho^2)}} ds \, dt
\]

- from which the corresponding Gaussian copula density is given by

\[
c^{\text{Gaussian}}(u, v, \rho) = \frac{f^{\text{Gaussian}}(\Phi^{-1}(u), \Phi^{-1}(v), \rho)}{f^\Phi(\Phi^{-1}(u)) f^\Phi(\Phi^{-1}(v))}
\]
Constructing a **Gaussian Slug copula function**

- requires modification only w.r.t. the bivariate integrand

\[
C_{\text{Gaussian Slug}}(u, v, \rho) = \int_0^1 \int_0^1 \frac{1}{2\pi \sqrt{1-\rho^2}} \exp\left(-\frac{u^2 + v^2 - 2\rho uv}{2(1-\rho^2)}\right) \, du \, dv
\]

- from which the corresponding **Gaussian Slug copula density** is given by

\[
C_{\text{Gaussian Slug}}(u, v, \rho) = \frac{1}{2\pi \sqrt{1-\rho^2}} \exp\left(-\frac{u^2 + v^2 - 2\rho uv}{2(1-\rho^2)}\right)
\]
Figure 3: Standard Gaussian vs. ‘Gaussian Slug’ Copula Density – 3D Plots
Figure 4: Standard Gaussian vs. ‘Gaussian Slug’ Copula Density – Contour Plots
II. Devising a Copula Test Methodology

...
Surveyed Literature re: Copula Testing

- **Malevergne & Sornette (2003)**
  - adapted the Kolmogorov & Anderson-Darling distances as distributional metrics.

  - exploited the fact that the Student’s t distribution is a heavy-tailed generalisation of (and therefore embeds as a special case) the normal distribution.

- **Söderberg (2009)**
  - applied both methods to test the Gaussian copula on the Swedish stock mkt.

- **Patton (2006)**
  - investigated asymmetry, ‘joint appreciation’ vs. ‘joint depreciation’, in Forex mkt,
    - For equity, this is related to the well-known, post-Black Monday volatility skew phenomenon, where correlations rise dramatically during downturns, then revert to lower level once recovery takes place.
  - and whether dependency relation is time-conditional.
(1) Transform the original bivariate data
   - from its sample space within \( \mathbb{R}^2 \) to within \([0,1]^2\)
   - each axis via its own univariate empirical c.d.f.

(2) Partition each ‘transformed sample space’ into cells,
   - i.e. 25 square cells of equal size & perform sample counts within each.
   - Designate this as our observed frequency matrix.

(3) For any copula, integrate the copula density over cell boundaries
   - Multiply each cell by the total number of data points.
   - Designate this as our expected frequency matrix.

(4) Calculate the \( \chi^2 \) statistics:
\[
\chi^2 (\rho) = \sum_{j=1}^{3} \sum_{k=1}^{3} \frac{(\text{Observed}_{j,k} - \text{Expected}_{j,k})^2}{\text{Expected}_{j,k}} = \chi^2_{\rho, (3-1)(3-1)-3} \]
Figure 5: Scatter Plots of "NYSE Index" vs. "Coca Cola" returns – ‘actual’ (left) & ‘[0,1]’ (right)

Figure 6: Scatter Plots of "SET Index" vs. "Siam Cement" returns – ‘actual’ (left) & ‘[0,1]’ (right)
### Table 1: Uniformed "NYSE Index" & "Coca Cola" Returns — Observed Frequencies

<table>
<thead>
<tr>
<th>Oij</th>
<th>[0,0.2]</th>
<th>(0.2,0.4]</th>
<th>(0.4,0.6]</th>
<th>(0.6,0.8]</th>
<th>(0.8,1]</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.8,1]</td>
<td>9</td>
<td>18</td>
<td>14</td>
<td>18</td>
<td>41</td>
</tr>
<tr>
<td>(0.6,0.8]</td>
<td>12</td>
<td>15</td>
<td>16</td>
<td>32</td>
<td>25</td>
</tr>
<tr>
<td>(0.4,0.6]</td>
<td>14</td>
<td>21</td>
<td>27</td>
<td>24</td>
<td>14</td>
</tr>
<tr>
<td>(0.2,0.4]</td>
<td>19</td>
<td>27</td>
<td>24</td>
<td>18</td>
<td>12</td>
</tr>
<tr>
<td>[0,0.2]</td>
<td>46</td>
<td>19</td>
<td>19</td>
<td>8</td>
<td>8</td>
</tr>
</tbody>
</table>

### Table 2: Uniformed "SET Index" & "Siam Cement" Returns — Observed Frequencies

<table>
<thead>
<tr>
<th>Oij</th>
<th>[0,0.2]</th>
<th>(0.2,0.4]</th>
<th>(0.4,0.6]</th>
<th>(0.6,0.8]</th>
<th>(0.8,1]</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.8,1]</td>
<td>1</td>
<td>5</td>
<td>12</td>
<td>24</td>
<td>58</td>
</tr>
<tr>
<td>(0.6,0.8]</td>
<td>3</td>
<td>9</td>
<td>28</td>
<td>35</td>
<td>24</td>
</tr>
<tr>
<td>(0.4,0.6]</td>
<td>14</td>
<td>27</td>
<td>25</td>
<td>21</td>
<td>14</td>
</tr>
<tr>
<td>(0.2,0.4]</td>
<td>19</td>
<td>37</td>
<td>24</td>
<td>16</td>
<td>4</td>
</tr>
<tr>
<td>[0,0.2]</td>
<td>63</td>
<td>22</td>
<td>11</td>
<td>4</td>
<td>0</td>
</tr>
</tbody>
</table>
$\chi^2$ Error Criterion and Goodness-of-Fit Test for Copulas (3)

Table 3: Consistent with the Gaussian copula density with $\rho = 0.5$ — Expected Frequencies

<table>
<thead>
<tr>
<th>Eij</th>
<th>[0,0.2]</th>
<th>(0.2,0.4]</th>
<th>(0.4,0.6]</th>
<th>(0.6,0.8]</th>
<th>(0.8,1]</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.8,1]</td>
<td>4.2</td>
<td>10.2</td>
<td>16.6</td>
<td>25.4</td>
<td>43.6</td>
</tr>
<tr>
<td>(0.6,0.8]</td>
<td>10.2</td>
<td>17.4</td>
<td>21.9</td>
<td>25.2</td>
<td>25.4</td>
</tr>
<tr>
<td>(0.4,0.6]</td>
<td>16.6</td>
<td>21.9</td>
<td>22.9</td>
<td>21.9</td>
<td>16.6</td>
</tr>
<tr>
<td>(0.2,0.4]</td>
<td>25.4</td>
<td>25.2</td>
<td>21.9</td>
<td>17.4</td>
<td>10.2</td>
</tr>
<tr>
<td>[0,0.2]</td>
<td>43.6</td>
<td>25.4</td>
<td>16.6</td>
<td>10.2</td>
<td>4.2</td>
</tr>
</tbody>
</table>

Table 4: Consistent with the Gaussian Slug copula density with $\rho = 0.5$ — Expected Frequencies

<table>
<thead>
<tr>
<th>Eij</th>
<th>[0,0.2]</th>
<th>(0.2,0.4]</th>
<th>(0.4,0.6]</th>
<th>(0.6,0.8]</th>
<th>(0.8,1]</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.8,1]</td>
<td>2.8</td>
<td>6.4</td>
<td>10.5</td>
<td>16.7</td>
<td>63.7</td>
</tr>
<tr>
<td>(0.6,0.8]</td>
<td>14.0</td>
<td>20.3</td>
<td>22.8</td>
<td>23.3</td>
<td>19.6</td>
</tr>
<tr>
<td>(0.4,0.6]</td>
<td>16.6</td>
<td>21.9</td>
<td>23.0</td>
<td>21.9</td>
<td>16.6</td>
</tr>
<tr>
<td>(0.2,0.4]</td>
<td>19.6</td>
<td>23.3</td>
<td>22.8</td>
<td>20.3</td>
<td>14.0</td>
</tr>
<tr>
<td>[0,0.2]</td>
<td>63.7</td>
<td>16.7</td>
<td>10.5</td>
<td>6.4</td>
<td>2.8</td>
</tr>
</tbody>
</table>
(5) For the purpose of **fitting** the copula density parameter,

- this $\chi^2(\rho)$ defines our error criterion to minimise.

- Thus for our Gaussian Slug copula:

$$
\rho^* = \min_{\rho} \left\{ \chi^2(\rho) = \sum_{i=1}^{5} \sum_{j=1}^{5} \frac{(\text{Observed}_i - \text{Expected}_i(\rho))^2}{\text{Expected}_i(\rho)} \right\},
$$

$$
\text{Expected}_i(\rho) = \int_{-0.2}^{0.2} \int_{-0.2}^{0.2} c_{\text{Gauss}}(u, v, \rho) \, du \, dv, \quad i, j = 1, ..., 5
$$

(6) For the purpose of **testing** the goodness-of-fit of the copula density,

- this $\chi^2$ “hat” is precisely our test statistics, as per Pearson’s $\chi^2$ test.

- For example, with 16 d.f. at the 99% confidence level, reject the copula whenever $\chi^2$ “hat” $\geq$ CHIINV(0.01,16) = 32.
III. Experiments
Results – (example) US Equity Data

500 weekly returns (January 2\textsuperscript{nd}, 2000 – August 2\textsuperscript{nd}, 2009)

Figure 7: "NYSE Index" vs. "Coca Cola" Returns – Chi-Square Test Statistics for Different Copulas
Figure 8: "SET Index" vs. "Siam Cement" Returns – Chi-Square Test Statistics for Different Copulas
Results – 8 US & 2 Thai Equity Data
500 weekly returns (January 2\textsuperscript{nd}, 2000 – August 2\textsuperscript{nd}, 2009)

<table>
<thead>
<tr>
<th>#</th>
<th>Equity Market/Index vs. Individual-Stock Paired Returns - Weekly Data (2Jan00-2Aug09)</th>
<th>Pearson’s correlation</th>
<th>Chi-Square (p-Value)</th>
<th>Independent Copula</th>
<th>Gaussian Copula’s rho (best fit)</th>
<th>Chi-Square (p-Value)</th>
<th>Gaussian Copula</th>
<th>Chi-Square (p-Value)</th>
<th>Gaussian Slug Copula</th>
<th>Chi-Square (p-Value)</th>
<th>Gaussian Slug Copula</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>NYSE-ATT</td>
<td>0.51</td>
<td>157.3 (0%)</td>
<td>0.49</td>
<td>29.9 (1.83%)</td>
<td>0.3</td>
<td>42.8 (0.03%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>NYSE-CocaCola</td>
<td>0.45</td>
<td>105.9 (0%)</td>
<td>0.41</td>
<td>21.4 (16.49%)</td>
<td>0.22</td>
<td>28 (3.18%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>NYSE-ExxonMobil</td>
<td>0.63</td>
<td>180.5 (0%)</td>
<td>0.55</td>
<td>19.7 (23.39%)</td>
<td>0.38</td>
<td>34.4 (0.49%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>NYSE-GE</td>
<td>0.68</td>
<td>254.4 (0%)</td>
<td>0.62</td>
<td>28.8 (2.5%)</td>
<td>0.51</td>
<td>41.6 (0.05%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>NYSE-IBM</td>
<td>0.59</td>
<td>246.5 (0%)</td>
<td>0.59</td>
<td>44 (0.02%)</td>
<td>0.43</td>
<td>72.9 (0%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>NYSE-Merck</td>
<td>0.43</td>
<td>122.5 (0%)</td>
<td>0.4</td>
<td>39.7 (0.09%)</td>
<td>0.22</td>
<td>42.5 (0.03%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>NYSE-Microsoft</td>
<td>0.49</td>
<td>163.7 (0%)</td>
<td>0.48</td>
<td>37.4 (0.19%)</td>
<td>0.29</td>
<td>49.7 (0%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>NYSE-WalMart</td>
<td>0.52</td>
<td>155.9 (0%)</td>
<td>0.52</td>
<td>16.1 (44.78%)</td>
<td>0.33</td>
<td>32 (1.01%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>SET-BangkokBank</td>
<td>0.79</td>
<td>394.4 (0%)</td>
<td>0.77</td>
<td>15.4 (49.69%)</td>
<td>0.74</td>
<td>57.5 (0%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>SET-SiamCement</td>
<td>0.70</td>
<td>307 (0%)</td>
<td>0.72</td>
<td>10.5 (83.66%)</td>
<td>0.65</td>
<td>42.8 (0.03%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5: Chi-Square statistics, \(*\) indicates that the copula cannot be rejected at 99% confidence level.
Conclusions

- The Goods 😊
  - Motivation
  - Mathematical simplicity
  - Non-parametric copula test methodology

- The Bads 😞
  - Empirical results show no improvement over standard Gaussian copulas.
IV. Q&A (please!)

...