# 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

<sup>1</sup>Poomjai Nacaskul & Worawut Sabborriboon

Abstract - A bivariate normal distribution, with the attendant nonanalytically integrable p.d.f., lies at the hearts of many financial theories. Its derived Gaussian copula ostensibly does away with the normality assumptions, only to retain the linear (Pearson's) correlation measure implicit to said bivariate normal p.d.f. In financial modelling context, the Gaussian copula suffer from at least three setbacks, namely its inability to capture (extreme) tail, asymmetric (upside vs. downside), and nonlinear (diminishing) dependency structures. Noting that various fixes have been proposed w.r.t. the former two issues, (i) this paper attempts to address the nonlinearity with the proposal of a bivariate 'Gaussian Slug' distribution (ii) from which a derived copula density function quite naturally and parsimoniously captures a particular nonlinear dependency structure. In addition, (iii) this paper devises a simple, intuitive formulation of copula parameter estimation as a minimisation of a chi-square test statistics, (iv) whose resultant value readily lends itself to the widely popular statistical goodness-of-fit testing. Tests were performed comparing independent vs. Gaussian vs. 'Gaussian Slug' copulas on weekly US and Thai equity market index and individual stock returns data, all available on *Reuters*<sup>TM</sup>.

# 1. Introduction

A bivariate normal distribution, with the attendant non-analytically integrable p.d.f., lies at the hearts of many financial theories: from the Treynor/Sharpe/Lintner/Mossin *Capital Asset Pricing Model (CAPM)* and subsequent 1-factor market/credit risk models, which presuppose *joint normality* of market

<sup>&</sup>lt;sup>1</sup> Quantitative Models & Financial Engineering team, Financial Institutions Strategy Department, the Bank of Thailand, Correspondence: <u>PoomjaiN@bot.or.th</u> & <u>WorawutS@bot.or.th</u>.

and individual risk factors, to the Vasicek/Gordy *Asymptotic Single Risk Factor (ASRF)* model underlying the Basel II *Internal-Ratings Based (IRB)* approach to calculating minimal regulatory creditrisk capitals, to the Duffie-Singleton/Li CDO pricing methodology, which by utilising *Gaussian copula* [2][6][10] ostensibly does away with the normality assumptions for marginal loss distributions, only to retain the *linear (Pearson's) correlation measure* implicit and integral to said bivariate normal p.d.f.

Indeed, from the advent of Markowitz *Modern Portfolio Theory (MPT)* of Pareto optimal asset allocation down to contemporary *Value-at-Risk (VaR)* measure of market risks in the aggregate, questions have been raised—and widely controverted since—about the assumed normality of the marginal distributions of asset returns. By and large, this line of inquiries has been superseded by misgiving w.r.t. the *dependence structure* between risk factors as implied by the widely adopted Gaussian copula.<sup>2</sup>

One line of inquiry is whether the Gaussian copula can adequately capture dependency in the *(extreme) tails* of the marginal distributions. In their seminal paper on testing the Gaussian copula hypothesis for financial-asset dependencies, Malevergne & Sornette (2003) [4] adapted the Kolmogorov as well as Anderson-Darling distances as their distributional test metrics.<sup>3</sup> In particular, Mashal & Zeevi(2002) [5], and similarly Chen, Fan, Patton(2004) [1], exploited the fact that the Student's *t* distribution is effectively a *heavy-tailed* generalisation (and therefore embeds as a special case) of the normal distribution. Söderberg (2009) [9] then applied both methods to good effect in testing the Gaussian copula on the Swedish stock market.<sup>4</sup> Another line of inquiry is whether 'upside' dependency needs be symmetric with the 'downside' leg. 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. A less obvious investigation in the foreign exchange market is whether

<sup>&</sup>lt;sup>2</sup> In other words, joint normality has two components, normal marginals and Gaussian copula, so going beyond normal marginals whilst retaining the Gaussian copula constitutes but a *partial* generalisation.

<sup>&</sup>lt;sup>3</sup> They expressly highlighted how the Kolmogorov measure is "more sensitive to the deviations occurring in the bulk of the distributions", while the Anderson-Darling measure is "more accurate in the tails of the distributions".

<sup>&</sup>lt;sup>4</sup> Although such tail dependency issues arose vis-à-vis market returns, i.e. random variables with distributional support over the real number line, similar issues may arise w.r.t. random variables with distributional support over the positive interval, i.e. dollar values of operational/credit losses.

'joint appreciation' and 'joint depreciation' are indeed asymmetric, and whether dependency relation is time-conditional Patton (2006) [8].

Our line of inquiry is whether the inherently *linear* dependence structure imposed by the Gaussian copula is appropriate for financial modelling. For instance, in an emerging equity market dominated by a few large stocks, it is anecdotally observed how initially these leading stocks would be the ones that move very quickly in response to (perhaps because they are leading) the general market trends, but as the overall index moves deeper in the negative territory, these core stocks seem to hold value better as long-term bargain hunters come in, dismissing what they perceive as 'overshot' market sentiments. Conversely, in a strong rally, international capital funds may opt to take early profits in the core stocks they principally built their cross-country investment diversification positions on. In a sense, *betas* for these lead stocks peak in the middle and taper off at either end. This paper is motivated initially from this nonlinearity consideration in particular. Although not discussed further in this paper, should one take up the demarcation between *high-frequency/low-impact vs. low-frequency/high-impact* operational losses and pursue an extended "diminishing dependency" hypothesis, i.e. distinguishing *high-frequency/low-impact/low-dependency* operational losses, a similar nonlinear extension to copula modelling would be just as essential, if not more so.<sup>5</sup>

In particular, we go back to the bivariate normal distribution in an attempt to endow it with nonlinear dependency. First, we write out a 5-parameter a bivariate normal p.d.f. (1):

$$f_{XY}(x, y) = \frac{\exp\left(-\frac{1}{2(1-\rho^{2})}\left(\frac{(x-\mu_{X})^{2}}{\sigma_{X}^{2}} + \frac{(y-\mu_{Y})^{2}}{\sigma_{Y}^{2}} - \frac{2\rho(x-\mu_{X})(y-\mu_{Y})}{\sigma_{X}\sigma_{Y}}\right)\right)}{2\pi\sigma_{X}\sigma_{Y}\sqrt{1-\rho^{2}}}$$
(1)

The conditional expectation of *Y* is a linear function of *x* thus:

<sup>&</sup>lt;sup>5</sup> The authors are pursuing such investigation, especially in relations to the Basel II *Advanced* 

*Measurement Approach (AMA)* to calculating minimal regulatory operational-risk capitals, where different types of operational-risk events and/or different business lines tend to exhibit clear 'regime preference', i.e. weak dependency amongst low-frequency, high-impact natural catastrophes, in contrast with stronger dependency amongst high-frequency, low-impact retail transaction errors and/or frauds, which could be traced back to poor control, training, manpower issues, etc.

$$E[Y|x] \equiv \int_{-\infty}^{\infty} y \cdot f_{Y|X}(y \mid x) \, dy = \frac{\overline{\mu_Y \sigma_X - \rho \mu_X \sigma_Y}}{\sigma_X} + \frac{\overline{\rho \sigma_Y}}{\sigma_X} x \tag{2}$$

The aim of this paper is simply to find an (analytically simple) alternative to, and possibly a generalisation of, (1) for which an expression derived analogously to (2) is nonlinear. The resulting bivariate distribution is called a 'Gaussian Slug', for a reason which will become intuitively/visually obvious. This is done in section 2, on top of which section 3 then derives a corresponding copula density function. Section then 4 formulates and proposes a simple, intuitive method for optimising a copula fit over the relevant *parameter space*, one which readily lends itself to *Pearson's chi-square (goodness-of-fit) test*.<sup>6</sup> Section 5 demonstrates how the proposed copula is used and tested against weekly US as well as Thai equity market index and individual stock returns data downloaded from *Reuters*<sup>TM</sup>. In summary, our paper introduces innovations in all four areas: distribution, copula, estimation, and testing.

## 2. The 'Gaussian Slug' Distribution

This paper is motivated by the inability of linear dependence structure to implement *marginally decreasing* dependency, i.e. whereby a response to stimulus is positive throughout, but the sensitivity peaks and wanes, in essence a mixture between linear and *sigmoidal* responses.<sup>7</sup> In other words, we want conditional expectation of *Y* to be a nonlinear function of *x* of the form:

$$E[Y|x] = \overbrace{\alpha + \beta x}^{linear} + \gamma \overbrace{\beta 0}^{Sigmoidal} (3)$$

Indeed there are a number of alternatives available:

<sup>&</sup>lt;sup>6</sup> Incidentally, Fermanian (2005) [3] and Panchenko (2005) [7] both proposed a goodness-of-fit test for copulas, the former non-parametrically, the latter parametrically. We believe our approach is much simpler than either one. And unlike Malevergne & Sornette (2003) [4], there is no simulation involved. <sup>7</sup> Another motivation could be to model the "run-away" effect, i.e. whereby a response to stimulus is positive throughout, but this time the sensitivity keeps increasing at both ends, essentially reversing the roles between the response and the stimulus, i.e. instead of  $\overline{y}(x) \propto \sqrt[3]{x}$ , it would be  $\overline{y}(x) \propto x^3$ , etc.

$$E[Y|x]_{i}(x) = \alpha + \beta x + \frac{\gamma}{1+e^{-x}}$$

$$E[Y|x]_{ii}(x) = \alpha + \beta x + \gamma \tanh(x), \quad \tanh(z) = \frac{e^{2z} - 1}{e^{2z} + 1}$$

$$E[Y|x]_{iii}(x) = \operatorname{arcsinh}(\alpha + \beta x), \quad \operatorname{arcsinh}(z) = \ln(z + \sqrt{z^{2} + 1})$$

$$E[Y|x]_{iv}(x) = \sqrt[3]{\alpha + \beta x}$$
(4)

It turns out that in order to introduce such nonlinearity while keeping closest to the original Gaussian functional form, an expedient choice is to modify the bivariate normal p.d.f. *via* a simple change of variable, again, for which a number of alternatives are available:

$$f_{XY}(x, y) \mapsto f_{XY}(x, h(y)), \begin{cases} h_i(y) = y^3 \\ h_{ii}(y) = y^{2k+1}, k \in \{0, 1, 2, ...\} \\ h_{iii}(y) = y + \frac{y^3}{3!} + \frac{y^5}{5!} + \dots = \sinh(y) \\ h_{iv}(y) = \beta_1 y + \beta_3 y^3 + \beta_5 y^5 + \dots \end{cases}$$
(5)

For simplicity, this paper considers the cubic function, hence the following *kernel function*:

$$f_{XY}(x, y^{3}) = \frac{\exp\left(-\frac{1}{2(1-\rho^{2})}\left(\frac{(x-\mu_{X})^{2}}{\sigma_{X}^{2}} + \frac{(y^{3}-\mu_{Y})^{2}}{\sigma_{Y}^{2}} - \frac{2\rho(x-\mu_{X})(y^{3}-\mu_{Y})}{\sigma_{X}\sigma_{Y}}\right)\right)}{2\pi\sigma_{X}\sigma_{Y}\sqrt{1-\rho^{2}}}$$
(6)

For a bivariate distribution proper, a normalising term must be found. Unfortunately, although indeed analytically expressible<sup>8</sup>, the expression is rather messy, involving *gamma* and so-called *(Kummer's) confluent hypergeometric function of the first kind*:

<sup>&</sup>lt;sup>8</sup> We used *Mathematica*® throughout.

$$\int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} f_{XY}(x, y^{3}) dx dy$$

$$= \frac{\Gamma(\frac{7}{6})}{\sqrt{\pi}} \left(\frac{2}{\sigma_{Y}}\right)^{2/3} H\left(\frac{1}{3}, \frac{1}{2}, -\frac{1}{2}\left(\frac{\mu_{Y}}{\sigma_{Y}}\right)^{2}\right),$$

$$H(a, b, x) \equiv 1 + \frac{a}{b}x + \frac{a(a+1)}{b(b+1)}\frac{x^{2}}{2!} + \frac{a(a+1)(a+2)}{b(b+1)(b+2)}\frac{x^{3}}{3!} + \cdots$$
(7)

Writing it all out yields the following p.d.f. proper:

$$g_{XY}(x,y) = \frac{f_{XY}(x,y^{3})}{\int_{-\infty-\infty}^{\infty} f_{XY}(x,y^{3}) dx dy}$$

$$= \frac{\exp\left(-\frac{1}{2(1-\rho^{2})}\left(\frac{(x-\mu_{X})^{2}}{\sigma_{X}^{2}} + \frac{(y^{3}-\mu_{Y})^{2}}{\sigma_{Y}^{2}} - \frac{2\rho(x-\mu_{X})(y^{3}-\mu_{Y})}{\sigma_{X}\sigma_{Y}}\right)\right)}{2^{5/3}\sigma_{X}\sigma_{Y}^{1/3}\sqrt{\pi-\pi\rho^{2}}} \cdot \Gamma(\frac{1}{6}) \cdot H\left(\frac{1}{3}, \frac{1}{2}, -\frac{1}{2}\left(\frac{\mu_{Y}}{\sigma_{Y}}\right)^{2}\right)$$
(8)

Unfortunately, we have moved from a distribution which is *location-scale invariant* w.r.t. both random variates *x* and *y*, to a distribution which is merely location invariant w.r.t. the random variates *x*. The trick then is to first reduce this 5-parameter p.d.f. into a 1-parameter *standard Gaussian Slug distribution*:

$$g_{XY}^{std.}(x,y) = \frac{\exp\left(-\left(\frac{x^2 + y^6 - 2\rho x y^3}{2(1-\rho^2)}\right)\right)}{2^{5/3} \Gamma(\frac{7}{6}) \sqrt{\pi - \pi \rho^2}}$$
(9)

Then reintroduce the conditional means and variance parameters as a simple linear transformation to obtain the 5-parameter *Gaussian Slug distribution*. In summary:

$$g_{XY}(x,y) = g_{XY}^{std.} \left( \frac{x - \mu_X}{\sigma_X}, \frac{y - \mu_Y}{\sigma_Y} \right)$$
$$= \frac{\exp\left(-\left( \frac{\left(\frac{x - \mu_X}{\sigma_X}\right)^2 + \left(\frac{y - \mu_Y}{\sigma_Y}\right)^6 - 2\rho \left(\frac{x - \mu_X}{\sigma_X}\right) \left(\frac{y - \mu_Y}{\sigma_Y}\right)^3\right)}{2(1 - \rho^2)}\right)}{2^{5/3} \Gamma(\frac{7}{6}) \sqrt{\pi - \pi \rho^2}}$$
(10)

We now have one working bivariate distribution whose conditional expectation has the 'S' shape resembling common *gastropod mollusc*, hence a 'Slug' (Figure 1, Figure 2).



Figure 1: Bivariate Normal vs. 'Gaussian Slug' p.d.f. - 3D Plots



Figure 2: Bivariate Normal vs. 'Gaussian Slug' p.d.f. - Contour Plots

Whereas with a bivariate normal distribution, the conditional expectation of Y is, as noted in (2), a linear function of x, with our Gaussian Slug distribution, the conditional expectation of Y, i.e. the value of y that maximises the conditional distribution given x, is a cubic root of x as desired thus:

$$\mathbf{E}[Y|x] \equiv y^* \,\mathfrak{s} \,\frac{\partial g_{XY}(x, y^*)}{\partial y} = 0 \quad \Rightarrow \quad \mathbf{E}[Y|x] = \mu_Y + \sigma_Y \cdot \sqrt[3]{\frac{\rho(x - \mu_X)}{\sigma_X}} \tag{11}$$

Note how it would now be incorrect to refer to  $\rho$  as the correlation parameter, as per normal distribution. So when they appear together, it would be advisable to write  $\rho^{Gaussian}$  for the old correlation parameter and  $\rho^{GaussianSlug}$  for the new (Gaussian Slug) dependency parameter.

In terms of estimation, the advantage to forcing our new distribution to remain location-scale invariant in this form is that the means and variance estimates are as before, which leaves only the question of how to estimate the dependency parameter. Here, in analogy with simple regression analysis<sup>9</sup>, we can use (11) as the fitting function, then define an *Ordinary Least Square (OLS)* estimate<sup>10</sup> of  $\rho$  as  $\hat{\rho}$  that, given the data set  $\{(x_i, y_i)\}_{i=1}^n$ , minimises the fit error:

$$\hat{y}_{i}(\rho) = \hat{\mu}_{Y} + \hat{\sigma}_{Y} \cdot \sqrt[3]{\frac{\rho(x_{i} - \hat{\mu}_{X})}{\hat{\sigma}_{X}}} \implies \hat{\rho}_{OLS} = \min_{\rho} \left\{ \sum_{i=1}^{n} (\hat{y}_{i}(\rho) - y_{i})^{2} \right\}$$
(12)

#### 3. The 'Gaussian Slug' Copula

Recall that a *Gaussian copula function* is constructed as a bivariate standard normal c.d.f. of the inverses of univariate normal c.d.f. thus:

equating the result,  $\frac{\partial}{\partial \rho} \left( \sum_{i=1}^{n} \log \left( g_{XY}^{std.} (x_i, y_i) \right) \right) = \sum_{i=1}^{n} \frac{x_i y_i^3 + \rho - \rho^3 - \rho x_i^2 + \rho^2 x_i y_i^3 - \rho y_i^6}{1 - 2\rho^2 + \rho^4}, \text{ to zero}$ 

does not yield a simple solution for an arbitrary sample size n.

<sup>&</sup>lt;sup>9</sup> However, the analogy is not perfect, as we are not assuming uniform noise over a cubic root of x, as would be the case of a *nonlinear regression*, hence the un-weighted error terms assign less weights than perhaps ideal to 'fitting error' as x goes away from its average, i.e. as  $|x - \mu_x|$  increases.

<sup>&</sup>lt;sup>10</sup> Alternatively, given an *n*-pair sample data,  $\{(x_i, y_i)\}_{i=1}^n$ , a *Maximum Likelihood Estimator (MLE)* could be found by taking the partial derivative of the log-likelihood function w.r.t.  $\rho$ . Unfortunately, equating the result  $\frac{\partial}{\partial n} \left( \sum_{i=1}^n \log(q^{std_i}(x_i, y_i)) \right) = \sum_{i=1}^n \frac{x_i y_i^3 + \rho - \rho^3 - \rho x_i^2 + \rho^2 x_i y_i^3 - \rho y_i^6}{\rho + \rho^3 - \rho x_i^2 + \rho^2 x_i y_i^3 - \rho y_i^6}$  to zero.

$$C^{Gaussian}(u,v,\rho) = \Phi(\Phi^{-1}(u), \Phi^{-1}(v), \rho)$$
  
=  $\int_{-\infty}^{\Phi^{-1}(v)} \left( \int_{-\infty}^{\Phi^{-1}(u)} \frac{f_{XY}^{std.}(s,t)ds}{f_{XY}^{std.}(s,t)ds} \right) dt$   
=  $\int_{-\infty}^{\Phi^{-1}(v)} \left( \int_{-\infty}^{\Phi^{-1}(u)} \frac{\exp\left(-\left(\frac{s^{2} + t^{2} - 2\rho st}{2(1 - \rho^{2})}\right)\right)}{2\pi\sqrt{1 - \rho^{2}}} ds \right) dt$  (13)

From which the corresponding *Gaussian copula density* is given by:

$$c^{Gaussian}(u,v,\rho) = \frac{f_{XY}^{std.}(\Phi^{-1}(u),\Phi^{-1}(v),\rho)}{f_X(\Phi^{-1}(u))f_Y(\Phi^{-1}(v))}$$
(14)

Constructing a *Gaussian Slug copula* requires modification only w.r.t. the bivariate integrand:

$$C^{Gaussian}(u,v,\rho) = G_{XY} \left( \Phi^{-1}(u), \Phi^{-1}(v), \rho \right) = \int_{-\infty}^{\Phi^{-1}(v)} \left( \int_{-\infty}^{\Phi^{-1}(u)} g_{XY}^{std.}(s,t) ds \right) dt = \int_{-\infty}^{\Phi^{-1}(v)} \left( \int_{-\infty}^{\Phi^{-1}(u)} \frac{\exp\left( -\left(\frac{s^2 + t^6 - 2\rho st^3}{2(1-\rho^2)}\right)\right)}{2^{5/3} \Gamma(\frac{7}{6}) \sqrt{\pi - \pi\rho^2}} ds \right) dt$$
(15)

From which the corresponding *Gaussian Slug copula density* is given by:

$$c_{Slug}^{Gaussian}(u,v,\rho) = \underbrace{\left(\frac{2^{2/3}\Gamma(\frac{7}{6})}{\sqrt{\pi}}\right)}^{normalising} \cdot \frac{g_{XY}^{std.}(\Phi^{-1}(u),\Phi^{-1}(v),\rho)}{f_X(\Phi^{-1}(u))f_Y(\Phi^{-1}(v)^3)}$$
(16)

We now have one working bivariate copula whose dependency relation has the 'S' shape resembling common *gastropod mollusc*, hence a 'Slug' (Figure 3, Figure 4).



Figure 3: Standard Gaussian vs. 'Gaussian Slug' Copula Density - 3D Plots



Figure 4: Standard Gaussian vs. 'Gaussian Slug' Copula Density - Contour Plots

The next section details our simple, intuitive approach to constructing and utilising the chi-square statistics: firstly as the error criterion for optimising a fit over the copula density function's parameter space, and secondly as a goodness-of-fit statistics, as per the widely popular Pearson's chi-square test.

# 4. Chi-Square Error Criterion and Goodness-of-Fit Test for Copulas

Our methodology<sup>11</sup> is as follows.

Transform the original bivariate data from its *sample space* within R<sup>2</sup> (e.g. Figure 5, Figure 6 – *left* plots) to within [0,1]<sup>2</sup> (e.g. Figure 5, Figure 6 – *right* plots), each axis *via* its own univariate *empirical c.d.f.*



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)

Partition each [0,1]<sup>2</sup> 'transformed sample space' into cells, i.e. 25 square *cells* of equal size and perform sample counts within each cell (e.g. Table 1 & Table 2, each with 500 points). Call this the *observed frequency* matrix.

<sup>&</sup>lt;sup>11</sup> Although the method generalises to any copulas, our exposition is restricted to bivariate ones.

Oij	[0,0.2]	(0.2,0.4]	(0.4,0.6]	(0.6,0.8]	(0.8,1]
(0.8,1]	9	18	14	18	41
(0.6,0.8]	12	15	16	32	25
(0.4,0.6]	14	21	27	24	14
(0.2,0.4]	19	27	24	18	12
[0,0.2]	46	19	19	8	8

Table 1: Uniformed "NYSE Index" & "Coca Cola" Returns - Observed Frequencies

Oij	[0,0.2]	(0.2,0.4]	0.4] (0.4,0.6] (0.6,0.8]		(0.8,1]
(0.8,1]	1	5	12	24	58
(0.6,0.8]	3	9	28	35	24
(0.4,0.6]	14	27	25	21	14
(0.2,0.4] 19		37	24	16	4
[0,0.2]	63	22	11	4	0

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

3. For any bivariate copula, integrate the copula density over the areas corresponding to the cell boundaries. Multiply each cell by the total number of data points. Call this the *expected frequency* matrix (i.e. expected cell populations consistent with the specified copula density). For example, performing 25 double integrations w.r.t. the Gaussian copula density parameterised by *ρ* = 0.5 yields Table 3 (note: a *symmetric* matrix); whereas, performing 25 double integrations w.r.t. the Gaussian Slug copula density parameterised by *ρ* = 0.5 yields Table 3 (note: a *symmetric* matrix); whereas, performing 25 double integrations w.r.t. the Gaussian Slug copula density parameterised by *ρ* = 0.5 yields Table 3 (note: a *symmetric* matrix); whereas, performing 25 double integrations w.r.t. the Gaussian Slug copula density parameterised by *ρ* = 0.5 yields Table 3 (note: an *asymmetric* matrix). Note how linearity of dependency structure now translates to *symmetry* of the corresponding expected frequency matrix; whereas, nonlinearity of dependency structure now translates to *asymmetry* of the corresponding expected frequency matrix. Incidentally, the un-parameterised *independent copula* density translate into uniform expected frequencies, i.e. 500 points divide evenly into 20 data points for each of the 25 cells.

Eij	[0,0.2]	(0.2,0.4]	(0.4,0.6]	(0.6,0.8]	(0.8,1]
(0.8,1]	4.2	2 10.2 16.6		25.4	43.6
(0.6,0.8]	10.2	17.4 21.9 25.2		25.4	
(0.4,0.6]	16.6	21.9 22.9		21.9	16.6
(0.2,0.4]	25.4	25.2	21.9	17.4	10.2
[0,0.2]	43.6	25.4	16.6	10.2	4.2

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

Eij	[0,0.2]	(0.2,0.4]	(0.4,0.6]	(0.6,0.8]	(0.8,1]
(0.8,1]	2.8	6.4 10.5		16.7	63.7
(0.6,0.8]	14.0	20.3	22.8 23.3		19.6
(0.4,0.6]	16.6	21.9	21.9 23.0 21.9		16.6
(0.2,0.4]	19.6	23.3	22.8	20.3	14.0
[0,0.2]	63.7	16.7	10.5	6.4	2.8

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

4. Calculate the chi-square statistics. Recall that it is given by:

$$\chi^{2}(\rho) = \sum_{j=1}^{5} \sum_{i=1}^{5} \frac{\left(Observed_{ij} - Expected_{ij}\right)^{2}}{Expected_{ij}} = \hat{\chi}^{2} \sim \chi^{2}_{d.f.=(5-1)\times(5-1)=16}$$
(17)

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:

$$\rho^{*} = \min_{\rho} \left\{ \chi^{2}(\rho) = \sum_{j=1}^{5} \sum_{i=1}^{5} \frac{\left(Observed_{ij} - Expected_{ij}(\rho)\right)^{2}}{Expected_{ij}(\rho)} \right\},$$

$$Expected_{ij}(\rho) = \int_{0.2j-0.2}^{0.2j} \left( \int_{0.2i-0.2}^{0.2i} C^{Gaussian}_{slug}(u, v, \rho) du \right) dv, \quad i, j = 1, \dots, 5$$
(18)

6. For the purpose of testing the goodness-of-fit of the parameterised copula density, this  $\hat{\chi}^2$  is precisely our test statistics, as per Pearson's chi-square test. Thus, with 16 *degrees of freedom*, at the  $(1 - \alpha)100\% = 99\%$  confidence level, reject the copula whenever  $\hat{\chi}^2$  reaches or exceeds 32 in value<sup>12</sup>.

<sup>&</sup>lt;sup>12</sup> Where the *Excel*® function "CHIINV(0.01,16)" yields 31.99993.

### 5. Testing 'Gaussian Slug' Copula for Nonlinear Dependencies in Equity Returns

For our purpose<sup>13</sup>, the *x*-data will represent some kind of market/index returns, while the *y*-data will represent some kind of individual-stock returns. Our US & Thai data consist of 500 weekly (January  $2^{nd}$ , 2000 – August  $2^{nd}$ , 2009) equity market/index and individual-stock returns.

Three alternative copula density functions available for testing are thus: (i) the un-parameterised independent copula density, (ii) the  $\rho^{Gaussian}$ -parameterised Gaussian copula density, and (iii) our  $\rho^{GaussianSlug}$ -parameterised Gaussian Slug density. Indeed for both Gaussian and Gaussian Slug copulas, the parametric search/optimisation is rather restricted, i.e.  $\rho^{Gaussian}$ ,  $\rho^{GaussianSlug} \in (0,1)$ , so for most cases, a simple bisection would suffice as an optimisation routine. For illustration, we simply ran through nine values each, i.e. calculating the chi-square test statistics for  $\rho^{Gaussian} \in \{0.1, 0.2, ..., 0.9\}$  and again for  $\rho^{GaussianSlug} \in \{0.1, 0.2, ..., 0.9\}$ .

In the case of US equity, let's now consider the "NYSE Index"/"Coca Cola" pair in particular (Figure 7). Right away, the (i) independent copula (uniform expected frequency matrix) is emphatically rejected, as its chi-square test statistics is extremely high, at 105.9. The (ii) Gaussian copula, with only an exception when  $\rho^{Gaussian} = 0.9$ , yields much better results, and at  $\rho^{Gaussian} \in \{0.3, 0.4, 0.5\}$  cannot be rejected (at 99% confidence). The (iii) Gaussian Slug copula, contrary to what we had hoped, does not improve over the standard Gaussian copula, with an only redeeming fact being at  $\rho^{GaussianSlug} = 0.2$  it cannot be rejected (at 99% confidence) either.

In the case of Thai equity, let's now consider the "SET Index"/"Siam Cement " pair in particular (Figure 8). Right away, the (i) independent copula (uniform expected frequency matrix) is emphatically rejected, as its chi-square test statistics is extremely high, at 307. The (ii) Gaussian copula, with no exception, yields much better results, and at  $\rho^{Gaussian} \in \{0.6, 0.7, 0.8\}$  cannot be rejected (at 99% confidence). The (iii) Gaussian Slug copula, contrary to what we had hoped, does not improve over the standard Gaussian copula, and in fact is rejected for all values of  $\rho^{GaussianSlug} \in \{0.1, 0.2, ..., 0.9\}$  tried.

<sup>&</sup>lt;sup>13</sup> Again, should we wish to capture/test for a "run-away" effect, the roles would be effectively reversed, i.e. the *x*-data would instead represent some kind of individual-stock returns, while the *y*-data would instead represent some kind of market/index returns.

In fact, similar results were observed with a number of other pair returns, prompting us to preliminarily concede that the case for enhancing the standard Gaussian copula models in finance with the kind of nonlinearity we envisioned is yet unproven. Nonetheless, we believe the test devised and used in this paper, while negating our contribution in terms of adding realism to copula modelling, is appropriate for testing the goodness-of-fit of any copula model and provides a transparent benchmark by which to evaluate any candidate copulas against established ones.



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

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

Detailed results of our experiments is given in Table 5

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

### 6. Concluding remarks

We set out to construct, without introducing additional fit parameters, nonlinear dependency structure, while keeping as close as possible to the widely popular normal distribution and Gaussian copula functions. Our 'Gaussian Slug' distribution appears functionally similar to the normal distribution, and retains the location-scale invariant property, yet manages to transform the linear expression of the conditional expectation into one that expresses a nonlinear, diminishing sensitivity response. Although this is initially motivated by stylised observation that, especially in emerging market equities, individual 'core' stocks tend to react sharply to general market/index initially, but retain value better as the market slumps, conversely get capped by profit taking when the overall market rallies, preliminary evidences from weekly US and Thai equity data do not support our position. Notwithstanding, the parametric search and goodness-of-fit methodologies devised and demonstrated in this paper are simple, intuitive, and transparent. Put in another way, of the four areas we attempted to innovate (distribution, copula, estimation, and testing), the latter two remain useful despite the former two having yet to prove their worth.

# Bibliography

 [1] Chen, Fan, Patton (2004), "Simple Tests for Models of Dependence Between Multiple Financial Time Series, with Applications to U.S. Equity Returns and Exchange Rates", SSRN: http://papers.ssrn.com/sol3/papers.cfm?abstract\_id=513024.

[2] Cherubini, Luciano, Vecchiato (2004), Copula Methods in Finance, Chichester: John Wiley & Sons.

[3] **Fermanian, Jean-David (2005)**, "Goodness-of-Fit Tests for Copulas", *Journal of Multivariate Analysis*, vol. 95, no.1, pp.119-152.

[4] Malevergne, Y. & Sornette, D. (2003), "Testing the Gaussian Copula Hypothesis for Financial Assets Dependences", *Quantitative Finance*, vol. 3, pp. 231-250.

 [5] Mashal, R. & Zeevi, A. (2002), "Beyond Correlation: Extreme Co-Movements between Financial Assets", *Working Paper – Columbia University*, SSRN:

http://papers.ssrn.com/sol3/papers.cfm?abstract\_id=317122.

- [6] Nelsen, Roger B. (1999), An Introduction to Copulas, New York: Springer.
- [7] Panchenko, Valentyn (2005), "Goodness-of-fit Test for Copulas", Physica A, vp;. 355. pp. 176-182.
- [8] Patton, Andrew J. (2006), "Modelling Asymmetric Exchange Rate Dependence", *International*

Economic Review, vol. 47, no. 2, pp. 527-556.

[9] Söderberg, Jonas (2009), "Test of the Gaussian Copula on the Swedish Stock Market", *Working Paper Series – Centre for Labour Market Policy Research*, no. 9,

http://vxu.se/ehv/filer/forskning/cafo/wps/Nek\_wp9\_09.pdf.

[10] Wikipedia, http://en.wikipedia.org/wiki/Copula\_(statistics).