Systemic risk and contagion effects in Australian financial institutions and sectors

Piet de Jong, Geoff Loudon and Weihao Choo

Department of Applied Finance and Actuarial Studies
Macquarie University, Sydney, NSW 2109.
piet.dejong@mq.edu.au

1. Literature review

The starting point for the proposed research is the recent literature and the CIFR targeted areas and APRA aims and functions. This recent literature includes the following Adrian and Brunnermeier (2011), Acharya et al. (2012), Acharya et al. (2012) and Brownlees and Engle (2010). The proposed research aims to extend and apply these techniques particularly in relation to the entities regulated by APRA. Thus our broad aim is to develop, implement and bring to bear recent developments in stress testing on the aims of APRA and the CIFR targeted research areas detailed above.

2. Improved measures of contagion and systematic risk

CoVaR_q as proposed in Adrian and Brunnermeier (2011) is a basis for proposed measures contagion, exposure and systemic risk. It suffers from a number of drawbacks:

- Couched in terms of VaR_q containing the scale of the original measurements. It is worthwhile to have measures and techniques robust to scale.
- Conditioning on $\text{VaR}_{0.5}$ is undesirable and relatively intractable. In our proposal we reference stress with respect to the unconditional VaR_q. This permits a more transparent analysis and estimation.
- Our proposed approach separates out scale effects and interdependence effects and aims to relates these separately to external variables including shocks and drivers of systemic risk. Thus VaR_q movements due to scale are disentangled from movements due to codependence with separate driver responses.

3. Significance of the project and policy implications

Understanding the impact of external shocks and their propagation through the financial system is vital for managing and remediating systemic risk. Effective regulation is dependent upon the development of a robust and reliable set of appropriate risk measures. We propose new measures of systemic risk that relate marginal and joint distributions separately to external drivers. This allows for more cogent and coherent stress testing as it includes the estimation of contagion effects, exposure effects and systemic risk across related entities and different financial sectors. Improved stress testing, estimation of risk effects and transmission of shocks through the financial system will make for more cogent prudential policy, prudential margin setting and better identify sources of risk to the financial system.

4. Percentile sensitivity and contagion

4.1. Theorem

Suppose x is a random vector with marginal distributions[1]

$$F(x) \equiv \{F(x_1), \ldots, F(x_p)\} \equiv (u_1, \ldots, u_p) \equiv u \ .$$

Further suppose $0 \leq q \leq 1$ is given and $\mathrm{Q}(x)$ is the vector of q–quantiles

$$F\{\mathrm{Q}(x)\} = q1 = \mathrm{Q}(u) \ ,$$

where 1 is a vector of p ones. Define the stress vector with respect to x_j as[2]

$$\frac{\mathrm{dQ}(x)}{\mathrm{d}x_j} \equiv \mathrm{Q}(x|u_j > q) - \mathrm{Q}(x) \ , \tag{1}$$

where $\mathrm{Q}(x|u_j > q)$ is the vector of q–quantiles of x given $u_j > q$. Then if $\mathrm{Q}(x)$ is linear in q then

$$\frac{\mathrm{dQ}(x_i)}{\mathrm{dQ}(x_j)} \equiv \frac{\mathrm{dQ}(x_i)/\mathrm{d}x_j}{\mathrm{dQ}(x_j)/\mathrm{d}x_j} = \frac{f_j}{f_i}\frac{q_{ij}}{q(1-q)} \ , \qquad q_{ij} = C_{ij}(q + q_{ij}, q) - q^2 \ , \tag{2}$$

where f_i and f_j are the densities of x_i and x_j evaluated at $\mathrm{Q}(x_i)$ and $\mathrm{Q}(x_j)$ and C_{ij} is the copula of (u_i, u_j). Further if

$$s_{ij} \equiv \frac{q_{ij}}{q(1-q)} \ ,$$

then $-1 \leq s_{ij} \leq 1$ with $s_{ij} = \pm 1$ if x_i and x_j are comonotonic and counter monotonic, respectively. If x_i and x_j are independent then $s_{ij} = 0$. If u_i and u_j are exchangeable then $s_{ji} = s_{ij}$.

---

[1] To economise on notation, write $F_j(x) \equiv F_j(x_j) \equiv F(x_j)$ and similar for other vector functions.

[2] In Adrian and Brunnermeier (2011) $\Delta CoVar_q \equiv q_{y|x=q_x} - q_y$. Variable y is generally the "financial system" and hence considered is the change in the VaR_q of the system when institution x stressed, with stress interpreted as $x = q_x$. On page 10 of their paper they incorrectly state "... $CoVaR$ conditions on the event that [an] institution is at its VaR level, which occurs with probability q."

4.2. Proof

By definition

$$\frac{\mathrm{dQ}(u_i)}{\mathrm{d}u_j} \equiv \mathrm{Q}(u_i|u_j > q) - q \equiv q_{ij} \ ,$$

where q_{ij} is such that

$$q = \frac{\mathrm{P}(u_i \leq q + q_{ij}, u_j > q)}{1-q} = \frac{q + q_{ij} - C_{ij}(q + q_{ij}, q)}{1-q} \ , \tag{3}$$

Rearranging yields the second equation in (2). Now

$$\frac{\mathrm{dQ}(x_i)}{\mathrm{d}x_j} \equiv \mathrm{Q}_{q+q_{ij}}(x_i) - \mathrm{Q}(x_i) \approx \mathrm{Q}'(x_i)q_{ij} = \frac{q_{ij}}{f_i} \ , \tag{4}$$

where $'$ denotes differentiation with respect to q and the subscript on Q indicates the revised q for the quantile calculation. The approximation follows from a first order Taylor expansion and is exact if the quantile is linear in q. Similarly

$$\frac{\mathrm{dQ}(x_j)}{\mathrm{d}x_j} \equiv \mathrm{Q}_{q+q(1-q)}(x_j) - \mathrm{Q}(x_j) \approx \frac{q(1-q)}{f_j} \ . \tag{5}$$

Dividing (4) by (5) yields the first equation in (2) and completes the proof.

4.3. Discussion

The critical result is that sensitivities factor into contributions from the ratios of the marginal densities and quantities calculated from the pairwise copulas. The via the implicit equation for q_{ij} in (2), solved using a root finding algorithm. The quantities q_{ij} are implicitly defined from the pairwise copulas. In summary the "sensitivity" matrix

$$S \equiv \frac{\mathrm{d}Q(x)}{\mathrm{d}Q'(x)} \ ,$$

where $'$ denotes transposition has ones on the diagonal quantities between ± 1 off the diagonal. The matrix S is called the sensitivity matrix and displays the sensitivity of the q–quantile of each component of x to stress in the same or other components.

Since f_i and f_j are the densities evaluated at $\mathrm{Q}(x_i)$ and $\mathrm{Q}(x_j)$ it follows that the ratio f_i/f_j can be replaced by the ratio of the hazards. This is useful in cases were it is practical to model the hazard rather than the density.

4.4. Implementation

Figure 1 displays the empirical copulas calculated from four weekly closing stock prices labelled anz, cba, mcq and wbc for $n = 761$ weeks from 2000 April 12 through to 2014 October 29. The empirical copulas are calculated by converting each observation to an empirical percentile and plotting the same against each of the other series percentiles.

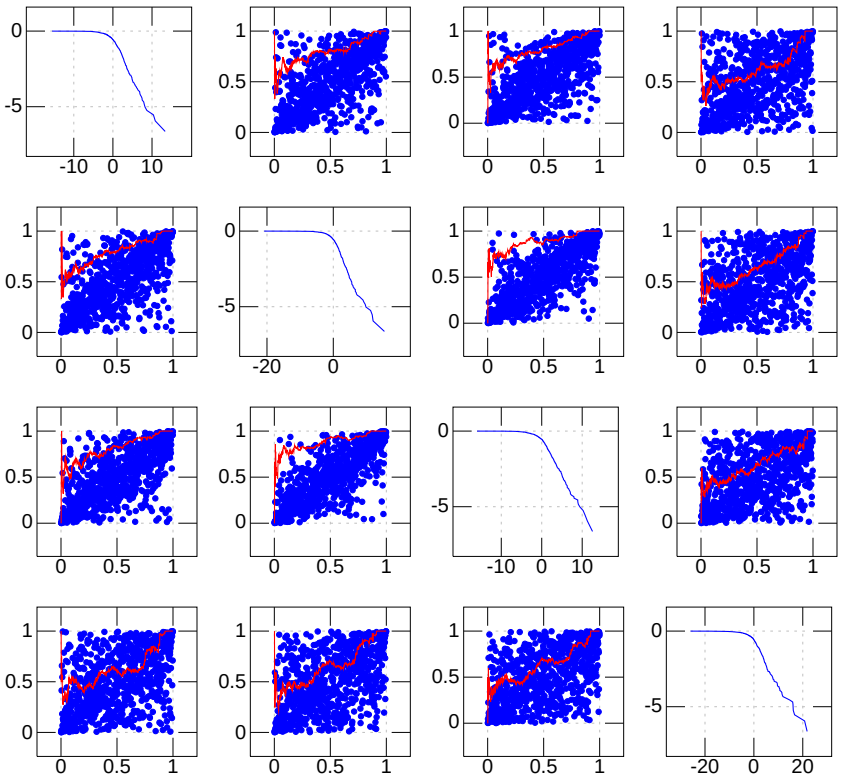

Figure 1: Pairwise copulas of bank stocks cba, anz, mqg, wbc, and the overall bank index. Red lines plot sensitivity s_{ij} as a function of q.

To estimate q_{ij} at a particular q, the second equation in (2) is iterated[3] starting from $q_{ij} = 0$ where copulas are estimated as

$$\hat{C}_{ij}(u_i, u_j) \equiv \hat{\mathrm{E}}\{(p_{ik} \leq u_i)(p_{jk} \leq u_j)\} \ .$$

Here $\hat{\mathrm{E}}$ computes the empirical mean over the cases $k = 1, \ldots, n$ and p_{ik} and p_{jk} are the empirical percentiles of case k of x_i and x_j, respectively.

4.5. Generalisations

Similar results apply when $\mathrm{Q}(x)$ is replaced by other risk measures such as $\mathrm{R}(x) \equiv \mathrm{E}\{xr(u)\}$ where r is a given function which acts componentwise and $xr(u)$ denotes componentwise multiplication. For example if $r(u) = mu^{m-1}$ then $\mathrm{R}(x) = \mathrm{E}\{\max(x^1, \ldots, x^m)\}$ where $x^1, \ldots, x^m$ are m independent copies of x and the risk measure is the expected worst outcome in m independent trials.

With $\mathrm{R}(x)$, the analogue of (1) is

$$\frac{\mathrm{dR}(x)}{\mathrm{d}x_j} \equiv \mathrm{R}(x|u_j > r_j) - \mathrm{R}(x_j) \ , \qquad r_j \equiv \mathrm{P}\{x_j \leq \mathrm{R}(x_j)\} \ . \tag{6}$$

This differs from (1) in that a different risk measure is used and r_j replaces q. If the components of $\mathrm{R}(x)$ are linear in the r_i then

$$\frac{\mathrm{dR}(x_i)}{\mathrm{dR}(x_j)} = \frac{f_j}{f_i} \frac{q_{ij}}{q(1-q)} \ , \qquad q_{ij} = C_{ij}(r_i + q_{ij}, r_j) - r_i r_j \ , \tag{7}$$

where f_i and f_j are the x_i and x_j densities at the r_i and r_i quantiles, respectively.

5. Partially Gaussian copulas

Suppose $(u, v) = F_*(x, y)$ where x and y are scalar random variables. Suppose $0 < u_1 < \cdots < u_n < 1$ are ordered observed values of u and $q_t = \Phi^-(u_t)$, $t = 1, \ldots, n$. The correspondingly ordered values of v and $z = \Phi^-(v)$ are denoted v_t and z_t. Note $\mathcal{E}(z_t) = \mathcal{E}(q_t) = 0$ and $\mathcal{E}(z_t^2) = \mathcal{E}(q_t^2) = 1$ where $\mathcal{E}$ is the empirical mean.

Smoothing methods are here proposed to smooth and simulate y values using a copula on (u, v) constructed via a cubic spline model linking z and u. Of particular concern are simulated z when u is near 0 or 1 corresponding to x extremes. Given a simulated u and in turn z, $(x, y) = F_*^-\{u, \Phi(z)\}$, a draw from the joint. Constraining u draws to say the upper tail, yields corresponding upper tail values for y. The challenge is to produce cogent z's or y's that properly reflect dependence in different parts of the distribution unencumbered by inappropriate constraints, and in particular tail constraints, inherent in say the Gaussian copula setup. Practical methods admit fitting on the basis of observed

[3]The second equation in (2) is a contraction mapping and hence has a fixed point.

data, and testing procedures for departures or otherwise from Gaussianity. The method outlined below achieves this.

The partially Gaussian (PG) copula for (u, v) is stated in terms of $q \equiv \Phi^-(u)$ and $z \equiv \Phi^-(v)$ on the grid $u_t = t/(n+1)$, $t = 1, \ldots, n$ with

$$z_t = \alpha_t + \beta q_t + s h_t \epsilon_t \ , \qquad \alpha_{t+1} = \alpha_t + \delta_t + c\eta_t \ , \qquad \delta_{t+1} = \delta_t + \eta_t \ . \tag{8}$$

Here ϵ_t and η_t are serially and contemporaneously uncorrelated zero mean disturbances with common variance σ^2 and $c = 3-\sqrt{3}$. The q_t term is the Gaussian component and α_t a nonparametric deviation.

The parameter s controls overall smoothness while $h_t \geq 0$, controlling the variability at different t determines local smoothness. In particular if $s = 0$ then the signal $\alpha_t + \beta q_t$ is forced to reproduce every z_t while $h_t = 0$ implies reproduction at the particular t. If h_t is large then little notice is taken of z_t and the fitted signal at t is relatively smooth. Thus s controls overall smoothness of the signal and h_t the smoothness at a particular t. Comparability between models and fits is facilitated by normalising $h_t > 0$ such that the average is 1.

The Gaussian copula is where α_t and h_t are identically 0 and 1, respectively, implying (z_t, q_t) with t uniform on $1, \ldots, n$ is bivariate normal with correlation $\beta = \pm\sqrt{1-(s\sigma)^2} \equiv \rho$.

The starting condition for the recursion in (8) are the second two equations with $t = 0$ and where α_0 and δ_0 are unknown parameters, similar to β. If $s \to \infty$ then $\sigma \to 0$, $(\alpha_t, \delta_t) \to 0$ for $t = 0, \ldots n$, and (8) describes a Gaussian copula.

The function α_t is thought of as an intercept function and provides for a nonparametric extension to the Gaussian copula. If $\sigma \to 0$, equivalent to $s \to \infty$, the first equation in (8) is $z_t = \alpha_0 + \delta_0 t + \beta q_t + s\epsilon_t$ and $\beta = \rho$ if $\alpha_0 = \delta_0 = 0$. Model (8) with $\beta = 0$ is the nonparametric model since z_t is varies smoothly with t, unrelated to q_t. If $\beta = 0$ and $\sigma \to 0$ there is a probit relationship between z_t and u_t. Finally $s = 0$ implies $\alpha_t + \beta q_t = z_t$ for any β. The full range of possibilities is set out in Table 1 discussed in more detail below.

The $\alpha_t = 0$ and $\alpha_t = \alpha_0 + \delta_0 t$ specifications are generalised by supposing α_t is twice differentiable and minimising

$$\sum_{t=1}^{n} \left(\frac{z_t - \alpha_t - \beta q_t}{h_t} \right)^2 + s^2 \int_{-\infty}^{\infty} (\alpha_t'')^2 \, \mathrm{d}t \ . \tag{9}$$

with respect to α_t and β. In (9) the h_t are fitting weights: a large h_t implies error at t is largely ignored while a small h_t implies error at t is minimized at all cost. The function a_t and scalar b minimising (9) deliver a function $a_t + bq_t$ where a_t is piecewise twice differentiable and $a_t + bq_t$ passes closest to, in a mean square error sense, to z_t with a penalty s^2 for roughness measured in terms of the second derivative. At the minimum, the first term in (9) is $(s\sigma)^2$. Brown and de Jong (2001) demonstrate the equivalence of (8) and (9).

The function α_t minimising (9) is, at integer t, $a_t = \mathrm{E}(\alpha_t | z_1, \ldots, z_n)$ computed with (8) where ϵ_t and η_t are normal, 0 mean, variance σ^2 random variables

and q_t fixed. Between integer t, the minimum is quadratic in t. A useful feature is that β is simultaneously estimated alongside a_t as the generalised least squares estimate under (8). All calculations under (8) provide error covariance matrices as detailed and applied below.

6. Special cases of the PG model

Special cases of (8) are displayed in Table 1. In (8), if $s \to \infty$ then $\sigma \to 0$ since the variance of z_t is finite. Thus $s \to \infty$ implies the noise to signal ratio becomes large where noise refers to $s\epsilon_t$ and signal to η_t, the error term driving the smooth component α_t. As the noise to signal ratio s increases, less of the variation in z_t is explained by α_t. In the limit σ is zero implying $\alpha_t = \alpha_0 + \delta_0 t$, as straight line. The further restriction $\alpha_0 = \delta_0 = 0$ forces α_t identically zero.

The first row in Table 1 is the Gaussian copula. In this setup α_t is identically zero: there is no signal component. The additional restriction $\beta = 0$ eliminates the Gaussian component q_t. Co (or counter) monotonicity results in the Gaussian framework if $\rho = \pm 1$ as displayed in the third row. If $s = 0$ then there is no measurement noise and the combined signal $\alpha_t + \beta q_t$ must pass through the points z_t resulting in a perfect fit. In terms of (9), $s = 0$ implies there is no roughness penalty and hence the minimisation is solely focussed on minimising the first, squared error, term, achieved by setting $\alpha_t + \beta q_t = z_t$.

Table 1: Special cases of the PG copula; in all cases $\rho \equiv \pm\sqrt{1-(s\sigma)^2}$

case	α_t	β	s	σ	α_0	δ_0	other
Gaussian	0		∞	0	0	0	$\rho = \beta$
independence	0	0	∞	0	0	0	$\rho = 0$
$\pm$comonotonic	0	± 1	0	0	0	0	$z_t = \pm q_t$
empirical	z_t		0				$\rho = 1$
linear probit	$\alpha_0(1-2u_t)$	0	∞	0		$\frac{-2\alpha_0}{n+1}$	indep. if $\alpha_0 = 0$
nonparametric	$\mathrm{E}(z_t \vert u_t)$	0					non Gaussian
PG	$\mathrm{E}(z_t - \beta q_t \vert u_t)$						

The linear probit row in Table 1 eliminates the Gaussian component q_t by setting $\beta = 0$, and giving maximum penalty for rougness. The resulting prediction is a straight line $\hat{z}_t = \alpha_0 + \delta_0 t$. However the mean of the z_t and hence $\hat{z}_t$ is zero implying $0 = \alpha_0 + \delta_0(n+1)/2$. Solving for δ_0 gives the δ_0 constraint displayed in Table 1 and the resulting straight line signal α_t.

The bottom row of Table 1 displays the general partial Gaussian specification absent any constraints on the parameters. Eliminating the Gaussian component from the signal by setting $\beta = 0$ leads to the nonparametric, non Gaussian specification in the second last row of Table 1.

7. Assessing goodness of fit

A plot of $a_t + bq_t$ versus rq_t is useful for assessing Gaussianity – that is whether the copula is Gaussian. If $a_t + bq_t \approx rq_t$ then the quantiles predicted under the G and PG models are approximately the same and a 45 degree line plot suggests a Gaussian copula is adequate. This analog of the qq–plot is converted to the pp–plot analog by transforming each coordinate using Φ. Using the inverse marginal distribution of each of the variables displays the copula fit information in the original scale.

For the Gaussian copula the single parameter is $0 \leq \rho \leq 1$, estimated by r. The analogous estimate for the PG copula is $R = \pm\sqrt{1-(s\hat{\sigma})^2}$ signed according to b. A measure of the contribution of the nonparametric component α_t to the PG fit is $R^2 - r^2 = 1 - r^2 - (s\hat{\sigma})^2$. A more formal analysis is based on the log–likelihood where in (8) the disturbance terms ϵ_t and η_t are assumed normally distributed. Then minus twice the log–likelihood[4] of the Gaussian and PG model are computed and their difference compared to a chi–squared distribution with 2 degrees of freedom for significance.

Local measures of departure from Gaussianity are

$$\frac{z_t - rq_t}{h_t\sqrt{1-r^2}}\ , \qquad \frac{z_t - a_t - bq_t}{s\hat{\sigma}h_t}\ . \tag{10}$$

The standardised residuals provide insight into areas of lack of fit, with mean and standard deviation, across t, of 0 and 1, respectively.

8. Tuning PG to particular parts of the joint distribution

The function h_t is used to assign importance to parts of the joint of distribution. If all parts are equally important then $h_t = 1$ for all t. If $h_t \approx 0$ when $t \approx n$ then fidelity is enforced in the upper tail. Alternatively if $h_t \approx 0$ for $t \approx n/2$ then there is an attempt to accurately reproduce z_t in the middle of the distribution with smoothness in both tails.

9. Overall and local correlation

Bjerve and Doksum (1993) state "The local regression slope, which focuses on change in expected y as x changes, may be of greater interest than expected y for a given x." They go on to argue for a "scale–free standardised version of the local regression slope."[5] In terms of the current notation, their suggested

[4] This is the log–likelihood $z_t \sim N(\rho q_t, 1-\rho^2)$ and differs from the bivariate Gaussian log–likelihood. It appears as $s \to \infty$, $s\hat{\sigma} \to \sqrt{1-b^2}$ and so the G copula appears enforced.

[5] The scale free feature conflicts with the calculations below where slope differs according to whether it is with respect to t compared to q_t. Hence the noise terms must also be affected by the transformation in a way not apparent in the expressions below.

measure is

$$\rho_t \equiv \pm \left\{ 1 + \left(\frac{\psi_t}{s\sigma h_t} \right)^{-2} \right\}^{-\frac{1}{2}} , \tag{11}$$

where s, σ and h_t are as in (8) or (9) and ψ_t is the regression slope at t as set out below. The ratio $\psi_t/(s\sigma h_t)$ is a signal to noise ratio and a high value implies $\rho_t \approx \pm 1$. In general the signal ψ_t is the local slope of the relationship while $s\sigma h_t$ is a measure of local noise. The sign of ρ_t is discussed in the context of particular choices of signal ψ_t.

With the Gaussian copula

$$z_t = \beta q_t + (\sqrt{1-\rho^2})\varepsilon_t , \qquad \epsilon_t \sim N(0,1) .$$

In this case the nonparametric component α_t is absent. The slope with respect to q_t is constant $\psi_t = \beta = \rho$ as is the noise variance: $s\sigma h_t = \sqrt{1-\rho^2}$. Further (11) reduces to $\rho_t = \rho$ for all t. The Gaussian copula case is displayed in the first row of the of Table 2.

Table 2: Overall and local correlations. In all cases $R = \sqrt{1 - \mathcal{E}\{(z_t - \hat{z}_t)^2\}}$

case	ψ_t	q_t	α_t	estimates
Gaussian				
– overall, local	$\rho = \beta$	✓	×	$r = \mathcal{E}(z_t q_t)$
nonparametric				
– overall	ρ	×	✓	r from $\widehat{\text{PG}}(\beta = 0)$
– local	$n\phi(q_t)\delta_t$	×	✓	$\hat{\delta}_t$ from $\widehat{\text{PG}}(\beta = 0)$
PG				
– overall	ρ	✓	✓	r from $\widehat{\text{PG}}$
– local	$n\phi(q_t)\delta_t + \beta$	✓	✓	$\hat{\beta}$, $\hat{\delta}_t$ from $\widehat{\text{PG}}$

The second set of row in Table 2 is the nonparametric Gaussian copula where, in (8), $\beta = 0$. Hence q_t is removed and the relationship between z_t and t or u_t is modelled in terms of the nonparametric component α_t. The slope $\alpha_{t+1} - \alpha_t$ at t is, apart from error, δ_t. However δ_t is the slope with respect to t and required is the slope with respect to q_t:

$$\psi_t = \frac{\mathrm{d}\alpha_t}{\mathrm{d}q_t} = \left(\frac{\mathrm{d}u}{\mathrm{d}t} \frac{\mathrm{d}q_t}{\mathrm{d}u} \right)^{-1} \frac{\mathrm{d}\alpha_t}{\mathrm{d}t} \approx \left\{ \frac{1}{n\phi(q_t)} \right\}^{-1} \delta_t = n\delta_t \phi(q_t) , \tag{12}$$

where ϕ is the standard normal density.

The final set of rows in Table 2 deals with the general PG case. As in the other situations, the overall correlation ρ is derived from the overall fit. The overall correlation exceeds that derived from both the specialised Gaussian and nonparametric cases. The higher overall correlation is partialed out locally

according to both nonparametric and parametric components. In any PG fit $\hat{\beta}$ will differ from the Gaussian copula estimate $\mathcal{E}(z_t q_t)$.

If $s = 0$ then $s\epsilon_t = 0$ and the signal to noise ratio is infinite implying $\rho_t = 1$ for all t. Thus the empirical and comonotonic cases imply, as expected, $\rho_t = \pm 1$. With the linear probit model $\sigma = 0$ and $\alpha_t = \alpha_0(1-2u_t)$ with slope with respect to u_t of $-2\alpha_0$ and

$$\psi_t = \left(\frac{\mathrm{d}q_t}{\mathrm{d}u_t}\right)^{-1} \frac{\mathrm{d}\alpha_t}{\mathrm{d}u_t} = -2\alpha_0\phi(q_t) \;, \qquad \rho_t \equiv \pm\left[1 + \left\{\frac{2\alpha_0\phi(q_t)}{s\sigma h_t}\right\}^{-2}\right]^{-\frac{1}{2}} \;, \quad (13)$$

If h_t is constant then local correlation is highest if $q_t = 0$ and monotonically declines towards zero as q_t moves away from zero.

The above ψ_t and implied correlations are for the (q, z) scale. However the local correlations also apply on the (u, v) scale since the transformation from u to q and v to z are the same inverse normal probability transform and slopes are identical[6]

$$\frac{\mathrm{d}v}{\mathrm{d}u} = \left(\frac{\mathrm{d}z}{\mathrm{d}v}\right)^{-1} \frac{\mathrm{d}z}{\mathrm{d}q}\frac{\mathrm{d}q}{\mathrm{d}u} = \frac{\mathrm{d}z}{\mathrm{d}q} \;.$$

10. Simulating from the PG copula

To explain PG simulation, consider initially the Gaussian model. Given data (q_t, z_t), $t = 1, \ldots, n$ where $q_t = \Phi(u_t)$ and $u_t = t/(n+1)$ suppose

$$\tilde{z}_i \sim N(r\tilde{q}_i, 1 - r^2) \;, \qquad r = \mathcal{E}(z_t q_t) = \mathrm{E}(\rho|z) \;, \qquad \tilde{q}_i \sim N(0, 1) \;, \qquad (14)$$

where $z \equiv (z_1, \ldots, z_n)$ contains the observed values of z_t. Simulating $\tilde{q}_i$ multiple times produces multiple $\tilde{z}_i$ while r remains fixed. To ensure coverage, $\tilde{q}_i = q_i$, $i = 1, \ldots, n$. The n pairs $(u_i, \tilde{v}_i)$ where $\tilde{v}_i = \Phi(\tilde{z}_i)$, $i = 1, \ldots, n$ provide n independent draws with uniform empirical distribution for the u_i and approximately uniform empirical distribution for the $\tilde{v}_i$. Repeatedly cycling over $i = 1, \ldots, n$ produces an unlimited supply of draws.

The PG generalisation of (14) is to repeatedly cycle over $i = 1, \ldots, n$ with

$$\tilde{z}_i \sim N(a_i + bq_i, s^2\hat{\sigma}^2) \;, \qquad a_i = \mathrm{E}(\alpha_i|z) \;, \qquad b = \mathrm{E}(\beta|z) \;. \qquad (15)$$

The $a_i \equiv \mathrm{E}(\alpha_i|z)$ and $b \equiv \mathrm{E}(\beta|z)$ and estimate of σ are calculated with the Kalman Filter Smoother (KFS) (De Jong, 1991) while s is set to achieve desired smoothness. Alternatively s is estimated as discussed in §**??**. In (**??**) the a_i, q_i and b remain fixed across cycles, analogous to fixing r in the Gaussian copula simulation. If $a_i = 0$ for all i then (15) reduces to (14) since in this case $b = r$ and $s^2\hat{\sigma}^2 = 1 - r^2$.

[6] This also follows since local correlation is independent of scale.

Iterations (14) and (15) ignore estimation variability in r or a_i and b. This shortcoming is addressed by varying a_i and b across each cycle

$$\tilde{a}_i \sim N\{a_i, \mathrm{cov}(\alpha_i|z)\} \ , \qquad i = 1, \ldots, n \ , \qquad \tilde{b} \sim N\{b, \mathrm{cov}(\beta|z)\} \ ,$$

and using these in the first relation of (15) instead of a_i and b. The conditional covariance matrices are recursively calculated with the KFS alongside a and b, while consistent draws across i are made with the simulation smoother extension (De Jong and Shephard, 1995).

11. Ordering

Suppose z_{xt} is the percentile based z–score of series x at index t. In this section two common indexes are discussed: time t and ordering with respect to some other variable. A first order approximation is $z_{xt} = \beta_x q_t + \epsilon_{xt}$ or in vector form $z_t = \beta q_t + \epsilon_t$ where z_t, b and ϵ_t are vectors and q_t scalar. For the Gaussian copula $\beta = \rho$.

12. Factor PG copula model

The PG model is stated in terms of $q_t = \Phi^-(u_t)$ where $u_t = t/(n+1)$, $t = 1, \ldots, n$. The role of x is to order z: the variable x is permuted and z is permuted using the same permutation. This suggests using a composite variable for ordering.

For example suppose x_t is a vector of variables initially ordered in terms of time t. The associated percentile based z–scores are $z_t = (\Phi^{-1} \circ \mathrm{P})x_t$ where P is the generic percentile operator.[7] Write $z_t = Uc_t$ where $(c_1, \ldots, c_n) = DV'$ derived from the singular decomposition $UDV' = (z_1, \ldots, z_n)$ arranged so that the singular values in the diagonal matrix D are in decreasing order. Use the first component in c_t, corresponding to the largest singular value, to permute $x_1, \ldots, x_n$ and fit PG copulas

$$x_t = \alpha_t + \beta q_t + sh_t\epsilon_t \ ,$$

where α_t and β are vectors and h_t is a diagonal matrix. The components of α_t are individual departures from the Gaussian component and $\beta\beta'$ models the joint Gaussian structure.

Suppose y_t is a p–vector with components y_{it}, $i = 1, \ldots, p$ with y_t is observed for $t = 1, \ldots n$. Suppose each component series in y_t is ordered according to x_t, a scalar series observed or inferred along with the y_{it}. Without loss of generality assume y_t, $t = 1, \ldots, n$ denotes the component series after ordering with respect to x_t and define v_t and z_t from the componentwise ordered y_t as, for $t = 1, \ldots, n$,

$$z_t = \Phi_*^-(v_t) \ , \qquad v_t = F_*(y_t) \ , \qquad q_t = \Phi^-(u_t) \ , \qquad u_t = \frac{t}{n+1} \ .$$

[7] A final adjustment is made to ensure the z–scores have mean zero and variance 1.

Using this notation the factor PG model is as in (8) except that z_t, α_t, β and ϵ_t are vector and h_t is a diagonal matrix. The diagonal matrix h_t distributes the quantum of error σ across component series i and percentiles t. The matrix h_t is standardised such that $1'h_t1 = p$. Further α_t is formed as in the second and third equations in (8) expect that both α_t and δ_t are vector. More specific structure is:

- If all components of α_t are identical then α_t is written as $1\alpha_t$ where in the last expression α_t is scalar. In this formulation deviations from Gaussianity are the same for each series. This is testable in the current framework.
- If all components of β are the same then β is written as 1β where β is scalar. In this case all components in z_t respond in the same way to the Gaussian component q_t but deviations may occur.
- A generalisation is where $z_t = B_t\alpha_t + Q_t\beta + sh_t\epsilon_t$.

Replacing $1\alpha_t$ in (8) by a vector α_t and making $s\sigma$ component specific yields the PG model. Variants include forcing a common response to q_t by requiring β to be scalar.

The one–factor PG model is estimated in the same way as the PG model, using vector z_t. In practice x_t, the scalar variable serving to order the components of y_t, is either an external driver or a composition from y_t. At present q_t is structured in terms of the normal score derived from the percentile of x_t. More extreme tail behavior is modelled by for example replacing the normal score q_t by say χ_t the chi–squared score derived from the percentile. This is explained in detail in the next section.

13. Extreme tail copulas

14. Below here is review/junk

14.1. A model for monitoring slope

Suppose

$$y_t = \beta m_t + \mu_t + s\epsilon_t \ , \qquad \mu_{t+1} = \mu_{t-h+1} + hr_t + c\eta_t \ , \qquad r_{t+1} = r_t + \eta_t \ , \tag{16}$$

where y_t is a stock price, m_t is a market index, μ_t is the price level and r_t is the return per period and h is a positive integer. Note r_t measures the expected per period change in the level μ_t. Thus

14.2. Kelly criterion

Suppose r is a vector of stock returns with mean μ and covariance matrix C. Suppose proportion π_i of total wealth w is invested in stock i. Then the rate of return on the portfolio is $\pi'r$. The expected log wealth after one period is

$$\mathrm{E}[\ln\{w(1+\pi'r)\}] \approx \ln(w) + \pi'\mathrm{E}(r) - \frac{\pi'\mathrm{E}(rr')\pi}{2} = \ln(w) + \pi'\mu - \frac{\pi'(C+\mu\mu')\pi}{2} \ .$$

To maximise log wealth (and hence wealth), maximise this expression with respect to π. Setting up the Lagrangian, differentiating and equating to zero yields

$$\pi = (C + \mu\mu' - \lambda I)^{-1}\mu \ , \qquad 1'\pi = 1 \ .$$

Given the optimal π the growth rate is

$$\mathrm{E}\ln(1 + \pi' r) \approx \mu'(C + \mu\mu' - \lambda I)^{-1}\mu - \frac{\mu'(C + \mu\mu' - \lambda I)^{-1}}{2}$$

The factor $1 + \pi' r$ is the growth factor. Over n periods the growth factor is

$$\mathrm{e}^{\ln(1+\pi' r_1)+\ln(1+\pi' r_2)+\cdots \ln(1+\pi' r_n)} \to \mathrm{e}^{n\mathrm{E}\ln(1+\pi' r_t)} = \mathrm{e}^{nm}$$

where $r_t \sim (\mu, C)$ are the vectors of growth rates for each period and π is the fixed allocation. Hence the rate of growth is

$$m = \mathrm{E}\ln(1 + \pi' r_t)$$

14.3. Continuous formulation

For a single stock we have $\mathrm{d}s/s = \mu \mathrm{d}t + \sigma \mathrm{d}b$ where b is brownian motion. Alternatively $\mathrm{d}s = \mu s \mathrm{d}t + \sigma s \mathrm{d}b$. For a portfolio π of stocks this becomes

$$\frac{\pi' \mathrm{d}s}{\pi s} = \pi' \mu \mathrm{d}t + \pi' \mathrm{d}b$$

$$\pi = -C^{-1}(1 r_f - \mu)$$

Note $C^{-1}(r - \mu)$ is the standardised prediction error when predicting each component of r using all the other components and supposing r is realised. Since $\mu > 1 r_f$, outcomes $r = 1 r_f$ are downside surprises. Thus π is the negative of the standardised prediction error when predicting each component of r using all other components and assuming the downside surprise $r = 1 r_f$. A high value π_i occurs if there is huge downside surprise if $r_i = r_f$ when all other components downward surprise at r_f. Hence the Kelly criterion loads the portfolio with stocks i which have large downward surprise at $r_i = r_f$ if all other stocks downward surprise with outcomes r_f.

14.4. Markowitz

Suppose we wish to maximise the expected return $\pi'\mu$ subject to a constraint on the volatility, $\pi' C \pi \le c$. Then the Lagrangian is $\pi'\mu + \lambda(c - \pi' C \pi)$. Differentiating and equating to zero yields the first order conditions

$$\mu = 2\lambda C \pi \qquad \Rightarrow \qquad \pi = \frac{1}{2\lambda} C^{-1} \mu$$

where λ is chosen so that $\pi' C \pi = c$.

If $C = c 1 1'$ then

To maximise the expected growth rate requires π to be chosen as above. Can we write a SSM for this?

$$1 = 1'\pi \ , \qquad s_t = 1\mu + r_t + \epsilon_t$$

Multiplying the first equation by $1'$ yields

$$\pi = n \leq \ , \qquad \lambda = n1'\mu - n1'(C + \mu\mu')\pi \ ,$$

$$\mu = (C + \mu\mu')\pi \qquad \Rightarrow \qquad \pi = (C + \mu\mu')^{-1}\mu$$

Maximising the expected log–accumulation yields the first order conditions $\mathrm{E}(r_t) = \mathrm{E}(r_t r_t')\pi$ or $\pi = \{\mathrm{E}(r_t r_t')\}^{-1}\mathrm{E}(r_t)$. If one the components of r_t is constant (the risk free rate of return δ say), then (using partitioned inverses)

$$\pi = \{\mathrm{cov}(r_t)\}^{-1}\{\mathrm{E}(r_t) - \delta 1) \ .$$

14.5. Bjerve and Doksum (1993) review

1. Points out $\mathrm{cov}(z|t = t^*) \neq 0$ while $\mathrm{cov}\{\mathrm{E}(z)|t = t^*\} = 0$. So cannot form a correlation from the variation at t compared to the variation in the mean – there is no variation in the latter.
2. Conditioning t on a neighbourhood of t^* and calculating the correlation in neighbourhood will lead to low correlation even if relationship is strong. Not sure I understand this.
3. Distinguishes between $\mathrm{cov}\{\mathrm{E}(z|t)\}/\mathrm{cov}(z)$ as a measure of correlation and ordinary correlation. The former he calls Pearson correlation while Gallton–Pearson is the usual correlation.
4. States that to the first order both correlations are such that

$$\mathrm{cor}\{z, t|t \in N_h(t^*)\} = \frac{h\beta_{t^*}}{\sqrt{3}\sigma_{t^*}}$$

   where β_t is the local regression and σ_t the local residual? variation.
5. Then considered is

$$\xi_{t^*} = \lim_{h\to 0} \frac{\mathrm{cor}\{z, t|t \in N(t^*)\}}{h/\sqrt{3}} = \frac{\beta_{t^*}}{\sigma_{t^*}}$$

   that is the local slope devided by the local standard deviation. Note I assume the overall correlation is 1 and so σ_1 does not appear in the numerator.
6. Maps ξ_{t^*} into the unit interval

$$\rho_{t^*} = \pm\left(1 + \xi_{t^*}^{-2}\right)^{-\frac{1}{2}} = \pm\left(1 + \frac{\sigma_{t^*}^2}{\beta_{t^*}^2}\right)^{-\frac{1}{2}} \ ,$$

15. Fitting PG copulas

The PG is readily fit to percentile data, and provides a practical platform for copula simulation. This is discussed, implemented, and used in subsequent sections. The parameters defining a PG copula are s, β, σ and starting conditions α_0 and δ_0. Given s, the other parameters have closed form expressions calculated with the Kalman filter. A best estimate of s is derived by iterating over s so as to minimise an appropriate objective function such as least squares.

To compute local correlation, write $\Delta z_t \equiv z_{t+1} - z_t$ and similarly for other quantities. Assume t is given and note that

$$\Delta z_t = \delta_t + c\eta_t + s(\epsilon_{t+1} - \epsilon_t) + \beta \Delta q_t \ , \qquad \Delta q_t \approx \frac{q_t'}{n+1} = \frac{1}{(n+1)\phi(q_t)}$$

Hence

$$\mathrm{E}(\Delta z_t) \approx \mathrm{E}(\delta_t) + \frac{\mathrm{E}(u_t)}{\phi(q_t)}$$

Then q_t is a random variable and

$$\mathrm{cov}(\Delta z_t, \Delta q_t | t) = \beta^2 \mathrm{cov}(\Delta q_t | t) \approx (\beta q_t')^2 \mathrm{cov}(u_t) = \frac{(\beta\sigma)^2 \pi \mathrm{e}^{q_t^2}}{6}$$

$$\mathrm{cov}(\Delta z_t | t) = (1 + c^2 + 2s^2)\sigma^2 + \beta^2 \mathrm{cov}(\Delta q_t | t)$$

This final expression assumes $\mathrm{cov}(\delta_t) = (n+1)\sigma^2/2$: a random t is chosen to observe the random walk. Hence the local correlation is

$$\frac{\mathrm{cov}(\Delta z_t, \Delta q_t)}{\sqrt{\mathrm{cov}(\Delta z_t)\mathrm{cov}(\Delta q_t)}} = \left\{ 1 + \frac{A}{\beta^2 \mathrm{cov}(\Delta q_t)} \right\}^{-1/2}$$

The local correlation between z_t and q_t is

$$\rho_t \equiv \mathrm{cor}(\Delta z_t, \Delta q_t | t) = \frac{\beta \mathrm{cov}(\Delta q_t)}{\sqrt{\mathrm{cov}(\Delta z_t | t)\mathrm{cov}(\Delta q_t)}} = \pm \left\{ \frac{\mathrm{cov}(\Delta z_t | t)}{\beta^2 \mathrm{cov}(\Delta q_t)} \right\}^{-1/2}$$

where "$|t$" indicates conditioning on information up to and including time t. A calculation shows writing $c = 4 - \sqrt{3}$,

$$\mathrm{cov}(\Delta z_t | t) = (2s^2 + c)\sigma^2 + \beta^2 \mathrm{cov}(\Delta q_t) \ , \qquad \mathrm{cov}(\Delta q_t) \approx (q_t')^2 \mathrm{cov}(u_t) = \frac{\sigma^2 \pi \mathrm{e}^{q_t^2}}{6} \ .$$

Hence

$$\emptyset(\rho_t) = \frac{\beta}{s_t \sigma} \ , \qquad s_t = \sqrt{\frac{6(2s^2 + c)}{\pi}} \mathrm{e}^{-q_t^2/2} \ .$$

Divisor s_t is small if $|q_t|$ is large indicating the model posits higher tail correlation than normal if $\beta \neq 0$. The ratio

$$\frac{\emptyset(\rho_t)}{\emptyset(\rho)} = \frac{s}{s_t} = \sqrt{\frac{\pi}{6(2 + c/s^2)}} \ \mathrm{e}^{q_t^2/2} \to \frac{\sqrt{2\pi} \ \mathrm{e}^{q_t^2/2}}{\sqrt{24}} \ ,$$

compares local to overall correlation. If $s \to \infty$ local correlation in the tails will far exceed overall correlation. This is obvious given $s \to \infty$ implies the model will exactly reproduce the data. If $s \to 0$

15.1. Alternative view of local correlation

In (8), a large $|\beta|$ suggests more extreme tail behaviour than in the Gaussian copula. To see this note

$$\Delta \mathrm{E}(\alpha_t + \beta q_t) \approx \frac{-2\delta_0}{n+1} + \frac{\beta}{(n+1)\phi(q_t)} = \frac{\beta/\phi(q_t) - 2\delta_0}{n+1} \ , \tag{17}$$

where Δ denotes change with respect to t. Change (17) is relatively large if q_t^2 is large, that is t is near 1 or n. Higher than normal tail dependence implies $|\beta|$ large. For a Gaussian copula, $\beta = \pm\sqrt{1-(s\sigma)^2}$ and comparing the ratio of estimates of the two sides to 1 is a test for Gaussianity.

15.2. Correlation between z_t and α_t

The term $\alpha_t + \beta q_t$ is the signal and ϵ_t, noise. The signal is relatively smooth and contains both a nonparametric and Gaussian component q_t. Departures of z_t from the Gaussian component are the sum of a persistent component α_t and transient component, $s\epsilon_t$. Since ϵ_t and the driver of persistence η_t share σ, the smoothing parameter s measures relative importance of transience as opposed to persistence. The scale parameter σ relates to the degree of roughness since it is the variance of the changes in the slope of α_t.

Since $\mathcal{E}(q_t) = \mathcal{E}(z_t) = \mathcal{E}(\epsilon_t) = \mathcal{E}(\eta_t) = 0$ it follows

$$0 = \mathcal{E}(\alpha_t) = \alpha_0 + \delta_0\mathcal{E}(t) = \alpha_0 + \delta_0\mathcal{E}(t) = \alpha_0 + \delta_0\frac{n+1}{2} \qquad \Rightarrow \qquad \delta_0 = \frac{-2\alpha_0}{n+1} \ .$$

Further $\mathcal{V}(q_t) = \mathcal{V}(z_t) = 1$ implies an analysis of variance type decomposition

$$1 = \mathcal{V}(\alpha_t) + \beta^2 + (s\sigma)^2 \ . \tag{18}$$

Decomposition (18) splits the total variance of 1 into that due to the nonparametric component α_t, Gaussian component β^2 and residual noise, $(s\sigma)^2$. Alternatively (18) states the variance of the nonparametric component is $1 - \beta^2 - (s\sigma)^2$.

An example PG copula with $s\sigma = xx$, $\sigma = 0$, $\beta = xx$ and $\alpha_0 = xx$ is displayed in Figure **??**.

16. Correlations with the PG copula

With (8) and t uniform on $1, \ldots, n$

$$\mathcal{V}\begin{pmatrix} z_t \\ \alpha_t \\ q_t \end{pmatrix} = \begin{pmatrix} 1 & 1-(s\sigma)^2 & \beta \\ 1-(s\sigma)^2 & 1-\beta^2-(s\sigma)^2 & 0 \\ \beta & 0 & 1 \end{pmatrix} \ .$$

This covariance matrix serves to define a variety of ordinary and partial correlations.

The correlation between z_t and q_t is β. The partial correlation between z_t and the Gaussian component q_t, after removing the effect of the nonparametric component α_t, is called the Gaussian correlation:

$$\rho_q \equiv \frac{\mathrm{cov}(q_t, z_t - \alpha_t)}{\sqrt{\mathrm{cov}(q_t)\mathrm{cov}(z_t - \alpha_t)}} = \frac{\beta}{\sqrt{\beta^2 + (s\sigma)^2}} = \pm \left\{ 1 + \left(\frac{s\sigma}{\beta} \right)^2 \right\}^{-1/2} , \quad (19)$$

where the sign is that of β. Hence $\rho_q = 1$ or 0 if $s\sigma = 0$ or $\beta = 0$, respectively. Using (18), ρ_q^2 equals the proportion of residual variance $\mathcal{V}(z_t) - \mathcal{V}(\alpha_t) = \beta^2 + (s\sigma)^2$ explained by the Gaussian component q_t.

and ρ_a is 1 or 0 according to whether $s\sigma$ equals 0 or $\sqrt{1-\beta^2}$, respectively. Finally the overall correlation is the correlation between z_t and the signal, consisting of both nonparametric and Gaussian component:

$$\rho \equiv \mathrm{cor}(z_t, \alpha_t + \beta q_t) = \mathrm{cor}(z_t, z_t - s\epsilon_t) = \frac{1 - (s\sigma)^2}{\sqrt{\mathrm{cov}(z_t - s\epsilon_t)}} = \sqrt{1 - (s\sigma)^2} \ .$$

with $\rho = \rho_a$ if $\beta = 0$. If $s = 0$, there is no measurement error in (8) or, equivalently, there is no roughness penalty in (9) and $\rho_q = \rho_a = \rho = 1$.

17. Special cases of the PG copula

Table **??** sets out possible choices for s and β. These combinations are discussed in the further subsections. Note that $s = 0$ does not imply $s\sigma = 0$ as $s \to 0$ may imply $s\sigma$ converges to a positive number.

17.1. Empirical copula

The empirical copula results in (8) when $s = 0$. In this case $z_t = \alpha_t + \beta q_t$. The deviation of $z_t - (\alpha_t + \beta q_t)$ can be made zero by selecting appropriate η_t. Hence $s = 0$ implies the fitted copula is the empirical copula. Note $\mathcal{V}(\alpha_t) = 1 - \beta^2$ and since β is arbitrary, the correlation between z_t and the nonparametric component is 1.

17.2. Large local variation

The other extreme $s \to \infty$ is displayed in the bottom rows of Table **??**. Large s forces negligible σ and straight line behaviour for α_t yielding

$$z_t = \alpha_0(1 - 2u_t) + \beta q_t + s\epsilon_t \ ,$$

where $s\epsilon_t$ has finite variance $(s\sigma)^2$.

17.3. Gaussian copula

If $s \to \infty$ and $\alpha_0 = 0$ then (q_t, z_t) is bivariate Gaussian with correlation β and (8) reduces to a Gaussian copula with correlation β and $\rho = \rho_q$. In this case $s\sigma = \sqrt{1 - \beta^2}$.

17.4. Probit copula

If $s \to \infty$ and $\beta = 0$ then (8) defines a probit copula. From (**??**), $\alpha_0 = \pm\sqrt{3\{1-(s\sigma)^2\}}$.

17.4.1. Independence

If $s \to \infty$ and $\alpha_0 = \beta = 0$ then the variables are independent. The situation is achieved either letting $\beta \to 0$ with a Gaussian copula or letting $\alpha_0 \to 0$ with a Probit copula. Under independence $s\sigma = 1$.

17.5. Conditionally Gaussian copula

This is the intermediate situation where $0 < s < \infty$. Testing $\alpha_0 = 1/s = 0$ is a test for a Gaussian copula. Goodness of fit is measured with

$$\hat{\rho}^2 = 1 - (s\hat{\sigma})^2 = 1 - \mathcal{E}\{(z_t - \hat{z}_t)^2\} \ , \tag{20}$$

where $\hat{z}_t$ is the estimate of z_t based on (8) using estimates of α_0, β and $s\sigma$. Values of $\hat{\rho}^2$ lie between 0 and 1 and have the usual interpretation: near 1 indicates the percentiles of u well explain z values of v. Note $R \to \beta$ if $\alpha_0 = 0$ and $s \to \infty$.

18. Estimation of the PG model

A convenient feature of the GC model is the straightforward estimation. The unknown parameters are α_0, β, σ, and s. Given s, estimates of the other parameters are in closed form – generalised least squares (GLS) estimators of α_0 and β, while σ is estimated from the fitted residuals. GLS estimates are derived using the Kalman filter (KF). The KF and the associated Smoothing filter (SF) also deliver diagnostics for assessing goodness of fit and estimates of sampling error. In these calculations s is either given or alternatively is iterated over to produce a quasi maximum likelihood estimate.

From (20)

$$\mathcal{E}\left\{\left(\frac{z_t - \hat{z}_t}{s\hat{\sigma}}\right)^2\right\} = 1$$

suggesting the comparison of standardised residuals to ± 2 and perhaps plotting the percentile rank of the standardised residuals against u_t to identify lack of fit.

The implied correlation between z_t and q_t is

$$\beta = \frac{\beta}{\sqrt{\beta^2 + (s\sigma)^2}} = \pm\left\{1 + \left(\frac{s\sigma}{\beta}\right)^2\right\}^{-1/2} \ .$$

The term $\beta/(s\sigma)$ is the signal to noise ratio. If $z_t, q_t \sim (0,1)$ then $\beta^2 + (s\sigma)^2 = 1$ and

$$\beta = \pm\sqrt{1-(s\sigma)^2} = R = \beta \ , \qquad \frac{\beta}{s\sigma} = \frac{\pm\sqrt{1-(s\sigma)^2}}{s\sigma} = \frac{\beta}{\sqrt{1-\beta^2}}$$

.

Thus the empirical correlation between z_t and q_t is

$$\sum_{t=1}^{n} z_t q_t = \beta + \delta_0 \sum_{t=1}^{n} t q_t$$

By construction both z_t and q_t have mean zero and variance 1 implying the correlation between z_t and q_t is

If $\beta = 0$ then

$$z_t = \alpha_0 + n\delta_0 u_t + \beta \Phi^-(u_t) + s\epsilon_t \ ,$$

Testing $\delta_0 = 0$ amounts to testing for independence. If $\beta = \pm 1$ then $z_t - (\pm q_t) = \alpha_0 + t\delta_0$ implying $\alpha_0 = n\delta_0 = 0$ and $z_t = \pm q_t$, that is the variables are co or counter monotonic. Finally if $\beta \neq 0, \pm 1$ then $z_t = \alpha_0 + t\delta_0 + \beta q_t + s\epsilon_t$ since $\alpha_0 = \delta_0 = 0$, implied by the fact that Since both z_t and q_t are, by construction, standard normal, this amounts to prescribing a Gaussian copula with correlation indicated in Table **??**. The ratio $\beta/(s\sigma)$ is the signal to noise ratio.

The middle column of the body of Table **??** displays alternatives for finite $s \neq 0$. If $\beta = 0$ the implied copula is nonparametric with α_t evolving smoothly as a function of t with the actual degree of smoothness increasing with s. A constraint is $\alpha_0 = \alpha_n = \sum_t \delta_t = 0$ and hence α_t forms a bridge. If $\beta = \pm 1$ then the signed difference $z_t - (\pm q_t)$ is modelled as a smooth. In a fit, the bridge structure is expected to be enforced by the data.

The scale σ quantifies roughness in the signal α_t but its size is determined by s with $s\sigma \leq 1$ implied by the standard normality of z_t. In fitting, s is either chosen a–priori, or estimated together with the other parameters. Given s, estimates of β, α_0, δ_0 and σ are closed form generalised least squares calculations. Thus s can be estimated using a one dimensional search. Given the values of the parameters, standard diagnostics are used to assess the fit and reveal areas of inadequate fit. These techniques are used and explained in the applications.

Use of q_t permits a partially Gaussian explanation of the copula with Gaussianity increasing with s. In the limit the enforced Gaussian copula has correlation R signed according to $\hat{\beta}$. Co or countermonotoniticity implies $R = \pm 1$ with the sign indicating the direction of the relationship. Setting $\beta = 0$ excludes the Gaussian copula component.

The fitted values $\hat{z}_t$ stacked in vector $\hat{z}$ provide a basis for simulating all percentiles including extreme percentiles. In particular suppose $z \sim N(\hat{z}, \Sigma)$ where Σ is the covariance matrix implied by the model. Then $v \sim \Phi_*(z)$ is a simulation of the the v_t. Bringing v into the original order yields (u_t, v_t), a simulation from the copula. Simulated percentiles are mapped back into the original scale as $F_*^-(u_t, v_t)$ to provide simulated (x_t, y_t) outcomes.

19. Copula forcing

Even if data suggests a strict Gaussian copula it may be desirable to force stronger dependence, especially in the tails. This forcing is achieved with

Table 3: Predicted z_t with conormal copula model

	$s=0$	$0<s<\infty$	$s\to\infty$
$\beta=0$	z_t empirical	α_t non parametric probit	$\alpha_0(1-2u_t)$ linear probit independence: $\alpha_0=0$
$\beta\neq 0,\pm 1$		$\alpha_t+\beta q_t$ conormal	$\alpha_0(1-2u_t)+\beta q_t$ Gaussian if $\alpha_0=0$
$\beta=\pm 1$		$\alpha_t\pm q_t$ cointegrated departure	$\alpha_0(1-2u_t)\pm q_t$ co/counter monotonic if $\alpha_0=0$

20. Application to copula fitting

The following example illustrates the copula fitting technique.

21. Financial sensitivity and contagion

Suppose the singular value decomposition of the matrix S of sensitivities at a particular q is $S=UDV'$. Here D a diagonal matrix of singular values, arranged in decreasing order along the diagonal and U and V are orthonormal: $U'U=V'V=I$. If the series are independent then $U=D=V=I$. If the series are comonotonic then $U=V=(1,0)$ while $D=\mathrm{diag}(1,0,\ldots,0)$. Hence in the comonotonic case $S=11'$.

So the best thing is to compare

$$Q=s_{i1}c_{j1}+\cdots+s_{ip}c_{jp}$$

If $Q=I$ then $S=C=I$. If $Q=11'$, a matrix of ones, then $S=\sqrt{p}(1,0)$ and $C=(c,0)$ both have rank 1: here s and c are the leading columns of S and C, respectively and 0 a matrix of zeros.

The approximation $Q\approx sc'$ is exact if Q has rank 1 implying $Q=11'$. The approximation $Q\approx sc'$ is inadequate if $Q=I$. Inadequacy can be judged from the largest singular value in D, as a proportion of p, the number of variables or rows in Q.

$$Q_*\equiv\frac{1}{p-1}(Q-I)=UDV'\ ,\qquad s\equiv Q_*1\ ,\qquad c\equiv Q_*'1\ .$$

$$Q=I+s1'+UDV'\ ,\qquad s\equiv\frac{1}{p-1}(Q-I)1$$

Then s is the vector of average VaR_q sensitivity of each variable to all others. Further c is the average VaR_q impact of each variable on all others. The matrices U, D and V define the singular value (svd) decomposition of Q_*, arranged so that diagonal matrix D has the singular values on the diagonal in descending

order. If $b = d_1 u_1$ and $c = v_1$ where u_1, d_1 and v_1 are the first column, top diagonal entry, and first column of U, D and V respectively, then for two variables $y \neq x$ in Q,

$$\Delta_x q_y = s_y + b_y c_x + \epsilon_{yx}$$

This states that percentage sensitivities are, apart from the "error" ϵ_{yx}, an average sensitivity plus a scaled response to the contagious effect of the x variable. The contagious effects contained in c are estimated by maximising the explanation of Q.

The vector $1'Q$ sums the changes in VaR_q when each of the column variables is stressed, and writes this as a proportion of the change in the variable being stressed. These proportional sums, subtracting 1 and divided by $p-1$ where p is the number of variables, measures the average contagion of each variable on all others: $c' = (p-1)^{-1}1'(Q-I)$ or $c = (p-1)^{-1}(Q-I)'1$

Alternatively the vector $Q1$ sums the changes in VaR_q of each row variable when all the column variables are stressed. Again it is appropriate to remove the effect of a variable on itself and consider the average over the remaining variables: $s \equiv (p-1)^{-1}(Q-I)1$. If $Q_* \equiv (p-1)^{-1}(Q-I)$ then $s = Q_*1$ and $c = \dot{q}_*1$ are the vectors of sensitivities and contagions, respectively. If $Q = I$ then $s = c = 0$. If $Q = 11'$ then $s = c = 1$.

22. Systemic risk and causal chains

A rank one approximation to the matrix Q is $Q \approx sc'$. Vector c is an index of the contagious impact of each of variables on the others while s measures the sensitivity of each variable to each of the others. The vectors s and c are derived from the singular value decomposition $Q = UDV'$ where $U'U = V'V = 1$ and where s and c are the first column of UD and V respectively, assuming the svd is organised so that the singular values in the diagonal matrix D are organised from largest to smallest. The appropriateness of the summarisation sc' is measured with $\text{tr}(Q - sc')$.

If the variables are independent then $Q = I$ and both s and c equal a column of the identity matrix with sc' a matrix of 0's except in a single diagonal position where it is 1. Then all but one variable has a contagion effect and only that variable is sensitive to the contagion provided by the variable.

If the variables are comonotonic then $Q = 11'$ and $s = 1$ and $c = 1$ where 1 denotes a vector of ones. Thus the rank 1 approximation is exact and each variable is equally contagious and equally sensitive.

Note that

$$Q^n \approx (s'c)^{n-1} sc' = ,$$

If the random variables are independent the $Q - I = 0$ and $a = b = k = 0$ and there is no error in the first order svd approximation. If the random variables are comonotonic then $Q = 11'$ and $Q - I$ has ones everywhere except on the diagonal where it is zero. The vector of row means is then $p^{-1}(p-1)1$

Systemic risk in the system is measured with $b'k = d_1(u_1'v_1)$. In the case of comonotonic random variables $Q = 11'$, $a = 1$ and x, where p is the number of variables, $d_1 = 1$.

Furthermore we may define quantities such as $u^- \equiv \text{VaR}_q(v|u \leq q)$ measuring the impact of a non distressed state in v. For brevity we do not dwell on these constructs in this writeup although the ramifications and potential uses of these constructs will be investigated in the research.

23. Generalising Sklar's Theorem

Skalr's Theorem states that if $F(x)$ is a a joint then there exists a C such that $F = C \circ F_*$ where $C_F = F$

24. Literature

We propose a measure for systemic risk: CoVaR, the value at risk (VaR) of financial institutions conditional on other institutions being in distress. We define an institution?s (marginal) contribution to systemic risk as the difference between CoVaR and the financial system?s VaR. From our estimates of CoVaR for characteristic-sorted portfolios of publicly traded financial institutions, we quantify the extent to which characteristics such as leverage, size, and maturity mismatch predict systemic risk contribution. We argue for macro-prudential regulation based on the degree to which such characteristics forecast systemic risk contribution.

25. Econometric implementation

The above development sets out our proposed broad framework for linking bivariate copulas and marginals to external variables and shocks study the impact of the same on stresses within the system and the contagious effects of crises. Proposed econometric analysis will implement and extend Brownlees and Engle (2010).

26. Data

We will employ publicly available data as published by APRA and other regulators.

References

Acharya, V., R. Engle, and M. Richardson (2012). Capital shortfall: A new approach to ranking and regulating systemic risks. *The American Economic Review 102*(3), 59–64.

Acharya, V., L. Pedersen, T. Philippon, and M. Richardson (2012). Measuring systemic risk.

Adrian, T. and M. K. Brunnermeier (2011). Covar. Technical report, National Bureau of Economic Research.

Bjerve, S. and K. Doksum (1993). Correlation curves: measures of association as functions of covariate values. *The Annals of Statistics*, 890–902.

Brown, P. and P. de Jong (2001). Nonparametric smoothing with the state space model. *Canadian Journal of Statistics 29*(1), 37–50.

Brownlees, C. T. and R. Engle (2010). Volatility, correlation and tails for systemic risk measurement. *New York University, mimeo*.

De Jong, P. (1991). The diffuse Kalman filter. *Annals of Statistics 19*(2), 1073–1083.

De Jong, P. and N. Shephard (1995). The simulation smoother for time series models. *Biometrika 82*, 339–350.