GUY CARPENTER

Quantifying Correlation with Copulas
Multivariate t-Copulas

G Venter
CARe June 2003

Marsh & McLennan Companies

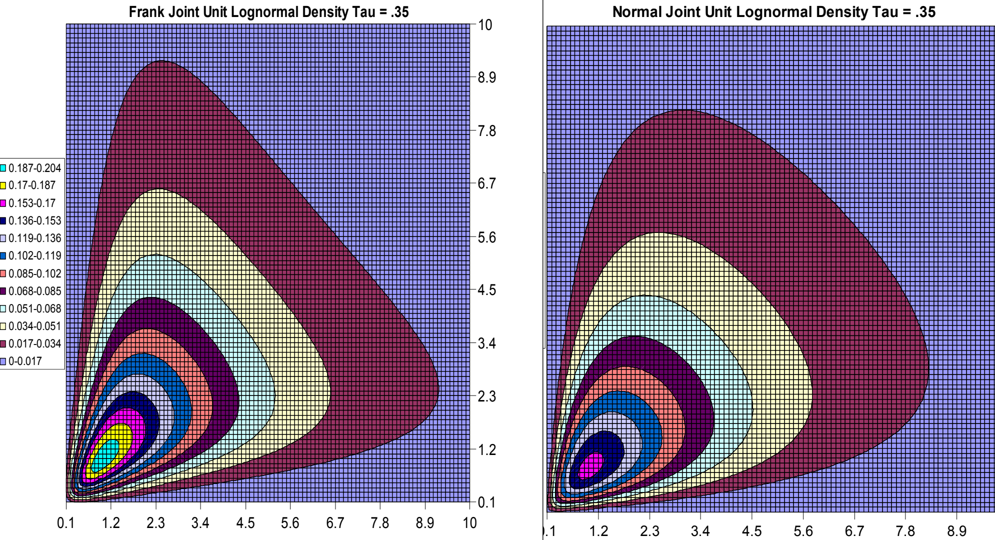
What is a copula?

- A way of specifying joint distributions when you know the individual marginals
- A way to specify what parts of the marginal distributions are most correlated
- Works by the copula correlating the probabilities, then applying inverse distributions to get the correlated marginal distributions
- Formally copulas are joint distributions of unit uniform variates, as probabilities are uniform on [0,1]

Formal Rules

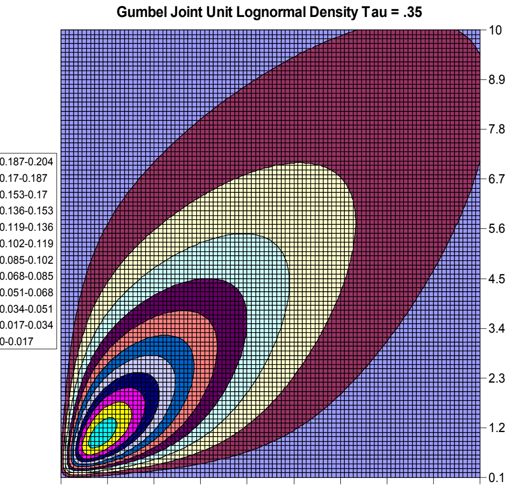
- $F(x,y) = C(F_X(x),F_Y(y))$
  - Joint distribution is copula evaluated at the marginal distributions
  - Expresses joint distribution as inter-dependency applied to the individual distributions
- $C(u,v) = F(F_X^{-1}(u),F_Y^{-1}(v))$
  - u and v are unit uniforms, F maps R^2 to [0,1]
- $F_{Y|X}(y) = C_1(F_X(x),F_Y(y))$
  - Derivative of the copula is the conditional distribution
- E.g., $C(u,v) = uv$, $C_1(u,v) = v = Pr(V<v|U=u)$
  - So independence copula

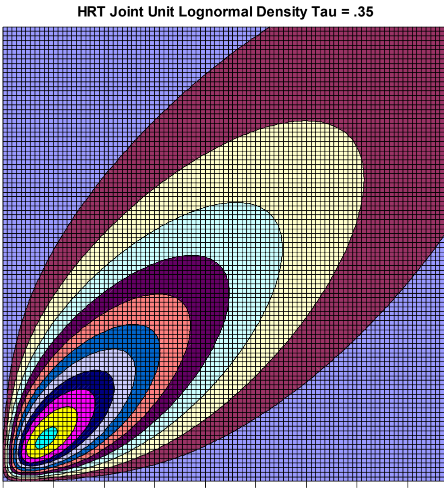
Copulas Differ in Tail Effects
Light Tailed Copulas Joint Lognormal



Copulas Differ in Tail Effects

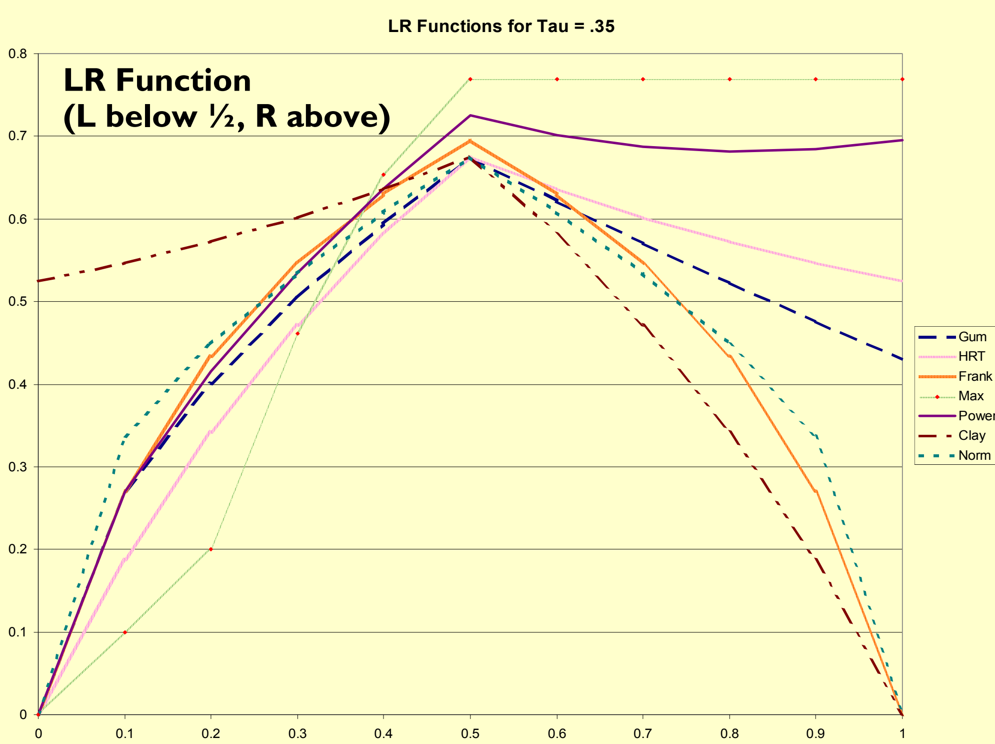
Heavy Tailed Copulas Joint Lognormal





Quantifying Tail Concentration

- $L(z) = Pr(U<z|V<z)$
- $R(z) = Pr(U>z|V>z)$
- $L(z) = C(z,z)/z$
- $R(z) = [1 - 2z + C(z,z)]/(1 - z)$
- $L(1) = 1 = R(0)$
- Action is in R(z) near 1 and L(z) near 0
- lim R(z), $z \to 1$ is R, and lim L(z), $z \to 0$ is L


LR Functions for Tau = .35
LR Function
(L below ½, R above)
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Gum
HRT
Frank
Max
Power
Clay
Norm

Example: ISO Loss and LAE

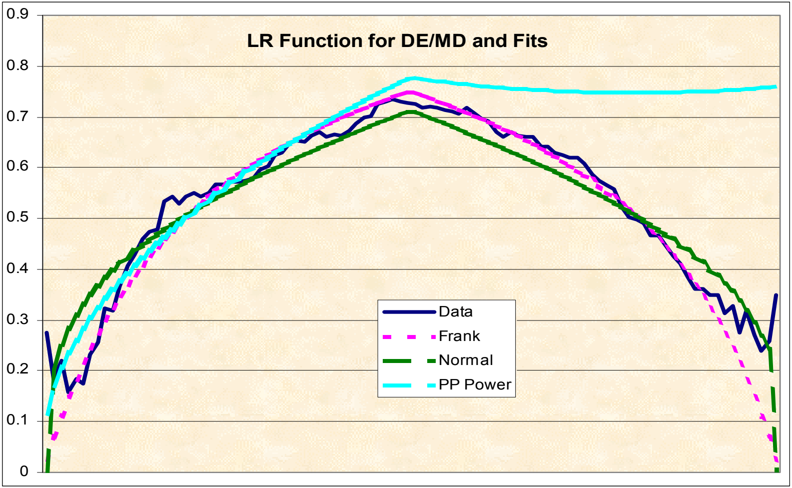
- Freez and Valdez find Gumbel fits best, but only assume Paretos
- Klugman and Parsa assume Frank, but find better fitting distributions than Pareto

	Loss Median	Loss Tail	Expense Median	Expense Tail
Frees & Valdez	12,000	1.12	5500	2.12
Klugman & Parsa	12,275	1.05	5875	1.58

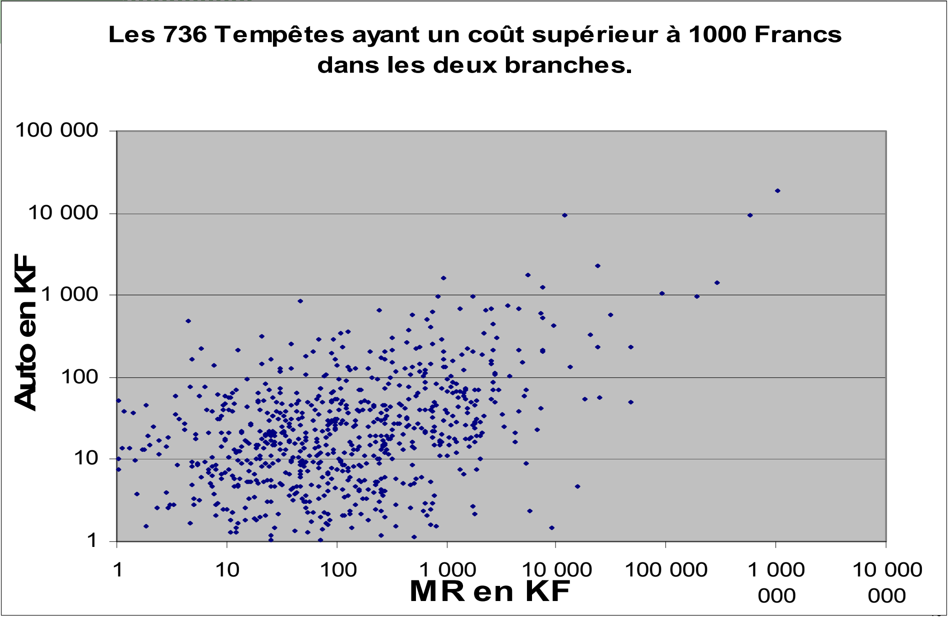
All moments less than tail parameter converge

Fitting Copulas to Cat Loss Data and Testing Fit by LR

	HRT	Gumbel	Frank	Normal	Flipped Gumbel
Parameter	0.968	1.67	4.92	0.624	1.68
Ln Likelihood	124	157	183	176	161
Tau	0.34	0.40	0.45	0.43	0.40

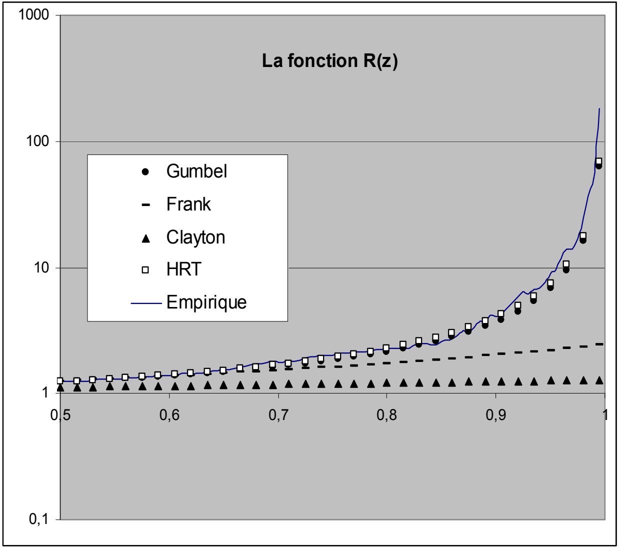


Auto and Fire Claims in French Windstorms



Modified Tail Concentration Functions

- Both MLE and R function show that HRT fits best



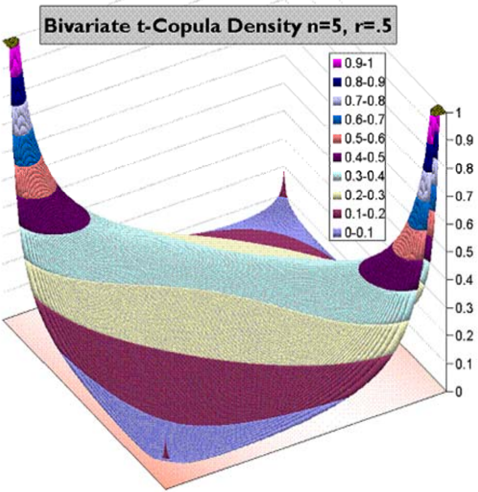
Extending to Multi-Variate Case

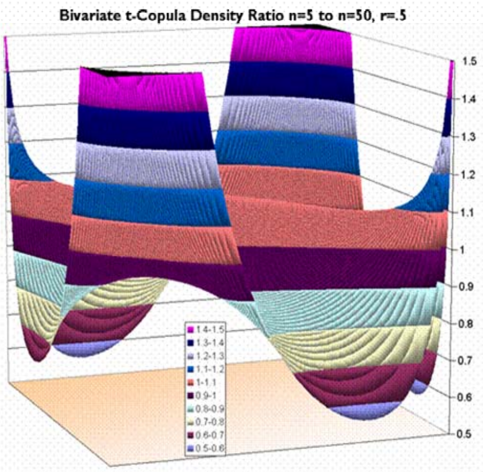
- Single parameter not enough – all variates would have same correlation
- You would like to have at least a parameter for each pair of variates to determine the strength of their dependency, and one overall for tail strength
- The t-copula has this minimum set
- With this minimum you can control all correlations but all tail strengths are the same

T- Distribution and t-copula

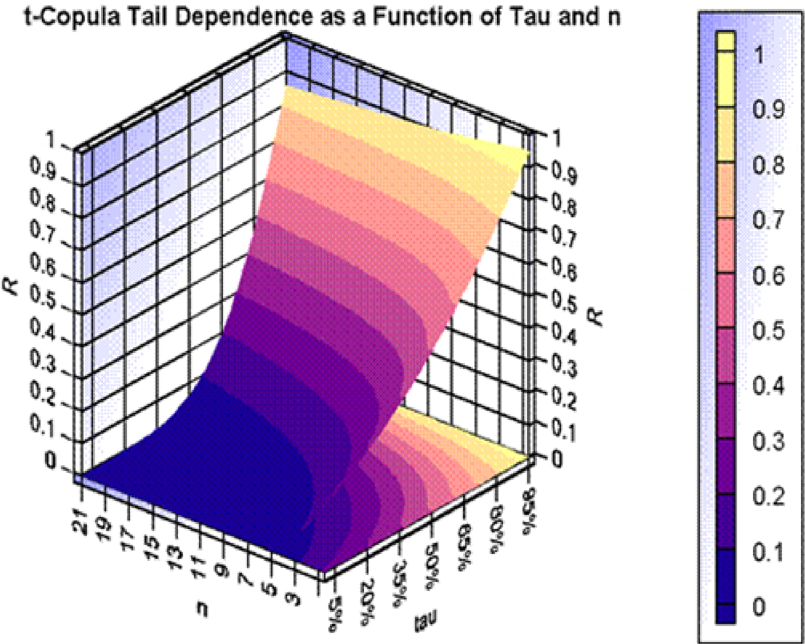
- T- distribution, n degrees of freedom (n=integer not necessary)
  - $f_n(x) = K_1(1+x^2/n)^{-(n+1)/2}$, with $K_1=\Gamma(1/2+n/2)(n\pi)^{-1/2}/\Gamma(n/2)$
  - $F_n(x) = ½ + ½\ \mathrm{sign}(x)\mathrm{betadist}[x^2/(n+x^2), ½, n/2]$
  - $s = F_n^{-1}(u) = \mathrm{sign}(u-½)n^{1/2}[-1+1/\mathrm{betainv}(|2u-1|,½, n/2)]^{-1/2}$ with Excel betainv
  - T ~ normal * {inverse gamma}$^{1/2}$
- Bivariate t-copula
  - $C(F_n(x), F_n(y))$ is bivariate t- distribution
  - Has common degrees of freedom n and a correlation parameter ρ
  - $c(u,v; n,\rho) = K_2[(1+s^2/n)(1+t^2/n)]^{(n+1)/2}\{1+[s^2-2\rho st+t^2]/[(1-\rho^2)n]\}^{-1-n/2}$
  - with $K_2 = ½[\Gamma(n/2)/\Gamma(0.5+n/2)]^2 n(1-\rho^2)^{-1/2}$ and $s=F_n^{-1}(u)$, $t=F_n^{-1}(v)$
  - C(F(x), G(y)) is bivariate distribution with t- copula for any distributions F and G
  - C ~ {normal copula with same ρ} *{inverse gamma}$^{1/2}$
  - Kendall's τ is related to ρ by $\tau = (2/\pi)\arcsin(\rho)$ (same as normal copula)
  - R, defined as limit $_{z\to 1}\Pr(U>z|V>z)$, is given by: $R/2 = 1-F_{n+1}\{[(n+1)(1-\rho)/(1+\rho)]^{0.5}\}$

Bivariate t-Copula and Ratio to Normal Copula





Tail Dependence R



Multivariate t-Copula

- Take m variates, and **u** a vector of m probability values (numbers in [0,1])
- Take **s** as the vector of univariate t-quantiles of **u** with n degrees of freedom, that is $s=F_n^{-1}(u)$ for each element of **s** and **u**.
- Take Σ as an m x m correlation matrix with determinant d.
- The m-dimensional t-copula has density:

$$c(\mathbf{u}; n, \Sigma) = K_m[\Pi_{i=1}^{m}(1+s_i^2/n)]^{(n+1)/2}(1+\mathbf{s}'\Sigma^{-1}\mathbf{s}/n)^{-(m+n)/2}$$

$$\text{where } K_m = \Gamma[(m+n)/2][\Gamma(n/2)]^{m-1}[\Gamma(½+n/2)]^{-m}d^{-1/2}.$$

- With Kendall's τ coefficient matrix T, the correlation matrix is $\Sigma = \sin(T\pi/2)$.
- C ~ {normal copula with same K}*{inverse gamma}$^{1/2}$

Loss Scenario Generation for t-Correlated Lines

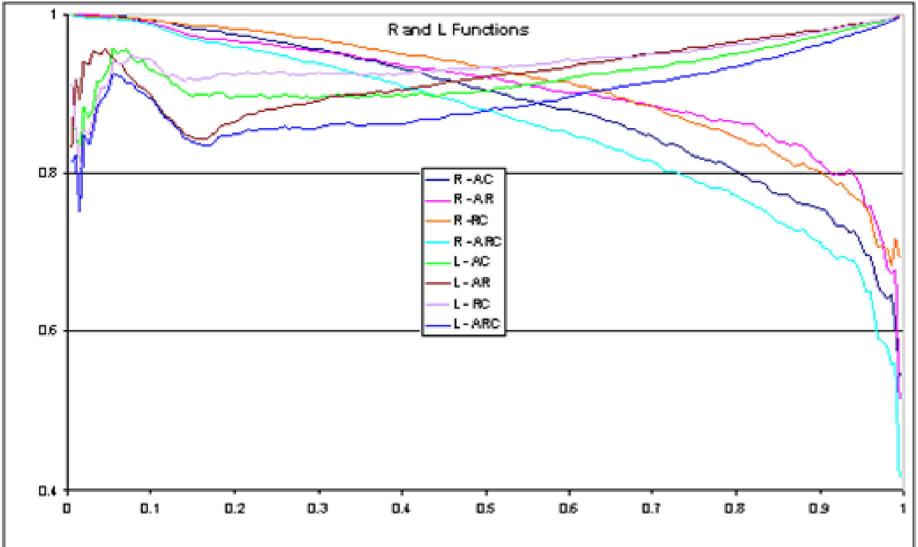
- Generate a multi-variate normal loss vector with the same correlation matrix.
- Divide each loss by $(y/n)^{0.5}$ where y is a number simulated from a chi-squared distribution with n degrees of freedom. This gives a t-distributed loss vector.
- Apply the t-distribution F_n to each loss to get the probability vector generated for the t-copula.
- The inverse severity distributions for each line can then be applied to get the by-line losses for the scenario.

Dependency Measures for Multiple Variates

- $\tau = 4E(C) - 1$, where E(C) is the expected value of the copula
- $\tau = 0$ for the independence case and $\tau = 1$ for perfect correlation
- For m-dimensions this is $\tau = [2^m E(C) - 1]/[2^{m-1} - 1]$.
- R(z)= Pr(U>z & V>z)/Pr(V>z).
  - Since Pr(V>z) = 1– z = Pr(U>z), U and V can be switched in the definition of R(z).
- A similar concept can be defined for the multivariate copula:
  - R(z)= Pr(U>z & V>z & W>z)/(1 – z) = Pr(U>z & V>z | W>z)
  - Because of the symmetry in the first equation, U, V, and W can be swapped around at will in the second equation.
  - This function provides a measure of the overall tail dependency of the three variates, and it can be generalized to higher dimensions.
- A similar tail dependency function can be defined for the left tail:
  - L(z) = Pr(U<z & V<z & W<z)/z = C(z,z,z)/z

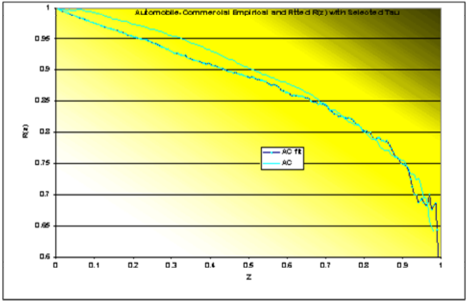
Sample Auto, Residential and Commercial Losses Hurricane Cat Model Data

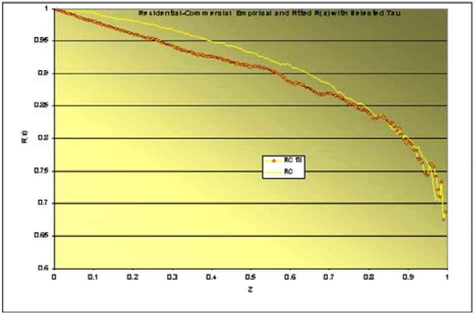
	AC	AR	RC	ARC
▪ τ	82.4%	84.4%	87.6%	84.8%

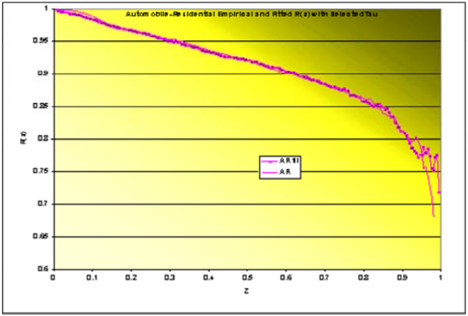


Fit to Right Side Only
Try n=20

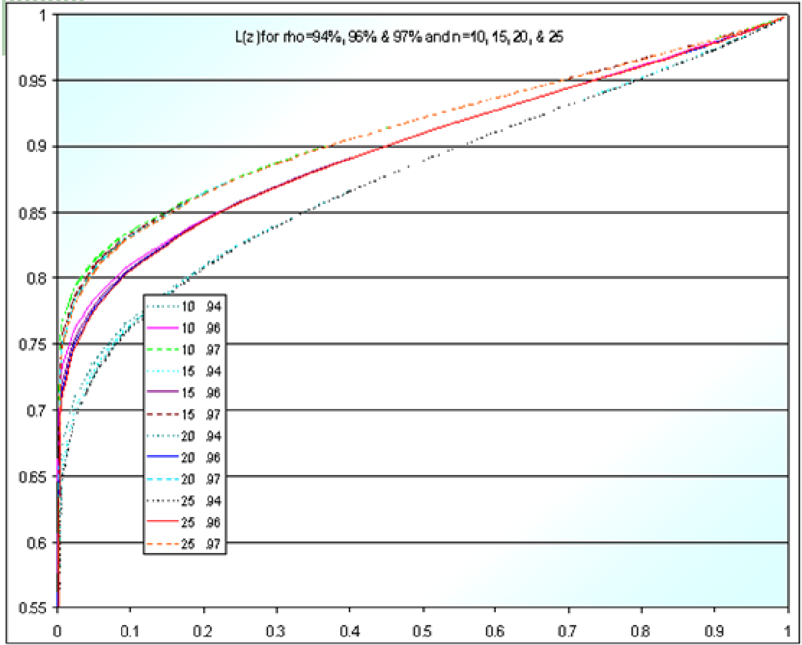
	AC	AR	RC
Sample ρ	.96	.97	.98
Selected ρ	.94	.97	.96







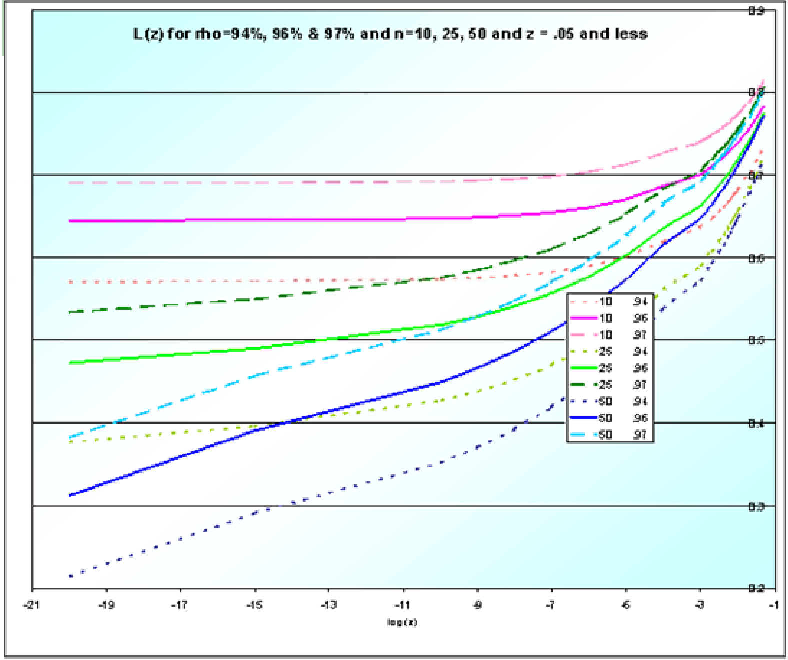
Effect of n on L (and so on R)



Guy Carpenter

Effect of n on L for Small z

L(z) still declining at $z=10^{-20}$ for n=50



Guy Carpenter

Compromising on n
Take n = 42

Target/Fit	94% (AC)	97% (AR)	96% (RC)
0.005	.54 / .61	.52 / .68	.693 / . 695
0.01	.61 / .64	.68 / .70	.718 / .715

Too heavy in extreme tail for AC and AR but close for RC

finis