A Linear Approximation To Copula Regression

Paul G. Ferrara
Tokio Marine

Rahul A. Parsa
Iowa State University

Bryce A. Weaver
James Madison University

14 Nov 2016

Special Thanks:

- This work was funded by:
  - The *Casualty Actuarial Society*
  - *The Actuarial Foundation*
  - *Committee on Knowledge Extension Research* (CKER) of the Society of Actuaries

- Special thanks to *Stuart A. Klugman*
  - for his help on this paper, and for sharing many prescient insights over the years.

Outline:

1. Review the *Parsa-Klugman* version of Copula Regression.
2. A linear approximation to Copula Regression.
3. Comparison of the linear approximation vs. Copula Regression estimates.
4. Convexity of Transmutation mappings.
5. New results:
   - General requirements for the convexity of *Transmutation maps.*
   - The 2-parameter Pareto Distribution
   - The Gamma Distribution
6. Next steps....

Parsa & Klugman (2011) *Copula Regression*

Parsa & Klugman describe the concept of Copula Regression under a *Multivariate Normal* Copula.

- i.e. the joint CDF of the variables $x_1, x_2, \ldots, x_{n-1}, y$ is:

$$F(x_1, x_2, \ldots, x_{n-1}, y) = G\Big(\Phi^{-1}[F_1(x_1)], \ldots, \Phi^{-1}[F_{n-1}(x_{n-1})], \Phi^{-1}[F_y(y)]\Big)$$

  - Where G is a *multivariate normal* cumulative dist. function(CDF).
  - y denotes the dependent variable.
  - Where F_y is the CDF of y
  - Where $F_1, F_2, \ldots, F_{n-1}$ are the CDFs of $x_1, x_2, \ldots, x_{n-1}$.
  - All variables $y, x_1, x_2, \ldots, x_{n-1}$, are assumed to be *continuous*

The *Parsa-Klugman* version of Copula Regression:

The copula regression *estimate* of Y given $\mathbf{x} = \{x_1, x_2, \ldots, x_{n-1}\}$, is:

- $\hat{y} = E_{f_{MVNC}}\left[Y \mid \mathbf{X} = \mathbf{x}\right]$
  - where the expected value, $E_{f_{MVNC}}$, is taken WRT the conditional density

$$f(y \mid x_1, x_2, \ldots, x_{n-1}) =$$

$$= \frac{1}{\sqrt{1-\vec{r}^T \cdot R_{n-1}^{-1} \cdot \vec{r}}} \cdot f_y(y) \cdot \exp\left\{ -\frac{1}{2}\left[\frac{\left(\Phi^{-1}[F_y(y)] - \vec{r}^T \cdot R_{n-1}^{-1} \cdot v^*\right)^2}{1-\vec{r}^T \cdot R_{n-1}^{-1} \cdot \vec{r}} - \left(\Phi^{-1}[F_y(y)]\right)^2 \right]\right\}$$

  - where $v^* = \{v_1, v_2, \ldots, v_{n-1}\}$ with $v_i = \Phi^{-1}[F_i(x_i)]$ for $i = \{1, 2, \ldots, n-1\}$ and $v_n = \Phi^{-1}[F_y(y)]$.
  - R_{n-1} is the correlation matrix of $x_1, x_2, \ldots, x_{n-1}$
  - $\vec{r} = (r_{y,x_1}, r_{y,x_2}, \ldots, r_{y,x_{n-1}})^T$

Comparison with other regression techniques:

- Similar to Ordinary Least Squares(OLS) regression, and GLMs:
  - The conditional mean of the response variable is some function of a *linear combination* of the covariates

$$\hat{y} = E[Y \mid \mathbf{X} = \mathbf{x}]$$

    - ... i.e. Identity link-function.

- Differs from OLS and GLMs:
  - Dependence between the independent variable and *each of* the dependent variables, is induced from a *MVN Copula*.
  - Each variable can be fitted to it's own *best-fitting* marginal distribution.
  - Moreover, *heavy-tailed* distributions may be used for the marginals.
  - No need to use distributions from the *Exponential Family*.

Motivation for: *A Linear Approximation To Copula Regression*, Variance 2015

a question posed to Dr. Parsa & Klugman (shortly after the debut of their *Copula Regression* paper) at the *Spring 2011 CAS meeting*!

Question?

Why don't you just do an OLS regression of v_n on $v_1, v_2, \ldots, v_{n-1}$ .....

.... and then *transform* the results back?

- where $v_i = \Phi^{-1}[F_i(x_i)]$ for $i = 1, 2, \ldots, n-1$
- and $v_n = \Phi^{-1}[F_y(y)]$

..... their initial reaction was, this cannot possibly work... *can it*?

Details: *proposed* linear approximation to Copula Regression:

1. Transform each of the n variables:
   - $U = \Phi^{-1}[F_y(y)]$ and $V_i = \Phi^{-1}[F_i(x_i)]$ for $i = \{1, 2, \ldots, n-1\}$
2. Perform an ordinary OLS of U on the V_i, to obtain $\hat{U}$:
   - $U = \beta_0 + \beta_1 \cdot V_1 + \ldots + \beta_{n-1} \cdot V_{n-1} + \epsilon$
   - where $\epsilon \propto N(0,1)$
3. Then transform the $\hat{U}$ back to the original scale, to obtain the estimate $\hat{Y}$:
   - $\hat{Y} = \left(F_y^{-1} \circ \Phi\right)(\hat{U})$

Some benefits of the approximation *would* be:

- easy to implement can be done in Excel.
- OLS is well understood.
- Transformations are common within OLS Regression (.... though, as we may see the repercussions may not always be considered...)

Initial investigation of the linear approximation:

Hence: Dr Parsa set out to confirm the initial scepticism of the approximation.

- Both copula regression, and the *linear approximation*, were fit to several datasets.
- ... the difference between the estimates (from the two models) was analyzed.

.... it turns out, that this *approximation* is, *often*, not that bad!

- the estimates from the *linear approximation* were *surprisingly close* to those from Copula Regression.
- Moreover, they seem to *consistently underestimate* (across the whole range of the independent variables $x_1, x_2, \ldots, x_{n-1}$) those from Copula Regression.

Comparison of estimates from Copula Regression and a linear approximation:

... so the question became, was this a coincidence?

Or, is there some *systematic bias* in the estimates from the approximate method, verses those from Copula Regression?

... as a new professor at Drake University, Dr. Parsa asked me to:

1. determine if this *bias* was just an artifact of the samples that he had examined?
2. (if not) if I could *prove* what conditions were causing this *systematic bias*?

Writing (regular) Copula Regression in terms of transformations:

Result 1:

If $F_y(y)$, and $F_i(x_i)$ for $i = \{1, 2, \ldots, n-1\}$ are continuous CDF's, corresponding to the RV's Y, $\vec{X}$, where $\vec{X} = \{X_1, X_2, \ldots, X_{n-1}\}$,

then:

$$E(Y|\,\vec{X}) = E\Big[(F_y^{-1} \circ \Phi)(U|\,\vec{V})\Big]$$

where

$U = \Phi^{-1}[F_y(y)]$ and $V_i = \Phi^{-1}[F_i(x_i)]$ for $i = \{1, 2, \ldots, n-1\}$

Note: $(F_y^{-1} \circ \Phi)(\,)$ can be viewed as a transformation of $U|\,\vec{V}$.

- Albeit, a non-trivial transformation . . . .
  - F_y^{-1} is the *quantile function* of the distribution of Y.
  - Φ is the *CDF* of the *Standard Normal* distribution.

Comparison: Copula Regression and the Linear Approximation:

So we have that:

- Estimates from *Copula Regression*:

$$E(Y|\,\vec{X}) \;=\; E\Big[(F_y^{-1} \circ \Phi)(U|\,\vec{V})\Big]$$

- The estimates from the *Linear Approximation* are:

$$\hat{Y} = (F_y^{-1} \circ \Phi)\Big(E(U|\,\vec{V})\Big)$$

So any bias can be ascribed to the *transformation* $(F_y^{-1} \circ \Phi)(\;)$...

- ... this has implications regarding the use of transformations, in general, within regression models...

Comparison of estimates from Copula Regression and a linear approximation:

... Seems like a natural candidate for *Jensen's Rule*:

Observation 1:

If:

1. $F_y(y)$, and $F_i(x_i)$ for $i = \{1, 2, \ldots, n-1\}$ are continuous CDF's, corresponding to the RV's Y, $\vec{X}$, where $\vec{X} = \{X_1, X_2, \ldots, X_{n-1}\}$, and:
2. and, the mapping $(F_y^{-1} \circ \Phi)(\cdot)$ is ***convex***,

Then:

$$E\left[(F_y^{-1} \circ \Phi)(U | \vec{V})\right] \geq (F_y^{-1} \circ \Phi)\left(E(U | \vec{V})\right)$$

where

$U = \Phi^{-1}[F_y(y)]$ and $V_i = \Phi^{-1}[F_i(x_i)]$ for $i = \{1, 2, \ldots, n-1\}$

Convexity of the transformation:

Hence, the *systematic* nature of the bias can be established ...
.... ***if*** it can be proven that $(F_y^{-1} \circ \Phi)(\cdot)$ is ***convex***.

- needs to be *convex* over the whole real line $\mathbb{R}$.

The mappings $(F_y^{-1} \circ \Phi)(\cdot)$ send the *percentiles* of Φ to the corresponding *percentile* of F_y. (coined *Transmutation mappings* by *Shaw et.al.*)

- were first investigated by *Cornish & Fisher* in 1937.
  - The origin of *Cornish-Fisher* (*C-F*) expansions
  - ... *approximate* method: estimates the quantiles of distributions, $F(x)$, from known moments.
- more recently studied in the *"Quantile Mechanics" (I, II, and III)*:
  - *Steinbrecher & Shaw* 2008 , *Shaw & Brickman* 2010, *Munir & Shaw* 2012
  - ... research pointed out (to authors) by *Vytaras Brazauskas*, from the University of Wisconsin.

Convexity of *Transmutation mappings*, cont....:

But the results of *Shaw et.al.* (*King's College*, London) do *not* help prove *convexity* of $(F_y^{-1} \circ \Phi)$:

- *Quantile Mechanics I, II, and III*, only *approximations* to $(F_y^{-1} \circ \Phi)(\cdot)$ are used.
- No (or very few) analytical proofs, and certainly not regarding *higher-order* properties such as convexity.

However, we need to prove that $(F_y^{-1} \circ \Phi)(\cdot)$ is convex, *analytically*:

- $\Phi(x)$, by itself, is equivalent to the (non-elementary) special function - the *Error function*.
- F_y, is often, also, a (non-elementary) special function....

  .. hence, dealing with a composition of two special functions.

In general, *no* (rigorous, analytical) proofs regarding *convexity* of $(F_y^{-1} \circ \Phi)$ exist in the literature.

General criterion: *convexity* of Transmutation maps:

Result 2:

Let:

- $f(x)$ be a continuous density, corresponding to $F(x)$, and:
- $\Phi(x)$ be the *CDF* of the *standard normal* distribution, and:
- $y(x) = (F^{-1} \circ \Phi)(x)$.

Then, the following (equivalent) conditions imply *convexity* of $y(x)$, for all x:

- $\dfrac{f'(y(x))}{f^2(y(x))} \leq \dfrac{\phi'(x)}{\phi^2(x)}$ for all x, $\left(\text{where } f'(y(x)) = \frac{d}{dz} f(z)\Big|_{z=y(x)} \right)$
- $\frac{d}{dx} \ln\left[f(y(x))\right] \leq \frac{d}{dx} \ln\left(\phi(x)\right)$ for all x.

Results for common loss distributions:

Result 3:

Let:

- $f(x)$ be a ***lognormal*** distribution, with parameters μ, σ
- $\Phi(x)$ be the *CDF* of the *Standard Normal* distribution

Then:

$$y(x) = (F^{-1} \circ \Phi)(x) \text{ is } \mathit{convex} \quad \text{for all } x, \text{ and all } \mu, \text{ and } \sigma$$

Result 4:

Let:

- $f(x)$ be a *two-parameter* ***Pareto*** *distribution* with parameters α, θ
- $\Phi(x)$ be the *CDF* of the *Standard Normal* distribution

Then:

$$y(x) = (F^{-1} \circ \Phi)(x) \text{ is } \mathit{convex} \quad \text{for all } x, \text{ and all } \alpha, \ \theta$$

The Gamma distribution:

Proving the convexity of $(F^{-1} \circ \Phi)(x)$ when $F(\cdot)$ the *Gamma distribution* (regularized incomplete gamma function), is *much more difficult*.

- The CDF of the Gamma is an *especially* intractable special function:
  - *Tricomi* fondly referred to it as "*the Cinderella of special functions*"
- .. can be represented in terms of various special functions:
  - *Confluent Hypergeometric function*, *Bessel functions*, etc..
- Related to a famous conjecture of *Ramanujan*'s (circa 1913):
  - ... that $\frac{1}{3} < \theta(n) < \frac{1}{2}$ (for any n) in the following equality $\frac{e^n}{2} = \sum_{k=0}^{n-1} \frac{n^k}{k!} + \theta(n) \frac{n^n}{n!}$
  - *Choi* (1994) proved that $1 - \theta(\alpha) = \left(\frac{e}{\alpha}\right)^{\alpha} \int_{\alpha}^{m(\alpha+1)} t^{\alpha} e^{-t} dt$

    where $m(\alpha+1)$ is the median of a *gamma* distribution with *shape* parameter $\alpha + 1$.

The Gamma distribution:

.. to make matters worse $(F^{-1} \circ \Phi)(x)$ involves the *quantile function* (inverse) of the *CDF* of the Gamma distribution..

$$F_{\alpha}^{-1}(z) = [-(z-1)\Gamma(\alpha+1)]^{\frac{1}{\alpha}} + \frac{\left([-(z-1)\Gamma(\alpha+1)]^{\frac{1}{\alpha}}\right)^2}{\alpha+1} + \frac{(3\alpha+5)\cdot\left([-(z-1)\Gamma(\alpha+1)]^{\frac{1}{\alpha}}\right)^3}{2(\alpha+1)^2(\alpha+2)} + O\left((z-1)^{\frac{4}{\alpha}}\right)$$

Result 5:

Let:

- $f(x)$ be a ***gamma*** *distribution* with shape parameter α, and scale parameter θ.
- $\Phi(x)$ be the *CDF* of the *Standard Normal* distribution.

Then:

$y(x) = (F^{-1} \circ \Phi)(x)$ is *convex* for all x, and all α, θ.

Systematic Bias:

... Back to the original question:

Is there some *systematic bias* in the estimates from the approximate method, verses those from Copula Regression?

Answer: Yes...

- If (all) the marginal distributions are modeled using one of the standard loss distributions (*Lognormal*, *Pareto*, or *Gamma*)

.... Further, this holds for any permissible parameter values of the *Lognormal*, *Pareto*, or *Gamma* distributions.

In this case, the estimates from the *Linear Approximation* to Copula Regression will *always*, at least, slightly underestimate the *true* values.

Thanks