A Linear Approximation to Copula Regression

by Paul G. Ferrara, Rahul A. Parsa, and Bryce A. Weaver

ABSTRACT

Recently, Parsa and Klugman (2011) proposed a generalization of ordinary least squares regression, which they called copula regression. Though theoretically appealing, implementation, especially calibration, of copula regression is generally more involved than for generalized linear models. In this paper a linear approximation to copula regression, for which implementation is similar to that for least squares regression, will be introduced. We proceed by investigating the connection between the proposed approximation to copula regression, and copula regression itself. In particular, we develop a set of criteria which ensure a predictable bias in the estimates from the linear approximation to copula regression.

1. Introduction

Over the past several decades, generalized linear models (GLMs) have become ubiquitous in actuarial science. Within the insurance industry, GLMs were first utilized in the property and casualty industry, due to its inherent need for custom modeling. However, GLMs and predictive modeling have recently found application in the life insurance industry as well. GLMs have enjoyed this popularity largely due to their ability to overcome the limitations of ordinary least squares regression (OLS) when applied to data which exhibit non-linear relationships, and non-normal distributions. In their traditional form, GLMs require that the dependent variable have a distribution from the *exponential family* of distributions. This requirement primarily ensures the existence of a desirable relationship between the variance and mean of dependent variable. In most situations this restriction is at most a minor inconvenience; however, there are circumstances in which data can be best described by distributions outside of the exponential family of distributions. In such cases, it can be beneficial to have access to an even more general class of models. In response to this need, Parsa and Klugman (2011) proposed a regression modeling framework which exploits the flexibility of copula techniques.

Not long after the introduction of copula regression, however, it was noticed that the formulation of copula regression as a conditional expected value could be approximated by an OLS regression on appropriately transformed data. Such an approximation would have the advantage of ease of implementation, using almost any statistical software or even Microsoft Excel. Moreover, upon investigating this linear approximation to copula regression, Parsa and Klugman (2011) noticed that the approximation often produced estimates which were quite close to, yet consistently underestimated those from exact copula regression. Further, there seemed to be a systematic deviation between the estimates from each method. These observations motivated the current research. This paper will proceed as follows: In section 2 we briefly describe copulas, and copula regression. In section 3 we describe the linear approximation to copula regression, which is the main subject of this research. In section 4 we investigate sufficient conditions which ensure a predictable pattern of bias in the estimates from the linear approximation to copula regression. In section 5 we introduce *transmutation mappings,* which play an integral role within the linear approximation to copula regression, and also within the theory of copulas in general. In section 6, we investigate the most commonly used loss distributions to see if they satisfy the sufficient conditions from section 4, and hence lead to predictably biased estimates (with respect to full copula regression) from the linear approximation to copula regression.

2. Copula regression

Copula methods are essentially just a particular way to construct multivariate distributions. Copula functions themselves are simply multivariate distributions whose domain has uniform marginal distributions. As described in Parsa and Klugman (2011), a copula model can be constructed for an n-dimensional set of data by first fitting n marginal distributions to the data $F_1(X_1), F_2(X_2), \ldots, F_n(X_n)$. Then the joint-distribution, $F_{x_1,\ldots,x_n}(X_1, X_2, \ldots, X_n)$, can be created by applying a copula function to $F_1(X_1), F_2(X_2), \ldots, F_n(X_n)$, specifically $F_{x_1,\ldots,x_n}(X_1, X_2, \ldots, X_n) = C[F_1(X_1), F_2(X_2), \ldots, F_n(X_n)]$. The copula function C induces a correlation structure between the variables $X_1, X_2, \ldots, X_n$. By construction, this correlation structure will be independent of the marginal distributions of the data. In order to ensure that the resulting multivariate function is in fact a distribution function, only certain functions can be used as copula functions. However, Sklar's theorem guarantees that, for any set of random variables (*RVs*) $X_1, X_2, \ldots, X_n$ with associated marginal distributions $F_1(X_1), F_2(X_2), \ldots, F_n(X_n)$, a copula function, C, exists s.t.; $F_{x_1,\ldots,x_n}(X_1, X_2, \ldots, X_n) = C[F_1(X_1), F_2(X_2), \ldots, F_n(X_n)]$. Moreover, if the marginal distributions $F_i(X_i)$ for $i \in \{1, 2, \ldots, n\}$ are continuous, then there exists a *unique* copula function, defined

on $F_1(X_1)$, $F_2(X_2)$, . . . , $F_n(X_n)$, such that $F_{x_1,...,x_n}(X_1, X_2, . . . , X_n) = C[F_1(X_1), F_2(X_2), . . . , F_n(X_n)]$. Further, in the case of continuous marginal distributions, the copula function allows the multivariate dependence structure to be separated from the marginal distributions of the data. At this point, it is important to point out that, although the mathematical theory of copulas is very rich, and many powerful results have been proven, not every multivariate distribution can be naturally, or "usefully", modeled using copulas. To quote Paul Embrechts: "copulas form a most useful concept for a lot of applied modeling, they do not yield, however, a panacea for the construction of useful and well-understood multivariate density functions" Embrechts (2009). While there are many copula models, most are only bivariate, and hence do not allow for fully independent correlation between the response variable and each of the covariates. For this reason, Parsa and Klugman assert that the multivariate normal and *t*-copula are the most useful for the purposes of copula regression. In basic OLS regression, the distribution of *Y* given the covariates is assumed to be normal, and the predicted values are specified by $E[Y|X_1, . . . , X_{n-1}]$. The predicted values, under copula regression, are defined analogously.

Definition. Given *N* observations of the RVs $X_1, . . . , X_{n-1}$, *Y,* copula regression, as defined by Parsa and Klugman (2011), is the process of estimating the observed values $\{y_i | i \in 1, 2, . . . , N\}$ of the RV *Y,* based on the corresponding values of the RVs $X_1, . . . , X_{n-1}$, namely $\{x_{1,i}, . . . , x_{n-1,i} | i \in 1, 2, . . . , N\}$.

$$\hat{y}_i = E[Y|X_1 = x_{1,i}, . . . , X_{n-1} = x_{n-1,i}], \qquad (1)$$

where the conditional expected value is computed with respect to the conditional density of the RVs $X_1, . . . , X_{n-1}$, *Y* induced from a multivariate copula function *C,* i.e.,

$$F_{x_1,...,x_n}(X_1, X_2, . . . , X_n, Y)$$

$$= C[F_1(X_1), F_2(X_2), . . . , F_n(X_{n-1}), F_Y(Y)] \qquad (2)$$

In this paper, we concentrate on the case where the copula function *C* is the multivariate normal copula, and reserve investigation of the *t*-copula for subsequent research. We now review the basics of copula regression, under the multivariate normal copula. For a more complete treatment, the interested reader is referred to Parsa and Klugman (2011). The distribution function induced by the multivariate normal copula has the form:

$$F_{x_1,...,x_{n-1},y}(X_1, . . . , X_{n-1}, Y)$$

$$= G\left(\Phi^{-1}[F_1(X_1)], . . . , \Phi^{-1}[F_{n-1}(X_{n-1})], \Phi^{-1}[F_y(Y)]\right) \quad (3)$$

where $F_1(X_1), . . . , F_{n-1}(X_{n-1})$, $F_y(Y)$ are the marginal distributions of the RV's $X_1, . . . , X_{n-1}$, *Y,* and $G(\cdot)$ is the multivariate normal distribution. In what follows we assume that all marginal distributions are continuous. Under this assumption, the density function induced by the multivariate normal copula can be written as follows:

$$f(X_1, X_2, . . . , X_{n-1}, Y) = f_1(X_1) . . . f_n(X_{n-1})f_y(Y)$$

$$\cdot \exp\left\{-\frac{\vec{v}^T(R^{-1} - I)\vec{v}}{2}\right\} \times |R|^{-\frac{1}{2}} \quad (4)$$

where:

i) $\vec{v} = (v_1, . . . , v_{n-1}, v_n)$, and $v_i = \Phi^{-1}[F_i(X_i)]$ for $i = \{1, 2, . . . , n - 1\}$, and $v_n = \Phi^{-1}[F_y(Y)]$.

ii)   I is the Identity matrix.

iii) $R = \begin{vmatrix} R_{n-1} & \vec{r} \\ \vec{r}^T & 1 \end{vmatrix}$ is the copula correlation matrix.

where

a) $R_{n-1} = \begin{vmatrix} 1 & r_{x_1,x_2} & \cdots & r_{x_1,x_{n-1}} \\ r_{x_2,x_1} & 1 & \cdots & r_{x_2,x_{n-1}} \\ \vdots & \cdots & \cdots & \vdots \\ \vdots & & \ddots & r_{x_{n-2},x_{n-1}} \\ r_{x_{n-1},x_1} & \cdots & r_{x_{n-1},x_{n-2}} & 1 \end{vmatrix}$, and

b) $\vec{r} = (r_{x_1,y}, r_{x_2,y}, \ldots, r_{x_{n-1},y})^T$, and:

c) r_{x_i,x_j} is the correlation between v_i and v_j, and $r_{x_i,y}$ is the correlation between v_i and v_n, for $i, j < n$.

Correspondingly, it is shown in Clement and Reilly (1999) that the conditional density of Y, given X_1, $X_2, \ldots, X_{n-1}$ has the following form:

$$f(Y|X_1, X_2, \ldots, X_{n-1}) = \frac{f_y(Y)}{\sqrt{1 - \vec{r}^T \cdot R_{n-1}^{-1} \cdot \vec{r}}} \times$$

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

$$\exp\left\{\frac{1}{2}(\Phi^{-1}[F_y(Y)])^2\right\} \qquad (5)$$

where $v^* = \{v_1, v_2, \ldots, v_{n-1}\}$ with $v_i = \Phi^{-1}[F_i(X_i)]$, and $\vec{r}$, R_{n-1}, and R are defined as above. Under the multivariate normal copula, $f(Y|X_1, X_2, \ldots, X_{n-1})$ specifies the predicted values $E[Y|X_1, \ldots, X_{n-1}]$, up to specification of the marginal distributions $F_1(X_1)$, $F_2(X_2), \ldots, F_{n-1}(X_{n-1})$, $F_Y(Y)$.

Ideally, to parameterize the copula regression model based on empirical data, maximum likelihood estimation is performed using the density function $f(X_1, X_2, \ldots, X_{n-1}, Y)$, above. If $n = 8$, and each marginal distribution has two parameters, it is necessary to estimate $8(2) + 8(8 - 1)/2 = 44$ parameters. Alternatively, one can use the marginal data to fit each of the marginal distributions $F_1(X_1), \ldots, F_{n-1}(X_{n-1})$, $F_y(Y)$, and then a second optimization can be performed to estimate the $n(n - 2)/2$ correlations, within the MVN copula. As pointed out in Parsa and Klugman (2011), this alternative will produce suboptimal results. The relative complexity of fitting the $n(m) + n(n - 1)/2$ parameters of a full multivariate normal copula regression model with n marginals, each with m parameters, is one of the motivations for the linear approximation to copula regression, which is the topic this paper. For more details on the parametrization of copula regression models, the reader is referred to Parsa and Klugman (2011).

3. A linear approximation to copula regression

Now that copula regression has been described, we introduce what the authors have dubbed the linear approximation to copula regression. The possibility of such an approximation to copula regression arose organically in response to a question posed by a practicing property and casualty (P&C) actuary at the 2011 CAS Spring meeting. During the question and answer section of their presentation on copula regression, Parsa and Klugman were asked the following question: "For a given set of *RVs* $X_1, \ldots, X_{n-1}$, Y, since the distribution induced by the multivariate normal copula is essentially the multivariate normal distribution applied to the *transformed RVs* $\Phi^{-1}[F_1(X_1)]$, $\Phi^{-1}[F_2(X_2)], \ldots, \Phi^{-1}[F_{n-1}(X_{n-1})]$, $\Phi^{-1}[F_Y(Y)]$, and since a fully probabilistic version of multivariate linear regression assumes that the *RVs* follow a multivariate normal distribution, *how does copula regression, under a multivariate normal copula, differ from simply applying OLS regression to the transformed RVs*:

$$\Phi^{-1}[F_1(X_1)], \Phi^{-1}[F_2(X_2)], \ldots,$$

$$\Phi^{-1}[F_{n-1}(X_{n-1})], \Phi^{-1}[F_Y(Y)]"$$

In other words, if Y is considered the response variable, what is the difference in the estimates from the following models?

Model 1: Perform OLS regression in the U, $\vec{V}$ space, and then transform back to the Y, $\vec{X}$ space:

(a) Transform each of the n variables $X_1, \ldots, X_{n-1}$, Y:

$$U = \Phi^{-1}[F_y(y)] \text{ and } V_i = \Phi^{-1}[F_i(x_i)]$$

$$\text{for } i = \{1, 2, \ldots, n-1\}$$

(b) Perform an ordinary OLS of U on the V_i, to obtain $\hat{U}$:

$$U = \beta_0 + \beta_1 \cdot V_1 + \cdots + \beta_{n-1} \cdot V_{n-1} + \varepsilon$$

$$\text{where } \varepsilon \propto N(0, 1)$$

(c) Then transform the $\hat{U}$ back to the original scale: $\hat{Y} = F_y^{-1}(\Phi(\hat{U}))$ to obtain the estimates

$$\hat{y}_i = F_y^{-1}\left(\Phi\left(\hat{u}_i \mid V_1 = v_{1,i}, \ldots, V_{n-1} = v_{n-1,i}\right)\right).$$

Model 2: Perform copula regression in the Y, $\vec{X}$ space:

$$\hat{y}_i = E[Y \mid X_1 = x_{1,i}, \ldots, X_{n-1} = x_{n-1,i}], \text{ where}$$

$f(Y \mid X_1, X_2, \ldots, X_{n-1})$ is defined as in equation (5).

Parsa and Klugman were immediately struck by the simplistic beauty of this reductive argument, and intrigued by the possibility that this approach could actually produce a close approximation to copula regression. Moreover, they soon realized that this question transcended copula regression, and actually addressed the larger issue of the effect of transformations of variables within regression modeling.

4. Sufficient conditions for bias

It seems likely that an understanding of the limitations of variable transformations, as a means to coerce data to fit the restrictive normal distribution assumption of linear regression, informed the development of GLMs. A simple example illustrates the difference between these two approaches: Suppose that it is known that the dependent variable Y follows a lognormal distribution. In introductory applied statistic classes, students are taught that in this case they should apply a *log* transformation to Y, and then perform OLS. The underlying model is that $\log(Y) = aX + b + \varepsilon$ where $\varepsilon \sim N(0, \sigma^2)$. In other words, the assumption is that $\log(Y) \mid X$ has a normal distribution with mean $aX + b$, and constant variance. Then, one simply applies the inverse of the *log* transform to obtain the predicted values $\hat{Y} = E(Y \mid X) = \exp(aX + b)$. However, if Y in fact follows a lognormal distribution, then $E(Y \mid X) = \exp(aX + b + 1/2\sigma^2)$. Hence, by applying the *log* transformation and assuming a normal distribution, the resulting estimate differs from the theoretically correct estimate by a factor of $\exp(1/2\sigma^2)$. The disparity

between these two estimates of $E(Y \mid X)$ can be viewed as representing a bias in the estimates from the model involving transformations. Although it may escape consideration, such a disparity arises whenever transformations of variables are used within OLS. In light of this observation, it can be seen that the question at hand is fundamentally a question of the bias that is induced by applying the specific transformation $\Phi^{-1}[F_i(\cdot)]$ within OLS regression. Further, due to the general nature of the transformations $\Phi^{-1}[F_i(\cdot)]$, an investigation of the bias between copula regression and the linear approximation to copula regression may shed light on the bias induced by the use of transformations within OLS in general. As a first step towards investigating this bias, the linear approximation to copula regression was computed for the examples provided in Parsa and Klugman (2011). This resulted in several interesting observations. As anticipated by the motivating question, it was noticed that the estimates from the linear approximation to copula regression were in fact often very close to the estimates based on copula regression. Moreover, a consistent relationship between the two sets of estimates was observed. Specifically, the estimates from the linear approximation to copula regression were noticed to be consistently slightly lower than those from copula regression. This piqued the interest of the authors, making further investigation irresistible. The following results pave the way toward a more quantitative understanding of the relationship between the estimates from copula regression and the linear approximation to copula regression. Lemma 4.1 provides a connection between the copula regression of Y on $\vec{X}$, and the transformed variables U, $\vec{V}$. In what follows, the authors sometimes use the slightly more suggestive notation $(F_y^{-1} \circ \Phi)(x)$, in place of $F_y^{-1}(\Phi(x))$.

Lemma 4.1

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

$$E(Y \mid \vec{X}) = E\left[(F_y^{-1} \circ \Phi)(U \mid \vec{V})\right] \qquad (6)$$

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

Proof. Let U and $\vec{V}$ be defined as $U = (\Phi^{-1} \circ F_y)(Y)$ and $\vec{V} = (\Phi^{-1} \circ F_i)(\vec{X})$, where $\Phi^{-1} \circ F_i$ is applied component-wise to $\vec{X}$. Let $f(U|\vec{V})$ be the conditional multivariate normal distribution of U given $\vec{V}$, then:

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

$$= \int_{u=-\infty}^{u=\infty} \left(F_y^{-1} \circ \Phi\right)(U) \cdot f\left(U \big| \vec{V} = \vec{v}\right) dU$$

$$= \int_{u=-\infty}^{u=\infty} Y \cdot f\left(U \big| \vec{V} = \vec{v}\right) dU$$

$$= \int_{y=h^{-1}(-\infty)}^{y=h^{-1}(\infty)} Y \cdot f\left(Y \big| \vec{V} = \vec{v}\right) \left|\frac{dY}{dU}\right| dU$$

$$= \int_{y=0}^{y=\infty} Y \cdot f\left(Y \big| \vec{X} = \vec{x}\right) dY = E\left(Y \big| \vec{X}\right)$$

where $h(y) = (\Phi^{-1} \circ F_y)(y)$. The 2nd to last equality holds since $V_i = v_i \rightarrow \Phi^{-1}[F_i(X_i)] = \Phi^{-1}[F_i(x_i)] \rightarrow X_i = x_i$, and since $h'(y) > 0$ for all y by the inverse function theorem, and the last equality holds since $h(0) = -\infty$, and $h(\infty) = \infty$.

Next, Theorem 4.2 quantifies when the estimates from the linear approximation to copula regression will underestimate those from copula regression, by providing a sufficient condition for the under-estimation.

Theorem 4.2 *If the CDF $F_y(Y)$ is continuous, and the mapping $g(U) = (F_y^{-1} \circ \Phi)(U)$ is convex, then:*

$$E\left(Y \big| \vec{X}\right) \geq F_y^{-1}\left(\Phi\left(\hat{U}\right)\right) \tag{7}$$

where $E[Y|\vec{X}]$ is the conditional expectation used to define copula regression, and $F_y^{-1}(\Phi(\hat{U}))$ defines the linear approximation to copula regression.

Proof. Let $F_y(Y)$, and $F_i(X_i)$ for $i \in \{1, 2, \ldots, n-1\}$ be the continuous CDF's corresponding to the RV's Y, and $\vec{X}$, respectively, where $\vec{X} = \{X_1, X_2, \ldots, X_{n-1}\}$. Let U and $\vec{V}$ be defined as $U = (\Phi^{-1} \circ F_y)(Y)$ and $\vec{V} = (\Phi^{-1} \circ F_i)(\vec{X})$, where $\Phi^{-1} \circ F_i$ is applied

component-wise to $\vec{X}$. If $E[Y|\vec{X}]$ is the conditional expected value with respect to the density induced by the multivariate copula given in equation (5), then by Lemma 4.1 we have that: $E(Y|\vec{X}) = E[(F_y^{-1} \circ \Phi)(U|\vec{V})]$, and since $g(U) = (F_y^{-1} \circ \Phi)(U)$ is convex, Jensen's inequality gives that:

$$E\left[\left(F_y^{-1} \circ \Phi\right)\left(U \big| \vec{V}\right)\right] \geq \left(F_y^{-1} \circ \Phi\right)\left(E\left[U \big| \vec{V}\right]\right) \tag{8}$$

Further, since $F_y^{-1}(\Phi(\hat{U}))$ is the linear approximation to copula regression, $\hat{U}$ is obtained from the OLS regression of U on the V_i. So, by definition $\hat{U} = E[U|\vec{V}]$, we have that:

$$F_y^{-1}\left(\Phi\left(\hat{U}\right)\right) = \left(F_y^{-1} \circ \Phi\right)\left(E\left[U \big| \vec{V}\right]\right) \tag{9}$$

Hence, in conclusion, $E[Y|\vec{X}] \geq F_y^{-1}(\Phi(\hat{U}))$.

So, the question is, for what *CDFs* F_y is the mapping $(F_y^{-1} \circ \Phi)(\cdot)$ *convex*? Armed with Theorem 4.2, we now know that convexity of the mapping $(F_y^{-1} \circ \Phi)(\cdot)$ is sufficient to ensure that the estimates from the linear approximation to copula regression will be bounded above, by the corresponding estimates from exact copula regression. We now present a set of criteria which ensure the convexity of the mappings $(F_y^{-1} \circ \Phi)(\cdot)$, and hence help quantify the bias in the estimates from the linear approximation to copula regression.

Lemma 4.3 *If $F_y(Y)$ is a continuous CDF, and $y(x) = (F_y^{-1} \circ \Phi)(x)$, then $y(x)$ is convex, for all x IFF:*

$$\frac{\phi'(x)}{\phi^2(x)} \geq \frac{f_y'(y(x))}{f_y^2(y(x))} \tag{10}$$

for all x where $f_y'(y(x))$ denotes the derivative of $f_y(y(x))$ WRT y, evaluated $y(x)$, and $\Phi(x)$ is the standard normal CDF, and correspondingly, $\phi(x)$ is the standard normal density.

Proof. Since, $y(x) = (F_y^{-1} \circ \Phi)(x)$, we have that $F_y(y(x)) = \Phi(x)$. Taking the derivative of both sides WRT x, we have that: $f(y(x)) \cdot \dfrac{dy(x)}{dx} = \phi(x)$, which implies:

$$\frac{dy(x)}{dx} = \frac{\phi(x)}{f(y(x))} \tag{11}$$

Once again, taking the derivative of both sides WRT *x,* re-arranging terms, and substituting in the equality from equation (11) for $\frac{dy(x)}{dx}$, we have:

$$\frac{d^2 y(x)}{dx^2} = \frac{\phi'(x) - f'(y(x))\left(\frac{\phi(x)}{f(y(x))}\right)^2}{f(y(x))}. \text{ Hence, we}$$

have:

$$\frac{d^2 y(x)}{dx^2} \geq 0 \Leftrightarrow \frac{\phi'(x)}{(\phi(x))^2} \geq \frac{f'(y(x))}{(f(y(x)))^2} \tag{12}$$

The next corollary gives an equivalent condition for convexity of the mappings $(F_y^{-1} \circ \Phi)(\cdot)$, and is an immediate consequence of Lemma 4.3.

Corollary 4.4 If $F_y(Y)$ is a continuous CDF, and $y(x) = (F_y^{-1} \circ \Phi)(x)$, then $y(x)$ is *convex*, for all *x* IFF:

$$\frac{d}{dx}\log[\phi(x)] \geq \frac{d}{dx}\log[f(y(x))] \tag{13}$$

for all *x* where $f_y'(y(x))$ denotes the derivative of $f_y(y)$ WRT *y,* evaluated at $y(x)$, and $\Phi(x)$ is the standard normal CDF, and correspondingly, $\phi(x)$ is the standard normal density.

Proof. Since $\phi(x) \geq 0$ we have that: $\frac{\phi'(x)}{\phi^2(x)} \geq$
$\frac{f_y'(y(x))}{f_y^2(y(x))} \Leftrightarrow \frac{\phi'(x)}{\phi(x)} \geq \frac{f_y'(y(x))}{f_y(y(x))} \frac{\phi(x)}{f_y(y(x))} \Leftrightarrow$
$\Leftrightarrow \frac{d}{dx}\log[\phi(x)] \geq \frac{d}{dx}\log[f(y(x))]$, where the last inequality follows, since:

$$\frac{f_y'(y(x))}{f_y(y(x))} \frac{\phi(x)}{f_y(y(x))} = \frac{d\log[f(y(x))]}{dy(x)} \cdot \frac{dy(x)}{dx}$$

$$= \frac{d}{dx}\log[f(y(x))]$$

which follows from equation (12). So, in summary, if $y(x) = (F_y^{-1} \circ \Phi)(x)$, where $F_y(Y)$ is a continuous CDF, and $\Phi(x)$ is the standard normal CDF, then either of the following equivalent conditions imply the convexity of $y(x)$, for all *x*:

1. $\dfrac{\phi'(x)}{\phi^2(x)} \geq \dfrac{f_y'(y(x))}{f_y^2(y(x))}$, or

2. $\dfrac{d}{dx}\log[\phi(x)] \geq \dfrac{d}{dx}\log[f(y(x))]$

Before investigating the convexity of the mappings $y(x) = (F_y^{-1} \circ \Phi)(x)$, when *F* is a common loss distribution, we pause to consider the interpretation of the preceding results. We focus on the 2nd condition, since it appears to be more amenable to interpretation. By integrating from $x = 0$ to $x = x'$, and recalling that $y(x)$ sends the percentiles of $\phi(x)$ to the corresponding percentiles of $f(y)$, and in particular, that $y(0)$ is the median of $f(y)$, the 2nd condition implies:

$$f_y(y(x')) \leq f_y(y_m^f)e^{-\frac{x'^2}{2}} \qquad \text{for all } x' > 0. \tag{14}$$

Similarly,

$$f_y(y(x')) \geq f_y(y_m^f)e^{-\frac{x'^2}{2}} \qquad \text{for all } x' < 0, \tag{15}$$

where y_m^f denotes the median of $f(y)$. The first condition shows that $f_y(y(x))$, as a function of *x,* decays, at least, proportionally to $e^{-\frac{x^2}{2}}$, as *x* grows large. Regarding the 2nd condition, corresponding to the left-hand tail, we consider the case when *F* has positive support. The condition given in equation (15) is surely satisfied if $f(y)$ is monotonically decreasing, as is the case for the exponential distribution. However, even if $f(y)$ is uni-modal, with $f(0) = 0$, as is the case for the Gamma distribution with scale parameter $\alpha > 1$, the 2nd condition requires that, if $y(x) = c \approx 0$ $(c > 0)$, then the corresponding $x = y^{-1}(c) < 0$ is sufficiently small that $f_y(y_m^f) \exp\left[-\frac{(y^{-1}(c))^2}{2}\right]$ is, nonetheless, less than $f_y(c)$.

5. Transmutation mappings

In the previous section it was seen that convexity of the mappings; $(F_1^{-1} \circ \Phi)(\cdot), \ldots, (F_{n-1}^{-1} \circ \Phi)(\cdot)$, $(F_Y^{-1} \circ \Phi)(\cdot)$ ensures that the linear approximation to copula regression will underestimate exact copula regression. In fact, the investigation of the mappings $(F^{-1} \circ \Phi)(\cdot)$ dates back at least to 1937 when Cornish-Fisher (C-F) formed polynomial expansions to approximate the quantiles of a given non-normal distribution function, F, in terms of the quantiles of the standard normal distribution. More specifically, if Y is a RV with distribution F_y, and X_α is the αth quantile of the standard normal distribution, then the αth quantile of Y can be approximated by:

$$Y_\alpha = m + \sigma \left(X_\alpha + \frac{1}{6} \frac{\kappa_3}{\sigma_3}(X_\alpha^2 - 1) + \frac{1}{24}\frac{\kappa_4}{\sigma_4}(X_\alpha^3 - 3X_\alpha) \right.$$

$$\left. - \frac{1}{36}\left(\frac{\kappa_3}{\sigma_3}\right)^2 (2X_\alpha^3 - 5X_\alpha) + \cdots \right) \quad (16)$$

where m, σ and κ_i denote the mean, standard deviation, and i^{th} order cumulant of the RV Y. It is important to note that C-F expansions are only approximations. Moreover, there are problems with C-F expansions, including the fact that introduction of additional terms does not always lead to a more accurate approximation. In fact, the likelihood of negative density values increases as higher-order terms are added to the series. As a result, C-F expansions are not useful for investigating the convexity of $(F_y^{-1} \circ \Phi)(\cdot)$. However, during the 2013 Actuarial Research Conference it was pointed out by Vytaras Brazauskas, from the University of Wisconsin at Milwaukee, that the mappings $(F_y^{-1} \circ \Phi)(\cdot)$ were recently studied by Shaw and *Buckley* (2007), and later by Steinbrecher and Shaw (2008), Shaw and Brickman (2010), and Munir and Shaw (2012). Shaw and Buckley (2007) dubbed the mappings $(F_y^{-1} \circ \Phi)(\cdot)$ transmutation mappings, and noted that these mappings, essentially, turn samples from one distribution (Φ), into samples from another (F_y). Hence, a practical and precise representation of trans-

mutation mappings has the potential to produce very efficient sampling algorithms, via the leveraging of existing samples to create samples from the desired distribution(s). In addition to having obvious application to sampling theory, especially copula-based sampling, Shaw and Buckley also note that transmutation mappings have utility within hypercube-filling quasi-Monte-Carlo (QMC) methods.

More relevant to our purposes, Shaw and Buckley claim that, prior to their research, there had been few, or no, analytical results published on transmutation maps, outside the asymptotic domain. Shaw and Buckley ended this drought through an investigation of the differential equations that transmutation mappings obey. They then used these results to form new transmutation map expansions, based on the power series solutions to these non-linear ODEs. Though the results of Shaw and Buckley represent a step forward in the analytical analysis of transmutation mappings, and are based on differential equations, their results are not, however, sufficiently strong enough to quantify properties of transmutation mappings, such as convexity. In fact, to the authors knowledge, there has been no analytical investigation into the higher-order mathematical properties, such as convexity, of transmutation mapping. This statement is consistent with those of Shaw and Buckley, who go on to, quite rightly, point out that the dearth of such research is likely due to the fact that many common statistical distributions, and their inverses, including the ubiquitous normal distribution, involve rather intractable mathematical special functions. As a result, an attempt to investigate the convexity of transmutation mappings is tantamount, in many cases, to proving results about the composition of two mathematical special functions, both of which may be very intractable, on their own. This is especially the case when one of the distributions is the Gamma distribution, or the *regularized incomplete gamma,* as it is known within mathematical physics. Regardless, the final section of this paper contains a proof of the convexity of this particular transmutation mapping.

6. Results under specific loss distributions

We now investigate the convexity of transmutation mappings when the distribution F is among those commonly used for severity, or size-of-loss, modeling. The authors suspect that the results of this section will be useful, even outside the statistical community. In particular, we feel it is likely that researchers in finance and even applied mathematics will find interest in these results. We first present an example where convexity of the transmutation mapping, almost trivially, follows. Since each distribution F under consideration is a loss, or severity, distribution, each distribution F has support on the positive reals.

Proposition 6.1 *(lognormal distribution) If $F_y(y)$ is a lognormal distribution, with parameters (μ,σ), and $\Phi(x)$ $(\phi(x))$ is the CDF (density) of the standard normal distribution, then $y(x) = (F_y^{-1} \circ \Phi)(x)$ is convex, for all x.*

Proof. Since $(F_y^{-1} \circ \Phi)(x)$, and its inverse $(\Phi^{-1} \circ F_y)(y)$ are both increasing functions, the convexity of $(F_y^{-1} \circ \Phi)(x)$ is equivalent to the concavity of $(\Phi^{-1} \circ F_y)(y)$. Hence, it suffices to show that $(\Phi^{-1} \circ F_y)(y)$ is concave. Since $Y \sim LN(\mu, \sigma)$, we have that:

$$F_y(y) = \Phi\left(\frac{\ln(y) - \mu}{\sigma}\right)$$

Hence: $(\Phi^{-1} \circ F_y)(y) = \Phi^{-1}\left[\Phi\left(\frac{\ln(y) - \mu}{\sigma}\right)\right] = \frac{\ln(y) - \mu}{\sigma}$

and so, for all y:

$$\frac{d^2}{dy^2}(\Phi^{-1} \circ F_y)(y) = -\frac{1}{\sigma y^2} \leq 0 \qquad (17)$$

Next, we consider another distribution that is commonly used for severity, or loss-size, modeling: the Pareto distribution. In particular, we consider the Pareto Type II distribution, or Lomax distribution, to which it is sometimes referred, with CDF parameterized as follows, for $y > 0$, and $\alpha, \theta > 0$:

$$F(y) = 1 - \left(\frac{\theta}{y + \theta}\right)^\alpha \qquad (18)$$

Unlike the lognormal case, convexity of the transmutation mapping is not as easily verified when F is a Pareto distribution.

Proposition 6.2 *(Pareto distribution) If $F_y(y)$ is a Pareto distribution, with shape parameter α, and scale parameter θ, and $\Phi(x)$ $(\phi(x))$ is the CDF (density) of the standard normal distribution, then $y(x) = (F_y^{-1} \circ \Phi)(x)$ is convex, for all x.*

Proof. First, consider the case where $x < 0$. Note that, since $\phi(x)$ is the normal density, the sufficient condition for convexity in Corollary 4.4, namely, equation (13), can be written: $-x \geq \frac{d}{dx}\log[f(y(x))]$, or:

$$-x - \frac{d}{dx}\log[f(y(x))] \geq 0 \qquad (19)$$

However, since $f(y)$ is monotonically *decreasing,* for all y, and $y(x)$, and $\log(x)$ are both *increasing,* we have that: $\frac{d}{dx}\log[f(y(x))] < 0$, for all y, and hence equation (19) is satisfied, for $x < 0$. Now, consider the case where $x > 0$. Since $F_y^{-1}(\tau) = \theta(1 - \tau)^{-\frac{1}{\alpha}} - \theta$, we have that:

$$y(x) + \theta = F_y^{-1}(\Phi(x)) + \theta = \theta(1 - \Phi(x))^{-\frac{1}{\alpha}} \qquad (20)$$

so we have that $f_y(y(x)) = \frac{\alpha}{\theta}[1 - \Phi(x)]^{\frac{\alpha+1}{\alpha}}$. Therefore, after some routine computation and simplification, and using equation (20) for the last equality, we have:

$$-\frac{d}{dx}\log[f_y(y(x))] = \frac{\alpha + 1}{y(x) + \theta} \cdot \frac{\phi(x)}{f_y(y(x))}$$

$$= \frac{\alpha + 1}{\alpha}\phi(x)[1 - \Phi(x)]^{-1} \qquad (21)$$

Again, by equation (19), we only need to show that:
$x \leq -\frac{d}{dx}\log[f(y(x))]$, for $x > 0$, which by equation (21)

is equivalent to: $x \leq \dfrac{\alpha+1}{\alpha} \, \phi(x)[1 - \Phi(x)]^{-1}$, or for $x > 0$:

$$\frac{\alpha+1}{\alpha} \cdot \frac{1}{x} \geq \frac{1 - \Phi(x)}{\phi(x)} \qquad (22)$$

But, we note that the quantity $m(x) = (1 - \Phi(x))/\phi(x)$ is the well-known *Mill's Ratio* for the normal distribution, which has been the focus of a good amount of research. In particular, Baricz (2008, 2010, and 2012) points out that in 1941 R. D. Gordon proved that the Mill's ratio, for the normal distribution obeys the following inequalities, for $x > 0$:

$$\frac{x}{x^2+1} < \frac{1 - \Phi(x)}{\phi(x)} < \frac{1}{x} \qquad (23)$$

Hence, since $\dfrac{\alpha+1}{\alpha} > 1$ we have, for $x > 0$, and for all $\alpha > 0$, that:

$$\frac{\alpha+1}{\alpha} \cdot \frac{1}{x} \geq \frac{1}{x} \geq \frac{1 - \Phi(x)}{\phi(x)} \qquad (24)$$

We now turn our attention to the Gamma distribution, or the regularized incomplete gamma function, as it is known within the applied mathematics and physics communities. The incomplete gamma function appears ubiquitously within applied mathematics, and is related to many other mathematical special functions, including the Confluent Hypergeometric functions, Bessel functions Tricomi (1950a, 1950b), the Legendre and Laguerre polynomials, Kummer's functions, the Gaussian error function, as well as the exponential integrals. For this reason, the theory of the incomplete gamma function, has been (and continues to be) of keen interest to many mathematicians. The Italian mathematician Tricomi is one such example. In fact, in the paper *The Incomplete Gamma Function Since Tricomi,* Gautshi (1998) remarks that "the incomplete gamma function held a special fascination for him (Tricomi), as he was fond of calling it,

affectionately, the *Cinderella of special functions.*" More relevantly, Gautshi goes on to state that "Monotonicity, convexity, and higher monotonicity results abound for the gamma function, but seem to be scarce for the incomplete gamma function." Moreover, despite the author's assiduous attempt to find inequality results which are tight enough across a sufficiently broad range of the domain, to facilitate a direct proof of the convexity of the incomplete gamma function composed with the inverse error function, the breadth of published inequalities was, once again, found to be lacking. Some relevant investigations of the regularized incomplete gamma, and other special functions, include Alm (2003), Alzer (1997, and 2005), Berg (2008), Carlitz (1963), Cerone (2008), Gautschi (1998), Short (2013), and Strecok (1968).

The authors did have some success using various approaches, including the use of approximations to $(\Phi^{-1} \circ F_y)(y)$ involving inverse hyperbolic trigonometric functions, but this necessitated the breaking-up of the domain, and additional work would have been necessary to make the result rigorous over the full domain, *Y.* More importantly, this approach required a constructive argument, which was quite protracted. As a result, the focus of the authors returned to obtaining the result through contradiction. Finally, due to an éclair de génie, experienced by the third author, a surprisingly concise and elegant version of the proof was made possible. To the author's knowledge, this analytical result is unique within the study of special functions.

Proposition 6.3 *(Gamma distribution) If $F_y(y)$ is a Gamma distribution, with shape parameter α, and scale parameter θ, and $\Phi(x)$ (resp. $\phi(x)$) is the CDF (resp. density) of the standard normal distribution, then $y(x) = (F_y^{-1} \circ \Phi)(x)$ is convex, for all x, and for all values of the shape parameter α.*

Due to the complexity of the proof of *Proposition 6.3,* especially when $\alpha > 1$, the proof is relegated to the appendix, and is proceeded by several supporting lemmas.

7. Conclusion

Parsa and Klugman (2011) proposed a generalization of ordinary least squares regression which is better suited to the modeling of actuarial data sets, which often possess heavy-tailed marginal distributions and non-linear relationships between variables. However, not long after the introduction of copula regression, a surprisingly simple approximation was suggested. In this paper we have presented and investigated this approximation, which takes the form of an OLS regression, under particular transformations of the variables. Next, we described how this linear approximation to copula regression can produce estimates which are close to those from exact copula regression. However, we present a set of criteria which guarantee that the estimates from the linear approximation will underestimate those from exact copula regression. Further, it is described how the main driver of the discrepancy, or bias, between copula regression and its linear approximation is a consequence of the use of transformations of the variables. Moreover, we explain how this discrepancy is not due to the form of the transformations used within the linear approximation to copula regression, but rather such a discrepancy will likely arise whenever transformations of the variables are used within OLS regression. This realization has consequences well beyond the specific models investigated in this paper, and serves as a salient reminder of the dangers of using transformations of variables within OLS regression. Finally, armed with the aforementioned sufficient conditions for underestimation, we continued by investigating which of the common loss distributions satisfy these criteria. In particular, we were able to prove that the lognormal, Pareto, and gamma distributions all satisfy these criteria and in particular, satisfy these criteria *for all parameter values*. Hence, these results allow the partitioner to determine when the estimates from the linear approximation to copula regression will underestimate the true values. Further, if OLS regression involving transformations is used within the reserving, capital modeling, or even pricing processes of a firm, these results can aid the practitioner in avoiding the understatement of reserves and even insolvency.

Appendix A

Lemma A.1 *Let $F_y(y)$ denote the CDF of the Gamma distribution, with shape parameter α, and let $\Phi^{-1}(p)$ denote the quantile function corresponding to the standard normal distribution. Let $g(y) = (\Phi^{-1} \circ F_y)(y)$, then:*

1. $\lim_{y \to 0^+} g'(y) = +\infty$ for all α
2. $\lim_{y \to \infty} g'(y) = 0$ for all α
3. $g''(y)$ is continuous on $(0,\infty)$

where $g'(y) = \dfrac{d}{dy} g(y)$.

Proof. We first prove condition (1), for $\alpha \le 1$. By a straightforward calculation, we have that:

$$g'(y) = \frac{d}{dy} g(y) = \frac{d}{dy} (\Phi^{-1} \circ F_y)(y) = \frac{f_y(y)}{\phi(g(y))} \quad (25)$$

Hence, we immediately have the result, since when $\alpha \le 1$, we have that:

- $\lim_{y \to 0^+} f_y(y) = \begin{cases} \infty & \text{for } \alpha < 1 \\ 1 & \text{for } \alpha = 1 \end{cases}$, and:

- $\lim_{y \to 0^+} \phi(g(y)) = 0.$

Now consider the case when when $\alpha > 1$. Note that $g'(y) > 0$, for $y \in \mathbb{R}^+$, since $f_y(y)$, and $\phi(x)$, are always positive. Further:

$$\frac{d}{dy} \phi(g(y)) = \phi'(g(y)) g'(y)$$

$$= [-g(y) \cdot \phi(g(y))] \frac{f_y(y)}{\phi(g(y))}$$

$$= -g(y) \cdot f_y(y) \quad (26)$$

which follows from (25), and the identity:

$$\phi'(x) = -x \cdot \phi(x) \quad (27)$$

Next, we recall the well-known recurrence relation for $f_y(y)$, the density of the gamma distribution:

$$f_y'(y) = \left(\frac{\alpha - 1}{y} - 1 \right) f_y(y) \quad (28)$$

Hence, by l'Hôpital's rule, together with this relation, and equation (26), we have:

$$\lim_{y \to 0^+} g'(y) = \lim_{y \to 0^+} \frac{f'_y(y)}{\frac{d}{dy}\phi(g(y))}$$

$$= \lim_{y \to 0^+} \frac{f'_y(y)/f_y(y)}{-g(y)} = \lim_{y \to 0^+} \frac{\frac{\alpha-1}{y}-1}{-g(y)} \quad (29)$$

And, by another application of l'Hôpital and equation (25), we have:

$$\lim_{y \to 0^+} g'(y) = \lim_{y \to 0^+} \frac{-(\alpha-1)/y^2}{-g'(y)}$$

$$= \lim_{y \to 0^+} \frac{(\alpha-1)\phi(g(y))}{y^2 f_y(y)}.$$

And finally, using yet another application of l'Hôpital and equation (26), we have that:

$$\lim_{y \to 0^+} g'(y) = \lim_{y \to 0^+} \frac{(\alpha-1)\phi'(g(y))g'(y)}{2yf_y(y)+y^2 f'_y(y)}$$

$$= \lim_{y \to 0^+} \frac{-(\alpha-1)g(y)\cdot f_y(y)}{2yf_y(y)+y^2 f'_y(y)}$$

$$= \lim_{y \to 0^+} \frac{-(\alpha-1)g(y)}{2y+y^2\left(\frac{\alpha-1}{y}-1\right)}. \quad (30)$$

Since the denominator can be written $[2 + (\alpha - 1) - y]y$, the denominator of equation (30) goes to 0, from the right, for all $\alpha > 0$. Hence, the above limit equals $+\infty$, for $\alpha > 1$.

Result (2) can be seen to be true, by appealing to equation (29), which can be seen to be valid for $\alpha \leq 1$, as well as for all $\alpha > 1$, as $y \to \infty$.

Finally, using equation (26) within a derivation analogous to that which led to equation (12), it can be shown that:

$$g''(y) = \frac{d^2}{dy^2}g(y) = \frac{f'_y(y)-\phi'(g(y))\left[\frac{f_y(y)}{\phi(g(y))}\right]^2}{\phi(g(y))}$$

$$= \frac{\phi^2(g(y))f'_y(y)-f_y^2(y)\phi'(g(y))}{\phi^3(g(y))} \quad (31)$$

Hence, we see that $g''(y): [0, \infty) \to \mathbb{R}$ is continuous, at least on $(0,\infty)$, which gives 3.

Lemma A.2 *Let $f(x)$: $\mathbb{R}^+ \to \mathbb{R}$ be any function, s.t.:*

1. $f'(x) > 0$, for $x \in \mathbb{R}^+$
2. $\lim_{x \to 0^+} f'(x) = \infty$
3. $\lim_{x \to \infty} f'(x) = 0$
4. $f''(x)$ exists $\forall x \in \mathbb{R}^+$

Then, for any point $x \in \mathbb{R}^+$, $\exists\, x^*, x^{**} \in \mathbb{R}^+$ s.t. $0 < x^{**} < x < x^* < \infty$, and $f''(x^*) < 0$, and $f''(x^{**}) < 0$.

Proof. The result follows from a straightforward application of the *mean-value theorem*. Let $x \in \mathbb{R}^+$. By assumption 1, $f'(x) = \varepsilon > 0$. Further, by assumption 3, $\exists\, x' \in \mathbb{R}^+$ s.t. $x < x' < \infty$, and $f'(x') = \varepsilon/2$. Hence, by the mean-value theorem, applied to $f'(x)$, $\exists\, x^*$ s.t. $x < x^* < x' < \infty$ and $f''(x^*) = \dfrac{f'(x')-f'(x)}{x'-x} = \dfrac{-\varepsilon}{2(x'-x)}$

< 0, since $x' > x$.

The result for $f''(x^{**})$ follows from an analogous argument.

Definitions A.3
Define an auxiliary equation to $g''(y)$, as:

1. $\rho(y) = \dfrac{g''(y)}{g'(y)} = \left(\dfrac{\alpha-1}{y}-1\right) + g(y)\cdot g'(y),$ *and note that:*
   a) $\rho(y)$ *has the same sign as $g''(y)$.*
   b) $\rho(y)$ *is continuous on $(0,\infty)$.*

2. $\rho'(y) = -\dfrac{\alpha-1}{y^2} + [g'(y)]^2 + g(y)\cdot g''(y) = \xi(y) + g(y)\cdot g''(y)$ *where:* $\xi(y) = -\dfrac{\alpha-1}{y^2} + [g'(y)]^2$ *and note that*: $\rho'(y)$ *is continuous on $(0, \infty)$.*

3. $\xi'(y) = 2\left(\dfrac{\alpha-1}{y^3}\right) + 2\cdot g'(y)\cdot g''(y)$

Proof of Proposition 6.3
Let $F_y(y)$ denote the gamma distribution, with shape parameter α, and let $g(y): [0, \infty) \to \mathbb{R}$ be defined by $g(y) = (\Phi^{-1} \circ F_y)(y)$. Further, denote $\dfrac{d}{dy}g(y)$ by $g'(y)$,

and $\frac{d^2}{dy^2} g(y)$ by $g''(y)$. Since $(F_y^{-1} \circ \Phi)(x)$, and its inverse $(\Phi^{-1} \circ F_y)(y)$, are both increasing functions, in order to prove that $(F_y^{-1} \circ \Phi)(x)$ is convex, it suffices to show that $(\Phi^{-1} \circ F_y)(y)$ is concave. Further, to prove that $g(y)$ is concave, we show that *no* closed interval $[y_1, y_2]$ exists, s.t. $y_2 > y_1$ and $g''(y) \geq 0$ for all $y \in [y_1, y_2]$. Which, by the continuity of $g''(y)$ (*Lemma A.1*), in turn, implies that $g''(y) < 0$, *except* on, at most, a set of points that contains no interval, where $g''(y) = 0$, and therefore that $g(y)$ is *concave.*

Hence, assume that $\exists$ some interval $I = [y_1^{(I)}, y_2^{(I)}]$, with $y_1^{(I)} < y_2^{(I)}$, on which $g''(y) \geq 0$, for $y \in I$. By appealing to *Definitions A.3,* we have that $I \subseteq J = \rho^{-1}([0,\infty))$, and since $\rho(y)$ is continuous, this implies that J is a *closed* set *WRT* to the *relative* topology on $(0,\infty)$. Now, denote by J_I, the *maximal* connected component of J s.t. $I \subseteq J_I \subseteq J$. By *Lemma A.2*, J_I cannot be of the form $(0, \delta]$, or $[\delta, \infty)$, for any $\delta > 0$, hence, by an elementary topological argument, we have that J_I is closed *WRT* the *usual* topology on $\mathbb{R}$. Hence, we can write $J_I = [y_1, y_2]$. Further, by the continuity of $\rho(y)$, we have that $\rho(y_1) = \rho(y_2) = 0$, and moreover that $\rho'(y_1) \geq 0$ and $\rho'(y_2) \leq 0$. Also, by the continuity of $g''(y)$, we have that $g''(y_1) = g''(y_2) = 0$. Again, by the definitions in *A.3* we have that $\xi(y_1) = \rho'(y_1) \geq 0$, and $\xi(y_2) = \rho'(y_2) \leq 0$.

First, assume $\alpha > 1$. By the definition of J_I, we have that $g''(y) \geq 0$, for all $y \in [y_1, y_2]$, and hence that:

$$\xi'(y) = 2\left(\frac{\alpha - 1}{y^3}\right) + 2 \cdot g'(y) \cdot g''(y)$$

$$\geq 2\left(\frac{\alpha - 1}{y^3}\right) > 0 \text{ for all } y \in [y_1, y_2] \quad (32)$$

since $g'(y) > 0$, for all y. But this implies that $\xi(y_2) > \xi(y_1) \geq 0$, which contradicts that $\rho'(y_2) = \xi(y_2) \leq 0$.

Conversely, if $\alpha \leq 1$, then again by the definitions in *A.3,* we have the stronger condition that $\xi(y) > 0$, for all $y \in \mathbb{R}^+$. In particular, we have that $\rho'(y_2) = \xi(y_2) > 0$ which is, again, a contradiction.

Acknowledgments

References

Alm, S. E., "Monotonicity of the Difference between Median and Mean of Gamma Distributions and of a Related Ramanujan Sequence," *Bernoulli* 9: 2, 2003, pp. 351–371.

Alzer, H., "Proof of the Chen-Rubin Conjecture," *Proceedings of the Royal Society of Edinburgh,* Section A 135, no. 4, 2005, pp. 677–688.

Alzer, H., "On Some Inequalities for the Incomplete Gamma Function," *Mathematics of Computation* 66: 218, 1997, pp. 771–778

Baricz, Á., "Mills' Ratio: Monotonicity Patterns and Functional Inequalities," *Journal of Mathematical Analysis and Applications* 340: 2, 2008, pp. 1362–1370.

Baricz, Á., "Geometrically Concave Univariate Distributions," *Journal of Mathematical Analysis and Applications* 363, 2010, pp. 182–196.

Baricz, Á., "Mills' Ratio: Reciprocal Convexity and Functional Inequalities," *Acta Universitatis Sapientiae, Mathematica* 4: 1, 2012, pp. 26–35.

Berg, C., and H. L. Pedersen, "Convexity of the Median in the Gamma Distribution," 2008, arXiv.org/pdf/math/0609442v1.pdf.

Carlitz, L., "The Inverse of the Error Function," *Pacific Journal of Mathematics* 13, 1963, pp. 459–470.

Cerone, P., and S. Dragomir (eds.), *Advances in Inequalities for Special Functions,* New York: Nova Science, 2008.

Clement, R.T., and T. Reilly, "Correlations and Copulas for Decision and Risk Analysis," *Management Science* 45: 2, February 1999, pp. 208–224.

Embrechts, P., "Copulas: A Personal View," *Journal of Risk and Insurance* 76, 2009, pp. 639–650.

Gautschi, W., "The Incomplete Gamma Function since Tricomi," *Tricomi's Ideas and Contemporary Applied Mathematics, Atti dei Convegni Lincei* 147, *Accademia Nazionale dei Lincei,* 1998, pp. 203–237.

Munir, A., and W.T. Shaw, "Quantile Mechanics 3: Series Representations and Approximation of some Quantile Functions appearing in Finance," Papers, 2012, arXiv.org, http://EconPapers.repec.org/RePEc:arx:papers:1203.5729.

Parsa, R., and S.A. Klugman, "Copula Regression," *Variance* 5, 2011, pp. 45–54.

Shaw, W.T., and N. Brickman, "Quantile Mechanics II: Changes of Variables in Monte Carlo Methods and a GPU-Optimized Normal Quantile." http://econpapers.repec.org/paper/arxpapers/0901.0638.htm, August 2010.

Shaw, W.T., and I. Buckley, "The Alchemy of Probability Distributions," presented at the First IMA Conference on Computational Finance, March 2007. http://arxiv.org/abs/0901.0434v1.

Short, M., "Improved Inequalities for the Poisson and Binomial Distribution and Upper Tail Quantile Functions," *ISRN Probability and Statistics,* vol. 2013, article ID 412958, Hindawi Publishing.

Steinbrecher, G., and W.T. Shaw, "Quantile Mechanics," *European Journal of Applied Mathematics* 19: 2, 2008, pp. 87–112.

Strecok, A.J., "On the Calculation of the Inverse of the Error Function," *Mathematics of Computation* 22: 101, 1968, pp. 144–158.

Tricomi, F.G., "Asymptotische Eigenschaften der unvollständigen Gammafunktion," *Mathematische Zeitschrift* 53: 2, 1950a, pp. 136–148.

Tricomi, F.G., "Sulla Funzione Gamma Incompleta," *Annali di Matematica Pura ed Applicata* 31: 1, 1950b, pp. 263–279.