Copula Regression

by Rahul A. Parsa and Stuart A. Klugman

ABSTRACT

Regression analysis is one of the most commonly used statistical methods. But in its basic form, ordinary least squares (OLS) is not suitable for actuarial applications because the relationships are often nonlinear and the probability distribution of the dependent variable may be non-normal. One approach that has been successful in overcoming these challenges is the generalized linear model (GLM), which requires that the dependent variable have a distribution from the exponential family. In this paper, we present copula regression as an alternative to OLS and GLM. The major advantage of a copula regression is that there are no restrictions on the probability distributions that can be used. In this paper, we will present the formulas and algorithms necessary for conducting a copula regression analysis using the normal copula. However, the ideas presented here can be used with any copula function that can incorporate multiple variables with varying degrees of association.

KEYWORDS

Regression, least squares, multivariate normal copula, copula regression, generalized linear models

1. Introduction

Regression analysis is one of the most widely used statistical methods, dating back to Francis Galton's 1875 discovery of the regression slope (Stanton 2001). The key concept behind this methodology is identifying a relationship between a set of variables. The premise is that a dependent or response variable, labeled Y, is assumed to be related in some functional form to several covariates (also referred to as explanatory or independent variables), labeled $X_1, X_2, \ldots, X_k$. The response variable, Y, measures a random quantity of interest. Examples include the annual claims on an automobile insurance policy and the age at death of an insured life. The covariates, $X_1, X_2, \ldots, X_k$, could be a set of characteristics of the insured. For example, in automobile insurance the variables might be age, gender, accident history, credit score, and so on. In life insurance the variables might be age, gender, smoking status, blood pressure, height, weight, and so on. If it is possible to express how the distribution of the response variable is related to the covariates, the result is often a reduction in the uncertainty of insured events that depend on Y.

The basic OLS regression model presents a specific model for the relationship. The distribution of Y given the covariates is assumed to be normal with a variance that is constant (that is, not related to the covariates) and a mean that is related to the covariates as $E(Y \mid X_1 = x_1, \ldots, X_k = x_k) = \beta_0 + \beta_1 x_1 + \cdots \beta_k x_k$. The equivalent (in this case) techniques of maximum likelihood and least-squares are used to estimate the unknown coefficients. The covariates can be considered as random observations (in which case it is assumed the complete set of variables has a multivariate normal distribution) or as fixed quantities, for example, where the levels are set in a designed experiment. A measure of the success of the model is the reduction in the variance of Y through the use of the covariates. Basic OLS regression is covered in most introductory statistics texts.

Insurance losses rarely follow a normal distribution. Models for the age at death are usually skewed to the left, while models for casualty losses are usually skewed to the right. Furthermore, there is no logical argument, other than mathematical simplicity, to support a linear relationship between the covariates and the mean of the response variable. Embrechts, McNeil, and Straumann (2002) show how the Pearson correlation coefficient can be misleading when the underlying distributions are not normal. They advise using copulas to model data that are not normal because such models capture a greater variety of relationships (essentially being nonparametric).

A significant advance was provided by the development of the generalized linear model (GLM). The standard reference is McCullagh and Nelder (1989). The generalization comes in two places. First, the distribution of Y must be a member of the exponential family (which includes many commonly used distributions such as binomial, Poisson, negative binomial, normal, gamma, inverse Gaussian, and log-normal, see pages 26-28 of McCullagh and Nelder). Second, the relationship between the mean of Y and the covariates is expressed as $E(Y \mid X_1 = x_1, \ldots, X_k = x_k) = g^{-1}(\beta_0 + \beta_1 x_1 + \cdots \beta_k x_k)$ where $g(y)$ is called the link function. The conditional variance of Y is no longer constant but will depend on the conditional mean. The nature of the dependence depends on the choice of distribution for Y. Modern statistical software includes GLM analysis for specific combinations of distributions and link functions. Estimation is done by maximum likelihood. This model has been widely used in automobile pricing and there are numerous articles providing actuarial applications (Aitkin 1989, Mosley 2004).

The next step in generalizing this model is to remove the restriction that the distribution of Y be from the exponential family. There is nothing inherent in the GLM that demands the use of the exponential family, although some of the calculations become more difficult and there is no convenient expression for the conditional variance. This generalization is briefly discussed by Klugman, Panjer, and Willmot (2008) and is further expanded in Venter (2007).

We are proposing an alternative approach to generalizing the OLS regression model. We are not

claiming that it is superior to GLM or OLS, but rather that it provides an alternative that has the possibility of providing a better fit to observed data. Begin by assuming that all the variables (response and covariates) are random observations from some joint probability distribution. Because some of the variables are dependent (if there are no dependencies, then possessing the covariates provides no useful information about the response variable) a model with dependence must be used. A popular method for creating a multivariate model with dependence is the copula model. There are numerous papers with actuarial applications of copulas. A good introduction is provided by Frees and Valdez (1998). In this paper, we will use the normal copula for the reasons described in Section 2.

This paper will develop copula regression as follows. Section 2 introduces copula functions and the normal copula. Section 3 covers parameter estimation and computation of the conditional mean. Section 4 concludes the paper with comprehensive examples.

2. Copulas

Copulas are useful for describing multivariate non-normal distributions. They describe the dependence structure between the variables. Marginal distribution functions are used as inputs to the copula and these can be any set of disparate distributions. Thus, a copula is a realistic way of describing the multivariate distributions, as the analyst normally has a good idea about each marginal distribution and seldom has a good idea about the joint distribution of these variables.

A copula model can be constructed as follows. Begin with n random variables with distribution functions $F_{X_1}(x_1), F_{X_2}(x_2), \ldots, F_{X_n}(x_n)$. The joint distribution function is created by applying the copula function, C, as

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = C[F_{X_1}(x_1), \ldots, F_{X_n}(x_n)].$$

Only certain functions can be used as copula functions in order to ensure that regardless of the n input distribution functions, the resulting function is a legitimate multivariate distribution function. The key is that the copula creates the dependence structure independently of the marginal distributions. This concept of defining a multivariate dependence structure (and hence copula) is based on Sklar's theorem, which states:

Sklar's Theorem (Sklar 1959)**:** Let F be an n-dimensional joint cumulative distribution function for the random variables $X_1, X_2, \ldots, X_n$ with marginal distribution functions $F_1(x_1), F_2(x_2), \ldots, F_n(x_n)$. Then there exists a Copula function, C, such that

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = C[F_1(x_1), \ldots, F_n(x_n)].$$

If the marginal distributions are continuous, then C is unique. If any distribution is discrete, then C is uniquely defined on the support of the joint distribution.

In the case of continuous distributions, the multivariate dependence structure and marginal distributions can be separated and the copula can be considered "independent" of the margins (Joe 1997, pp. 12-13). The copula thus allows arbitrary continuous marginal distributions to be combined and describes their dependence structure.

There are several bivariate and multivariate copula distributions discussed in the literature. Hutchinson and Lai (1990), Nelson (1999), Joe (1997) and Klugman, Panjer, and Willmot (2008) are excellent texts that discuss various properties of bivariate copula distributions. Joe (1997) also discusses the properties of several multivariate copula distributions. Embrechts, McNeil, and Straumann (2002) use copulas to show how the Pearson correlation coefficient can be misleading. Klugman and Parsa (1999), and Frees and Valdez (1998) use copulas in modeling insurance data. Others who have contributed to the copula literature are Genest and MacKay (1986) and Miller and Liu (2002).

While many copula functions have been identified, we believe only two are useful for building a regression model with several covariates. One of the features of any useful regression model is that the

degree of association (e.g., Spearman's Rank Correlation) between the response variable and each covariate need not be constant. We are aware of only two copula models that allow for this, the normal copula and its generalization, the *t*-copula (which is based on the multivariate Student's *t* distribution. For example, see the list of copulas in Klugman, Panjer, and Willmot (2008, Chapter 7), where it can be seen that the other copulas do not allow for variations in the association measure. In a recent paper by Crane and Van Der Hoek (2008) they use arbitrary bivariate copulas for building regression models. Because their methods do not extend to multivariate situations, their approach is not useful for most applications. Also, they do not address the case of discrete variables, which are common in practice.

Computing predicted values of Y using copula regression is a three-step process.

1. Assume a model for the joint distribution of all the variables (response and covariates),
2. Estimate the parameters of the model (the parameters for the selected marginal distributions and the parameters of the copula), and
3. Compute the predicted values of Y given a set of covariates by using the conditional mean of Y given the covariates.

In this paper, we assume a normal copula model for the joint distribution of the variables and maximum likelihood for parameter estimation. For the normal copula explicit formulas for the likelihood function and conditional distribution are available. The conditional mean will usually need to be obtained by numerical integration. Because this is a one-dimensional integration, there are well-established accurate algorithms for obtaining the answer. Thus, this ease of use is one of the reasons we have selected the normal copula. The methodology presented in this paper can be used with other multivariate copulas (such as *t* or elliptical), but they would require extensive numerical analysis.

Note that if all the variables are assigned the normal distribution and the normal copula is used, the results of maximum likelihood estimation will match those from OLS. Thus, like GLM, this method is a (different) generalization of the OLS model.

3. Parameter estimation

The normal (multivariate normal) copula function is defined as (Song 2000)

$$C_R(u_1, \ldots, u_n) = G[\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_n)], \quad (3.1)$$

where $\Phi(u)$ is the standard normal cumulative distribution function and G is the multivariate normal cumulative distribution with zero means, unit variances, and correlation matrix R. The matrix R has ones on the diagonal and the off diagonal elements are correlations. However, it is important to note that when the normal copula is used these are not the correlations of the modeled variables.

Thus, for arbitrary marginals, the distribution function induced by the normal copula is

$$F(x_1, \ldots, x_n) = G\{\Phi^{-1}[F_1(x_1)], \ldots, \Phi^{-1}[F_n(x_n)]\}. \quad (3.2)$$

If the marginal distributions are continuous then the density function is given by (Clemen and Reilly 1999, Nelson 1999)

$$f(x_1, \ldots, x_n) = f_1(x_1) f_2(x_2) \cdots f_n(x_n) \exp\left\{-\frac{v^T (R^{-1} - I) v}{2}\right\} \times |R|^{-0.5}, \quad (3.3)$$

where v is a vector with ith element $v_i = \Phi^{-1}[F_i(x_i)]$ and I is the identity matrix. Note that if the correlations are all zero, then R is the identity matrix and the joint density is the product of the marginal densities and all the variables are independent.

This density function can then be used to determine the likelihood function.

Since we will be estimating predicted values using the conditional mean of Y given the independent variables, we will need the conditional distribution of x_n given $x_1, \ldots, x_{n-1}$. It is

$$f(x_n|x_1,\ldots,x_{n-1}) = f_n(x_n)\exp$$

$$\left\{-0.5\left[\frac{\{\Phi^{-1}[F_n(x_n)]-r^T R_{n-1}^{-1}v^*\}^2}{1-r^T R_{n-1}^{-1}r}-\{\Phi^{-1}[F_n(x_n)]\}^2\right]\right\}$$

$$\times\,(1-r^T R_{n-1}^{-1}r)^{-1/2} \qquad (3.4)$$

where $v^* = (v_1, \ldots, v_{n-1})$ and $R = \begin{bmatrix} R_{n-1} & r \\ r^T & 1 \end{bmatrix}$

where r is an $n-1 \times 1$ vector that is the right-most column of R with the last element removed.

In the regression context, X_n will be the response variable (Y). Thus, this is the density function of the response variable given the covariates. The expected value is the predicted value (the median could also be used) and the standard deviation is equivalent to the standard error in OLS regression. However, the standard deviation will not be constant as it is in OLS regression.

Maximum likelihood estimates are asymptotically normal with covariance matrix given by the inverse of Fisher's information matrix $(I(\theta)^{-1})$. To use this result, we would have to calculate the second derivatives of the likelihood function and take their expected values. In this problem, it is difficult to calculate second derivatives and their expected values, so we instead use the observed information matrix (Klugman, Panjer, and Willmot 2008). Once again, this is a well-understood problem. Most commercially available numerical analysis software provides the Hessian matrix, which is used to estimate the asymptotic covariance matrix (inverse of the Hessian matrix) for the estimates. We present this for Example 1 in Section 4.

When all the marginal distributions are continuous, maximum likelihood estimation is performed by using the density function in (3.3). Note that if there are, for example, eight marginal distributions, each with two parameters, there will be 8(2) + 8(8 – 1)/2 = 44 parameters to estimate. Often it is easier to use the marginal data to obtain estimates of the parameters of the marginal distributions and use the joint data to estimate the correlation parameters in the normal copula. These can be used as is, producing a suboptimal result. On the other hand, one can use the estimates from the marginal distributions as starting values for the global maximization (which is optimal). We used the latter procedure for estimating the parameters. This was done numerically. One author used the solver tool in Excel and the other author used SAS IML and the NLPTR optimization procedure (NLPFDD to compute the Hessian matrix).

When any of the marginal distributions are discrete, there is an additional problem. This is illustrated with an example. Let X take on the values 0, 1, and 2 with probabilities 0.3, 0.4, and 0.3 respectively. Let Y be the other marginal with a uniform (0,1) distribution. Consider the independence copula. Now consider the contribution to the likelihood function for the observation $x = 1$ and $y = 0.6$. It is the product of the probability function of X at 1 and the density function of Y at 0.6, which is 0.4(1) = 0.4. But now suppose the normal copula is being used. This requires calculation of cumulative probabilities from the multivariate normal distribution (in order to obtain the discrete probability), a non-trivial task. In addition, if there are several discrete variables, the probabilities for higher dimensional cubes will be required.

In OLS and GLM regressions, distributions for the covariates are usually not specified. In a copula model these distributions must be specified. However, there is an easy way to replicate this situation—for the marginal models, use the empirical distribution. Thus, in many applications all of the covariates will be modeled as discrete variables. The same problem affects conditional distributions.

Our recommended solution is to replace each discrete distribution with a continuous distribution. The simplest choice is to use a kernel density with a uniform kernel and a small bandwidth. In particular, select the bandwidth so that it is less than half the distance separating the two closest discrete probability points.

The likelihood function requires the kernel density pdf and cdf at each data point. Let b be the bandwidth. For any observed value, let $p(x)$ be the probability function and $P(x)$ be the distribution function of the discrete distribution. For writing the likelihood function, for an observation at x, the kernel density

pdf is $p(x)/(2b)$ and the cdf is $P(x) - p(x)/2$. Note that the value of b has no impact on the MLE.

For the conditional distribution, the only case that needs further study is when the response variable is discrete. The formula for the conditional density function (for continuous variables) can be written

$$f(x_n|w) = f_n(x_n)g[F_n(x_n), w],$$

where w represents all the other variables, and g is defined by removing $f_n(x_n)$ from (3.4). To obtain the discrete conditional probability it makes sense to integrate the continuous conditional probability over the bandwidth:

$$\begin{aligned} p(x_n|w) &= \int_{x_n-b}^{x_n+b} p(x_n)(2b)^{-1} \\ &\times g[P_n(x_n) + p(x_n)(t - x_n - b)/(2b), w]dt \\ &= \frac{p(x_n)}{2b}\int_{-b}^{b} g[P(x_n) + p(x_n)(z/2b - 0.5), w]dz. \end{aligned}$$

The second line follows from the transformation $z = t - x_n$.

The integral must be evaluated numerically; the value of b has no impact on the result.

4. Examples

To illustrate our methodology, we present six examples using simulated data. In all of the examples, we have three variables—one dependent and two independent and the sample size is 50 in all the cases. We use the same correlation matrix in all the examples as given below.

$$R = \begin{bmatrix} 1 & 0.7 & 0.7 \\ 0.7 & 1 & 0.7 \\ 0.7 & 0.7 & 1 \end{bmatrix}$$

To simulate the 50 trivariate data points (see the Appendix for the simulated data), we used the methodology described in Clemen and Riley (1999). It uses the following process:

Let $F_i(x_i)$, $i = 1,2,3$ be specified along with the correlation matrix.

1. First generate a vector (u_1, u_2, u_3) from a multivariate normal distribution with zero means and unit variances and the specified correlation matrix.
2. Calculate $s_i = \Phi(u_i)$ for each of the three variables.
3. Calculate $F_{X_i}^{-1}(s_i)$, $i = 1,2,3$.

The resulting vector from step (3) will have the specified marginals and copula correlation structure (see Appendix 1 for one of the data sets).

We measured the error in estimation using $\Sigma(Y_i - \hat{Y}_i)^2$ and compared the results to OLS and GLM (where appropriate). We chose OLS and GLM for comparison as they are the procedures commonly used in practice. Nowadays, GLM is commonly used whenever the distribution of the dependent variable is non-normal.

Example 1: The distributions of the three variables[1] are: $X_1 \sim$ Pareto ($\alpha = 3$, $\theta = 100$), $X_2 \sim$ Pareto (4, 300), $X_3 \sim$ Gamma (3,100) and X_3 is the dependent variable. The parameters were estimated in two stages, using the method of maximum likelihood. In stage 1, only the data from the marginal distribution was used to estimate the parameters. In the second stage, the joint density was maximized using estimates from stage 1 as starting values. In this example, parameter estimates for X_2 did not converge, so we approximated it using a gamma distribution. The copula regression model performed much better than OLS. The MLEs and errors are given in Table 1.

Table 1. Parameter estimates and errors for Example 1

Variable	X_1	X_2	X_3	Sum of squared errors	
Model	Pareto	Gamma	Gamma	Copula	OLS
MLEs/Errors	3.44, 161.11	1.04, 112.003	3.77, 85.93	590,000.54	637,172.84

[1]All variables are parameterized as in Klugman, Panjer and Willmot (2008).

Table 2. Asymptotic standard deviations

	X_1 Pareto		X_2 Gamma		X_3 Gamma		Correlations		
	$Alpha_1$	$Theta_1$	$Alpha_2$	$Theta_2$	$Alpha_3$	$Theta_3$	R_{21}	R_{31}	R_{32}
$Alpha_1$	0.266606	0.966067	0.359065	–0.33725	0.349482	–0.33268	–0.42141	–0.33863	–0.29216
$Theta_1$	0.966067	15.50974	0.390428	–0.25236	0.346448	–0.26734	–0.37496	–0.29323	–0.25393
$Alpha_2$	0.359065	0.390428	0.025217	–0.78766	0.438662	–0.35533	–0.45221	–0.30294	–0.42493
$Theta_2$	–0.33725	–0.25236	–0.78766	3.558369	–0.38489	0.464513	0.496853	0.35608	0.470009
$Alpha_3$	0.349482	0.346448	0.438662	–0.38489	0.100156	–0.93602	–0.34454	–0.46358	–0.46292
$Theta_3$	–0.33268	–0.26734	–0.35533	0.464513	–0.93602	2.485305	0.365629	0.482187	0.481122
R_{21}	–0.42141	–0.37496	–0.45221	0.496853	–0.34454	0.365629	0.010085	0.457452	0.465885
R_{31}	–0.33863	–0.29323	–0.30294	0.35608	–0.46358	0.482187	0.457452	0.01008	0.481447
R_{32}	–0.29216	–0.25393	–0.42493	0.470009	–0.46292	0.481122	0.465885	0.481447	0.009706

In Table 2 are the asymptotic standard deviations of the estimates and correlations among them (diagonal terms are standard deviations and off-diagonal terms are correlations). The SAS IML procedure NLPFDD was used to calculate the Hessian matrix. It uses a finite difference approximation procedure to calculate the Hessian. We used the Hessian to calculate the observed information matrix (asymptotic variance matrix).

From this it is possible to construct confidence intervals. For example, the 95% confidence interval for alpha of the Pareto variable (X_1) is given by 3.44 ± 1.96*0.266606.

The maximum likelihood estimate of the correlation matrix is given below.

$$\hat{R} = \begin{bmatrix} 1 & 0.711 & 0.699 \\ 0.711 & 1 & 0.713 \\ 0.699 & 0.713 & 1 \end{bmatrix}.$$

As a reminder, while this is called a correlation matrix, it is merely the parameters of the normal copula. These values do not represent the correlations of the marginal variables.

Example 2: Once again, the distributions of the three variables were: $X_1 \sim$ Pareto (α = 3, θ = 100), $X_2 \sim$ Pareto (4, 300), $X_3 \sim$ Gamma (3,100) and X_3 was the dependent variable. This time, instead of assuming a distribution for X_1 and X_2 and estimating the parameters using MLE, we estimate the distributions empirically. This is similar to OLS and GLM wherein we make no assumption on the distribution of the independent variables. Since the empirical distribution is discrete, we approximated it by a continuous distribution and in particular we used the kernel density with uniform kernel as previously described. We then took the limiting case of the uniform kernel as band width goes to zero, resulting in the following equations for the distribution and density functions (where x is one of the data points).

$$F(x) = \frac{\#obs \le x}{n} - \frac{\#obs = x}{2n}$$

and

$$f(x) = \frac{1}{n}$$

We also analyzed the data using a generalized linear model using the gamma distribution and log link. Once again, the copula regression model performed better than OLS and, much better than GLM which performed very poorly. The results are given in Table 3.

The number of parameters in the various models is five for the copula model (two gamma parameters and three correlation coefficients, four for OLS (three regression parameters and the standard error), and four for GLM (three regression parameters and the gamma shape parameter). The large differences in the sums of squares are not likely due to the one parameter difference. A more formal comparison might use the log-likelihood functions and an information criterion.

Table 3. Parameter estimates and errors for Example 2

Variable	X_1	X_2	X_3	Sum of squared errors		
Model	Empirical	Empirical	Gamma	Copula	OLS	GLM
MLEs/Errors	N/A	N/A	4.0245, 81.0415	595,947.47	637,172.84	814,264.754

Table 4. Parameter estimates and errors for Example 3

Variable	X_1	X_2	X_3	Sum of squared errors	
Model	Poisson	Exponential	Gamma	Copula	OLS
MLEs/Errors	5.65	119.38907	3.67, 88.98	574968	582459.5

Table 5. Parameter estimates and errors for Example 4

Variable	X_1	X_2	X_3	Sum of squared errors		
Model	Empirical	Empirical	Gamma	Copula	OLS	GLM
MLEs/Errors	N/A	N/A	3.96, 82.48	559,888.8	582,459.5	652,708.98

Table 6. Parameter estimates and errors for Example 5

Variable	X_1	X_2	X_3	Sum of squared errors	
Model	Poisson	Exponential	Gamma	Copula	OLS
MLEs/Errors	5.65	119.38907	3.67, 88.98	109.072	114.6649

Example 3: The three variables are: $X_1 \sim$ Poisson ($\lambda = 5$), $X_2 \sim$ Pareto (4, 300), $X_3 \sim$ Gamma (3,100) and X_3 is the dependent variable. We estimated the parameters using the method of maximum likelihood in two stages as in Example 1. The distribution of X_2 was approximated by an exponential distribution. Here again, the copula regression model did better than OLS. The results are given in Table 4.

Example 4: The situation is same as in Example 3 but the distributions of X_1 and X_2 were estimated empirically as in Example 2. Since X_1 is discrete, its distribution was approximated by a continuous distribution. Once again, we used the uniform kernel as in Example 2.

We also analyzed the data using GLM using gamma distribution and log link. The copula regression model performed better than OLS and GLM. Surprisingly, GLM performed the worst. The results are given in Table 5.

Example 5: The distribution of the three variables are $X_1 \sim$ Poisson ($\lambda = 5$), $X_2 \sim$ Pareto (4, 300), $X_3 \sim$ Gamma (3,100), and X_1 is the dependent variable. In this example, the dependent variable is discrete. The distribution of X_2 was approximated by an exponential distribution. In this example, the copula regression model performed better than OLS. The results are given in Table 6.

Example 6: The situation is same as in Example 5 but the distributions of X_2 and X_3 are estimated empirically as in Example 2. The data were also analyzed using GLM using the Poisson distribution and log link. The copula regression model performed the best and surprisingly, GLM performed the worst. The results are given in Table 7.

Table 7. Parameter estimates and errors for Example 6

Variable	X_1	X_2	X_3	Sum of squared errors		
Model	Empirical	Empirical	Gamma	Copula	OLS	GLM
MLEs/Errors	5.67	N/A	N/A	110.379	114.6649	119.518

Summary

In this paper, we have provided an alternative approach to generalizing OLS regression using a multivariate copula. When compared to GLM, the approach seems to be different rather than better or worse. Its strength lies in the ability to choose distributions for dependent variables that are not members of the exponential family. Also, it allows the researcher to arbitrarily choose distributions for the marginals (GLM only requires specification of the distribution of the dependent variable). As skewed heavy-tailed distributions are common in insurance, this method provides a way to incorporate arbitrary distributions into regression models. Like GLM, this method allows nonlinear dependence to be modeled. Because correlation is not a useful measure of dependence in a non-normal world (Embrechts 2002), it is vital to describe the relationship between variables appropriately. In this regard, copula regression provides a good alternative to OLS and GLM.

References

Aitkin, M., D. Anderson, B. Francis, and J. Hinde, *Statistical Modeling in GLIM,* Oxford: Oxford Science Publications, 1989.

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

Crane, G. J., and J. Van Der Hoek, "Conditional Expectation Formulae for Copulas," *Australian and New Zealand Journal of Statistics* 50, 2008, pp. 53–67.

Embrechts, P., A. McNeil, and D. Straumann, "Correlation and Dependence in Risk Management: Properties and Pitfalls," *Risk Management: Value at Risk and Beyond,* ed. M. Dempster, pp. 176–223, New York: Cambridge University Press, 2002.

Frees, E., and E. Valdez, "Understanding Relationships Using Copulas," *North American Actuarial Journal* 2, 1998, pp. 1–25.

Genest, C., and J. MacKay, "The Joy of Copulas: Bivariate Distributions with Uniform Marginals," *American Statistician* 40, 1986, pp. 280–283.

Hutchinson, T. P., and C. D. Lai, *Continuous Bivariate Distributions, Emphasizing Applications,* Adelaide: Rumsby Scientific Publishing, 1990.

Joe, H., *Multivariate Models and Dependence Concepts,* London: Chapman and Hall, 1997.

Klugman, S., H. Panjer, and G. Willmot, *Loss Models: From Data to Decisions* (3rd ed.), New York: Wiley, 2008.

Klugman, S. and R. A. Parsa, "Fitting Bivariate Loss Distributions with Copulas," *Insurance: Mathematics and Economics* 24, 1999, pp. 139–148.

McCullagh, P. and J. Nelder, *Generalized Linear Models,* New York: Chapman and Hall, 1989.

Miller, D. J., and W. Liu, "On the Recovery of Joint Distributions from Limited Information," *Journal of Econometrics* 107, 2002, pp. 259–274.

Mosley, R. C., "Estimating Claim Settlement Values Using GLM," Casualty Actuarial Society, *Discussion Paper Program,* 2004, pp. 291–314.

Nelson, R. B., *An Introduction to Copulas,* New York: Springer, 1999.

Sklar, A., "Functions de Repartition a n dimensions et Leurs Marges," *Publications de l'Institut Statistisque de l'Universite de Paris* 8, 1959, pp. 229–231.

Song, P. X.-K., "Multivariate Dispersion Models Generated from Gaussian Copula," *Scandinavian Journal of Statistics* 27, 2000, pp. 305–320.

Stanton, J. M., "Galton, Pearson, and the Peas: A Brief History of Linear Regression for Statistics Instructors," *Journal of Statistics Education* 9:3, 2001.

Venter, G., "Generalized Linear Models Beyond the Exponential Family with Loss Reserve Applications," Casualty Actuarial Society *E-Forum*, Summer 2007, http://www.casact.org/pubs/forum/07sforum/07s-venter1.pdf.

Appendix 1

Data set used in Example 1

X_1 - Pareto (3,100)	X_2 - Pareto (4,300)	X_3 - Gamma (3,100)
49.19615	168.9541	339.0285
59.79256	52.22341	239.234
69.32701	428.0111	507.1756
21.82025	26.6339	232.2198
80.69061	130.7855	391.3814
66.54305	168.3623	752.3682
75.11587	172.878	433.767
16.98576	52.90305	148.1844
29.07292	48.31109	159.9216
1.974758	9.093995	90.26908
50.54979	122.4136	161.8736
27.88491	258.5495	381.7534
8.246412	18.45999	206.3309
71.70904	82.3842	371.1022
13.82497	13.57812	185.7246
27.61229	16.27457	117.5737
37.53664	51.9417	168.9898
441.9211	265.2885	696.5949
50.09159	150.5714	550.9105
47.06166	171.4217	224.7604
3.752047	2.998249	191.9266
40.14482	69.7926	407.7342
1.809855	46.71836	215.6133
7.254417	42.41679	193.4215
1.075879	7.642631	150.1132
22.45518	73.72295	334.2683
113.8167	84.35639	483.9228
37.75271	251.8315	319.4582
70.74558	39.54529	450.789
55.58916	154.0807	476.7827
191.8165	198.5934	258.4653
35.7774	29.98192	231.878
21.72335	155.0001	367.1031
43.71393	197.3517	487.4687
18.35105	256.9912	278.2539
85.07306	334.1087	532.6584
262.8237	78.27474	484.2632
12.34403	26.53437	98.95171
7.146642	25.92019	341.1421
32.92591	154.3408	260.0865
12.74421	7.199771	257.0993
179.6528	234.2539	463.4871
351.5504	405.2429	672.6177
53.38737	87.35799	208.709
65.25637	130.8566	608.7428
146.9476	190.0862	321.3671
22.77773	20.92645	122.6571
42.13773	27.02435	209.0603
26.2721	38.75412	278.6517
9.12805	56.23531	227.4984