Using Copulas to Model Dependency Structures in Econometrics

Donald J. Erdman, Arthur Sinko, SAS Institute Inc., Cary NC

ABSTRACT

Modeling the correlation structures of economic variables is an important part of managing financial risk. Pearson correlation matrices fully characterize joint distribution when the underlying economic variables follow a multivariate normal distribution. When the economic variables are nonnormal, copulas are needed to model the correlation structure. In fact, given a copula and the marginal distributions, you can recover the joint probability density function. This fact can be used to estimate models separately and combine them for simulation. This paper introduces advanced copula modeling capabilities in the MODEL procedure. We also show how insight into the correlation structure of the copulas can be obtained by using animations produced by SAS.

INTRODUCTION

Modeling a system of variables accurately is a difficult task. The underlying, ideal, distributional assumptions for each variable are usually different from each other. An individual variable might be best modeled as a *t* distribution or as a Poisson process. The correlation of the various variables is very important to estimate as well. A joint estimation of a set of variables would make it possible to estimate a correlation structure but would restrict the modeling to single, simple multivariate distribution (for example, the normal). Even with a simple multivariate distribution, the joint estimation would be computationally difficult and would have to deal with issues of missing data.

You can separate a multivariate distribution into a set of marginal distributions and the correlation structure between the marginals. Copula theory is the formalization of the separation of the correlation structure of a multivariate distribution from the marginal distributions that make up the multivariate distribution.

A copula is a joint cumulative density function (CDF) of the d, independent, $U[0, 1]$ distribution functions. Another way to say this is that a copula maps independent uniform distributions to $U[0, 1]$. Mathematically that is:

$$C(u_1, u_2, u_3, \ldots u_d) \sim U[0, 1]$$

The copula framework can be generalized for any collection of marginal distributions and joint distributions. Sklar's theorem states: if H is a joint distribution function with marginal distribution functions $F_1, F_2, F_3, \ldots, F_d$ then there exists a copula C such that

$$H(x_1, x_2, x_3, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d))$$

If $F_1, F_2, F_3, \ldots, F_d$ are continuous, the copula function is unique. It is a convention that the inputs to the copula function are random variables with a uniform distribution. This convention is not a requirement for the theory to hold.

The beautiful part of copula theory is: once you have determined the marginal distributions, you can choose any copula C you desire. Different copulas induce different dependency structures, but the resulting multivariate distributions have the same marginals.

This paper introduces several common copulas and demonstrates their properties for modeling mulivariate distributions.

GAUSSIAN COPULAS

One of the simplest copulas to construct and estimate is the Gaussian copula. To create a bivariate distribution that uses a Gaussian copula and two marginal distributions, F_1 and F_2, you need a correlation matrix, **R**. Once you have a correlation matrix, draws from the multivariate distribution can be obtained as follows:

$$\mathbf{x} = \mathbf{Az}$$

where $\mathbf{z} \in R^d$ is a vector of independent random standard normal draws, $\mathbf{A} \in R^{d \times d}$ is the square root of the correlation matrix, **R**, and **x** is a correlated normal vector.

To obtain the marginals of the distribution, first you transform **x** into a multivariate uniform distribution:

$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} \Phi(x_1) \\ \Phi(x_2) \end{bmatrix}$$

where Φ is the normal CDF function. To produce the original marginals of the multivariate distribution, you compute the following:

$$\mathbf{y} = \left[\begin{array}{c} F_1^{-1}(u_1) \\ F_2^{-1}(u_2) \end{array} \right]$$

Now $\mathbf{y}$ is a draw from a multivariate distribution with marginals F_1 and F_2 and a correlation matrix $\mathbf{R}$.

There are several ways to use SAS to create correlation matrices, but for a normal copula you ultimately need a Pearson correlation matrix. Other correlation estimation methods, such as Kendall's tau and Spearman's, are usually preferred. McNeil, Frey, and Embrechts (2005) have a good discussion on the importance of using a Spearman or Kendall's tau correlation matrix to generate a Pearson correlation matrix.

If a Spearman or Kendall's tau correlation matrix is estimated, then a Pearson correlation matrix is easily calculated as

$$A_{ij} = 2\sin\left(\frac{\pi}{6}\rho_{i,j}^{spearman}\right)$$

for Spearman's correlation and

$$A_{ij} = \sin\left(\frac{\pi}{2}\rho_{i,j}^{tau}\right)$$

for Kendall's tau.

The following SAS statements perform a transformation of a Kendall's tau correlation matrix to a Pearson correlation matrix:

```
proc corr data=resid kendall outk=covken;
var y z;
run;
```

```
data pearson;
   set covken;
   y=sin(constant('pi')*y/2);
   z=sin(constant('pi')*z/2);
run;
```

OTHER COPULAS

Gaussian copulas are easy to compute and estimate but their linear correlation structure and thin tail structure make them less suitable for certain tail dependencies. Accurate tail dependency modeling is especially important for financial modeling—in areas like operational, market, or credit risk modeling. In the following subsections, the *t*, normal mixture, and archimedian copulas are introduced.

t COPULA

The copula based on the multivariate *t* distribution is better able to model extreme joint events. The following statements use a copula based on a *t* distribution to create a bivariate distribution with uniform marginals:
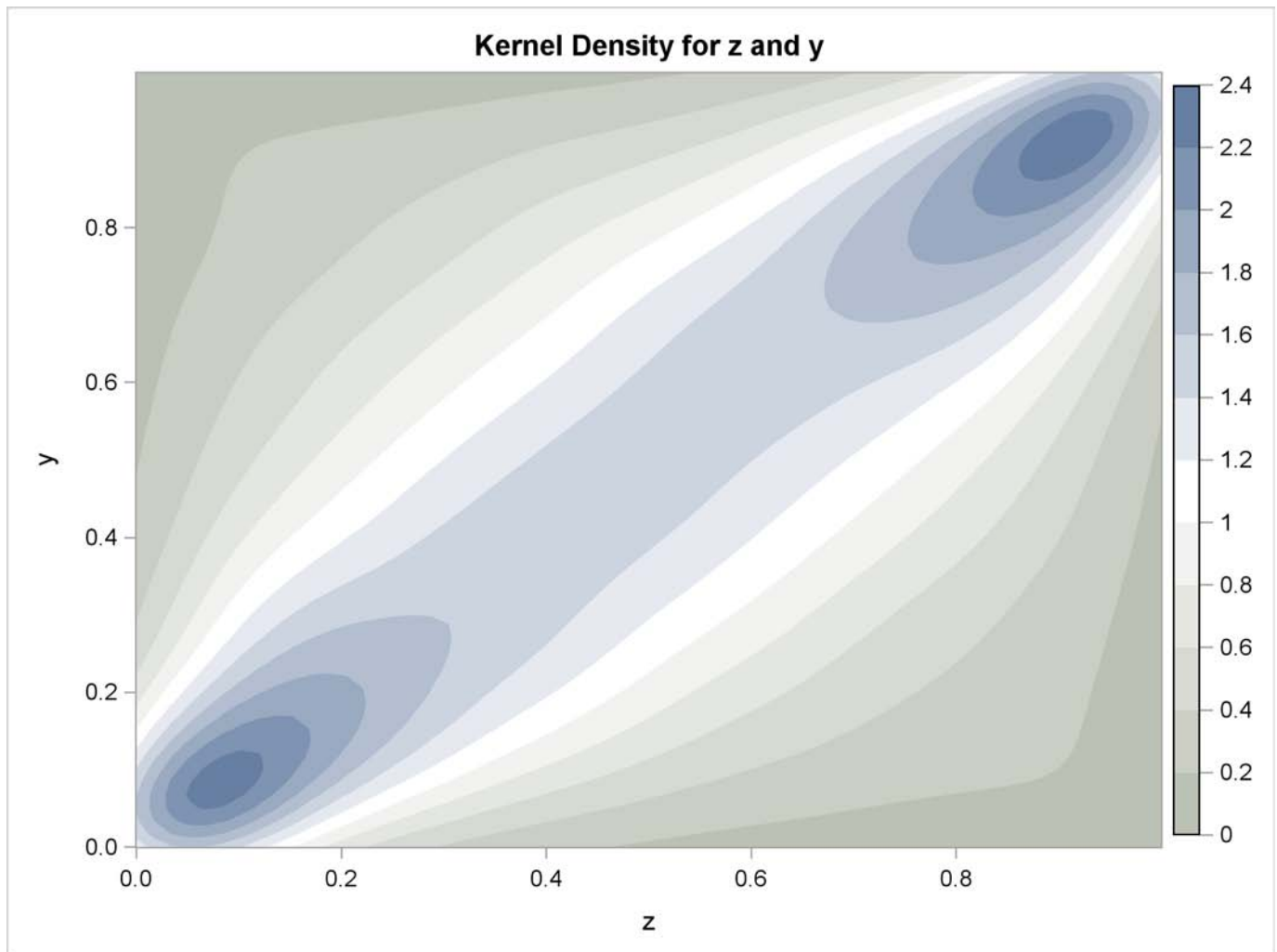
```
data histdata;
      y = .5; z = .5;
run ;
data sdata;
      _type_ = "cov";
      y = 1; z = .65;
      _name_ = "y";
      output;
      y = .65; z = 1;
      _name_ = "z";
      output;
run;

proc model out=sim data=histdata sdata=sdata;
   y = 0;
```

2

```
    errormodel y ~ Uniform(0,1);

    z = 0;
    errormodel z ~ Uniform(0,1);

    solve y z / random=10000 seed=12345 copula=(t(5));
run;
```
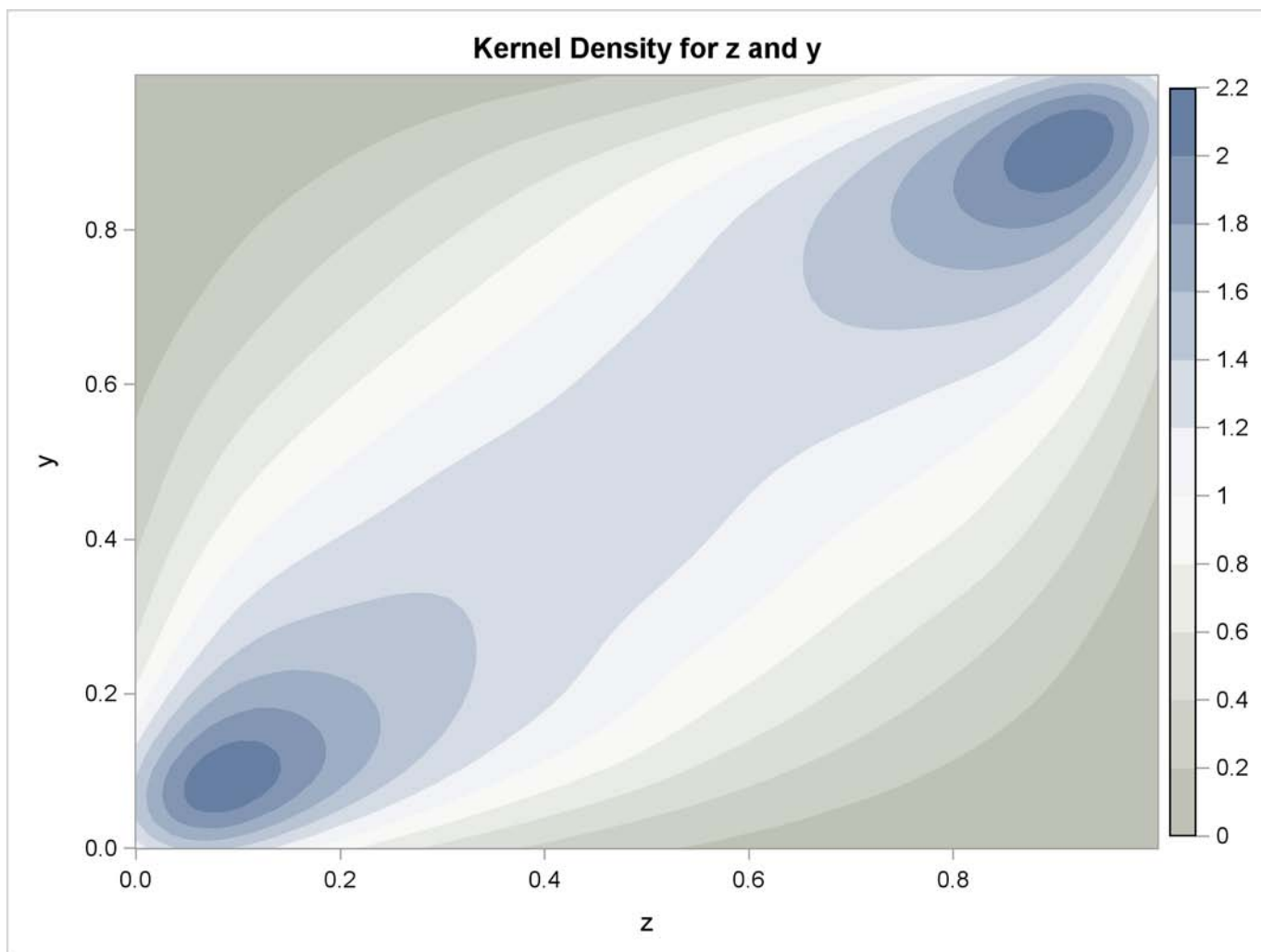
Figure 1 shows the contour plot for a bivariate unform with a correlation structure specified by a *t* copula.

Figure 1 *t* Copula

Replacing COPULA=(T(5)) with COPULA=NORMAL in the SOLVE statement produces a Gaussian copula. The contour plot for the Gaussian copula is shown in Figure 2:

Figure 2 Gaussian Copula

Note that the marginals for both distributions are uniform and the only difference is the structure of the correlation. The *t* copula produces more extreme events. The statements to generate the graphs are provided in the appendix.

NORMAL MIXTURE COPULAS
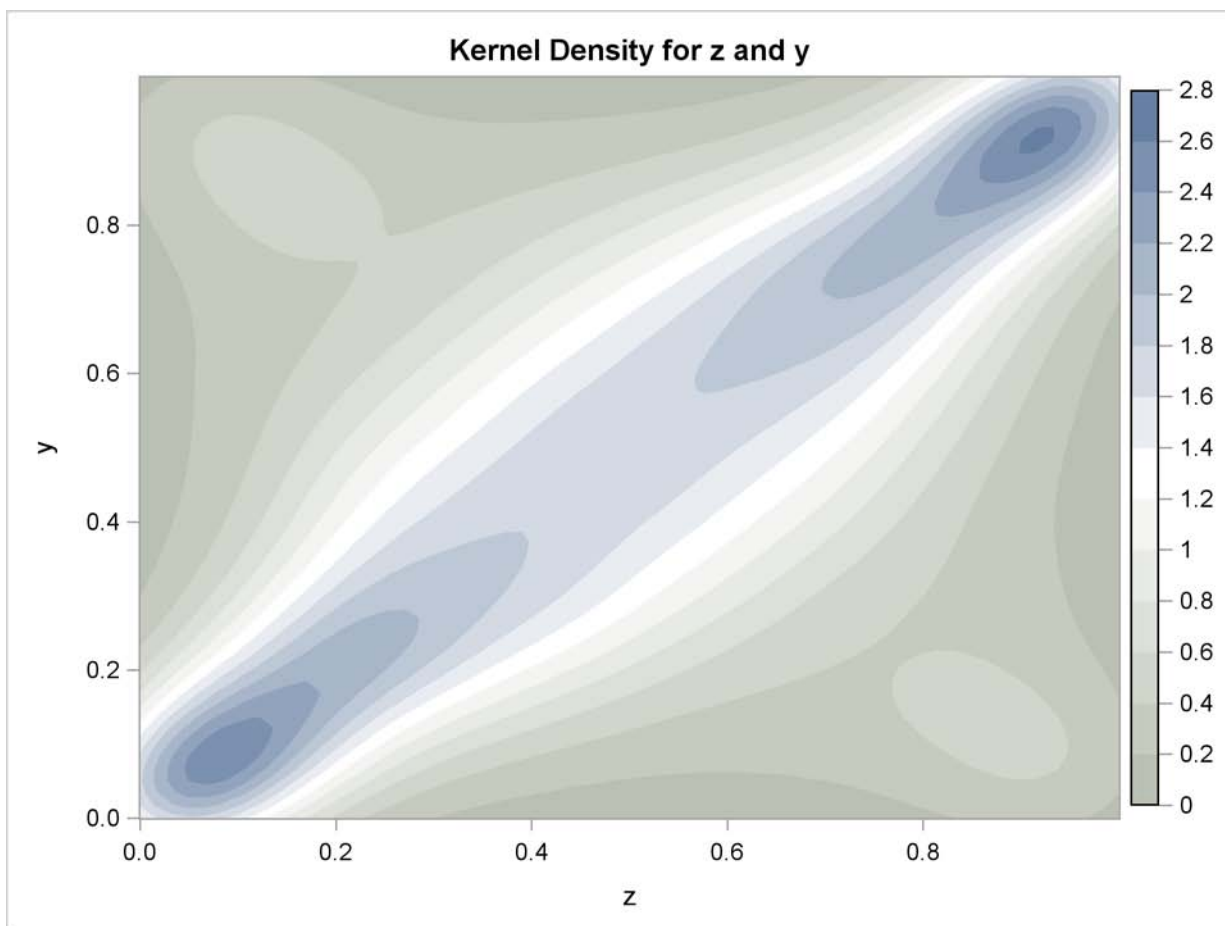
For the normal mixture and *t* copula, a draw is created as

$$\mathbf{x} = \sqrt{w}\mathbf{Az}$$

where w is a scalar random variable. If $W \sim$ inverse gamma$(df/2, df/2)$, then $\mathbf{x}$ is multivariate *t*. When COP-ULA=NORMALMIX is specified, w is distributed as a step function with each of the n positive variances, $v_1 \ldots v_n$, having probability $p_1 \ldots p_n$.

The correlation matrix $\mathbf{R} = \mathbf{A}'\mathbf{A}$ is specified with the SDATA= option.

Figure 3 shows a surface plot from a multivariate normal mixture copula.

Figure 3 Normal Mixture Copula

ARCHIMEDIAN COPULAS

So far, the copulas discussed are based on distributions with explicit covariance matrices. There are many other copulas discussed in the literature that are parameterized by coefficients rather than a covariance matrix. The class of Archimedian copulas have simple parametric forms. The bivariate Clayton copula is an example of an Archimedian copula and has the following form:

$$C_{Clayton}(u_1, u_2) = \exp\left(-\left((-\ln(u_1))^\theta + (-\ln(u_2))^\theta\right)^{1/\theta}\right)$$

where $1 \leq \theta \leq \infty$. Other copulas in this class include the Gumbel and Frank copulas. There are multivariate forms of these copulas as well.

The following subroutine computes a draw from a Clayton copula:

```
proc fcmp outlib=sasuser.copula.copula;

subroutine clayton_copula(theta,u[*]);
   outargs u;

   Dimension=dim(u);
   if theta > 1 or theta <= 0 then do;
      put "ERROR: Bad theta value to Clayton_Copula" theta=;
      return;
   end;

   r1 = rand("GAMMA", 1/theta);
```
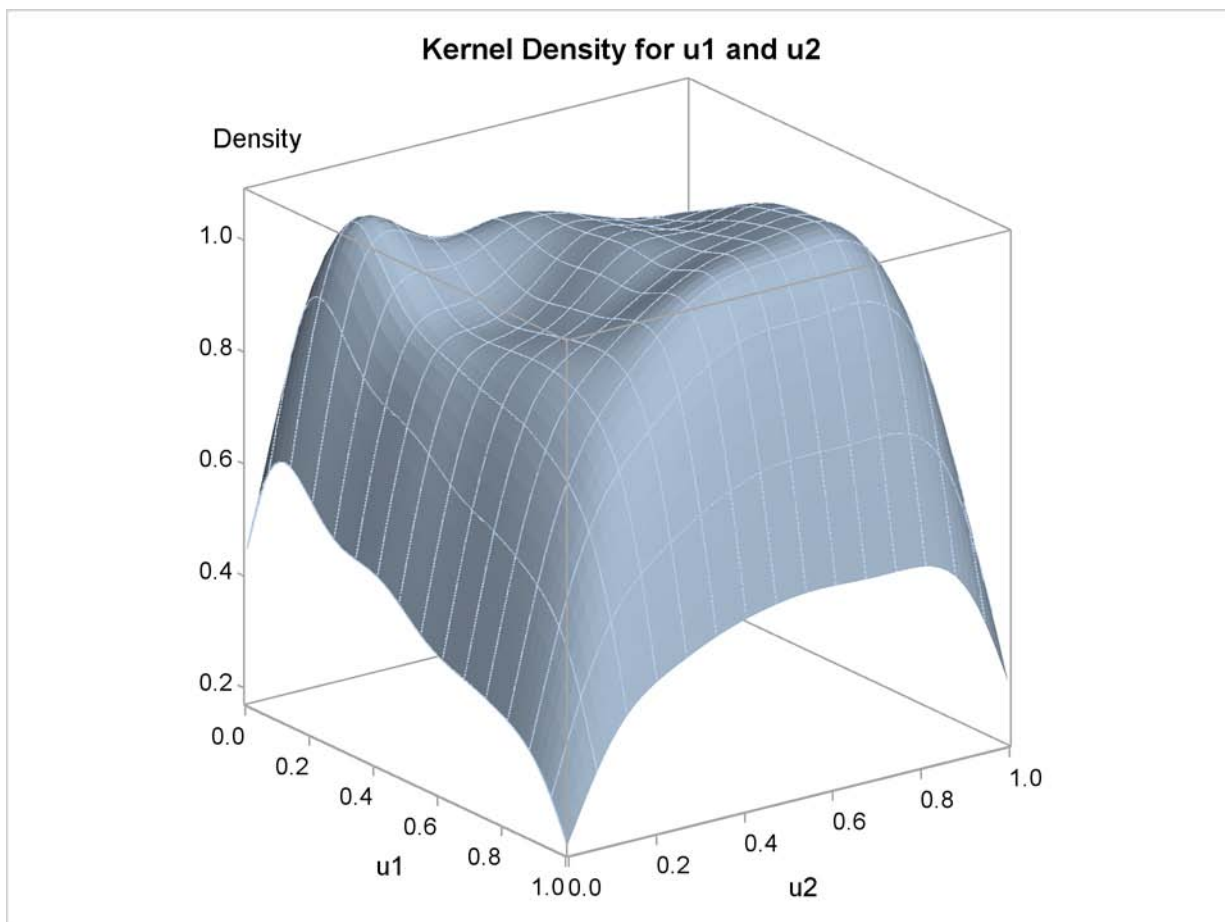
5

```
    do i=1 to Dimension;
        u[i] = (rand("EXPONENTIAL")/r1 + 1)**(-1/theta);
    end;
endsub;
```

This subroutine was used to produce the surface plot shown in Figure 4.

Figure 4 Clayton Copula

ASYMMETRICAL COPULAS

Asymmetrical parameters can be added to the Gaussian, normal mixture, and *t* copulas to further modify the correlation structure. Asymmetrical parameters enter into the generator as follows:

$$\mathbf{x} = w\boldsymbol{\gamma} + \sqrt{w}\mathbf{Az}$$

where w is a scalar random variable and $\boldsymbol{\gamma} \in R^d$ is a vector of asymmetry parameters. $\boldsymbol{\gamma}$ is specified in the SDATA= data set with _TYPE_=ASYM. The ASYM option specifies that the nonzero asymmetry vector, $\boldsymbol{\gamma}$, is to be used. The Gaussian, *t*, and the normal mixture copula and their corresponding asymetrical forms are available in the MODEL procedure.

In the following statements, an asymmetrical *t* copula is used to correlate two uniform distributions. The asymmetrical parameter is varied over a range of values to demonstrate its effect. The resulting graphs are produced by using ODS graphics and are shown in Figure 5.

```
data histdata;
    do asym = -1.3 to 1.1 by .3;
        date='01aug2007'd;
```

```
      y = .5;
      z = .5;
      output;
   end;
run ;


   /* Add the asymmetric parameter to cov mat */
data asym;
   do asym = -1.3 to 1.1 by .3;
      y = asym;
      z = 0;
      _name_ = "     ";
      _type_ = "asym";
      output;
      y = 1;
      z = .65;
      _name_ = "y";
      _type_ = "cov";
      output;
      y = .65;
      z = 1;
      _name_ = "z";
      _type_ = "cov";
      output;
   end;
run;


proc model out=sim(where=(_REP_ > 0)) data=histdata sdata=asym;
   y = 0;
   errormodel y ~ Uniform(0,1);

   z = 0;
   errormodel z ~ Uniform(0,1);

   solve y z / random=500 seed=12345 copula=(t(5) asym );
   by asym;
run;
```

To produce a panel plot of this joint distribution, use the following ODS Graphics statements.
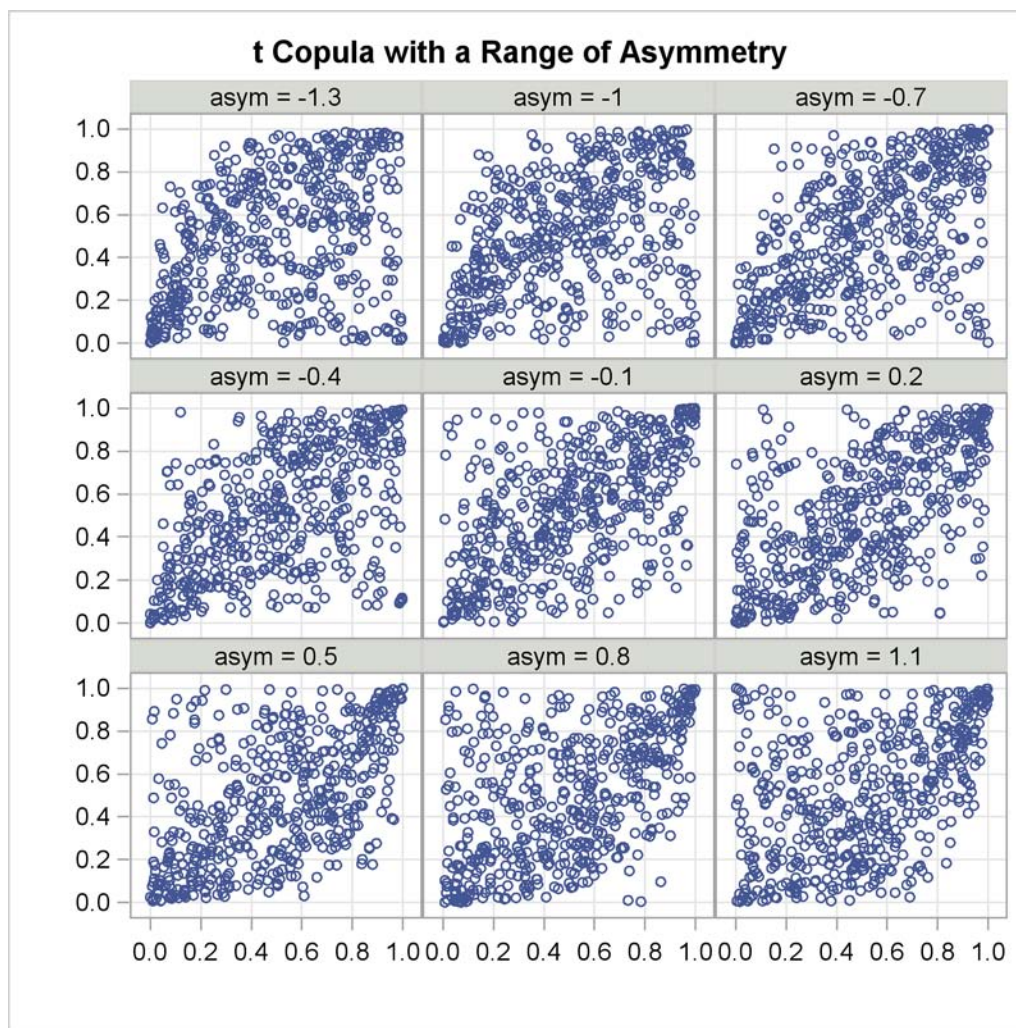
```
ods graphics on / height=800 width=800;
proc template;
  define statgraph myplot.panel;
  BeginGraph;
    entrytitle halign=left halign=center
          textattrs=GRAPHTITLETEXT "t Copula with a Range of Asymmetry";

    layout datapanel classvars=(asym) / rows=3 columns=3
          order=rowmajor height=1024 width=1420
          rowaxisopts=(griddisplay=on label=' ')
          columnaxisopts=(griddisplay=on label=' ');
      layout prototype;
          scatterplot x=z y=y ;
      endlayout;
    endlayout;
  EndGraph;
  end;
run;


proc sgrender data=sim template='myplot.panel';
run;
```

Figure 5 *t* Copula with Asymmetry

APPLICATION

This section demonstrates the importance of selecting the correct copula. In the simplest case, the value of a financial portfolio is the sum of the values of the individual positions in the portfolio.

$$Value = \sum_i Price(position_i)$$

The price of a position is a function of market factors. For fund managers, diversification is very important. This means that the positions (stocks, bonds, options, and so on) that the fund contains are either independent of each other or move in different directions due to market changes. Diversification reduces the risk of a devastating loss as the result of market changes.

Given a portfolio with two positions, A and B, whose prices depend linearly on two market factors, a and b, you can compute the value of the portfolio as:

$$Value = \alpha a + \beta b$$

For this example, assume that α and β are 1 and the marginal distributions of a and b are $U(0, 1)$. Figure 6 and Figure 7 compare the distribution of $Value$ for a t and Gaussian copulas:

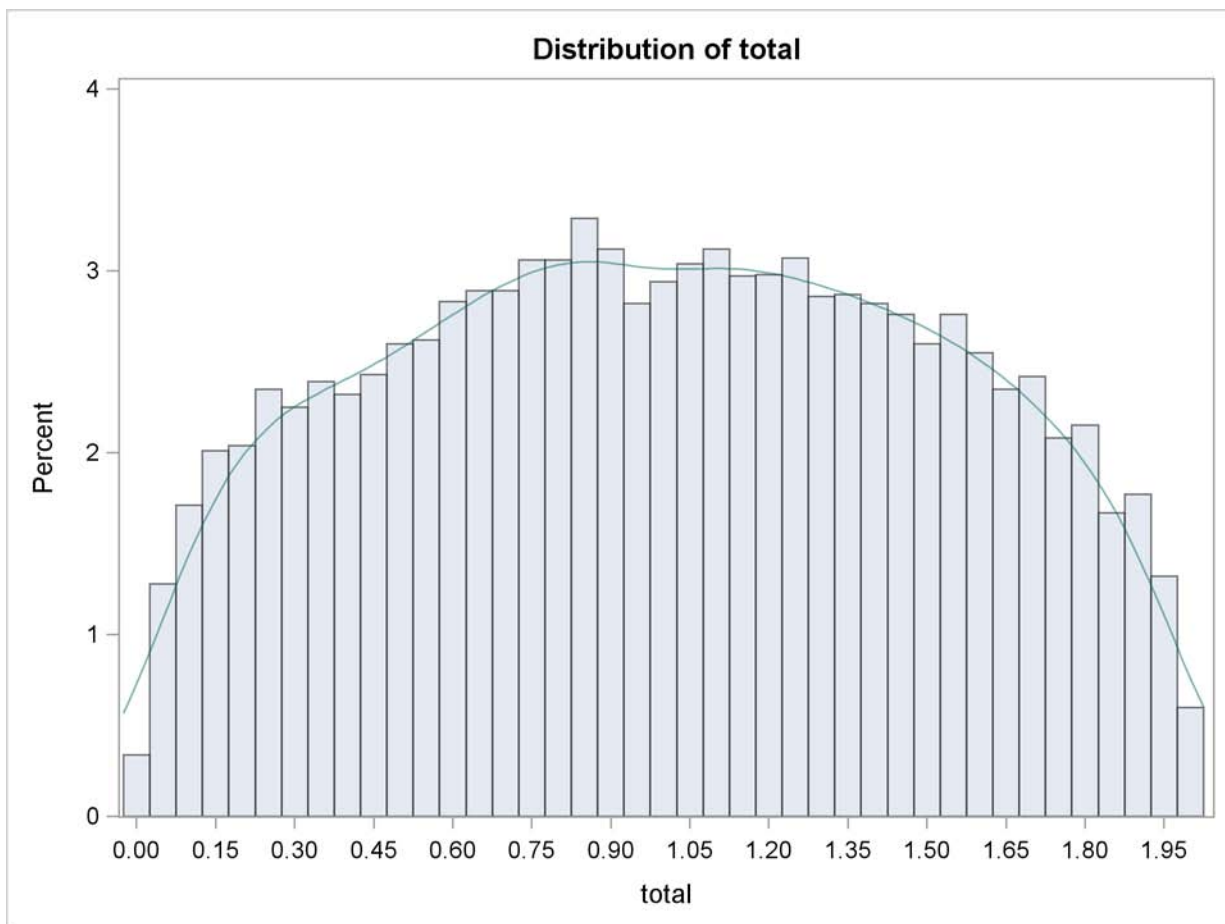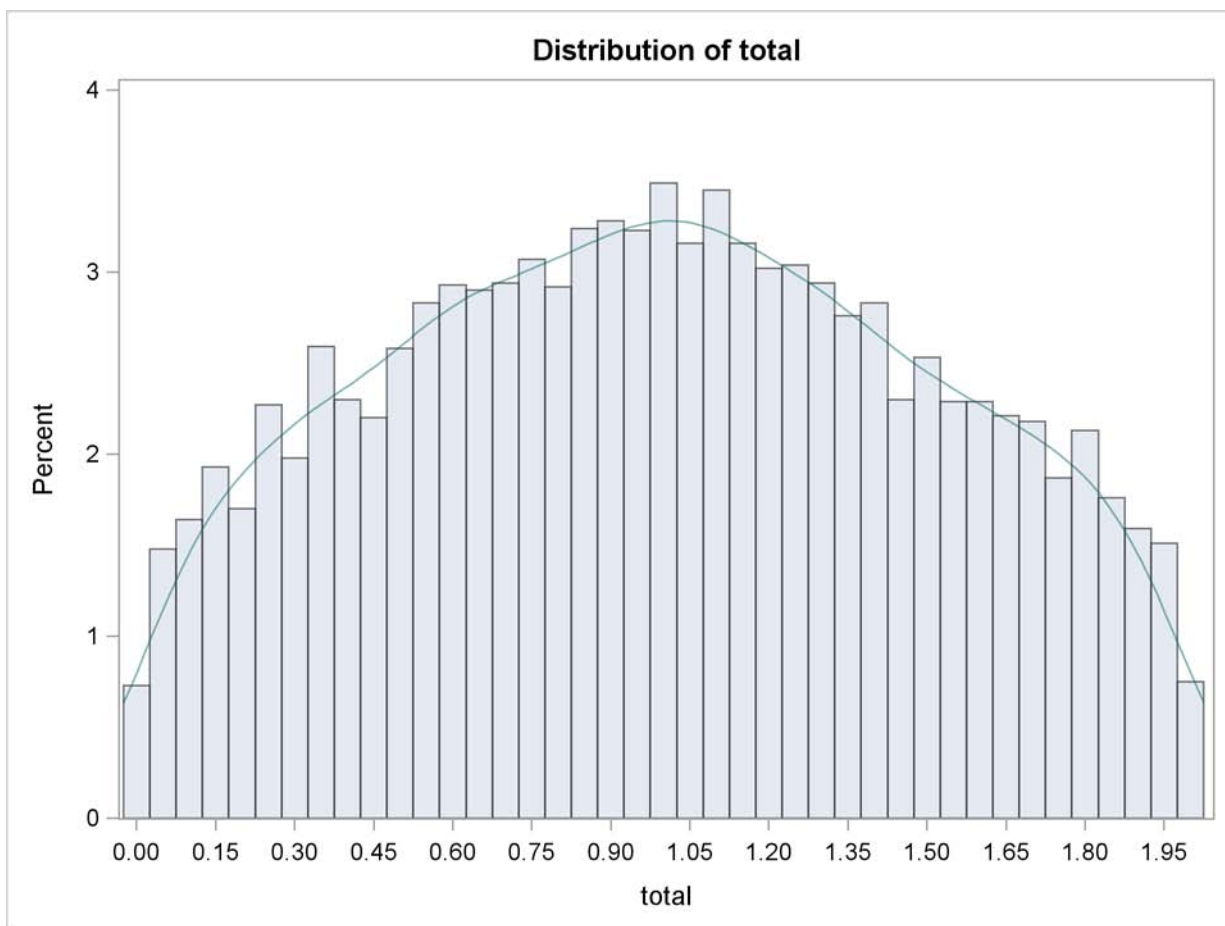**Figure 6** Portfolio with a Gaussian Copula

Figure 7 Portfolio with a *t* Copula

Fund managers are concerned about tail events, and the preceding graphs clearly show that the *t* copula reports many more tail events. Note that the correlation matrix was the same for both multivariate distributions and the marginal distributions are identical. The only thing that was changed was the chosen copula. If fund managers use the wrong copula, they could incorrectly forecast the amount of reserve capital they need by several millions of dollars.

SUMMARY

The theory of copulas is what enables the MODEL procedure to combine and simulate multivariate distributions with different marginals. Copulas form a useful concept for modeling many multivariate distributions. They cannot be used to model all multivariate distributions.

APPENDIX

The following statements were used to generate draws from a bivariate uniform distribution with three different copulas. The KDE procedure is used to display the contour plots of the distributions.

```
proc model data=histdata;
   y = 0;
   errormodel y ~ Uniform(0,1);

   z = 0;
   errormodel z ~ Uniform(0,1);

      /* t copula --------------------*/
   solve y z / random=10000 seed=12345 copula=(t(3))
         out=simt(where=(_REP_ > 0)) sdata=sdata;

      /* Gaussian copula --------------------*/
   solve y z / random=10000 seed=12345 copula=(normal)
         out=simn(where=(_REP_ > 0)) sdata=sdata;

      /* Normal mixture copula --------------*/
   solve y z / random=10000 seed=12345
         copula=(normalmix(3, .5, .3, .2, 1, 10, 100))
         out=simnm(where=(_REP_ > 0)) sdata=sdata;
run;

ods graphics on;
title "t Copula";
proc kde data=simt;
   bivar z y / levels bwm=1.5 gridl=0 gridu=1;
run;

title "Gaussian Copula";
proc kde data=simn;
   bivar z y / levels bwm=1.5 gridl=0 gridu=1;
run;

title "Normal Mixture Copula";
proc kde data=simnm;
   bivar z y / levels bwm=1.5 gridl=0 gridu=1;
run;
ods graphics off;
```

The following statements were used to create draws from a Clayton copula with parameter equal to 0.2:

```
options cmplib=sasuser.copula;
data clayton;

  array u[2] _temporary_;
  call streaminit(37583);

  do i = 1 to 10000;
     call clayton_copula(.2, u);
     u1 = u[1]; u2 = u[2];
     output;
  end;
  keep u1 u2;
run;
```

The following statements create a surface plot for the Clayton copula.

```
ods graphics on;
title "Clayton Copula";
proc kde data=clayton;
   bivar u1 u2 / levels bwm=1.5 gridl=0 gridu=1 plots=surface;
run;
ods graphics off;
```

The following statements were used to make one of the movies in the presentation:

```
    data histdata;
        do i = 1 to 1000;
            date=intnx("Month", '1jul90'd, i );
            y = .5;
            z = .5;
            output;
        end;
        drop i;
    run;


      /* Add the asymmetric parameter to cov mat */
    data asym;
        do asym = -1 to 1 by .1;
            y = asym;
            z = 0;
            _name_ = "     ";
            _type_ = "asym";
            output;
            y = 1;
            z = .65;
            _name_ = "y";
            _type_ = "cov";
            output;
            y = .65;
            z = 1;
            _name_ = "z";
            _type_ = "cov";
            output;
        end;
    run;


proc model out=sim(where=(_REP_ > 0)) data=histdata sdata=asym;
    y = 0;
    errormodel y ~ Uniform(0,1);

    z = 0;
    errormodel z ~ Uniform(0,1);

    solve y z / random=5000 seed=12345 copula=(t(5) asym );
    by asym;
run;

/* assign the destination for the output */
filename out 'tmovie.gif';

/* set the graphics environment */
goptions reset=global gunit=pct border
    cback=ligr ctext=black colors=(blue green red) ftext=swissb
    ftitle=swissb htitle=6 htext=5;

goptions display dev=gifanim gsfname=out gsfmode=replace
        iteration=0 delay=100;

symbol value=dot;

proc gplot data=sim;

    title "t Copula with Asymmetry";
    plot y * z;
    by asym;
run;

data _null_;
    file out recfm=n mod;
    put '3B'x;
run;

goptions reset=all;
```

REFERENCES

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

Sklar, A. 1959. "Fonctions de Répartition à n Dimensions et Leurs Marges," *Publications de l'Institut de Statistique de L'Université de Paris* 8, 229–231.

McNeil, A.J., Frey, R., and Embrechts, P. 2005. *Quantitative Risk Management: Concepts, Techniques, Tools*. Princeton, NJ: University Press.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Don Erdman
SAS Institute Inc.
100 SAS Campus Drive, R5214
Cary, NC 27513
Work Phone: (919) 531-7685
E-mail: Donald.Erdman@sas.com
Web: www.sas.com