Agribusiness Analysis and Forecasting
Mixed Marginal Distributions

Henry Bryant
Texas A&M University
Critical to appropriately reflect dependence among multiple variables

MV normal is one approach to reflecting dependence

Under MV Normal, all marginal distributions in the system are normal (not ideal)

Under MV Normal, dependence among variables is strictly linear (not ideal)
This topic: alternative approach to modeling dependence

The key is separating the modeling of the individual marginal distributions and the modeling of the dependence

The dependence is modeled using MV normal
Overview

\[ Y \]

mixed marginals
Overview
Overview
Review of CDF and Inverse CDF

\[ Y \]

RV with CDF \( F(y) \)
Review of CDF and Inverse CDF

RV with CDF $F_Y$
Review of CDF and Inverse CDF

$U \sim \text{uniform}(0,1)$

$F(y)$

RV with CDF $F(y)$
Review of CDF and Inverse CDF

\[ U \sim \text{uniform}(0,1) \]

\[ \text{RV with CDF } F(y) \]
Preparation

historical data

\( Y \)
Preparation

- Historical data
  - Y
  - $F(y)$
  - U

Diagram: Preparation
Preparation

historical data → Y → F(y) → U → \Phi^{-1}(u) → \Theta → Z

standard uniform

standard normal
Preparation

historical data

standard uniform

standard normal

$F(y_1)$

$F_2(y_2)$

$\Phi^{-1}(u_1)$

$\Phi^{-1}(u_2)$

$z_1$

$z_2$
Preparation

historical data

standard uniform

standard normal

sample covariance
Preparation

- Marginal distributions for each variable $Y_i$ are reflected in the individual $F_i$ (equivalently $F_i^{-1}$), which were specified individually and separately from one another.

- The dependence among the variables is captured in the sample covariance matrix ($\Sigma$) for the $z_i$. 
Simulation

\[ z = \sqrt{\sum z^2} \]
Simulation

\[
\mathbf{z} = \sqrt{\Sigma} \mathbf{z}'
\]

\[\mathbf{z}_1 \quad \mathbf{z}_2\]
Simulation

\[ Z = \sqrt{\Sigma} \, Z' \]

\[ \Phi(2) \]

\[ U_1, \ U_2 \]
Simulation

\[ Z = \sqrt{\sum Z^2} \]

\[ Z_1 \]
\[ \Phi(2) \]
\[ U_1 \]
\[ F_{1}^{-1}(u_1) \]
\[ Y_1 \]

\[ Z_2 \]
\[ \Phi(2) \]
\[ U_2 \]
\[ F_{2}^{-1}(u_2) \]
\[ Y_2 \]
Simulation

\[ Z = \sqrt{\sum Z^2} \]

\[ \Phi(Z), \Phi(Z), F^{-1}_1(u_1), F^{-1}_2(u_2) \]

\[ Y_1, Y_2 \]
Dependent $U$ values

- The values we have generated using this process for the $U$ variables reflect dependence among our $Y$ variables.
Independent Bivariate $U$ Draws
Dependent Bivariate $U$ Draws

Gaussian Copula, $\rho = 0.9$
Inverse CDFs Using $U$ draws

- So far, we have mostly used Simetar’s functions to generate stochastic draws without using the arguments for specifying $U$ values.

- Invisibly in the background, Simetar generated the $U$ values automatically and *independently for each variable*.

- To implement the last step in the simulation of mixed marginals, we will need to manually pass our *non-independent* $U$ values to the inverse CDF functions.
Inverse CDFs Using $U$ draws

- $\text{NORM}(\mu, \sigma, u)$
- $\text{UNIFORM}(\text{min}, \text{max}, u)$
- $\text{BETAINV}(u, \alpha, \beta, \text{min}, \text{max})$
- $\text{EMPIRICAL}($historical sample$, u)$
Empirical CDF

- So far, we used Simetar’s EMPIRICAL function to generate a stochastic draw (a y value) using a u value (either implicitly or, now, explicitly).
- That is, we have been applying \( F^{-1}(u) \) for the empirical distribution.
- In the preparation phase of a mixed marginals analysis, (if we specify an empirical distribution for one or more of our variables), we need to apply \( F(y) \).
- This can *almost* be accomplished using an Excel function: \( u = \text{PERCENTRANK}(\text{historical sample, } y) \).
- Unfortunately this will generate values of exactly one and zero that cannot be used by the inverse standard normal CDF.
- Instead, use

\[
\text{MAX}(0.001, \text{MIN}(0.999, \text{PERCENTRANK}(\text{historical sample, } y)))
\]