More Eyeball Tests of Copula Fits

Several descriptive functions of copulas have previously been proposed for understanding the behaviors of the different copulas and for comparing fitted and empirical copulas. One more is discussed here as are ways to do PP-plots of copula fits.

Two criteria for eyeball tests are used:

1. You can readily tell a good fit from a bad fit
2. You can tell the difference between a good fit and a pretty good fit.

Median Regression

Let $C_1(v|U=u)$ be the conditional distribution function for v given u. This is the derivative of the copula wrt u. Then define the median regression P(z) by $C_1(P(z)|U=z) = ½$. This is the conditional median for V when U=z. If the conditional distribution is invertible, $P(z) = C_1^{-1}(½|z)$ can be calculated in closed form.

Although the P(z) functions look fairly different for different copulas, the values for the empirical copula can be quite jumpy, so it is not necessarily easy to tell which parameterized copula fits best. For any point u,v in the empirical copula, usually P(u) = v, as there are not usually repeats of u values.

PP-Plots

A direct PP-plot can be constructed by plotting the value of the fitted copulas as a function of the empirical copula for each data point.

A conditional PP-plot can be formed by calculating the conditional parametric copulas for each data point, sorting these, and plotting the j^{th} sorted point against j/(n+1) where n is the sample size.

PP-Error Plots

Sometimes PP-plots are hard to read because everything can be close to the diagonal. By subtracting the diagonal value from each point you get a graph around zero which then can be shown at a more magnified scale.

Data

First we test with the data of Belguise and Levi, which consists of property and auto insured event costs from 736 windstorms in France. Belguise and Levi

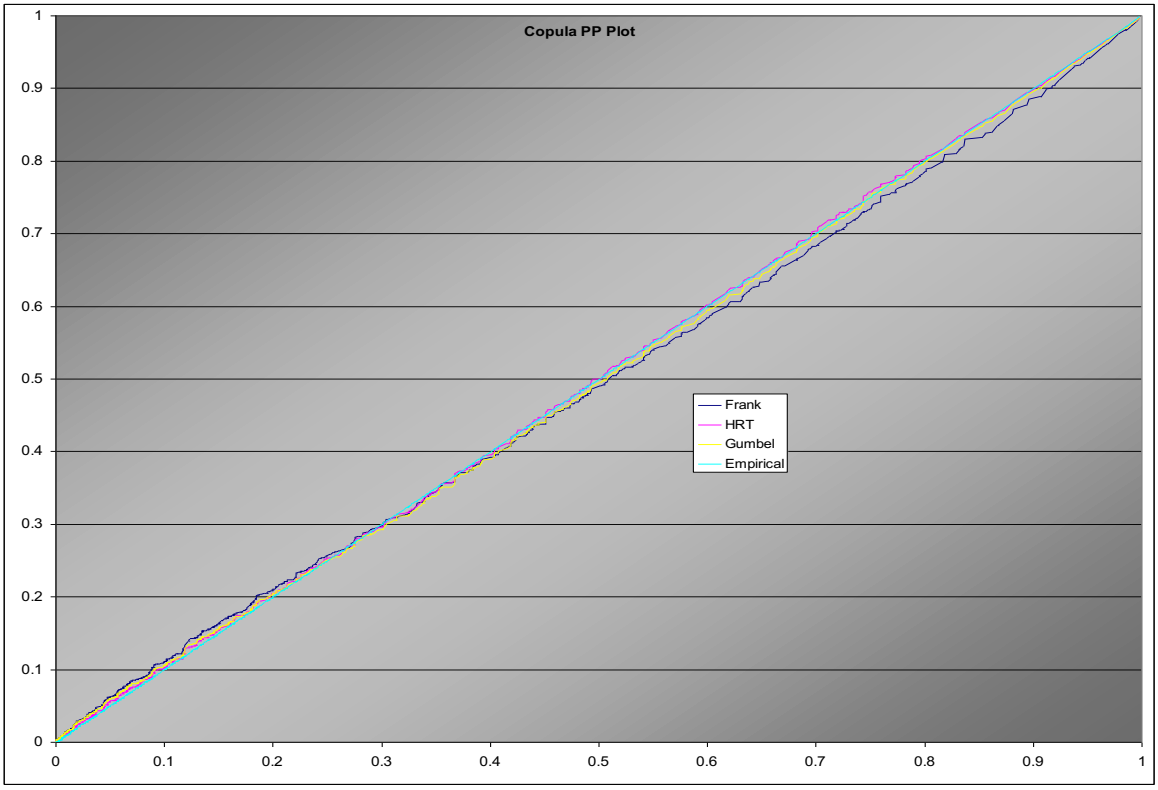
found that for this data the Frank copula was not a very good fit, the Gumbel copula gave a pretty good fit, and the HRT provided a good fit. Charpentier confirmed the latter. Thus we will graph these fits by the various methods.

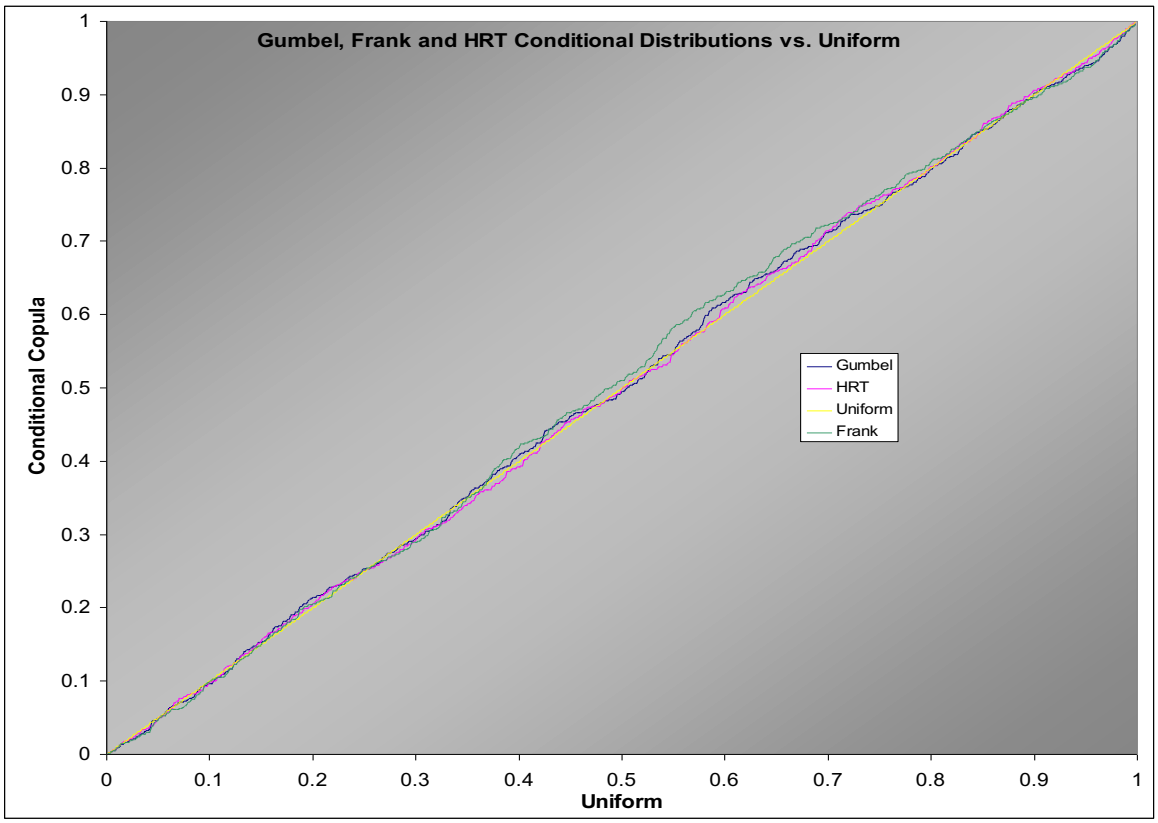
Median Regression

Two problems arise: first, the Gumbel conditional distribution is not invertible, so the medians have to be found numerically. Second, the empirical data jumps around a lot so the empirical medians are difficult to discern. For this problem the median of the 25 points centered at each data point was used (and so was not calculated for the first and last 12 observations).

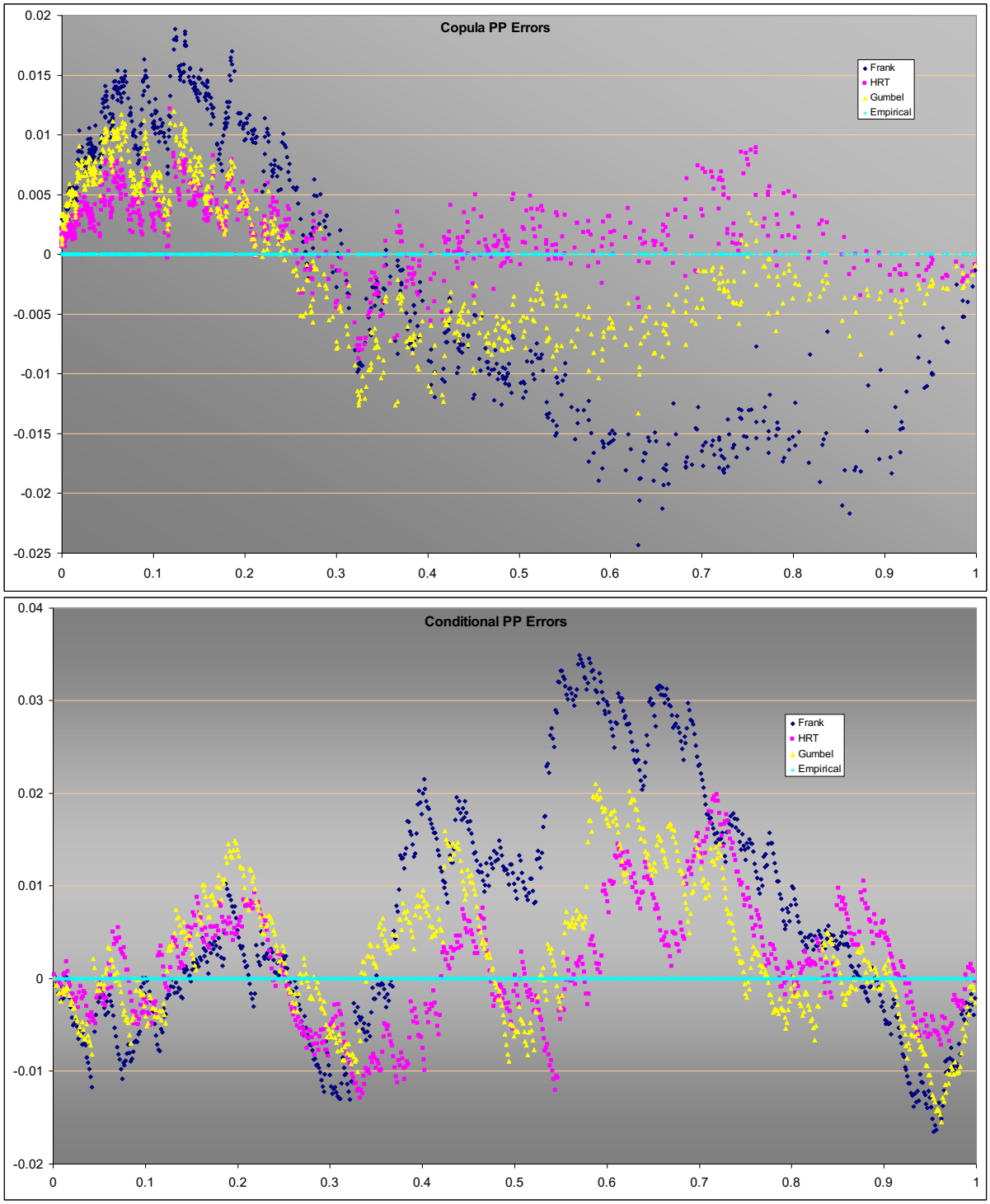
Clearly the Frank copula is a bad fit, so the first criterion is met. Picking between the HRT and Gumbel is not as easy, but in the left tail and extreme right tail the HRT fits better. It also looks to be somewhat closer to the middle of the observations in the middle of the range. So the second criterion seems to be ok.

PP-Plots


Copula PP Plot
Frank
HRT
Gumbel
Empirical
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1


Gumbel, Frank and HRT Conditional Distributions vs. Uniform
Conditional Copula
Uniform
Gumbel
HRT
Uniform
Frank
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

PP-Error Plots



On the PP-plots the Frank fit can be seen to be worse, but you have to look really hard to tell that the HRT fits better than the Gumbel. This is fairly apparent in the PP-error plots, however, more so for the copula than the conditional distribution. So these but not the PP-plots meet both criteria for this data set.