By Sami Umut Can and Roger J. A. Laeven

DETERMINING THE RIGHT TAIL DEPENDENCE MODEL USING R

An important problem in actuarial science and mathematical finance, as well as many other quantitative fields, is to model the statistical behavior of "extreme values" of random quantities. Consider, for example, the losses incurred in successive weeks by a certain portfolio of stocks. What can we say about the maximum weekly loss that will be incurred by a given stock in the portfolio, over a fixed time horizon? And, looking at the portfolio as a whole, what can we say about the *dependence* between the maximum weekly losses in the portfolio? If one stock yields an exceptionally high maximum loss over the coming year, say, should we expect the other stocks in the portfolio to co-move in a similar fashion? Or is it safe to treat the extreme losses by different stocks as independent occurrences? Clearly, questions like these are very relevant from a risk management point of view.

Dr. S. Umut Can (above) is Assistant Professor at Amsterdam University.

Prof. dr. R.J.A. Laeven is Professor at Amsterdam University.

Extreme Values and Tail Dependence

To make the setting more precise, let us consider a portfolio of two stocks only, with weekly losses $X_1,X_2,\dots$ for stock 1 and $Y_1,Y_2,\dots$ for stock 2. (Gains can be treated as "negative losses.") For the sake of simplicity, let's assume that the successive pairs of losses (X_1,Y_1), (X_2,Y_2),... are independent and identically distributed random vectors, with a common bivariate distribution function F. Suppose the X_i's have distribution F_1 and the Y_i's have distribution F_2. Note that we do not assume independence between X_i and Y_i; the joint distribution F determines their dependence structure.

The questions posed in the introduction can be answered by studying the statistical behavior of the pair of maximum losses $(\max\{X_1,\dots,X_n\},\max\{Y_1,\dots,Y_n\})$ over a long time horizon n. Now, the distribution of this random vector will not be very interesting without some form of normalization because each component will simply converge in probability to the right endpoint of the corresponding marginal distribution as n increases. So let us instead consider the pair of *normalized* maximum losses

(1) $$\left(\frac{\max\{X_1,\dots,X_n\}-a_1(n)}{b_1(n)},\ \frac{\max\{Y_1,\dots,Y_n\}-a_2(n)}{b_2(n)}\right)$$

where a_1,b_1,a_2,b_2 denote deterministic normalizing sequences.

Extreme value theory, the branch of statistics that deals with the behavior of sample maxima (among other things), tells us that when these normalizing sequences are chosen properly, the vector of the normalized maxima displayed above converges to a non-trivial distribution function G of a special type. (Note: Such a convergence is not guaranteed for *all* distribution functions F, but for a fairly large class of them. We assume that our underlying F belongs to that large class.) In particular, the marginal distributions of G have the so-called *generalized extreme value form*

(2) $$G_j(x)=\exp\{-(1+\gamma_j x)^{-1/\gamma_j}\},\quad 1+\gamma_j x>0,\quad j=1,2,$$

for some real numbers γ_1,γ_2, and the dependence structure between these marginal distributions is captured in full by the so-called *tail copula*, defined by

(3) $$R(x,y)=\lim_{t\to\infty} tP\,[F_1(X_1)>1-x/t,\ F_2(Y_1)>1-y/t],\quad (x,y)\in[0,\infty)^2.$$

We conclude that the asymptotic joint distribution G of the normalized maxima is completely determined by the marginal parameters γ_1,γ_2 and the tail copula R.

Goodness-of-Fit Testing for Tail Copulas

In practice, often a parametric model is used for the unknown tail copula R. That is, R is simply assumed to belong to some parametric family $\mathcal{R}$ of tail copulas (such as the logistic family), which reduces the problem of estimating R into a simpler problem of estimating a parameter. Consequently, the problem of testing whether a particular parametric tail copula model fits the data well or not becomes important.

In [1], we propose a novel approach for constructing goodness-of-fit tests for parametric tail copula models. To be precise, suppose that we want to test the goodness-of-fit of the parametric family $\mathcal{R}=\{R_\theta:\theta\in\Theta\}$ to the bivariate data at hand, where Θ denotes some (possibly multi-dimensional) parameter space. Assuming that the unknown tail copula R indeed belongs to $\mathcal{R}$ (the null hypothesis), we can estimate R in two different ways: parametrically, by estimating θ, and semi-parametrically, by estimating the marginal parameters a_j,b_j,γ_j for $j=1,2$ and using empirical distribution functions. Let's call these two estimators $R_{\hat\theta}$ and $\hat R_n$, respectively. Under the null hypothesis, both estimators are estimating the same object, and they should get "closer" to each other as n gets larger. So we consider the bivariate process η_n, which is simply the properly scaled difference between $R_{\hat\theta}$ and $\hat R_n$ We show that η_n converges weakly to some bivariate Gaussian process that we characterize explicitly. More importantly, we also show that a proper transformation of η_n, say $\tau(\eta_n)$, converges weakly to a *standard* bivariate Wiener process.

Thus we construct a bivariate process $\tau(\eta_n)$ from the data which is approximately a standard Wiener process if the null hypothesis is true, i.e. if $\mathcal{R}$ is a good model for the tail dependence of the data at hand. So we now have a nice way of assessing the goodness-of-fit of $\mathcal{R}$ to the data: check if the observed sample path for $\tau(\eta_n)$ is consistent with the statistical behavior of a standard Wiener process or not. There are standard test statistics one can use for such a comparison, such as Kolmogorov-Smirnov (KS), Cramér-von Mises (CvM) and Anderson-Darling (AD) test statistics.

A critical feature of this approach to goodness-of-fit testing is that it is "asymptotically distribution free," that is, the asymptotic distribution of a given test statistic is always the same when the null hypothesis is true, no matter which particular parametric family $\mathcal{R}$ we are testing for and what the true value of the parameter θ happens to be. So benchmark distribution tables for test statistics only need to be computed once and for all, as opposed to being recomputed again and again for different null hypotheses, which is the case in competitor approaches.

Implementation Using R

To assess the finite-sample performance of our convergence results, we have conducted a Monte Carlo simulation study using the statistical software R. We started by selecting a few different tail copula models to serve as our null hypotheses. For each of these models, we generated 300 data sets of size 1500 from a bivariate distribution F_0 for which the model is correct. From each generated data set, we constructed the process $\tau(\eta_n)$ on a suitable bivariate grid, and computed the values of three different test statistics (KS, CvM, AD) from this discretized sample path. This gave us an empirical distribution of 300 values for each of the three test statistics. We compared these empirical distributions with the empirical distributions of the same test statistics generated from 10,000 true standard Wiener process paths. We observed a very good match between the two sets of empirical distributions, as predicted by our convergence results. The PP-plots (in black) for the three test statistics, for one of the models we considered, are reprinted in Fig. 1 below. They stay very close to the 45° line, which verifies that the compared empirical distributions agree to a remarkable extent. We also generated 100 samples of size 1500 from a judiciously chosen alternative distribution F_a that does not satisfy the null hypothesis model, and computed the same test statistics from these samples. The resulting PP-plots (in red) can be seen to deviate quite significantly from the 45° line, suggesting that tests based on our approach have high power. The R script we used for the simulation study is available as an appendix to [1].

Conclusion

In [1], we propose a novel approach for determining the right tail dependence model, an important problem in actuarial and financial risk modeling. We hope practitioners will exploit it to improve their risk modeling techniques. ⏮

References

[1] Can, S.U., Einmahl, J.H.J., Khmaladze, E.V. and Laeven, R.J.A. (2015). Asymptotically Distribution-Free Goodness-of-Fit Testing for Tail Copulas. Ann. Statist. **43** 878-902. DOI: 10.1214/14-AOS1304.

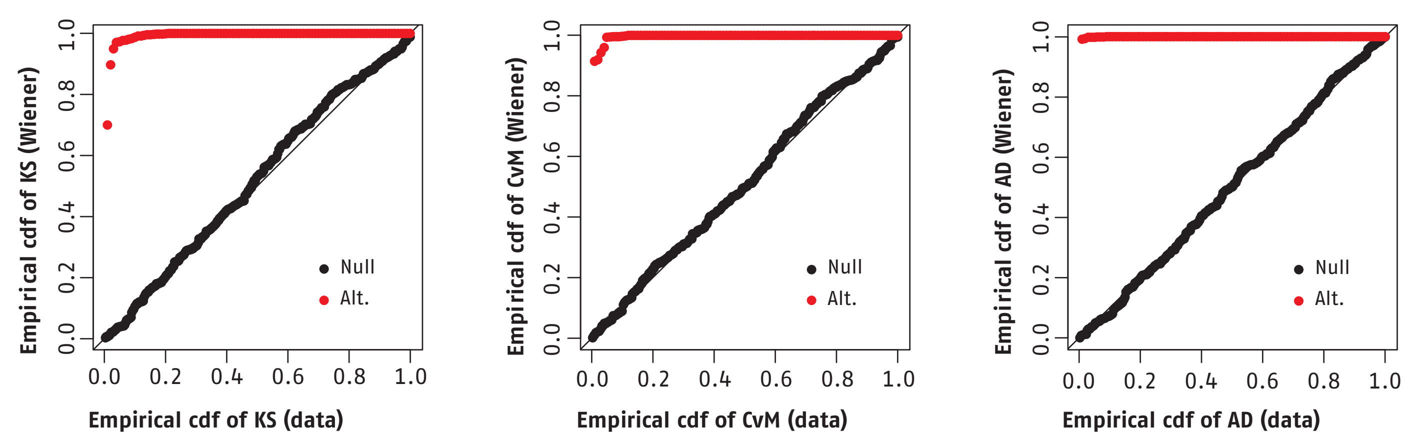

Figure 1. PP-plots for KS, CvM and AD test statistics, for one of the parametric models considered.