V Lecture

Correlation, Partial Correlations, Multiple Correlations. Copulae-Estimation and Testing

5.0 First of all, we would like to make some general comments on similarities and differences between correlations and dependencies.

Very often we are interested in correlations (dependencies) between a number of random variables and are trying to describe the "strength" of the (mutual) dependencies. For example, we would like to know if there is a correlation (mutual non-directed dependence) between the length of the arm and of the leg. **But** if we would like to get an information about (or to predict) the length of the arm by measuring the length of the leg, we are dealing with dependence of the arm's length on the leg's length. Both problems described in this example, make sense.

On the other hand, there are other examples/situations in which only one of the problems is interesting or makes sense. If we study the dependence between rain and crops, this makes a perfect sense but there is no sense at all to study the (directed) influence of crops on rain.

In a nutshell, we can say that when studying the mutual (linear) dependence, we are dealing with correlation theory whereas when studying directed influence of one (input) variable on another (output) variable, we are dealing with regression theory. It should be clearly pointed out though that correlation **alone**, no matter how strong, can not help us identify the direction of influence and can not help us in regression modelling. Our reasoning about direction of influence should come **outside** of Statistical theory, from another theory.

Another important point to always bear in mind is that, as already discussed in lecture 2, uncorrelated does not necessarily mean independent if the multivariate data happens to fail the multivariate normality test. Nonetheless, for multivariate **normal** data, the notions of "uncorrelated" and "independent" coincide.

5.1. Definition and estimation of partial correlation coefficients

In general, there are 3 types of correlation coefficients:

- The usual correlation coefficient between 2 variables
- Partial correlation coefficient between 2 variables after adjusting for the effect (regression, association) of set of other variables.
- Multiple correlation between a single random variable and a set of p other variables

For $\mathbf{X} \sim N_p(\mu, \mathbf{\Sigma})$ we defined the correlation coefficient $\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}}}, i, j = 1, 2, \dots, p$ and discussed the MLE's $\hat{\rho}_{ij}$ in (3.6). It turned out that they coincide with the sample correlations r_{ij} we introduced in the first lecture (formula (1.3)). To define *partial correlation coefficients*, recall the Property 4 of the multivariate normal distribution from II Lecture:

If vector $\mathbf{X} \in \mathbf{R^p}$ is divided into $\mathbf{X} = \begin{pmatrix} X_{(1)} \\ X_{(2)} \end{pmatrix}$, $X_{(1)} \in R^r, r < p, X_{(2)} \in R^{p-r}$ and according to this subdivision the vector means are $\mu = \begin{pmatrix} \mu_{(1)} \\ \mu_{(2)} \end{pmatrix}$ and the covariance matrix $\boldsymbol{\Sigma}$ has been subdivided into $\boldsymbol{\Sigma} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$ and the rank of $\boldsymbol{\Sigma_{22}}$ is full then the conditional density of $\mathbf{X_{(1)}}$ given that $\mathbf{X_{(2)}} = \mathbf{x_{(2)}}$ is

$$N_r(\mu_{(\mathbf{1})} + \boldsymbol{\Sigma_{12}\Sigma_{22}^{-1}}(\mathbf{x_{(2)}} - \mu_{(\mathbf{2})}), \boldsymbol{\Sigma_{11}} - \boldsymbol{\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}})$$

We **define** the partial correlations of $\mathbf{X_{(1)}}$ given $\mathbf{X_{(2)}} = \mathbf{x_{(2)}}$ as the usual correlation coefficients calculated from the elements $\sigma_{ij.(r+1),(r+2)...,p}$ of the matrix $\boldsymbol{\Sigma_{1|2}} = \boldsymbol{\Sigma_{11}} - \boldsymbol{\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}}$, i.e.

$$\rho_{ij.(r+1),(r+2),\dots,p} = \frac{\sigma_{ij.(r+1),(r+2),\dots,p}}{\sqrt{\sigma_{ii.(r+1),(r+2),\dots,p}}\sqrt{\sigma_{jj.(r+1),(r+2),\dots,p}}} \tag{5.1}$$

To find ML estimates for these, we use the translation invariance property of the MLE to claim that if $\hat{\boldsymbol{\Sigma}} = \begin{pmatrix} \hat{\Sigma}_{11} & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{21} & \hat{\Sigma}_{22} \end{pmatrix}$ is the usual MLE of the covariance matrix then $\hat{\boldsymbol{\Sigma}}_{\mathbf{1|2}} = \hat{\boldsymbol{\Sigma}}_{\mathbf{11}} - \hat{\boldsymbol{\Sigma}}_{\mathbf{12}}\hat{\boldsymbol{\Sigma}}_{\mathbf{22}}^{-1}\hat{\boldsymbol{\Sigma}}_{\mathbf{21}}$ with elements $\hat{\sigma}_{ij.(r+1),(r+2),\dots,p}, i,j = 1,2,\dots,r$ is the MLE estimate of $\boldsymbol{\Sigma_{1|2}}$ and correspondingly,

$$\hat{\rho}_{ij.(r+1),(r+2),\dots,p} = \frac{\hat{\sigma}_{ij.(r+1),(r+2),\dots,p}}{\sqrt{\hat{\sigma}_{ii.(r+1),(r+2),\dots,p}}\sqrt{\hat{\sigma}_{jj.(r+1),(r+2),\dots,p}}}, i,j = 1,2,\dots,r$$

will be the ML estimators of $\rho_{ij.(r+1),(r+2)\dots,p}, i,j = 1,2,\dots,r$. We call $\rho_{ij.(r+1),(r+2),\dots,p}$ **the correlation of the ith and jth component when the components $(r+1),(r+2)$, etc. up to the pth (i.e. the last $p-r$ components) have been hold fixed. The interpretation is that we are looking for the association (correlation) between the ith and jth component after eliminating the effect that the last $p-r$ components might have had on this association.**

5.1.1. Simple formulae For situations when p is not large, as a partial case of the above general result, simple plug-in formulae are derived that express the partial correlation coefficients by the usual correlation coefficients. We shall discuss such formulae now. (For higher dimensional cases computers need to be utilized- the procedure **CORR** in SAS is one good option). The formulae are given below:

i) partial correlation between first and second variable by adjusting for the effect of the third:

$$\rho_{12.3} = \frac{\rho_{12} - \rho_{13}\rho_{23}}{\sqrt{(1-\rho_{13}^2)(1-\rho_{23}^2)}}$$

ii) partial correlation between first and second variable by adjusting for the effects of third and fourth variable:

$$\rho_{12.3,4} = \frac{\rho_{12.4} - \rho_{13.4}\rho_{23.4}}{\sqrt{(1-\rho_{13.4}^2)(1-\rho_{23.4}^2)}}$$

5.1.2. Example. Three variables have been measured for a set of schoolchildren:

i) X_1- Intelligence

ii) X_2- Weight

iii)X_3 -Age

The number of observations was large enough so that one can assume the empirical correlation matrix $\hat{\rho} \in \mathscr{M}_{3,3}$ to be the true correlation matrix: $\hat{\rho} = \begin{pmatrix} 1 & 0.6162 & 0.8267 \\ 0.6162 & 1 & 0.7321 \\ 0.8267 & 0.7321 & 1 \end{pmatrix}$. This suggests there is a high degree of positive dependence between weight and intelligence. But (**do the calculation (!)**) $\hat{\rho}_{12.3} = 0.0286$ so that, after the effect of age is adjusted for, there is virtually no correlation between weight and intelligence, i.e. weight obviously plays little part in explaining intelligence.

5.2. Definition and estimation of the multiple correlation coefficient

Recall our discussion in the end of Section 2.2.2 for the best prediction in mean squares sense in case of multivariate normality: If we want to predict a random variable Y that is correlated with p random variables (predictors) $\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ . \\ . \\ X_p \end{pmatrix}$ by trying to minimize the expected value $E(Y - g(\mathbf{X}))^2$ the optimal solution (i.e. the regression function) was $g^*(\mathbf{X}) = E(Y|\mathbf{X})$. When the joint $(p+1)$-dimensional distribution of Y and $\mathbf{X}$ is **normal** this function was **linear** in $\mathbf{X}$. Given a specific realization $\mathbf{x}$ of $\mathbf{X}$ it was given by $b + \sigma_{\mathbf{0}}' \mathbf{C}^{-1}\mathbf{x}$ where $b = E(Y) - \sigma_{\mathbf{0}}' \mathbf{C}^{-1}\mathbf{E}(\mathbf{X})$, $\mathbf{C}$ is the covariance matrix of the vector $\mathbf{X}$, σ_0 is the vector of Covariances of Y with $X_i, i = 1, \ldots, p$. The vector $\mathbf{C}^{-1}\sigma_{\mathbf{0}} \in R^p$ was the *vector of the regression coefficients*.

Now, let us **define** the multiple correlation coefficient between the random variable Y and the random vector $\mathbf{X} \in \mathbf{R}^p$ to be the maximum correlation between Y and *any linear combination* $\alpha'\mathbf{X}, \alpha \in R^p$. This makes sense: to look at the maximal correlation that we can get by trying to predict Y as a linear function of the predictors. The solution to this which also gives us an algorithm to calculate (and estimate) the multiple correlation coefficient is given in the next lemma.

Lemma 5.2.1. The multiple correlation coefficient is the ordinary correlation coefficient between Y and $\sigma_0' \mathbf{C}^{-1}\mathbf{X} = \beta^{*'}\mathbf{X}$.

Proof. Note that for any $\alpha \in R^p : cov(Y, \alpha'\mathbf{X}) = \alpha'\mathbf{C}\beta^*$ and, in particular, $cov(Y, \beta^{*'}\mathbf{X}) = \beta^{*'}\mathbf{C}\beta^*$ holds.

Using *Cauchy-Bunyakovsky-Schwartz* inequality we have:

$$[cov(\alpha' X, \beta^{*'} X]^2 \le Var(\alpha' X).Var(\beta^{*'} X)$$

and therefore:

$$\sigma_Y^2 \rho^2(Y, \alpha' X) = \frac{(\alpha' \sigma_0)^2}{\alpha' C \alpha} = \frac{(\alpha' C \beta^*)^2}{\alpha' C \alpha} \le \beta^{*'} \mathbf{C} \beta^*$$

holds , σ_Y^2 denoting the variance of Y. In this last equality we can get the equality sign by choosing $\alpha = \beta^*$, i.e. the squared correlation coefficient $\rho^2(Y, \alpha' X)$ of Y and $\alpha' X$ is maximized over α when $\alpha = \beta^*$.

From Lemma 5.2.1 we see that the maximum correlation between Y and any linear combination $\alpha' X, \alpha \in R^p$ is $R = \sqrt{\frac{\beta^{*'} \mathbf{C} \beta^*}{\sigma_Y^2}}$. This is the multiple correlation coefficient.

Its square R^2 is called *coefficient of determination*. Having in mind that $\beta^* = C^{-1}\sigma_0$ we see that $R = \sqrt{\frac{\sigma_0'\mathbf{C}^{-1}\sigma_0}{\sigma_Y^2}}$. If $\mathbf{\Sigma} = \begin{pmatrix} \sigma_Y^2 & \sigma_0' \\ \sigma_0 & \mathbf{C} \end{pmatrix} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$ is the partitioned covariance matrix of the $(p+1)$-dimensional vector $(Y, \mathbf{X})'$ then we know how to calculate the MLE of Σ by $\hat{\Sigma} = \begin{pmatrix} \hat{\Sigma}_{11} & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{21} & \hat{\Sigma}_{22} \end{pmatrix}$ so the MLE of R would be $\hat{R} = \sqrt{\frac{\hat{\Sigma}_{12}\hat{\Sigma}_{22}^{-1}\hat{\Sigma}_{21}}{\hat{\Sigma}_{11}}}$

5.2.2. Interpretation of R. At the end of Section 2.2.2 we derived the minimal value of the mean squared error when trying to predict Y by a linear function of the vector $\mathbf{X}$. It is achieved when using the regression function and the value itself was $\sigma_Y^2 - \sigma_0'\mathbf{C}^{-1}\sigma_0$. The latter value can also be expressed by using the value of R. It is equal to $\sigma_Y^2(1-R^2)$. Thus, our conclusion is that when $R^2 = 0$ there is no predictive power at all. In the opposite extreme case, if $R^2 = 1$, it turns out that Y can be predicted without any error at all (it is a true linear function of $\mathbf{X}$).

5.2.3 Numerical example. Let $\mu = \begin{pmatrix} \mu_Y \\ \mu_{X_1} \\ \mu_{X_2} \end{pmatrix} = \begin{pmatrix} 5 \\ 2 \\ 0 \end{pmatrix}$ and $\Sigma = \begin{pmatrix} 10 & 1 & -1 \\ 1 & 7 & 3 \\ -1 & 3 & 2 \end{pmatrix} = \begin{pmatrix} \sigma_{YY} & \sigma_0' \\ \sigma_0 & \Sigma_{XX} \end{pmatrix}$. Calculate:

i) The best linear prediction of Y using X_1 and X_2.
ii) The multiple correlation coefficient $R^2_{Y.(X_1,X_2)}$.
iii)The mean squared error of the best linear predictor.

Solution.

$$\beta^* = \Sigma_{XX}^{-1}\sigma_0 = \begin{pmatrix} 7 & 3 \\ 3 & 2 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} .4 & -.6 \\ -.6 & 1.4 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ -2 \end{pmatrix}$$

and

$$b = \mu_Y - \beta^{*'}\mu_X = 5 - (1, -2)\begin{pmatrix} 2 \\ 0 \end{pmatrix} = 3.$$

Hence the best linear predictor is given by $3 + X_1 - 2X_2$. The value of:

$$R_{Y.(X_1,X_2)} = \sqrt{\frac{(1,-1)\begin{pmatrix} .4 & -.6 \\ -.6 & 1.4 \end{pmatrix}\begin{pmatrix} 1 \\ -1 \end{pmatrix}}{10}} = \sqrt{\frac{3}{10}} = .548$$

The mean squared error of prediction is: $\sigma_Y^2(1 - R^2_{Y.(X_1,X_2)}) = 10(1 - \frac{3}{10}) = 7$.

5.2.4. Remark about the calculation of R^2. Sometimes, the *correlation matrix only* may be available. It can be shown that in that case the relation

$$1 - R^2 = \frac{1}{\rho^{YY}} \tag{5.2}$$

holds. In (5.2) , ρ^{YY} is the upper left-hand corner of the inverse of the *correlation matrix*

$\rho \in \mathscr{M}_{\mathbf{p+1,p+1}}$ determined from Σ. We note that the relation $\rho = V^{-\frac{1}{2}}\Sigma V^{-\frac{1}{2}}$ holds with

$$V = \begin{pmatrix} \sigma_y^2 & 0 & 0 & \dots & 0 \\ 0 & c_{11} & 0 & \dots & 0 \\ 0 & 0 & c_{22} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & c_{pp} \end{pmatrix}$$

One can use (5.2) to calculate R^2 by first calculating the right hand side in (5.2). To show Equality (5.2) we note that

$$1 - R^2 = \frac{\sigma_Y^2 - \sigma_0' C^{-1}\sigma_0}{\sigma_Y^2} = \frac{|C|}{|C|}\frac{\sigma_Y^2 - \sigma_0' C^{-1}\sigma_0}{\sigma_Y^2} = \frac{|\Sigma|}{|C|\sigma_Y^2}$$

But $\sigma^{YY} = \frac{|C|}{|\Sigma|}$ is the entry in the first row and column of Σ^{-1}. Since $\rho^{-1} = V^{\frac{1}{2}}\Sigma^{-1}V^{\frac{1}{2}}$, we see that $\rho^{YY} = \sigma^{YY}\sigma_Y^2$ holds. Therefore $1 - R^2 = \frac{1}{\rho^{YY}}$.

5.3. Testing of correlation coefficients

5.3.1. Usual correlation coefficients

When considering the distribution of a particular correlation coefficient $\hat{\rho}_{ij} = r_{ij}$ the problem becomes bivariate because only the variables X_i and X_j are involved. Direct transformations with the bivariate normal can be utilized to derive the **exact** distribution of r_{ij} under the hypothesis $H_0 : \rho_{ij} = 0$. It turns out that in this case the statistic $T = r_{ij}\sqrt{\frac{n-2}{1-r_{ij}^2}} \sim t_{n-2}$ and tests can be performed by using the tables of the t-distribution. For other hypothetical values the derivations are more painful. There is one most frequently used **approximation** that holds no matter what the true value of ρ_{ij} is. We shall discuss it here. Consider **Fisher's z transformation** $Z = \frac{1}{2}\log[\frac{1+r_{ij}}{1-r_{ij}}]$. Under the hypothesis $H_0 : \rho_{ij} = \rho_0$ it holds:

$$Z \approx N(\frac{1}{2}\log[\frac{1+\rho_0}{1-\rho_0}], \frac{1}{n-3})$$

In particular, in the most common situation, when one would like to test $H_0 : \rho_{ij} = 0$ versus $H_1 : \rho_{ij} \neq 0$ one would reject H_0 at 5% significance level if $|Z|\sqrt{n-3} \geq 1.96$.

Based on the above, now you suggest how to test the hypothesis of equality of two correlation coefficients from two different populations(!).

5.3.2. Partial correlation coefficients

Coming over to testing *partial correlations* not much has to be changed. Fisher's Z approximation can be used again in the following way: to test $H_0 : \rho_{ij.r+1,r+2,\dots,p} = \rho_0$ versus $H_1 : \rho_{ij.r+1,r+2,\dots,p} \neq \rho_0$ we construct $Z = \frac{1}{2}\log[\frac{1+r_{ij.r+1,r+2,\dots,p}}{1-r_{ij.r+1,r+2,\dots,p}}]$ and $a = \frac{1}{2}\log[\frac{1+\rho_0}{1-\rho 0}]$. Asymptotically $Z \sim N(a, \frac{1}{n-(p-r)-3})$ holds. Hence, test statistic to be compared with significance points of the standard normal is now : $\sqrt{n-(p-r)-3}|Z-a|$.

5.3.3. Multiple correlation coefficients

It turns out that under the hypothesis $H_0 : R = 0$ the statistic $F = \frac{\hat{R}^2}{1-\hat{R}^2}.\frac{n-p}{p-1} \sim F_{p-1,n-p}$. Hence, when testing significance of the multiple correlation, the rejection region would be $\{\frac{\hat{R}^2}{1-\hat{R}^2}.\frac{n-p}{p-1} > F_{p-1,n-p}(\alpha)\}$ for a given significance level α.

Remark. It should be stressed that the value of p in Section 5.3.3 refers to the **total** number of all variables (the output Y and all of the input variables in the input vector X). This is different from the value of p that was used in Section 5.2. In other words, the p in Section 5.3.3 is the $p+1$ in Section 5.2.

5.4. Copulae. For the multivariate normal, independence is equivalent to absence of correlation between any two components. In this case the joint cdf is a product of the marginals. When the independence is violated, the relation between the joint multivariate distribution and the marginals is more involved. An interesting concept that can be used to describe this more involved relation is the concept of *copula.* We focus on the two-dimensional case for simplicity. Then the copula is a function $C : [0,1]^2 \to [0,1]$ with the properties:

i) $C(0,u) = C(u,0) = 0$ for all $u \in [0,1]$.

ii) $C(u,1) = C(1,u) = u$ for all $u \in [0,1]$.

iii) For all pairs $(u_1,u_2),(v_1,v_2) \in [0,1] \times [0,1]$ with $u_1 \le v_1, u_2 \le v_2$:

$$C(v_1,v_2) - C(v_1,u_2) - C(u_1,v_2) + C(u_1,u_2) \ge 0.$$

The name is due to the implication that the copula links the multivariate distribution to its marginals. This is explicated in the following celebrated **Theorem of Sklar :**

Let $F(.,.)$ be a joint cdf with marginal cdf's $F_{X_1}(.)$ and $F_{X_2}(.)$. Then there exists a copula $C(.,.)$ with the property

$$F(x_1,x_2) = C(F_{X_1}(x_1), F_{X_2}(x_2))$$

for every pair $(x_1,x_2) \in R^2$. When $F_{X_1}(.)$ and $F_{X_2}(.)$ are continuous the above copula is unique. Vice versa, if $C(.,.)$ is a copula and $F_{X_1}(.), F_{X_2}(.)$ are cdf then the function $F(x_1,x_2) = C(F_{X_1}(x_1), F_{X_2}(x_2))$ is a joint cdf with marginals $F_{X_1}(.)$ and $F_{X_2}(.)$.

Taking derivatives we also get:

$$f(x_1,x_2) = c(F_{X_1}(x_1), F_{X_2}(x_2)) f_{X_1}(x_1) f_{X_2}(x_2)$$

where

$$c(u,v) = \frac{\partial^2}{\partial u \partial v} C(u,v)$$

is the *density* of the copula. This relation clearly shows that the contribution to the joint density of X_1, X_2 comes from two parts: one that comes from the copula and is "responsible" for the dependence ($c(u,v) = \frac{\partial^2}{\partial u \partial v} C(u,v)$) and another one which takes into account marginal information only ($f_{X_1}(x_1) f_{X_2}(x_2)$).

It is also clear that the independence implies that the corresponding copula is $\Pi(u,v) = uv$ (this is called the independence copula).

These concepts are generalized also to p dimensions with $p > 2$.

An interesting example is the *Gaussian copula.* For $p = 2$ it is equal to:

$$C_\rho(u,v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} f_\rho(x_1,x_2) dx_2 dx_1.$$

Here $f_\rho(.,.)$ is the joint bivariate normal density with zero mean, unit variances and a correlation ρ and $\Phi^{-1}(.)$ is the inverse of the cdf of the standard normal. (This is "The formula that killed Wall street"). When $\rho = 0$ we see that we get $C_0(u, v) = uv$ (as is to be expected).

Non-Gaussian copulae are much more important in practice and inference methods about copulae are a hot topic in Statistics. The reason for importance of non-Gaussian copulae is that Gaussian copulae do not allow us to model reasonably well the tail dependence, that is, joint *extreme* events have virtually a zero probability. Especially in financial applications, it is very important to be able to model dependence in the tails. The Gumbel-Hougaard copula is much more flexible in modeling dependence in the upper tails. For an arbitrary dimension p is is defined as

$$C_\theta^{GH}(u_1, u_2, \dots, u_p) = \exp\{-[\sum_{j=1}^{p}(-\log u_j)^\theta]^{1/\theta}\}$$

where $\theta \in [1, \infty)$ is a parameter that governs the strength of the dependence. You can easily see that the Gumbell-Hougaard copula reduces to the independence copula when $\theta = 1$ and to the Fréchet-Hoeffding upper bound copula $\min(u_1, \dots, u_p)$ when $\theta \to \infty$.

The Gumbel-Hougaard copula is also an example of the so-called *Archimedean* copulae. The latter are characterized by their *generator* $\phi(.)$: a continuous strictly decreasing function from $[0, 1]$ in $[0, \infty]$ such that $\phi(1) = 0$. Then the Archimedean copula is defined via the generator as follows:

$$C(u_1, u_2, \dots, u_p) = \phi^{-1}(\phi(u_1) + \dots \phi(u_p)).$$

Exercise Show that the Gumbell-Hougaard copula is an Archimeden copula with a generator $\phi(t) = (-\log t)^\theta$.

The benefit of using the Archimedean copulae is that they allow for simple description of the $p-$dim dependence by using a function of one argument only (the generator). However it is seen immediately that the Archimedean copula is symmetric in its arguments and this limits its applicability for modelling dependencies that are not symmetric in their arguments. The so-called *Liouville* copulae are an extension of the Archiemedean copulae and can be used also to model dependencies that are not symmetric in their arguments.

Inference procedures about copulae are implemented in the procedure **copula** in the SAS/ETS package.