**A9Wg Standard error for the autocorrelation function (ACF)**

As we indicated in the textbook, the autocorrelation coefficients that constitute the autocorrelations function, should tend towards zero. As Part IV of the textbook shows, this is not always the case, but nevertheless some of these autocorrelations come very close to zero. The question is: are they really zero? To answer this question, we can calculate the standard error for the autocorrelation function (SEACF). The standard error for the autocorrelation function is calculated as follows:

$SE\_{ACF}=\frac{1}{\sqrt{n}}$ (1)

Where, n is the number of observations that were used to calculate autocorrelation coefficients.

If we multiply this value by ± 1.96 (recall from Chapter 4 that this is the z-value that covers 95% of all values in this distribution), then we get the 95% confidence interval around our autocorrelations.

95% confidence interval = ± 1.96 \* SEACF

**Example**

In this example, we will use the full set of UK visits abroad, as in textbook Example 8.1. The full set has 48 observations, and the standard error for the autocorrelations is:

$$SE\_{ACF}=\frac{1}{\sqrt{50}}=0.141$$

Figure 1 illustrates the Excel solution to calculate the autocorrelation function values, the standard errors, and associated confidence intervals (rows 20:40 are hidden).



Figure 1

**Excel solution for the calculation of SE and CI’s**

SE Cell F3 Formula: =1/SQRT(COUNT($C$3:$C$52))

Copy formula down F3:F18

+ CI Cell G3 Formula: =1.96\*F3

Copy formula down G3:G18

-CI Cell H3 Formula: =1.96\*F3

Copy formula down H3:H18

The confidence interval is calculated as ±z SEACF. As we were interested in 95% confidence interval, we used the value of z=1.96. If the time series was shorter (say, less than 30 observations), we would have used the t-value for the corresponding number of the degrees of freedom. Figure 2 illustrates the graphical relationship between the autocorrelation function and standard error at a particular lag value.



Figure 2

How do we interpret this chart? As we can see, the first 11 autocorrelation coefficients are outside the confidence interval. This means that we are 95% certain that these 11 coefficients are significantly different from zero. You might say, so what? Well, let’s answer this question. If you had a completely random series of numbers, then almost all the autocorrelation coefficients should be virtually zero. Figure 3 demonstrates this.



Figure 3

Truly random series always have almost all autocorrelation coefficients virtually zero. This fact will be very useful for measuring errors. As we have learned, errors are supposed to be completely random. If they are, then their autocorrelation coefficients should be virtually zero. This will be just one of many different applications that rely on autocorrelation function.

In our case the fact that the first 11 coefficients are “persistently” above zero has a special meaning. This means that the UK travel abroad time series is non-stationary, at least for the time interval explored here. The term non-stationary means that the variable is not horizontal, but it has an upward (or downward) trend. In other words, the time series does not have a constant mean value. In Part III of the textbook we have gone into greater depth and explained this.

In the section at the end of textbook Chapter 8, you can find the same example solved by using SPSS. Just to reconcile some differences between what we have done in Excel and what is done in SPSS, we modified our example. Figure 4 shows several new columns (J:M).



Figure 4

**Excel solution for the calculation of SE and CI’s**

K Cells J3:J18 Values

SE Cell K3 Formula:

=SQRT(1/COUNT($C$3:$C$52)\*((COUNT($C$3:$C$52)-J3)/(COUNT($C$3:$C$52)+2)))

Copy formula down F4:K18

+ CI Cell L3 Formula: =1.96\*K3

Copy formula down L3:L18

-CI Cell M3 Formula: =1.96\*K3

Copy formula down M3:M18

Column J lists the lag values k and column K shows modified Standard Error calculation that is identical to what is produced by SPSS. In our example the Standard Error is calculated as per textbook equation (9.3). SPSS uses a modified version, as per equation (2).

 $SE\_{ACF}=\sqrt{\frac{1}{n}(\frac{n-k}{n+2}})$ (2)

Where, n is the number of observations and k is the lag value.

According to SPSS printout, the graph in Figure 2 looks as the one in Figure 5.



Figure 5

As we can see the confidence interval is slightly modified and it gets narrower the more coefficients we calculate, although it might be difficult to see in this graph. In some other example, this might be more obvious. For longer data sets, the difference between the two methods becomes very small.