

# Advanced Analytics Group Roadmap to Chaotic Models in Auguri

## I. Definitions

### Independent Variables

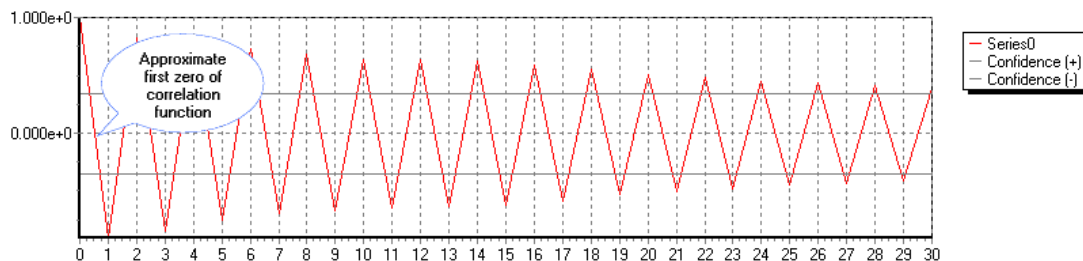
*Independent variables* are series whose measurements do not depend on the measurements of another series within a given system. For example, if we want to approximate the expansion (or contraction) of a certain metal alloy in a system defined by temperature, pressure and expansion measurements, we would select the measurements of temperature and pressure as our independent variables because in this system these measurements do not depend on any other measurements. These variables are often termed *inputs*, because they serve as an input to the system, which in turn generates a response.

### Dependent Variables

*Dependent variables* are series whose measurements do depend on the measurements of other series within a given system. In the previous example, we would select the measurements of expansion as our dependent variables because in this system this measurement depends on both measurements of temperature and pressure. *Note that there are instances where a variable may be both dependent and independent. In this case, all present states or measurements of the variable assume the rank of dependent, while the past of these states, the rank of independent.* Dependent variables are often termed *outputs*, because they represent the response generated by a system when excited with inputs.

### First Zero of the Auto- or Cross-Correlation Function

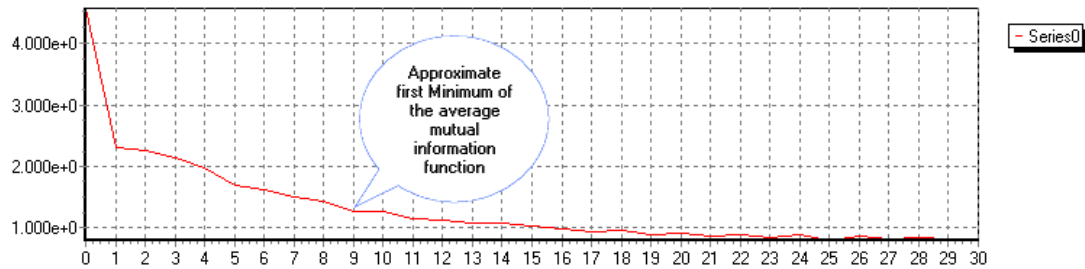
We refer to the first zero of the auto- or cross-[correlation function](#) of a series with itself (auto-) or another series (cross-), as the first lag at which the auto- or cross-[correlation function](#) of the series of interest crosses the abscissa, that is, reaches zero. When this point does not occur at an integer lag, we take the nearest integer as the first zero.



### First Zero or Minimum of the Auto- or Cross-Average Mutual Information

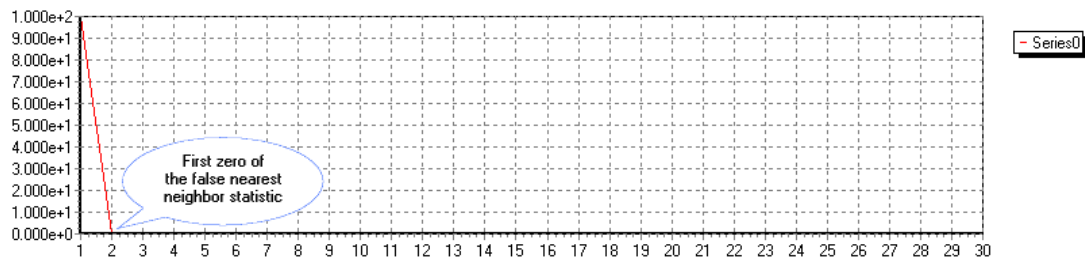
We refer to the first zero of the auto- or cross-[average mutual information](#) of a series with itself (auto-) or another series (cross-), as the first point in the abscissa at which the auto- or cross-[average mutual information](#) function reaches its first minimum.

# Advanced Analytics Group Roadmap to Chaotic Models in Auguri



## First Zero of the Auto- or Cross-False Nearest Neighbors Statistic

We refer to the first zero of the auto- or cross-[false nearest neighbor statistic](#) of a series with itself (auto-) or another series (cross-), as the first point at which the auto- or cross-[false nearest neighbor statistic](#) becomes zero.



## Embedding

An *embedding* is a vector arrangement of dependent and independent variables that considers present measurements of a dependent variable against present and past measurements of independent variables, and perhaps, past measurements of itself (refer to *Note* in the definition of *Dependent Variables* for additional information on this case.)

## II. Roadmap

- (1) In your worksheet, take note of those variables you are interested in forecasting or approximating (dependent variables), and those variables which you deem important to carry the forecast or approximation (independent variables).
- (2) **Carry** auto- and cross-correlation on the dependent series against itself and other series which may be of interest, up to a lag, say of 100 if your series is sufficiently long, otherwise, to about one half of the series length. (**Covariance and Correlation Function** in **Tools->Analysis**.) Save the results in a new worksheet.

**Advanced Analytics Group  
Roadmap to Chaotic Models in Auguri**

- (3) **Carry** auto- and cross-mutual information on the dependent series against itself and other series which may be of interest, up to a lag, say of 50. (**Average Mutual Information in Tools->Analysis.**) Save the results in a new worksheet.
- (4) **Carry** auto- and cross-false nearest neighbors on the dependent series against itself and other series which may be of interest, up to say 30 embedding steps using a time delay of whichever is lower between the auto- or cross-correlation function first zero and the first zero of the auto- or cross-mutual information. (**False Nearest Neighbors Statistic in Tools->Analysis.**) Save the results in a new worksheet.

(5) **Decide on an Embedding.**

- (a) **Linear dependency** — If the absolute values of the auto- or cross-correlation function remained high (above or below the 95% confidence interval) for lag values greater than 0, the series of interest (the dependent series) may have a linear dependency on its past (auto-) or the cross series (the independent series). In this case choose an embedding between the dependent, its past or the independent series to be as long as the lag until which the values remained above the 95% confidence level, if relatively small, or to some smaller amount if exceedingly large, with a delay equal to the cycle lag between the peaks of the correlation curve.
- (b) **Nonlinear dependency** — If the absolute values of the auto- or cross-correlation functions remained within the 95% confidence interval, then embed the dependent series to a dimension at least equal to the first zero value in the False Nearest Neighbor Statistic with a time delay equal to the time delay used to calculate the false nearest neighbor statistic. Later on you may experiment in increasing the dimension to higher values as in many cases this action may result in greater forecasting accuracy.
- (c) **Repeat** steps (a) and (b) for all dependent series against itself and all the independent series, taking note of the results.

**Note** that if your system consists of more than one dependent variable and if the dependencies for each of these variables are different, you may be required to copy the data over different sheets in order to create different models for each of the dependent variables.

- (6) **Define the System** — In **Tools->Model Specification**, indicate variable types and specify the dependencies for each series in terms of dimensions and delays as noted in step (4).
- (7) **Add Solutions to the System** — In **Solutions->Edit Solutions**, add a new **Global Least Square Solution**. Click **Class Definition** and in the **Global Least Squares** dialog box check the **Add Constant Term** box. Click **OK**. Click **Method Options** and in the

## Advanced Analytics Group Roadmap to Chaotic Models in Auguri

**Global Least Squares** Options dialog box, check the **Dynamic** radio button and click **OK**. Then click **OK** in *the* **New Solution** dialog box, and click **OK** again in the **Solutions Methods** dialog box.

- (8) **Test the Solutions** — In **Solutions->Simulation Test** type a start point that will test your method against, say, the last 50 points of your series. For example, select the name of the solution method you just created in the **Solution Method** box. On the bottom right corner of the **Simulation Test** dialog box, there is a set of 4 text boxes: **Total Available**, **Available for Use**, **Start Position** and **Test Size**. Subtract 50 from total available and enter this number in the **Start Position** box. The **Test Size** box should immediately reflect 50. What you are doing is running the solution against the last 50 points of the series and the test will give you an idea of the error you should expect from using this method to forecast and/or approximate the dependent series, though don't expect a substantial improvement from other methods unless these series are known to be chaotic or exhibit nonlinear dependencies.

**Note** that all these instructions assume the dependent series in question have been normalized to near stationarity, that is, trend and seasonality have been removed. If you plan to use **neural networks** you should consider scaling the normalized series in the range [-0.8 to 0.8], as it is known that neural networks converge faster when the targets, as opposed to the activation functions, are scaled.