Advanced Time Series Analysis

Instructor: Ma, Jun

School of Economics, Renmin University of China

March 6, 2018

Econometric and Statistical Analysis

- Statistics operates at two levels. At the first level you make an algorithm to estimate "something" using the data (sample). At the second level, you try to figure out how accurate those algorithms are.
- The second level uses "mathematics + computer" to answer the following question: if you have many many samples, how volatile your estimates are.
- In reality, we have only one sample. But the mathematical analysis could help finding hopefully the best algorithm.
- In this era, data become abundant and enormous.

Motivation for Time Series Analysis

- Suppose that one is given a data set of the last 50 years of annual GDP per capita series and other macroeconomic data (e.g., CPI, real money stock M1...), and tries to predict the next year GDP per capita. There are many ways to perform the prediction. One way is to rely on the researcher's subjective experiences and feelings and to come up with a number as a forecast.
- Another way is a model-based approach where one models how the GDP per capita series are generated, estimates the model using the data sets, and makes predictions based on the estimated model.
- The model-based approach has several advantages. The model-based approach shows transparently the way the predictions are made. The approach is objective in the sense that once the same model and the same method are used, the results generated will be the same regardless of who performs the prediction.

Motivation for Time Series Analysis

- However, it should be remembered that the model-based approach always has its own limitations.
- The model-based approach begins with a set of assumptions whose validity is sometimes hard to check due to lack of appropriate data sets. When the set of assumptions are no longer plausible, the validity of the results is also jeopardized.
- One of the most important assumptions is that a certain aspect of the reality will not change in the future. This forms a fundamental basis that makes it possible to forecast a future event by observing the history of observations in the past.
- However, in times of economic or financial crises, the prediction environment can be very different from one under economic or financial stability. Therefore, the model-based approach should be used along with other information such as past experiences of market participants and relevant news about historical or future events.

Motivation for Time Series Analysis

- Time series models are a particular class of statistical or econometric models that mainly involve time series observations.
- There are usually three components in econometric or statistical models.
- Suppose that Y_1, Y_2, ... represent random variables corresponding to the stock return of Amazon at period 1, the stock return of Amazon at period 2, etc. For these random variables, let us consider the following simple autoregressive model of order 1 (i.e. $\mathrm{AR}(1)$ model):

$$Y_t = \alpha Y_{t-1} + u_t,\ t = 1, 2, ..., T,$$

where $\{Y_t\}_{t=1}^{T}$ are observed random variables, $\{u_t\}_{t=1}^{T}$ are unobserved random variables, and α is a model parameter.

Observed Random Variables

- Every empirical analysis begins by viewing observed numbers as a particular realization of random variables.
- Observed random variables in the model are random variables that are thought to have generated the numbers that we observe. In the previous example of $\mathrm{AR}(1)$ model, $\{Y_t\}_{t=1}^{T}$ are observed random variables.

Unobserved Random Variables

- Models typically contain random variables that are not observed. That is, we do not observe any realizations of these random variables. It is almost impossible to fully model the mechanism in which the observed random variables are related to each other, in a manner that is consistent with what we observe.
- In reality, the way the relevant variables are related to each other can be very complicated, and we observe only part of those variables. Hence a model often contains unobserved random variables that account for certain aspects of the dynamics for which we do not have observations at all. In the case of $\mathrm{AR}(1)$ model, $\{u_t\}_{t=1}^{T}$ represent unobserved random variables.

Model Parameters

- Many models are not fully determined by the observed random variables and unobserved random variables. There are certain aspects of the model that are left to be unknown. For example, if one writes the $\mathrm{AR}(1)$ model in the following way:

$$Y_t = 0.5 Y_{t-1} + u_t,$$

  then the way Y_t's are generated is fully determined by observed random variables Y_t and unobserved random variables u_t. This model specifies the contribution of past variable Y_{t-1} to the next period to be precisely a factor of 0.5.
- Instead, one may make the specification more flexible by leaving this contribution unknown by putting:

$$Y_t = \alpha Y_{t-1} + u_t$$

  and estimate α later using the data set. An unknown constant such as α is called a model parameter.

Inference

- Inference refers to a procedure of learning about the model parameters from observed data sets.
- There are two problems of inference: estimation and hypothesis testing. Estimation is to make a conjecture about the parameter value and hypothesis testing is to answer the binary question of whether the parameter value is in a certain range.
- For example, in the case of the $\mathrm{AR}(1)$ model, if one asks the value of α, it is a question of estimation, and if one asks whether $\alpha < 1$, it is a question of hypothesis testing.

Prediction

- Once one estimates the model parameter, one has a model that fully specifies the way Y_t's are generated. Using this model, one can make forecast about the future values Y_t. Depending on the type of forecast, one can make point forecast or interval forecast. When one says the next period forecast of Y is 0.5, she is making a point forecast, but when one says the next period Y falls in the interval $[0, 0.7]$ with probability 90%, she is making an interval forecast.
- When one makes forecast based on an estimated model, it should be reminded that the quality of the forecast relies on the accuracy of the estimated parameters. When the parameters are not accurately estimated, the forecast obtained from the estimated model is not so reliable.

Model Selection and Evaluation

- In practice, there are many econometric models to choose from. Some models contain more parameters, having bigger flexibility, than other models. It is not true that it is always better to use a more flexible model. The reason is because one has to estimate more parameters using the given data sets when the model is highly flexible, or large. When too many parameters are estimated from the given data sets, the quality of the estimates deteriorates, worsening the forecast results.
- Therefore, it is a nontrivial matter to choose among models with varied flexibility. There are principled ways to perform model selection in this situation.