Sid Bhatia

FE670 – Algorithmic Trading Strategies

Professor Yang

14 September 2023

**Lecture #2 Notes – Basic Models and Empirics**

*Financial Modeling Fundamentals*

* **Financial Engineering** is a process of synthesis in the sense that the objective of the engineering process is to construct purposeful artifacts, such as portfolios, investment strategies, or derivative products.
* Constructive methodologies are available only when we arrive at the point where we can optimize, that is, codify our design in terms of variables and express the quality of our design in terms of a goal function defined on the design variables:
  + *Science is analytic:* We have the **models** to analyze a given system.
  + *Design is a constructive process:* We need tosynthesize a design starting from a high-level specification.
  + *Constructive design is performed iteratively:* We make an approximate design and analyze it.
  + *Design automation:* The process of design can be automated only when we arrive at the stage of expressing the design **quantitatively** in terms of a **goal function**.

*Learning, Theoretical, and Hybrid Approaches to Modeling*

* There are three basic approaches to financial modeling: **the learning approach**, **the theoretical approach,** and **the** **learning-theoretical approach**.
  + **Learning Approach** to financial modeling is in principle a consequence of the diffusion of low-cost high-performance computers. It Is based on using a family of models that 1) include an unlimited number of parameters and 2) can approximate a sample data with high precisions
    - **Neural networks** are a classical example. With an unrestricted number of layers and nodes, a neural network can approximate any function with arbitrary precision.
    - However, practice has shown that if we represent sample data with very high precision, we typically obtain poor forecasting performance.
  + **Theoretical Approach** to financial modeling is based on human creativity. In this approach, models are the result of new scientific insights that have been embodied in theories.
    - Laws such as the Maxwell equations of electromagnetism were discovered not through a process of learning but by a stroke of genius.
    - The **Capital Asset Pricing Modeling** (**CAPM**) is the most well-known example of a theoretical model in financial economics.
  + **Hybrid Approach** to financial modeling retains characteristics of both the theoretical and learning approaches. It uses a theoretical foundation to identify families of models but uses a learning approach to choose the correct model within the family.
    - **ARCH/GARCH** family of models is suggested by theoretical considerations while the right model is selected through a learning approach that identifies the model parameters.

*Biases*

* **Survivorship Bias** is exhibited by samples selected on the basis of criteria valid at the last point in the sample population.
  + In the presence of survivorship biases in our data, return processes relative to firms that ceased to exist prior to that data are ignored.
  + For example, while poorly performing mutual funds often down (and therefore drop out of the sample), better performing mutual funds continue to exist (and therefore **remain in the sample**).
  + In this situation, estimating past returns from the full sample would result in **overestimation** due to survivorship bias.
* **Selection Bias** is an error in choosing the individuals or groups to take part in a scientific study.
  + Intrinsic in common indexes such the Russell 1000 universe (large-cap stocks). In order to understand the selection bias, we can apply a selection rule to that of the Russell 1000 to artificially generate **random walks**.
  + Assume we have 100,000 **independent random walk price processes**, each representing the price of a company’s stock, over 1,000 periods using the recursive formula:
    - Pi(2) = (1 + Ri(2)) x Pi(1) = 1 + 0.008…

*Pitfalls in Choosing from Large Datasets*

* **STATEMENT**: Any statistical test, regardless of its complexity and power, will fail in a certain number of cases simply by chance.
  + For example, **pairs trading** is based on selecting pairs of stocks that stay close together. Suppose we know that the price paths of two stocks will stay close together. When they are at their maximum distance, we can go **long** in the stock with the highest value and short in the other stock. When their distance is **reduced** or **changes sign**, a **profit is realized**.
    - Given a large universe of stocks, a pairs trading strategy will look for cointegrated pairs. A typically approach will consist in running an cointegration test on each pair.
    - Actually test can consist of multiple tests that each pair has to pass in order to be accepted as **conintegrated**.
    - However, a price can appear cointegrated in a sample period purely by chance. Or a truly cointegrated pair may fail the test.

*Pitfalls in Selection in Data Frequency*

* In financial theory, we have both **discrete-time** and **continuous-time models**. For example, Black-Scholes option pricing equation, under certain assumptions, can be solved in a closed-form format. In other cases, we have to look for **numeric solutions**.
* Long-story short, we cannot assume a model based on **daily data** can be applied to **weekly/hourly/monthly** data. There may be differences in the data frequency.

*Stationarity and Ergodicity:*

* Much statistical inference relies on the **law of large numbers (LLN)** and **central limit theorem** (**CLT**).
  + **LLN** states that the average of the results obtained from a large number of trials should be close to the **expected value**, and will tend to become closer as more trials are performed.
  + **CLT** states that, given conditions, the mean of a sufficiently large number of independent and identically distributed random variables, with finite mean and variance, will be **approximately normally distributed**.
* Financial time series data by nature, **dependent**, therefore, we rely on the alternative: **stationarity** and **ergodicity**.

*Ergodicity:*

* A time series is **ergodic**, if its local stochastic behavior is (possibly in the limit) **independent** of the starting point (initial conditions).
  + Stochastic process is independent of the starting point.
  + When you look at stock returns, it doesn’t matter where you start your model (5-years of data, 2-years of data, etc..).
* An ergodic process eventually “**forgets**” where it started.

*Stationarity:*

* The distribution of your time series **today** will match the distribution **tomorrow, the day after, etc..**
* In practice, we use **weak stationarity**.
  + The first **two moments** are time-invariant.
  + First moment is E[SR] = mu.
* Weak stationarity means (in practice) the ability to make inferences concerning **future observations** (e.g., prediction).
  + Ensures that models have (some) predictive power.
* In practice, it is common to assume that an asset return series is **weakly stationary**.

*Autocorrelation Function (ACF):*

* **Serial (Auto) Correlations**