

# 1. Econometrics and Basic Statistics

## Introduction

Econometrics is the computational utilization and application of statistical and mathematical models using data to establish theories or test current economic hypotheses and predict future patterns from historical data. It submits statistical trials to real-world evidence, then compares and evaluates the findings against the theory or theories being tested.

Econometrics may be subdivided into two main groups, theoretical and applied, depending on whether you are interested in testing an existing theory or in using existing data to establish a new hypothesis based on those findings. Many who participate in this practice regularly are usually called econometricians.

## R language

R is just an implementation of the object-oriented programming language S. It is developed by statisticians, and is free, published under the GNU General Public License. Syntactically and functionally the common statistics kit, S+, is quite similar (if not identical).

R is much more versatile than most econometrician programs, since it is a modern language of mathematical programming, not just a program that does regressions and tests. This means we should not limit our analysis to the functions included in the default package. There is a vast and ever-growing online library collection for use in many disciplines. The corresponding libraries are posted on the R web site as researchers create new algorithms and processes. In this context R is still at the forefront of the awareness of statistics. It is easy to extend due to the facility and versatility of programming in R (Farnsworth, 2008).

## Distributions

**Normal distribution** is a bell-shaped, continuous probability density (PDF) function described by the mean and standard deviation parameters.

**Log normal distribution** is a continuous distribution where its log has a normal distribution.

**Uniform distribution** is a PDF in which all future occurrences-e.g. returns-are expected to occur similarly.

The following three distributions are particularly important, since they are used for many econometric procedures in hypothesis testing.

**t-distribution** is symmetrical, bell-shaped, and identical to the normal standard curve. One. The higher the degrees of freedom the more the t would match the normal standard distribution with a mean zero and a standard deviation of 1.

**Chi-squared distribution** is a distribution of the sum of normal squared standard deviates, where the degrees of distribution freedom are equal to the number of normal standard deviates being summed up.

**F-Distribution** is A PDF which is used in particular for variance analysis and is a function of the ratio of 2 independent random variables, say X and Y, each of which has a chi-square distribution and is divided by the number of degrees of freedom.

## Normality (graph, skewness)

Normality is a major feature of econometrics in many situations. A significant assumption frequently made in finance is that usually the returns are normally distributed. It is also used in modeling portfolio allocation, pricing choices, and calculation of Value-at - Risk (VaR). Analysts also note that there are many real-world distributions with fat (heavy) tails, and often high peaks.

The **skewness** is given by the following formula:

$$\frac{n}{(n-1)(n-2)} \Sigma \left( \frac{x_i - \bar{x}}{s} \right)^3$$

Interpretation:

- Skewness > 0 – Right skewed distribution
- Skewness < 0 – Left skewed distribution
- Skewness = 0 – symmetrical distribution

The **normality** (kurtosis) is given by the following formula:

$$k = \frac{\frac{\sum_{i=1}^n (X_i - X_{avg})^4}{n}}{\left( \frac{\sum_{i=1}^n (X_i - X_{avg})^2}{n} \right)^2}$$

- Kurtosis > 3 – Leptokurtic distribution
- Kurtosis < 3 – Platykurtic distribution
- Kurtosis = 3 – Mesokurtic distribution

## Periodic rate of return

The formula of PRR is given by:

$$rt = \ln \left( \frac{P_t}{P_{t-1}} \right)$$

- $P_t$  = Asset price at time  $t$ .
- $P_{t-1}$  = Asset price at time  $t - 1$
- $\ln$  = Natural logarithm
- $rt$  = Rate of return at time  $t$ .

## Types of Econometric Models

Econometric models include statistical models that are used in the economics field. An econometric model determines the statistical relationship that a particular economic phenomenon is assumed to have between the different economic quantities. An econometric model can be derived by allowing for uncertainty from a deterministic economic model, or

from an economic model that is stochastic itself. However, econometric models that are often not related to any particular economic theory may also be used.

Examples of common econometric models:

- Linear regression
- Generalized linear models
- Probit
- Logit
- Tobit
- ARIMA
- Vector Autoregression
- Cointegration
- Hazard

## Finance Topics and Econometrics

### Capital asset pricing model (CAPM)

The Capital Asset Pricing Model (CAPM) expresses the relation between systemic risk, especially stocks, and expected return on assets. CAPM is commonly used in finance to price volatile securities and to achieve projected returns on assets due to the risk of such assets and capital costs.

Formula:

$$ER_i = R_f + \beta_i(ER_m - R_f)$$

where:

$ER_i$  = expected return of investment

$R_f$  = risk-free rate

$\beta_i$  = beta of the investment

$(ER_m - R_f)$  = market risk premium

### Fama-French model (FFM)

The Fama and French Three-Factor Model is an asset pricing model built in 1992 that develops on the capital asset pricing model (CAPM) by incorporating size risk and value risk factors to CAPM's market risk factor. This model takes into account the fact that value and Small Cap stocks consistently outperform markets. The model accounts for this outperforming propensity by adding these two additional variables, which is thought to make it a stronger method for assessing manager performance.

Formula

$$R_{it} - R_{ft} = \alpha_{it} + \beta_1(R_{Mt} - R_{ft}) + \beta_2SMB_t + \beta_3HML_t + \epsilon_{it}$$

Where

$R_{it}$ : total return of a stock or portfolio  $i$  at time  $t$

$R_{ft}$ : risk free rate of return at time  $t$

$\beta_{1,2,3}$ : factor coefficients

$R_{Mt}$ : total market portfolio return at time  $t$

$R_{ft}$ : risk free rate of return at time  $t$

$SMB_t$  : size premium

$HML_t$ : value premium

### Arbitrage Pricing Theory (APT)

Arbitrage pricing theory ( APT) is a multifactor asset pricing model based on the notion that the returns of an asset can be calculated using the linear relationship between the expected return of the asset and a number of macroeconomic variables that capture systemic risk. It is a valuable method for evaluating portfolios from a value-investing perspective, to find securities that might be mispriced temporarily.

#### Formula

$$E(R)_i = E(R)_z + (E(I) - E(R)_z) \times \beta_n$$

$E(R)_i$  = Expected return on the asset

$Rz$  = Risk-free rate of return

$\beta_n$  =Sensitivity of the asset price to macroeconomic factor  $n$

$Ei$ =Risk premium associated with factor  $i$

Using linear regression, the beta coefficients in the APT model are calculated. APT factors are now the systematic risk that couldn't be minimized by the diversification of an investment portfolio (Investopedia, 2020).

### Yield curve (term structure of interest rates)

The term structure of interest rates, usually referred to as the yield curve, defines the interest rates of comparable quality bonds at different maturities.

#### Formula

$$r_{n,t} = P + \frac{[r_{1,t+1} + E_t(r_{1,t+2}) + \dots + E_t(r_{1,t+n-1})]}{n}$$

The estimated returns for one year are all  $t$  in time. Assume that the investor bases his / her assumption on the current one-year bond,  $r_{1,t}$ . The approximate equation for the long-term  $n$ -year bond is  $r_{n,t} = a + b r_{n,t}$  [approximate yield curve equation].

### Early warning systems

It is understood that financial imbalances typically lead to systemic financial stress that can cause banks or other businesses to fail, financial crises and recessions.

So detecting and anticipating future financial stress is of critical importance. A host of early warning systems exist for financial or credit crises.

One of those system is called MIMIC that consists of two sets of equations:

$$Y = bW + v \text{ and } W = cX + u$$

Where Y is an indication variable of crisis (crisis indicator), X is an observed possible cause of crisis and W is a latent, variable thinking reflecting the magnitude of the crisis. The words u and v for disturbance are later clarified. Macroeconomic variables which could be used as measures of crisis include:

- Credit growth
- Equity price growth
- Property price growth
- Credit to GDP gap

### Sharpe ratio

The Sharpe ratio is also used to denote the return offered taking into account the risk of the investment. When two assets are compared, the one with a higher Sharpe ratio shows a greater return for the same risk,  $\sigma$ .

### Stock valuation (present value model)

At time t, the price, P, of a share of company stock is equal to the estimated discounted amount of all future dividends (cash payments):

$$P_t = E_t \left\{ \frac{D_{t+1}}{(1 + \delta)} + \frac{D_{t+2}}{(1 + \delta)^2} + \dots \right\}$$

where:

$D$  = dividend expected at some time in the future

$\delta$  = the constant discount factor.

If  $D$  is constant then:

$$P_t = \frac{D_t}{\delta}$$

In log form:

$$\ln(P_t) = -\ln(\delta) + \ln(D_t)$$

### Parametric (analytical) Value-at-Risk (VaR)

Many companies have implemented the value-at-risk, or VaR, metric when calculating risk exposure, which is a predictive risk management methodology that estimates the potential loss that an investment portfolio is likely to face with a certain degree of confidence within a given time frame.

VaR modeling calculates the risk of loss in the assessed individual, and the likelihood of occurrence for the identified loss. One calculates VaR by determining the amount of possible loss, the likelihood of loss occurrence, and the timeline.