# Quantitative Methods

## Types of risk

1. Default risk
2. Liquidity risk
3. Maturity risk

Required interest rate on a security = nominal risk-free rate + default risk premium + liquidity risk premium + maturity risk premium

## Types of measures

1. Nominal scales
   1. 의미 없이 번호만 붙임
2. Ordinal scale
   1. 각 숫자의 우열은 있지만, 3과 4의 차이와 4와 5의 차이가 다름. 단지 다르다는 것만 알 수 있음
   2. Ex) Rank
3. Interval scale
   1. 각 숫자에 우열이 있고, 3과 4의 차이가 4와 5의 차이와 같음. 그러나 30이 10의 3배를 의미하는 것이 아님
   2. Ex) temperature
4. Ratio scale
   1. 각 숫자가 우열이 있고, 30이 10의 3배를 의미하고, 0은 진짜 0을 의미
   2. Ex) dollar

For values that are not equal: harmonic mean < geometric mean < arithmetic mean

## Chebyshev’s inequality

* 36%: 1.25 standard deviation
* 56%: 1.5 standard deviation
* 75%: 2 standard deviation
* 89%: 3 standard deviation
* 94%: 4 standard deviation

## Coefficient of variation (CV)

CV = standard deviation of x / average value of x

## Skewness

* Positively skewness – long right tail
  + Mode < median < mean
* Negatively skewness – long left tail
  + Mode > median > mean

## Kurtosis

* Leptokurtic
  + More peaked, kurtosis > 3, fat tail
* Platykurtic
  + Less peaked, kurtosis < 3, fat tail
* Mesokurtic
  + Kurtosis = 3

## Spurious correlation

* Correlation that is either the result of chance or present due to changes in both variables over time that is caused by their association with a third variable.

## Confidence interval

* 90%: 1.65
* 95%: 1.96
* 99%: 2.58

## Roy’s safety-first criterion

* The optimal portfolio minimizes the probability that the return of the portfolio falls below some minimum acceptable (threshold) level.
* Maximize the SFRatio, where SFRatio = (E(Rp) – threshold) / sd of portfolio

## Monte Carlo simulation

### Procedures

1. Specify the probability distributions of stock prices and of the relevant interest rate, as well as the parameters (mean, variance, possibly skewness) of the distributions.
2. Randomly generate values for both stock prices and interest rates.
3. Value the options for each pair of risk factor values.
4. After many iterations, calculate the mean option value and use that as your estimate of the option’s value.

### Usage

* Value complex securities
* Simulate the profits/losses from a trading strategy
* Calculate estimates of VaR to determine the riskiness of a portfolio of assets and liabilities.
* Simulate pension fund assets and liabilities over time to examine the variability of the difference between the two.
* Value portfolios of assets that have nonnormal returns distributions.

### Limitations

* Fairly complex
* Provide answers that are no better than the assumptions about the distributions of the risk factors and the pricing/valuation model that is used.
* Simulation is not an analytic method, but a statistical one, and cannot provide the insights that analytic methods can.

## Historical simulation

* Based on actual changes in value or actual changes in risk factors over some prior period.

### Advantages

* Using the actual distribution

### Disadvantages

* Past changes may not be a good indication of future changes.
* Event that occurs infrequently may not be reflected in historical simulation results unless the events occurred during the period from which the values for risk factors are drawn.
* Cannot address the sort of ‘what if’

Time-series data – observations taken over a period of time at specified and equally spaced time intervals.

Cross-sectional data – observations taken at a single point in time.

## Central limit theorem

* For simple random samples of size n from a population with a mean (mu) and a finite variance (sigma), the sampling distribution of the sample mean approaches a normal probability distribution with mean (mu) and variance equal to (sigma/sqrt(n)) as the sample size becomes large.

## Desirable properties

### Unbiased

* One for which the expected value of the estimator is equal to the parameter you are trying to estimate.
* An unbiased estimator is efficient if no other biased estimator of the same parameter has a sampling distribution with smaller variance.

### Efficient

* An unbiased estimator is also efficient if the variance of its sampling distribution is smaller than all the other unbiased estimators of the parameter you are trying to estimate.

### Consistent

* One for which the accuracy of the parameter estimate increases as the sample size increases.

## Appropriate test statistic

|  |  |  |
| --- | --- | --- |
|  | Small sample (n < 30) | Large sample (n >= 30) |
| Normal dist. & pop. Var. known | z-statistics | z-statistics |
| Normal dist. & pop. Var. unknown | t-statistics | t-statistics\* |
| Nonnormal dist. & pop. Var. known | Not available | z-statistics |
| Nonnormal dist. & pop. Var. unknown | Not available | t-statistics\* |

\*more conservative

## Bias

1. Data-mining bias
   1. Data mining – when analysts repeatedly use the same database to search for patterns or trading rules until one that ‘works’ is discovered.
   2. Results where the statistical significance of the pattern is overestimated because the results where found through data mining.
2. Sample selection bias
   1. When some data is systematically excluded from the analysis, usually because of the lack of availability.
3. Survivorship bias
4. Look-ahead bias
   1. When a study tests a relationship using sample data that was not available on the test date.
5. Time-period bias
   1. If the time period over which the data is gathered is either too short or too long.

## Hypothesis testing procedure

1. State the hypothesis
2. Select the appropriate test statistic
3. Specify the level of significance
4. State the decision rule regarding the hypothesis
5. Collect the sample and calculate the sample statistics
6. Make a decision regarding the hypothesis
7. Make a decision based on the results of the test

Type I error – the rejection of the null hypothesis when it is actually true

Type II error – the failure to reject the null hypothesis when it is actually false

Significance level = P(Type I error)

Power of test = 1 – P(Type II error)

## Hypotheses about two population means

Independent samples – difference-in-means test

Dependent samples – paired comparison test

Both of these are t-tests

## Hypotheses about variance

Variance of a single population – Chi-square test

Compare variance of two populations – F-test

Parametric tests rely on assumptions regarding the distribution of the population and are specific to population parameters.

Nonparametric tests either do not consider a particular population parameter or have few assumptions about the population that is sampled.

The Spearman rank correlation test can be used when the data are not normally distributed.

## Correlation t-test

## Exhaustive event

Events are exhaustive when they cover all possible outcomes.

## Mutually exclusive

Only one event can occur at a time.

## Dependent events

Two events are dependent if the occurrence of one event does affect the probability of occurrence of the other event.

## Tree diagram

A problem is worked backward to formulate an expected value as of today.

## Nonparametric test

Primarily concerned with ranks, signs, or groups, and they are used when numerical parameter are not known or do not meet assumptions about distribution.

## Odd

Odd = P(E) / (1 - P(E))

## Cross-sectional data

* Data on some characteristics of companies at a single point.

## Correlation coefficient

* The correlation coefficient for each pair of component securities are required to describe a distribution of returns for a portfolio.
* If a large number of outliers exists, the correlation is not reliable.

## Degree of freedom

The number of observations that can be chosen independent for the calculation of a sample variance.

## Forecasting future return

* Harmonic mean – not used to compute historical performance or forecast the expected performance of investment
* Geometric mean – best estimate of future multi-year annual compound returns
* Arithmetic mean – best estimate of a single year’s return