Quantitative

- Fintech in Investment Management
- Correlation and Regression
- Multiple Regression and Issues in Regression Analysis
- Time-series Analysis
- Probabilistic Approaches: Scenarios analysis, decision trees, and simulations

Fintech in Investment Management

Correlation and Regression

Summaries

- Correlation test
- Linear regression coefficient (t-test with $n-k-1$ degree of freedom)
    - Hypothesis test, confidence interval, p-value
- ANOVA
    - SSE, F(one-tailed), R^2, R^2_{ajust}

Covariance and Correlation

- Sample Variance 样本方差
    - $Cov(X,Y) = C_{XY} = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{n-1}$
- Sample Correlation Coefficient
    - $r_{XY} = \frac{C_{XY}}{s_X s_Y}$

Correlation Limitation

- Outliers
- Spurious Correlation
    - Correlated by chance with no economic explanation
- Nonlinear Relationships

Correlation Hypothesis Test

- $H_0: \rho = 0 \quad versus\ H_a: \rho \neq 0$
- Assume two populations are normally distributed
- T-test with degree of freedom $(df = n-2)$
    - $t = \frac{r}{\sqrt{1-r^2}} \times \sqrt{n-2}$
    - mean r
    - variance $1-r^2$, standard deviation$\sigma = \sqrt{1-r^2}$
    - degree of freedom $df = n-2$
    - standard error $\frac{\sigma}{\sqrt{df}} = \frac{\sqrt{1-r^2}}{\sqrt{n-2}}$

Linear Regression Variables

- dependent: explained/endogenous/predicted
- independent: explanatory/exogenous/predicting

Linear Regression Assumptions

- linear relationship exists between the dependent and independent variable
- independent variable is **uncorrelated** with the residuals
- the expected value of residual terms is **zero**
- the variance of the residual term is **constant** for all observations
    - Heteroskedasticity
- the residual term is **independently** distributed
    - Auto Correlation
- The residual term is **normally** distributed

Linear Regression Model

- $\mathrm{Y} = \mathrm{b}_0 + b_1 X + \epsilon$
  - b_0: intercept term
  - b_1: slope coefficient
  - ϵ: residual

Linear Regression Parameter Estimation

- Sum of squared errors (SSE) $\mathrm{SSE} = \sum_{\mathrm{i}} \epsilon_i^2$
  - Ordinary least squares (OLS) and least squares estimates
- slope coefficient $\mathrm{b}_1 = \frac{C_{XY}}{s_X^2}$
- intercept term $\mathrm{b}_0 = \bar{Y} - b_1 \bar{X}$ (using mean point)

Regression Coefficient

- **Distribution**
  - T-distribution with degree of freedom $\mathrm{n} - \mathrm{k} - 1 = \mathrm{n} - 2$
- **Confidence Interval**
  - $\widehat{\mathrm{b}_1} \pm \mathrm{t_c} \times \widehat{s_\iota}$
  - $\mathrm{t_c}$ is the critical two-tailed value
- **Test statistic**
  - $t = \frac{\widehat{b_\iota} - b_i}{\widehat{s_\iota}}$
  - Reject: t > critical value
- **p-value**
  - Smallest level of significance
  - Reject: p-value < significance level

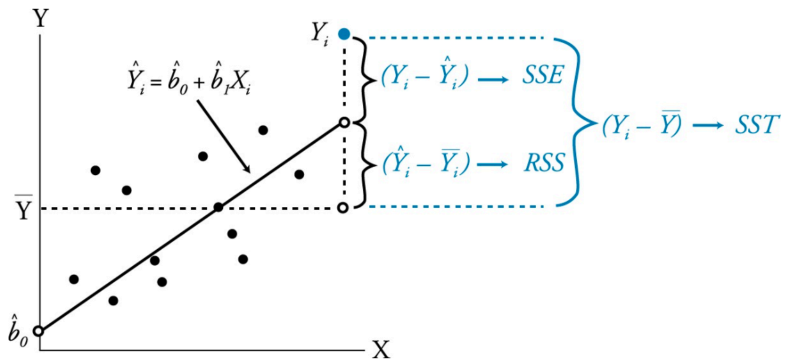
Predicting

- $\boldsymbol{\widehat{Y} = \widehat{b_0} + \widehat{b_1} X}$
- **Confidence interval** $\widehat{\mathrm{Y}} \pm \mathrm{t_c} \times \widehat{s_f}$
  - $\widehat{s_f}$: standard error of the forecast
  - $s_f^2 = SEE^2(1 + \frac{1}{n} + \frac{(X - \bar{X})^2}{(n-1)s_x^2})$
  - Degree of freedom $\mathrm{n} - \mathrm{k} - 1$

ANOVA

- **Total sum of squares (SST)**
  - $SST = \sum(Y_i - \bar{Y})^\wedge 2$
- **Regression sum of squares (RSS)**
  - $RSS = \sum(\widehat{Y_\iota} - \bar{Y})^\wedge 2$
- **Sum of squared errors (SEE)**
  - $SEE = \sum\left(Y_i - \widehat{Y_\iota}\right)^2$
- **Equation**
  - $SST = RSS + SEE$

Figure 7.8: Components of the Total Variation



Variation	Degree of freedom	Sum of squares	Mean sum of squares	Squared
Regression	k	RSS	$\mathrm{MSR} = \frac{\mathrm{RSS}}{\mathrm{k}}$	
Error	n-k-1	SSE	$\mathrm{MSE} = \frac{\mathrm{SSE}}{\mathrm{n-k-1}}$	**SEE** $= \sqrt{MSE}$
Total	n-1	SST	$\mathrm{MST} = \frac{\mathrm{SST}}{\mathrm{n-1}}$ (one-tailed test)	
		$\mathrm{R}^2 = \frac{RSS}{SST}$	$F = \frac{MSR}{MSE}$	
		$\mathrm{R}^2 = 1 - \frac{SSE}{SST}$	$R_a^2 = 1 - \frac{MSE}{MST} = 1 - \frac{\mathrm{SSE}}{\mathrm{SST}} \times \frac{n-1}{n-k-1}$	

Standard Error of Estimates (SSE) 标准误

- degree of variability of the actual and predicted Y-values
- SSE = **standard deviation** of error terms
- $\mathrm{SEE} = \sqrt{MSE}$
- **Standard error** of estimates/residual/regression

Coefficient of Determination ($\mathbf{R}^2$)

- Percentage of total **variation** in the dependent variable explained by the **variation** of independent variable
- $\mathrm{R}^2 = \frac{RSS}{SST} = 1 - \frac{SSE}{SST}$
- $\mathrm{R}^2 = \rho^2$ for on independent variable
- Increases as more variables are added

Multiple R

- Correlation between actual and predicted Y
- square root of R^2, that's **multiple R** = $\sqrt{\mathrm{R}^2}$
- The correlation between X and Y for only one independent variable

Adjusted $\mathbf{R}^2$

- $R_a^2 = 1 - \frac{MSE}{MST} = 1 - \frac{\mathrm{SSE}}{\mathrm{SST}} \times \frac{n-1}{n-k-1} = 1 - (1 - \mathrm{R}^2) \times \frac{n-1}{n-k-1}$

F-statistic

- Measures how well a set of independent variables (at least one independent variable) explains the variation in the dependent variable.
- $F = \frac{MSR}{MSE}$
- It is always a **one-tailed** test.
- Hypothesis: $\mathrm{H_0}: b_i = 0 \ \forall i \in 1 \cdots n$ (all **slopes** equal to zero)
- Degree of freedom (k, n-k-1)
- Decision rule: reject $\mathrm{H_0}$ if $\mathrm{F} > \mathrm{F_c}$
- F and t in simple regression
    - $\mathrm{F} = \mathrm{t}_{\mathrm{b}_1}^2$

Linear Regression Limitation

- Change over time
    - Parameter instability
- Assumptions may not hold
    - Heteroskedastic
    - Autocorrelation

Multiple Regression and Issues in Regression Analysis

Summary

- Qualitative Independent
  - n class with n-1 dummy variables
- Qualitative Dependent
  - Probit or logbit or discriminant variables
- Heteroskedasticity – error trend
  - standard errors are affected, t-test and F-test are **unreliable**
  - **Breusch-Pagan** Chi-square $test = n \times \mathrm{R}^2_{\mathrm{res}} \sim \mathrm{X}^2(k)$ **one-tailed** test
  - Correction: Robust standard errors or generalized least squares
- Auto/Serial Correlation – error correlation
  - H_0: no positive correlation
  - **Durbin Watson** test $DW = 2(1-r) \sim DW(n,k)$
  - less than lower -> reject H_0 and positive correlation
  - middle: inconclusive
  - larger than upper -> fail to reject
  - Correction: Hansen-Corrected errors or
- Multicollinearity
  - High R^2 and significant F-test but no t-tests are significant
  - Pairwise correlation
  - Correction: exclude one or more variables

Model

- Intercept term
- Partial slope coefficients
  - Holding others constant

Qualitative Independent Variables - Dummy Variable

- Dummy variable: value equals to 0 or 1
- n classes use $n-1$ dummy variables
- A model with four quarters
  - $Y = b_0 + b_1 Q_1 + b_2 Q_2 + b_3 Q_3 + \epsilon$
  - $Q_1 = 1$ if it is the first quarter and 0 otherwise
- Parameters
  - **Reference** point: the **omitted** class – the fourth quarter
  - Intercept: average value for the **fourth** quarter
  - Slope: **difference** between the current quarter and the fourth quarter
- Test
  - $b_i = 0$ means the current quarter = fourth quarter

Qualitative Dependent Variable – Other models

- **Probit and logit models**
  - Probit: normal distribution
  - Logit: logistic distribution
- **Discriminant models**
  - No assumptions about independent variables
  - A linear function similar to an ordinary regression

            - Score or rank -> classify

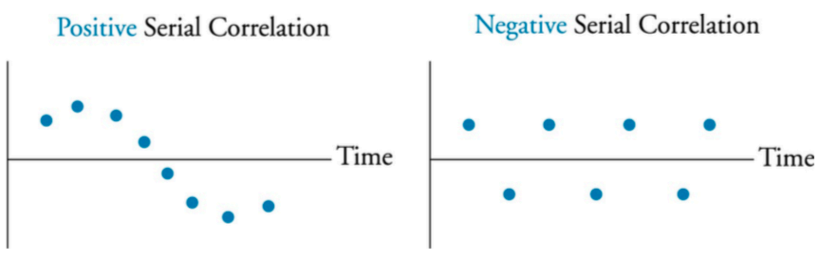
Linear Regression Assumptions

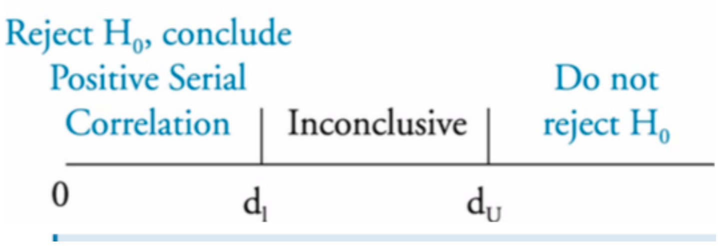
- linear relationship exists between the dependent and independent variable
- independent variable is **uncorrelated** with the residuals
- independent variables are not random
- independent variables have no exact linear relationship
    - **Multicollinearity (Correlation test)**
- the expected value of residual terms is zero
- the variance of the residual term is constant for all observations
    - **Heteroskedasticity (Chi-Square test)**
- the residual term is independently distributed
    - **Auto Correlation (DW test)**
- The residual term is normally distributed

Heteroskedasticity 异方差 – Error Trend

- **What is it?**
    - Variance across observations are not the same
- **Classification**
    - **Unconditional**
        - not related to level of independent variables
        - no major problems
    - **Conditional**
        - related to level of independent variables
        - cause **significant** problems
- **Effect**
    - Coefficient: **consistency** not affected
    - Standard errors: **unreliable**
        - Too small -> t large -> reject often -> type I error 拒真
        - Too large -> t small -> not reject -> type II error 受假
    - T-test: unreliable
    - F-test: unreliable
- **Detect**
    - Residual plot: residual vs independent variables
    - **Breusch-Pagan** Chi-square test (X^2)
        - Residual vs independent variables regression
            - The R-squared is $\mathrm{R}^2_{\mathrm{res}}$
        - X^2 with degree of freedom k (the number of independent variables)
        - $test = n \times \mathrm{R}^2_{\mathrm{res}} \sim \mathrm{X}^2(k)$
        - **One-tailed** test
- **Correct**
    - **Robust (White-corrected) standard errors**
        - Use them to recalculate the t-statistics with the original coefficients
    - Generalized least squares
        - Modify original equation

Serial Correlation – Error Correlation

- **What is it?**
  - Residual terms are correlated with one another
- Classification
  - Positive correlation
  - Negative correlation
- Effect
  - **Positive correlation** -> Type I error
    - Small standard errors -> reject more -> Type I error
  - Negative correlation -> type II error
    - Large standard errors -> reject less -> Type II error
- Detect
  - Residual plot: residual vs time
  - 

  - Durbin-Watson Statistic (DW)
    - $DW = \frac{\sum(\epsilon_t - \epsilon_{t-1})^2}{\sum \epsilon_t^2} \sim DW(n, k)$
  - **Positive** Correlation: When sample size is large
    - $DW \approx 2 \times (1 - r) \sim DW(n, k)$
    - $r = correlation\ between\ \epsilon_t\ and\ \epsilon_{t-1}$
  - DW
    - ≈ 2 if homoscedastic and no serially correlated
    - < 2 if positively serially corrected
    - > 2 if negatively serially corrected
  - DW has two values
    - Positive correlation d_{lower} and d_{upper}
    - Negative correlation $4 - d_{upper}$ and $4 - d_{lower}$
  - Hypothesis
    - H_0: no positive correlation
  - Decision rule
    - 

- Correct
  - **Hansen**-Corrected standard errors
    - Use White-corrected only when heteroskedasticity
    - Use Hanse for serial or both situations
  - Improve the specification of the model

            - Include time-series nature of the data

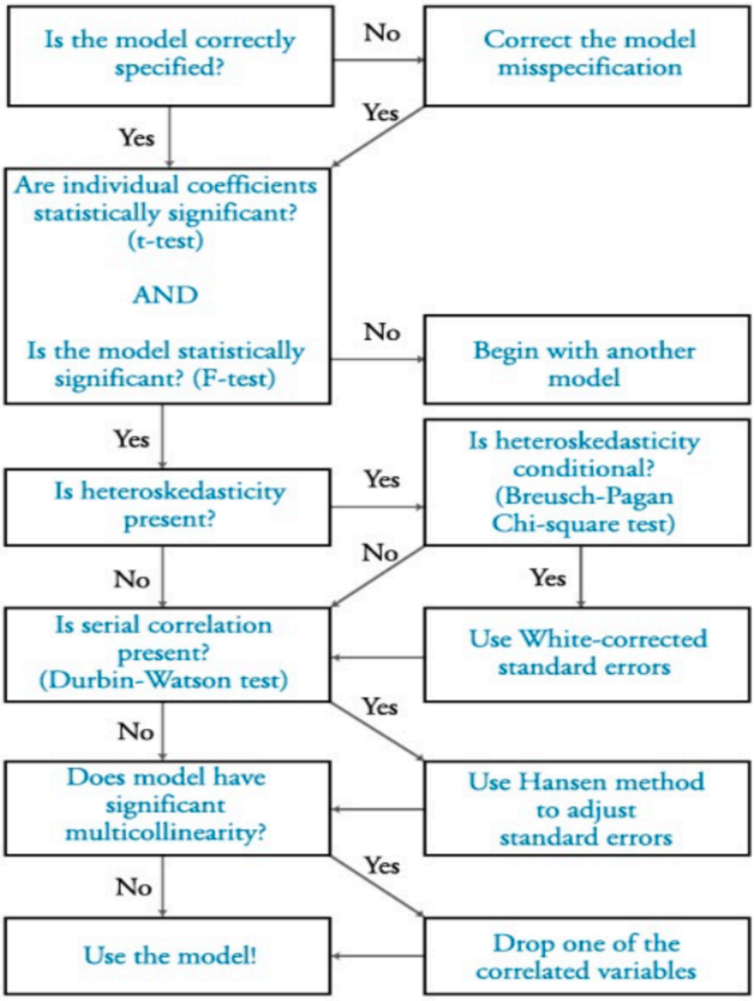
Multicollinearity

- What's it?
    - two or more variables or their combinations are highly correlated with each other
- Effect
    - Coefficient
        - Unreliable
        - Consistency
    - Standard errors
        - Inflated -> **type II error**
- Detect
    - **F-test and T-test**
        - R^2 is high
        - F-test is significant
        - But T-test indicate no coefficients are significantly different from zero
    - Pairwise Correlation
        - High correlation among independent variables is a sign
        - Two variables
            - If correlation > 0.7, is a potential problem
        - More than two
            - High -> possibility of multicollinearity
            - Low -> does not indicate it is not present
- Correct
    - Omit one or more variables
        - Stepwise regression

Violation	**Conditional Heteroskedasticity**	**Serial Correlation**	**Multicollinearity**
What is it?	Residual variance related to level of independent variables	Residuals are correlated	Two or more independent variables are correlated
Effect?	Coefficients are consistent. Standard errors are underestimated. Too many Type I errors.	Coefficients are consistent. Standard errors are underestimated. Too many Type I errors (positive correlation).	Coefficients are consistent (but unreliable). Standard errors are overestimated. Too many Type II errors.
Detection?	Breusch-Pagan chi-square test = $n \times R^2$	Durbin-Watson test $\approx 2(1 - r)$	Conflicting *t* and *F* statistics; correlations among independent variables if k = 2
Correction?	Use White-corrected standard errors	Use the Hansen method to adjust standard errors	Drop one of the correlated variables

Model Misspecification

- Classification
    - Function form
        - Important variables are **omitted**
        - Variables should be **transformed**
            - Z-score, log, square, squared
        - Data is improperly **pooled**
            - Relationship changes over time
    - Variables are **serial corrected** (time series models)
        - A **lagged dependent** variable is used as an independent variable
        - a **function** of dependent variables is used as an independent variable
            - predicting the past
        - independent variables are measured with **error**
    - **nonstationary** caused by other time-series misspecifications
- **Effects**
    - Misspecification -> coefficients are **biased and/or inconsistent** -> unreliable hypothesis testing and inaccurate predictions



Supervised Machine Learning

- Big data

- Structed and unstructured data

- Data analytics
    - Measure correlation
    - Make prediction
    - Make casual inferences
    - Classify
    - Clustering
    - Reduce dimension
- Classification
    - Supervised: classification, prediction
    - Unsupervised: clustering
- Supervised
    - Regression models
        - Penalized regression (overfitting)
    - Classification and Regression Trees (CART)
    - Random forest
    - Neural networks
        - Activation function (nonlinear)
- Unsupervised
    - Clustering
    - Dimension reduction
        - PCA

Time-series Analysis

Summary

- **Trend Model:** linear, log-linear 趋势模型
- **Autoregressive Model AR(p)**
- **Serial Correlation 残差自相关**
  - **Residual** Autocorrelations test **残差相关性检验**
    - degree of freedom $T-2$
    - $t = \rho_{\epsilon_t \epsilon_{t-k}} \sqrt{T} \sim T(T-2)$
    - Residual standard error $1/\sqrt{T}$
  - Correction: add lagged values
- **Conditional Heteroskedasticity (ARCH) 残差异方差**
  - $\epsilon_t^2 = a_0 + a_1 \epsilon_{t-1}^2 + u_t$
  - T-Test $H_0: \mathrm{a}_1 = 0$
  - Correction: generalized least squares
- Seasonality Model 季节性模型
  - Effects: AR model is mis-specified
  - Correction: add additional lag of dependent variable
- Covariance Stationary 协方差稳定
  - Mean constant, variance constant, covariance constant
- **Mean Reversion 均值回归**
  - covariance stationary times series -> mean reversion
  - AR(1) model with lag coefficient less than 1 -> mean reversion
  - $\mu = \frac{b_0}{1-b_1}$ if $|b_1| < 1$
- **Unit Root Time Series/ Random Walk Process -> Covariance Nonstationary 非稳态**
  - Test: **Dickey Fuller**-Test
    - $H_0: \mathrm{g} = 0 \rightarrow \mathrm{b}_1 = 1 \rightarrow \textit{unit root}$
  - Correction: first difference and model it with AR model
- Cointegration
  - Transform: linear regression error term
  - Hypothesis: error term timer series should be covariance stationary
  - Test: **DF-EG** test for unit root problem
- **Nonstationary Characteristics**
  - Non-constant mean: unit root
  - Non-constant variance: conditional heteroskedasticity
  - Non-constant correlation: serial correlation
  - Seasonality: lagged correlation
  - Structural change: different models

Trend Model

- **Linear trend**
  - $\mathrm{y_t} = b_0 + b_1 t + \epsilon_t$
  - Time begins with 1
  - Variable increases over time by a constant amount
- **Log-linear trend**
  - $\log y_t = b_0 + b_1 t + \epsilon_t \rightarrow y_t = \exp(b_0 + b_1 t)$

- Variable grows at a constant rate
- Exponential growth

- Limitations
  - **Autocorrelation**
    - use DW test
    - consider AR models

Autoregressive Model (AR)

- Model
  - $\mathrm{x_t} = b_0 + b_1 x_{t-1} + \epsilon_t$
- AR(p)
  - $\mathrm{x_t} = b_0 + b_1 x_{t-1} + \cdots + b_p x_{t-p} + \epsilon_t$
- Forecasting
  - Chain rule of forecasting
- **Serial Correlation**
  - the residuals should have no serial correlation
- **Conditional Heteroskedasticity**
  - The residuals should have no conditional heteroskedasticity
- **Covariance stationary**
  - Statistical Inferences based on OLS may be invalid unless the time series is Covariance stationary

Model Fit Steps – T-test

- **Estimate** the AR model using linear regression
  - Start with AR(1), and then increase it by 1
- **Calculate** the **autocorrelation** of residuals
- **Test** whether the **autocorrelations** are significantly different from zero

Residual Autocorrelations Test 残差的自相关性检验– T-test

- Residuals are correlated at different lags
- Idea
  - Calculate residual correlation for different lags 不同 lag 的相关系数
  - Test the correlation significance
- For each lag $k = 1,2,\cdots,p$
- $\mathrm{H_0}$: correlations of residuals are zero
- Mean $\rho_{\epsilon_t \epsilon_{t-k}}$, the correlation of error term t with the kth lagged error term
- Standard deviation of residual is 1 (normal distribution).
- Number of samples T
- **Residual** Standard error is $\frac{1}{\sqrt{T}}$
- T-test with degree of freedom $\mathrm{T} - 2$
- $\mathrm{t} = \frac{\rho_{\epsilon_t \epsilon_{t-k}}}{1/\sqrt{T}} = \boldsymbol{\rho_{\epsilon_t \epsilon_{t-k}} \sqrt{T}}$
- larger absolute t, reject the null hypothesis
- correction: add lagged values

Autoregressive conditional heteroskedasticity (ARCH) 残差平方的线性相关检验 – T-test

- Variance of residuals depends on the variance of residuals in a previous period
- Idea
    - Residual variance time series -> apply AR(1)
- ARCH(1) time series
    - $\epsilon_t^2 = a_0 + a_1\epsilon_{t-1}^2 + u_t$
    - u_t is an error term
- T-Test
    - If a_1 is statistically different from zero, it is ARCH(1) and heteroskedasticity
- Correction: generalized least squares
- Application: predict variance of future residuals

Seasonality

- A pattern tends to repeat from year to year
    - x_t is related to x_{t-12} for monthly data
    - The correlation between them is quite high
- Correcting
    - $x_t = b_0 + b_1x_{t-1} + \epsilon_t \rightarrow x_t = b_0 + b_1x_{t-1} + b_2x_{t-12} + \epsilon_t$
    - Add an additional lagged variable

Forecast Error

- In-sample forecast
- Out-of-sample forecast
- Root mean squared error (**RMSE**) for out-of-sample data
    - Square root of the average of the squared errors
- Lower RMSE for out-of-sample data will have more predictive power

Coefficients Stability

- Instability or nonstationary
- Dynamic conditions
- Shorter time series are more **stable** in coefficients
- Longer time series are more statistical **reliability**
- The underlying economic processes
    - Regulatory changes? Dramatic change in the underlying economic environment

Covariance Stationary

- Constant and finite **expected value** – mean reverting level
- Constant and finite **variance**
    - Volatility around its mean does not change over time
- Constant and finite **covariance** between values at any given lag
    - The covariance with leading or lagging values of itself is constant

Covariance Stationary -> Mean Reversion

- All covariance stationary time series will have a mean-reverting level
- $\mathrm{x_t} = b_0 + b_1x_t \rightarrow x_t = \frac{b_0}{1-b_1}$
    - It has mean-reverting level if $|b_1| < 1$

- Predicts the next value will be the same as its **current** value
  - $\widehat{x_t} = x_{t-1}$ 用当前值估计

Unit Root Time Series / Random Walk Process -> Covariance Nonstationary

- Unit root time series
  - The coefficient $b_1 = 1$
  - Least squares regression will not work without transforming the data
  - Cannot be fit using **AR** model
- Random walk $\mathrm{x_t} = x_{t-1} + \epsilon_t$
  - Best estimate of x_t is x_{t-1}
  - $E(\epsilon_t) = 0$: the expected value of each error term is zero
  - $E(\epsilon_t^2) = \sigma^2$: the variance of the error terms is constant
  - $E\left(\epsilon_i \epsilon_j\right) = 0\ if\ i \neq j$: no serial correlation in error terms
- Random walk with drift $\mathrm{x_t} = b_0 + x_{t-1} + \epsilon_t$
  - b_0: constant drift
- Covariance Nonstationary
  - Because $b_1 = 1 \rightarrow \frac{b_0}{1-b_1}\ undefined$

Covariance Nonstationary Test – DF Test

- $\mathrm{x_t} = b_0 + b_1 x_{t-1} + \epsilon_t$
  - $\rightarrow \mathrm{x_t} - x_{t-1} = b_0 + (\boldsymbol{b_1 - 1}) x_{t-1} + \epsilon_t$
  - $\rightarrow y_t = b_0 + \boldsymbol{g}\, x_{t-1} + \epsilon_t$
- Dickey and Fuller Test
  - Null hypothesis H_0: $g = 0$ (the time series has a unit root) 假定是非稳定的
  - $g = b_1 - 1$
  - If the model can be rejected, it does not have a unit root

Covariance Nonstationary Correction - First Differencing

- A random walk can be transformed into a covariance stationary time series using first differencing
- A new time series with $y_t = \mathrm{x_t} - x_{t-1} = \epsilon_t$
  - $y_t = b_0 + b_1 y_{t-1} + \epsilon_t = \epsilon_t$
  - $\rightarrow b_0 = b_1 = 0$
- Mean revering level is $\frac{b_0}{1-b_1} = 0$
- Steps: difference -> lag -> regression

Two Time Series – Linear Regression

- $y_t = b_0 + b_1 x_t + \epsilon$
- Both are covariance stationary
  - -> linear regression
- Dependent variable is covariance stationary
  - -> not reliable
- Independent variable is covariance stationary
  - -> not reliable
- Neither is covariance stationary

- Not cointegrated
- They are **cointegrated**
  - -> linear regression

Cointegration

- They are linked or follow the same **trend** and that relationship is not expected to change
- If cointegrated
  - Error term from regressing one on the other is **covariance stationary**
  - T-tests are reliable
- Test
  - $y_t = b_0 + b_1 x_t + \epsilon$
- DF-EG test
  - Residuals are tested for a **unit root using** Dickey Fuller test with critical t-values calculated by Engle and Granger
  - If rejected the unit root, then they are cointegrated

Nonstationary Characteristics

- Non-constant mean: unit root
- Non-constant variance: conditional heteroskedasticity
- Non-constant correlation: seasonality
- Seasonality: lagged correlation
- Structural change

Steps

- No seasonality or structural shift -> trend model (linear or log-linear)
- Residuals -> serial correlation with **Durbin Watson** test
  - No: use the trend model
  - Yes: use another model (AR)
- Check stationarity before running an AR model
- **If not stationary**
  - Linear trend -> first-difference the data
  - Log-Linear trend -> first-difference the log of the data
  - Structure change -> two separate models
  - Seasonal component -> add lagged variable
- After first-differencing
  - If no serial correlation and seasonality -> use the model
  - Otherwise, add seasonality
- Test for ARCH
  - Coefficient not significantly from zero -> use the model
  - Otherwise -> use generalized least squares
- Two models -> lower out-of-sample RMSE

Probabilistic Approaches: Scenarios analysis, decision trees, and simulations

Simulations

- Determine the probabilistic **variables**
- Define **probability distributions** for these variables
    - Historical data
    - Cross-sectional data
    - Pick a distribution and estimate the parameters
        - Subjective specification
- Check for **correlations** among variables
    - Use historical data to determine whether they are related
    - Solutions
        - Allow one variable to vary and others can be computed
        - Build the rules of correlation into simulation
- Run the simulation
    - Randomly draw variables
    - Use them to generate estimated values
    - Number of simulations
        - Number of uncertain variables
        - Types of distributions
        - The range of outcomes
- Advantages
    - Better input **quality**
    - Provides a **distribution** of expected value rather than a point estimate
- Constraints
    - Book Value constraints
        - Regulatory capital requirements
        - Negative equity
    - Earnings and cash flow constraints
        - Can be imposed internally to meet analyst expectations or to achieve bonus targets.
        - Can be imposed externally, such as a loan covenant.
    - Market value constraints
        - Minimize the likelihood of financial distress or bankruptcy
- Limitations
    - Input quality
    - Inappropriate statistical distributions
    - Non-stationary distributions
    - Dynamic correlation

Risk-Adjusted Value

- Cash-flow are not risk-adjusted, should not be discounted at risk-free rate
- Do not double count risk

Simulation, Scenario analysis, and Decision trees

- Simulation -> continuous risk
- Scenario analysis and decisions trees -> discrete risk

- Scenario analysis - **Correlation**
    - A finite set of scenarios (best, worst and most likely cases)
- Decision trees - **Sequential**
    - Discrete and sequential risks
    - Cannot include correlation

Appropriate method	Distribution of risk	Sequential?	Accommodates Correlated Variables?
Simulations	Continuous	Does not matter	Yes
Scenario analysis	Discrete	No	Yes
Decision trees	Discrete	Yes	No