FRM 一级 Quantitative Analysis 知识图谱

tell:400-600-8011 邮箱:frm@gaodun.com

① 提出模型

$Y = \alpha + \beta X + \varepsilon$

- α=intercept
- β=slope (regression coefficient)
- ε=shock/innovation/error

$Y = \alpha X^{\beta} \epsilon$ —— 转化 —— $\ln Y = \ln\alpha + \beta \ln X + \ln \epsilon$

Dummy variables —— 非黑即白，2选1（要么是0，要么是1）

Essential restrictions of linear regressions
- ❶ Linear relationship between Y and explanatory variables with unknown coefficients.
- ❷ Error must be additive.
- ❸ All explanatory variables must be observable.

②a 前提假设

Assumptions
- Homoskedasticity.
- Var(X) is positive.
- E[ε|X]=0 (mean independence) —— Violation
  - Sample Selection bias
  - Simultaneity bias
  - Omitted variables
  - Attenuation bias
- Data are realization from iid random variables.
- No outliers

③a 回归方法（OLS）

Unbiased and jointly normal distributed

Estimators
- $\hat{\beta} = \frac{Cov(X,Y)}{Var(X)}$
- $\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$

$\bar{X}, \bar{Y}$在回归线上

残差平方和最小

④a 假设检验（Inference and Hypothesis Testing）

t-test
- ❶ 提出假设 $H_0: \beta = \beta_0,\ H_1: \beta \neq \beta_0$
- ❷ 算出T统计量 $test\ statistic\ T = \frac{\hat{\beta} - \beta_0}{s.e.(\hat{\beta})}$
- ❸ 判断依据 $|t| > C_s$ / p-value≤α —— 拒绝原假设
- ❹ 得出结论

Confidence Interval —— 1-c confidence interval contains the set of null hypothesis values that are not rejected when using a test size of c.

$e.g.\ 90\%\ CI\ for\ \hat{\beta}\ is\ [\hat{\beta} - 1.645 \times s.e.(\hat{\beta}), \hat{\beta} + 1.645 \times s.e.(\hat{\beta})]$

① 建模

$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_K X_{Ki} + \varepsilon_i$

The intercept term is the expected value of Yi when X=0.

Partial effect $\beta_1 = \frac{\Delta Y}{\Delta X_1}$
- The coefficient are interpreted as holding all other values fixed. —— Stepwise regression
- e.g. β_1 is the effect of a small increase in X_1 holding all other variables constant.

② 前提假设

- All variables must have positive variances.
- $E[\varepsilon_i | X_{1i}, X_{2i}, \ldots, X_{Ki}] = 0$
- The random variables (Yi, X1i,…,Xki) are assumed to be iid.
- No outliers
- Constant variance of error conditional on the explanatory variables. —— $E[\varepsilon_i^2 | x_{1i}, x_{2i}, \ldots, x_{ki}] = \sigma^2$

Explanatory variables are not perfectly linearly dependent. (一元与多元假设的区别)

③ 基于OLS的思想，用stepwise regression 建模

④ 检验

t-test —— Same as t-test in single regression.

F-test
- Test whether at least one slope coefficient is significantly different from zero.
- $\rightarrow H_0: \beta_1 = \beta_2 = \cdots = \beta_k = 0$
  $H_1: \beta_j \neq 0$
  $F = \frac{(TSS - RSS_U)/K}{RSS_U/(n-K_U-1)} \sim F_{K,n-K-1}$
- Multivariate confidence intervals —— F-test → A region where the null hypothesis is not rejected.

思考：是否Explanatory (X) 越多越好

⑤ ANOVA

Analysis of Variance Table

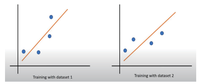
Source	Sum of square	Degrees of freedom	Mean square	F statistic
Explained sum of squares	$ESS = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2$	k	$MSE_{explained} = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}{k}$	$F = \frac{MSE_{explained}}{MSE_{residual}} = \frac{\frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}{k}}{\frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n-k-1}}$
Residual sum of squares	$RSS = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$	n-k-1	$MSE_{residual} = \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n-k-1}$	
Total sum of squares	$TSS = \sum_{i=1}^{n}(Y_i - \bar{Y})^2$	n-1		

R^2
- Formula —— $R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} \epsilon [0,1]$
- Properties —— Linear regressions with a single explanatory variable, $R^2 = \rho_{XY}^2$
- Limitations
  - Adding a new variable to the model always increases R^2.
  - R^2 can not be compared to across models.
  - There is no such thing as a "good" value for R^2.

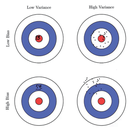
Adjusted R^2
- $\bar{R}^2 = 1 - \frac{RSS/(n-k-1)}{TSS/(n-1)} = 1 - \frac{n-1}{n-k-1}(1 - R^2)$
- Properties
  - Adjusted R^2≤R^2
  - Adjusted R^2 can be negative
  - Adding a regressor has two opposite effect, adjusted R^2 can increase or decrease.

① Bias-variance Tradeoff

Omitted Variable
- Definition —— A variable that has a non-zero coefficient but is not included in a model.
- Omitted variable bias
  - The omitted variable is correlated with the movement of the independent variable in the model
  - The omitted variable is a determinant of the dependent variable
- Effect
  - Biased and inconsistent estimates.
  - The estimated residuals are larger in magnitude than the true shocks.



Tradeoff

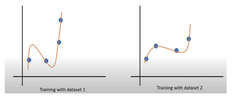


$E\left[(\hat{\theta} - \theta_0)^2\right] = Bias^2(\hat{\theta}) + Var(\hat{\theta})$ —— Methods of choosing model
- General-to-specific (GtS) model selection
- m-fold cross-validation

Smaller model → higher bias, less estimation error.

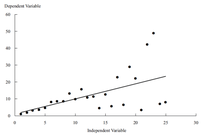
Larger model → lower bias, less precise estimated parameters;

Extraneous Included Variables
- Definition —— A variable that is included in the model but is not needed.
- Property —— True coefficient of 0 and consistent in large samples.
- Effect
  - Smaller adjusted R-square.
  - Larger standard error.



② Heteroskedasticity

Definition —— $V[\varepsilon_i | X_{1i}, X_{2i}, \ldots, X_{ki}] = \sigma^2(X)$ —— ϵ 残差项不稳定



Consequence
- 参数$\hat{b}_j$的有效性受到影响，但依旧是无偏的。
- t 检验失效，$\hat{b}_j$的标准误不准确。
- Different form of asymptotic distribution of estimated parameter.

Solution
- 1. Ignore the heteroskedasticity when estimating the parameters and then use heteroskedasticity-robust (White) covariance estimator in hypothesis tests.
- 2. Transform the Data
- 3. Use Weighted Least Square (WLS)

对残差项进行增强

④ Residual Plots and Outliers

Residual Plots
- Definition —— Compares residual against the realization of the explanatory variables x.
- Application —— It can be used to detect deficiencies in a model specification.

Outliers
- Definition —— If the values are removed, it will produce large changes in the estimated coefficients.
- Identification —— Cook's distance (Dj)
  - Measures the sensitivity of the fitted values in a regression to dropping a single observation.
  - Dj >1 → observation j has a large impact on the estimated model parameters.

③ Multicollinearity

Perfect collinearity
- Definition —— One of the variables perfectly explained by the others.
- Effect —— Produces division by zero in the OLS estimates.

Distinguish

Multicollinearity
- Definition —— One or more explanatory variables can be substantially explained by the others.
- Effect
  - The coefficients are jointly statistically significant and small individual t-statistic.
  - Effects estimator's efficiency
- Solution
  - Ignore the issue.
  - Remove multicollinear variables from the models.

*呈现内容仅为当前科目部分内容，内容解释权归高顿FRM学术部

扫描二维码，回复"全套知识图谱"领四科资料