第三章 经典单方程计量经济学模型：多元线性回归模型
Multiple Linear Regression Model

本章内容

- 多元线性回归模型概述/建模
- 多元线性回归模型的参数估计**
- 多元线性回归模型的统计检验
- 多元线性回归模型的预测
- 非线性回归模型*
- 受约束回归***

§3.1 多元线性回归模型概述
(Regression Analysis)

一、多元线性回归模型

二、多元线性回归模型的基本假设

一、多元线性回归模型

总体回归模型

- **总体回归模型：**总体回归函数的随机表达形式

$$\boxed{Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \mu_i} \qquad i=1,2\ldots,n$$

k为解释变量的数目；

习惯上，把常数项看成为虚变量的系数，该虚变量的样本观测值始终取1。于是，模型中解释变量的数目为（k+1）。

β_j称为回归参数（**regression coefficient**）。

总体回归函数

- **总体回归函数：**描述在给定解释变量X_i条件下被解释变量Y_i的条件均值。

$$E(Y_i \mid X_{1i}, X_{2i}, \cdots X_{ki}) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki}$$

β_j也被称为**偏回归系数(partial regression coefficients)**——表示在<u>其他解释变量保持不变的情况下</u>：X_j每变化1个单位时，Y的均值E(Y)的变化。

或者说β_j给出了X_j的单位变化对Y均值的“直接”或“净”（不含其他变量）影响。

总体回归模型的矩阵表示

$$\boxed{\mathbf{Y} = \mathbf{X}\beta + \mu}$$

$$\mathbf{X} = \begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{k1} \\ 1 & X_{12} & X_{22} & \cdots & X_{k2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{1n} & X_{2n} & \cdots & X_{kn} \end{bmatrix}_{n\times(k+1)}$$

$X_i = (1, X_{1i}, \cdots X_{ji} \cdots X_{ki})$ 可改写为:

$X_i = (1, X_{i1}, \cdots X_{ij} \cdots X_{ik})$

表示(k+1)个解释变量的一组样本观测值

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}_{n\times1} \qquad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}_{(k+1)\times1} \qquad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}_{n\times1}$$

样本回归函数与样本回归模型

- 从一次抽样中获得的总体回归函数的近似，称为**样本回归函数（sample regression function）**。
- 样本回归函数的随机形式，称为**样本回归模型（sample regression model）**。

$$\boxed{\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_{ki} X_{ki}}$$

$$\boxed{Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_{ki} X_{ki} + e_i}$$

样本回归函数的矩阵表示

$$\boxed{\hat{\mathbf{Y}} = \mathbf{X}\hat{\beta}}$$

$$\boxed{\mathbf{Y} = \mathbf{X}\hat{\beta} + \mathbf{e}}$$

$$\boxed{\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{pmatrix}}$$

$$\boxed{\mathbf{e} = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}}$$

二、多元线性回归模型的基本假设

说 明

- 为保证参数估计量具有良好的性质，通常对模型提出若干基本假设。
- 实际上这些假设与所采用的估计方法紧密相关。
- 以下假设，主要是针对采用**普通最小二乘法（Ordinary Least Squares, OLS）**估计而提出的；所以，有些教科书称之为**“The Assumption Underlying the Method of Least Squares”**。
- 不同教科书、关于基本假设的陈述略有不同，下面进行了重新归纳。

1、关于模型关系的1个假设

- 模型设定正确假设

  The regression model is correctly specified.

  比如，线性回归假设

  The regression model is linear in the parameters。

$$\boxed{Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \mu_i}$$

注意："**linear in the parameters**"的含义是什么？

2*、关于 解释变量*X* 的5项假设

- **确定性假设**

X values are fixed in repeated sampling. More technically, X is assumed to be non-stochastic.

注意："**in repeated sampling**"的含义是什么？

- **与随机项不相关假设**

The covariances between X_i and μ_i are zero.

$$\boxed{\begin{aligned}&\operatorname{cov}(X_i, \mu_i) = 0, i = 1, 2, \cdots, n \\ &E(X_i \mu_i) = 0, i = 1, 2, \cdots, n\end{aligned}}$$

由确定性假设可以推断。

- **观测值变化假设。** X values in a given sample must not all be the same.
- **无完全共线性假设。** There is no perfect multicollinearity among the explanatory variables.

  适用于多元线性回归模型。

时间序列数据作样本时，适用

- **样本方差有界假设。** 随着样本容量的无限增加，解释变量X的样本方差、趋于一有限常数

$$\boxed{\sum (X_i - \bar{X})^2 / n \to Q, \qquad n \to \infty}$$

3***、关于 随机项u 的3项假设

- **0均值假设**

The conditional mean value of u_i is zero.

$$\boxed{E(u_i \mid X_i) = 0, i = 1, 2, \cdots, n}$$

由模型设定正确假设推断。

- **同方差假设**

The conditional variances of u_i are identical. (Homoscedasticity)

$$\boxed{Var(u_i \mid X_i) = \sigma^2, i = 1, 2, \cdots, n}$$

是否满足需要检验。

- 序列不相关假设

The correlation between any two u_{i} and u_{j} is zero.

$$\boxed{Cov(u_i, u_j \mid X_i, X_j) = 0,\ \ i, j = 1, 2, \cdots, n, i \neq j}$$

是否满足需要检验。

4、随机项的正态性假设

- 采用**OLS**进行参数估计时，不需要正态性假设；
  但在利用参数估计量进行**统计推断**时，需要假设随机项的概率分布。

- 一般假设随机项***u***服从正态分布，可利用中心极限定理（central limit theorem, CLT）进行证明。
- 正态性假设

The *u*'s follow the normal distribution.

$$\boxed{\mu_i \sim N(0,\sigma^2) \rightarrow \mu_i \sim \mathbf{NID}(0,\sigma^2)}$$

5、CLRM 和 CNLRM

- 以上假设（正态性假设除外）也称为线性回归模型的“经典假设”或“高斯（Gauss）假设”，满足该假设的线性回归模型，也称为经典线性回归模型（Classical Linear Regression Model, CLRM）。

- 同时满足正态性假设的线性回归模型，称为经典正态线性回归模型（Classical Normal Linear Regression Model, CNLRM）。

§3.2 多元线性回归模型的估计**

一、普通最小二乘估计

二、最大或然估计**

三、矩估计*

四、参数估计量的性质**

五、样本容量问题

六、估计实例

说 明

估计目标：结构参数$\hat{\beta}_j$及随机误差项的方差$\hat{\sigma}^2$

估计方法：

- 3大类方法：**OLS、ML或者MM**
  - ➢ 经典模型中：多应用 **OLS**
  - ➢ 非经典模型中：多应用 **ML 或者 MM**

一、普通最小二乘估计(OLS)

1、普通最小二乘估计

- **最小二乘原理：**根据被解释变量的“所有观测值与估计值之差的平方和”最小的原则，求得参数估计量。

- **步骤：**

$$(Y_i, X_{ji}), i=1,2,\cdots,n, j=0,1,2,\cdots k$$

已知

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_{ki} X_{Ki}$$

假定

$$Q = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

$$= \sum_{i=1}^{n} \left(Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_k X_{ki})\right)^2$$

$$\mathrm{Min}\, Q \longrightarrow \begin{cases} \dfrac{\partial}{\partial \hat{\beta}_0} Q = 0 \\ \dfrac{\partial}{\partial \hat{\beta}_1} Q = 0 \\ \dfrac{\partial}{\partial \hat{\beta}_2} Q = 0 \\ \vdots \\ \dfrac{\partial}{\partial \hat{\beta}_k} Q = 0 \end{cases}$$

$$
\begin{cases}
\Sigma(\hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_k X_{ki}) = \Sigma Y_i \\
\Sigma(\hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_k X_{ki}) X_{1i} = \Sigma Y_i X_{1i} \\
\Sigma(\hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_{2i} X_{2i} + \cdots + \hat{\beta}_k X_{ki}) X_{2i} = \Sigma Y_i X_{2i} \\
\vdots \\
\Sigma(\hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_k X_{ki}) X_{ki} = \Sigma Y_i X_{ki}
\end{cases}
$$

$$
\boxed{\hat{\beta}_j ,\ j = 0,1,2,\cdots,k}
$$

- 正规方程组的矩阵形式

$$\begin{pmatrix} n & \sum X_{1i} & \cdots & \sum X_{ki} \\ \sum X_{1i} & \sum X_{1i}^2 & \cdots & \sum X_{1i}X_{ki} \\ \cdots & \cdots & \cdots & \cdots \\ \sum X_{ki} & \sum X_{ki}X_{1i} & \cdots & \sum X_{ki}^2 \end{pmatrix} \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \cdots \\ \hat{\beta}_k \end{pmatrix} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ X_{11} & X_{12} & \cdots & X_{1n} \\ \cdots & \cdots & \cdots & \cdots \\ X_{k1} & X_{k2} & \cdots & X_{kn} \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ \cdots \\ Y_n \end{pmatrix}$$

$$(\mathbf{X}'\mathbf{X})\hat{\beta} = \mathbf{X}'\mathbf{Y}$$

条件?

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

- **OLS估计的矩阵表示**

$$Q=\sum_{i=1}^{n}e_i^2=\mathbf{e}'\mathbf{e}=(\mathbf{Y}-\mathbf{X}\hat{\beta})'(\mathbf{Y}-\mathbf{X}\hat{\beta})$$

$$\boxed{\frac{\partial}{\partial\hat{\beta}}(\mathbf{Y}-\mathbf{X}\hat{\beta})'(\mathbf{Y}-\mathbf{X}\hat{\beta})=0}$$

$$\boxed{\frac{\partial}{\partial\hat{\beta}}(\mathbf{Y}'\mathbf{Y}-\hat{\beta}'\mathbf{X}'\mathbf{Y}-\mathbf{Y}'\mathbf{X}\hat{\beta}+\hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta})=0}$$

$$\boxed{-\mathbf{X}'\mathbf{Y}+\mathbf{X}'\mathbf{X}\hat{\beta}=0}$$

$$\boxed{\mathbf{X}'\mathbf{Y}=\mathbf{X}'\mathbf{X}\hat{\beta}}$$

$$\boxed{\hat{\beta}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}}$$

2、正规方程组的另一种表达

$$\boxed{\mathbf{X}'\mathbf{Y} = \mathbf{X}'\mathbf{X}\hat{\beta}}$$

将 $\mathbf{Y} = \mathbf{X}\hat{\beta} + \mathbf{e}$ 代入得

$$\boxed{\mathbf{X}'\mathbf{X}\hat{\beta} + \mathbf{X}'\mathbf{e} = \mathbf{X}'\mathbf{X}\hat{\beta}}$$

$$\boxed{\mathbf{X}'\mathbf{e} = \mathbf{0}}$$

$$\begin{cases} \sum\limits_i e_i = 0 \\ \sum\limits_i X_{ji} e_i = 0 \quad j = 1, 2, \cdots, k \end{cases}$$

该正规方程组成立的条件是什么？

3、随机误差项μ的方差 σ^2的无偏估计

$$\boxed{\mathbf{e} = \mathbf{Y} - \mathbf{X}\hat{\beta}}$$

$$\boxed{\begin{aligned} &= \mathbf{X}\beta + \mu - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \mu) \\ &= \mu - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mu \\ &= (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mu \\ &= \mathbf{M}\mu \end{aligned}}$$

M为等幂矩阵

$$\boxed{\mathbf{e}'\mathbf{e} = \boldsymbol{\mu}'\mathbf{M}'\mathbf{M}\boldsymbol{\mu} = \boldsymbol{\mu}'\mathbf{M}\boldsymbol{\mu}}$$

$$\begin{aligned} \mathbf{M}^2 &= [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'][\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] \\ &= \mathbf{I} - 2\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \\ &= \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \end{aligned}$$

$$
\boxed{\begin{aligned}
E(\mathbf{e}'\mathbf{e}) &= E[tr\ \boldsymbol{\mu}'(\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\ \boldsymbol{\mu})] \\
&= E[tr\ \boldsymbol{\mu}\ \boldsymbol{\mu}'(\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')] \\
&= \sigma^2 tr(\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') \\
&= \sigma^2 (tr\mathbf{I}-tr(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')) \\
&= \sigma^2 (n-(k+1))
\end{aligned}}
$$

$$
\boxed{\sigma^2 = \frac{E(\mathbf{e}'\mathbf{e})}{n-k-1}} \longrightarrow \boxed{\hat{\sigma}^2 = \frac{\mathbf{e}'\mathbf{e}}{n-k-1}}
$$

二、最大似然估计

1、最大似然法

- 最大似然法(**Maximum Likelihood,ML**)，也称最大或然法，是不同于最小二乘法的另一种参数估计方法，是从最大或然原理出发发展起来的其它估计方法的基础。

- 基本原理：当从模型总体随机抽取**n**组样本观测值后，最合理的参数估计量应该使得从模型中抽取该n组样本观测值的联合概率最大。

- **ML**必须已知随机项的分布。

2、估计步骤：以一元模型为例

$$Y_i \sim N(\hat{\beta}_0 + \hat{\beta}_1 X_i, \sigma^2)$$

Yi的分布

$$P(Y_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2}$$

Yi的概率函数

$$L(\hat{\beta}_0, \hat{\beta}_1, \sigma^2) = P(Y_1, Y_2, \cdots, Y_n)$$

$$= \frac{1}{(2\pi)^{\frac{n}{2}} \sigma^n} e^{-\frac{1}{2\sigma^2}\Sigma(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2}$$

Y的所有样本观测值的联合概率——似然函数

$$L^* = \ln(L)$$

$$= -n\ln(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2}\Sigma(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2$$

对数似然函数

$$\begin{cases} \dfrac{\partial}{\partial \hat{\beta}_0}\Sigma(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2 = 0 \\ \dfrac{\partial}{\partial \hat{\beta}_1}\Sigma(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2 = 0 \end{cases}$$

对数似然函数极大化的一阶条件

$$\begin{cases} \hat{\beta}_0 = \dfrac{\Sigma X_i^2 \Sigma Y_i - \Sigma X_i \Sigma Y_i X_i}{n\Sigma X_i^2 - (\Sigma X_i)^2} \\ \hat{\beta}_1 = \dfrac{n\Sigma Y_i X_i - \Sigma Y_i \Sigma X_i}{n\Sigma X_i^2 - (\Sigma X_i)^2} \end{cases}$$

结构参数的ML估计量

$$\frac{\partial}{\partial \sigma^2} L^* = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^2} \Sigma (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2 = 0$$

$$\hat{\sigma}^2 = \frac{1}{n} \Sigma (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2 = \frac{\sum e_i^2}{n}$$

分布参数的
ML估计量

3*、MLR的似然函数

$$\boxed{Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \mu_i}$$

$$\boxed{\boldsymbol{Y_i \sim N(\mathbf{X}_i\hat{\boldsymbol{\beta}}, \sigma^2)}} \longleftarrow \boxed{\mu_i \sim N(0, \sigma^2)}$$

$$\begin{aligned} L(\hat{\beta}, \sigma^2) &= P(Y_1, Y_2, \cdots, Y_n) \\ &= \frac{1}{(2\pi)^{\frac{n}{2}} \sigma^n} e^{-\frac{1}{2\sigma^2}\Sigma(Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_k X_{ki}))^2} \\ &= \frac{1}{(2\pi)^{\frac{n}{2}} \sigma^n} e^{-\frac{1}{2\sigma^2}(\mathbf{Y} - \mathbf{X}\hat{\beta})'(\mathbf{Y} - \mathbf{X}\hat{\beta})} \end{aligned}$$

4、ML估计量

- 由对数似然函数求极大，得到参数估计量

$$
\begin{aligned}
Max\, L^* &= Ln(L) \\
&= -nLn(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2}(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}})
\end{aligned}
$$

$$
\downarrow
$$

$$
Min\,(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}})
$$

$$
\downarrow
$$

$$
\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}
$$

结果与参数的OLS估计相同

- **分布参数σ²估计结果**与OLS不同

$$\boxed{\hat{\sigma}_{ML}^{2}=\frac{(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}})}{n}=\frac{\sum e_i^2}{n}}$$

$$\boxed{\hat{\sigma}_{OLS}^{2}=\frac{\sum e_i^2}{n-k-1}=\frac{\mathbf{e}'\mathbf{e}}{n-k-1}}$$

- 注意：
  - **ML估计，必须已知Y的分布。**
  - **只有在正态分布时，ML和OLS的结构参数估计结果相同。**
  - **如果Y不服从正态分布，不能采用OLS。例如：选择性样本模型、计数数据模型等。**

三、矩估计
Moment Method, MM

1、参数的矩估计

- 参数的矩估计，就是用<u>**样本矩**</u>去估计<u>**总体矩**</u>，如：

  用样本的<u>一阶原点矩</u>，作为期望的估计量；

  用样本的<u>二阶中心矩</u>，作为方差的估计量。

- 由样本观测值，计算样本一阶（原点）矩和二阶（原点）矩，以估计总体一阶矩和总体二阶矩；再进一步计算<u>总体参数</u>（期望和方差）的估计量。

$$X^{(1)}=\frac{1}{n}\sum_{i=1}^{n}y_i \qquad X^{(2)}=\frac{1}{n}\sum_{i=1}^{n}y_i^2$$

样本的一阶矩和二阶矩

$$\hat{M}^{(1)}=E(Y)=X^{(1)}=\frac{1}{n}\sum_{i=1}^{n}y_i$$

$$\hat{M}^{(2)}=E(Y^2)=X^{(2)}=\frac{1}{n}\sum_{i=1}^{n}y_i^2$$

总体一阶矩和总体二阶矩的估计量

$$\hat{\mu}=\hat{M}^{(1)}=E(Y)=X^{(1)}$$

$$\hat{\sigma}^2=\hat{M}^{(2)}-(\hat{M}^{(1)})^2=X^{(2)}-(X^{(1)})^2$$

总体参数（期望和方差）的估计量

2、多元线性计量经济学模型的矩估计

- 如果模型的设定是正确，则存在一些**为0的条件矩**；矩估计的基本思想，是利用矩条件估计模型参数。

$$\boxed{Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \mu_i} \quad \boxed{i = 1, \cdots n}$$

$$\boxed{\sum_{i=1}^{n} X_{ji} e_i = 0 \qquad j = 0, 1, 2, \cdots, k}$$

$$\sum_{i=1}^{n} X_{ji}(Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \cdots + \hat{\beta}_k X_{ki})) = 0, \; j = 0, 1, 2, \cdots, k$$

一组矩条件，等同于OLS估计的正规方程组。

四、参数估计量的性质

说 明

- 在满足基本假设的情况下，多元线性模型 “结构参数β的普通最小二乘估计、最大或然估计及矩估计” 具有线性性、无偏性、有效性。

- 同时，随着样本容量增加，参数估计量具有渐近无偏性、一致性、渐近有效性。

- 利用矩阵表达，可以很方便地证明以上性质；注意证明过程中，须利用基本假设。

1、线性性

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \mathbf{C}\mathbf{Y}$$

其中, **C=(X'X)**$^{-1}$ **X'** 为一仅与固定的**X**有关的(k+1)行的行向量

2、无偏性

$$\begin{aligned} E(\hat{\boldsymbol{\beta}}) &= E((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}) \\ &= E((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu})) \\ &= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}E(\mathbf{X}'\boldsymbol{\mu}) \\ &= \boldsymbol{\beta} \end{aligned}$$

3、有效性（最小方差性）

参数估计量 $\hat{\beta}$ 的方差-协方差矩阵

$$Cov(\hat{\beta}) = E(\hat{\beta} - E(\hat{\beta}))(\hat{\beta} - E(\hat{\beta}))'$$

$$= E(\hat{\beta} - \beta)(\hat{\beta} - \beta)'$$

$$= E((\mathbf{X}\mathbf{X}')^{-1}\mathbf{X}'\boldsymbol{\mu}\boldsymbol{\mu}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1})$$

$$= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\boldsymbol{\mu}\boldsymbol{\mu}')\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

$$= E(\boldsymbol{\mu}\boldsymbol{\mu}')(\mathbf{X}'\mathbf{X})^{-1}$$

$$\boxed{E(\mu\mu') = \sigma^2\mathbf{I}}$$

$$= \sigma^2\mathbf{I}(\mathbf{X}'\mathbf{X})^{-1}$$

$$= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

根据高斯—马尔可夫定理，$Cov(\hat{\beta}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ 在所有无偏估计量的方差中是最小的。

设另有一线性无偏估计量：

$$\hat{\beta}^* = C^* Y = (C + D)Y, \quad \text{其中 } D \neq 0$$

<u>再由无偏性</u>可得：

$$E(\hat{\beta}^*) = E(C^* X\beta) + E(C^* u) = \beta$$

$$\Rightarrow \quad C^* X = I \quad \Rightarrow \quad DX = 0$$

$$Cov(\hat{\beta}^*) = E(\hat{\beta}^* - \beta)(\hat{\beta}^* - \beta)'$$

$$= E\{[(X'X)^{-1} X' + D]\, u \cdot u'[X(X'X)^{-1} + D']\}$$

$$= Cov(\hat{\beta}) + \sigma^2 DD'$$

其中： $\hat{\beta}^* - \beta = C^* u = [C + D]\, u = [(X'X)^{-1} X' + D]\, u$

五、样本容量问题

1、最小样本容量

- 所谓"最小样本容量"，即从最小二乘原理和最大或然原理出发，欲得到参数估计量，不管其质量如何，所要求的样本容量的下限。

  - 样本最小容量必须不少于模型中解释变量的数目（包括常数项），即

$$n \geq k+1$$

为什么？

2、满足基本要求的样本容量

- 从统计检验的角度：

  $n>30$ 时，Z检验才能应用；

  $n-k\geq 8$时， t分布较为稳定。

- 一般经验认为：

  当$n\geq 30$或者至少$n\geq 3(k+1)$时，才能说满足模型估计的基本要求。

- **模型的良好性质，只有在大样本下才能得到理论上的证明。**

六、例题

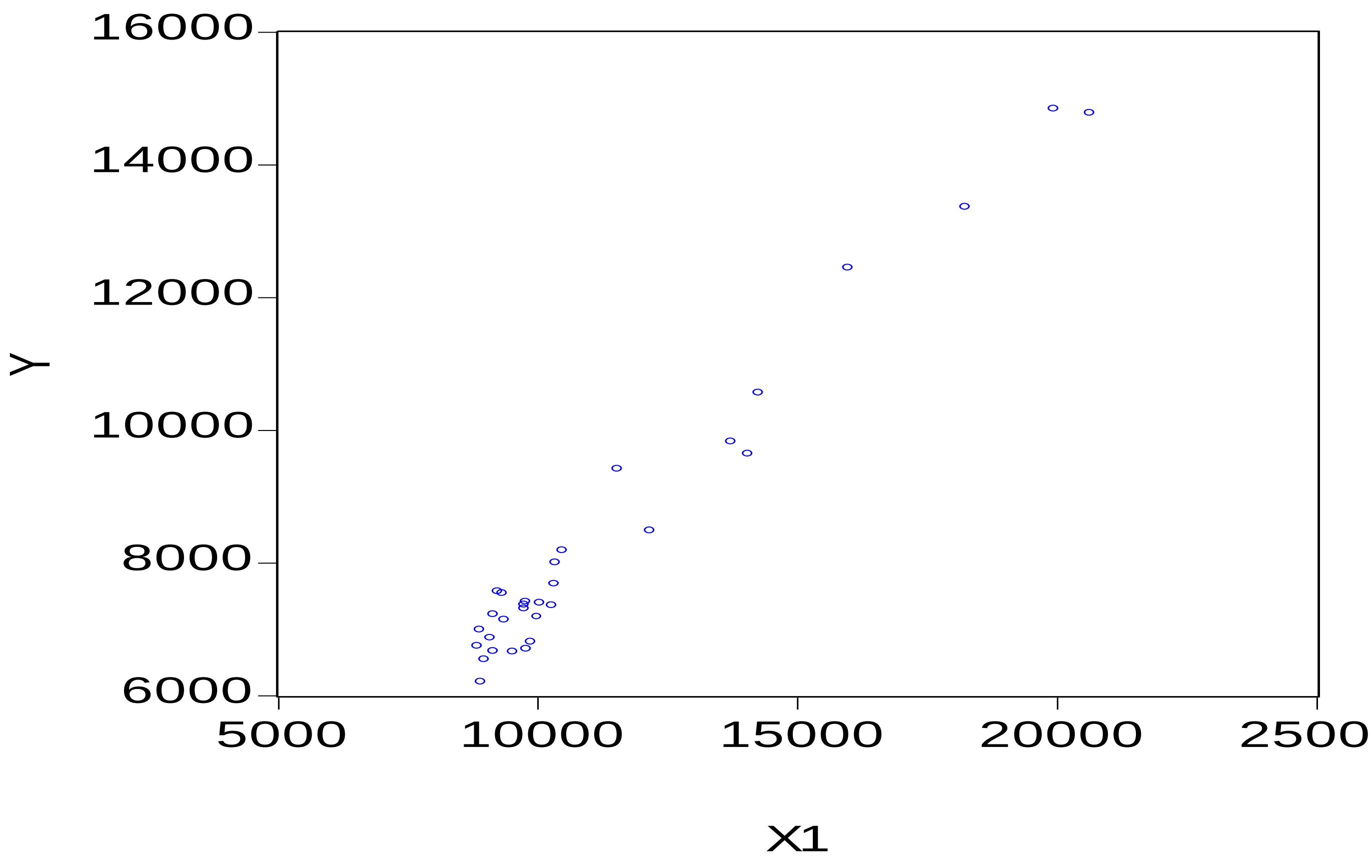
地区城镇居民消费模型

- 被解释变量：地区城镇居民人均消费Y
- 解释变量：
  - 地区城镇居民人均可支配收入X1
  - 前一年地区城镇居民人均消费X2
- 样本：2006年，31个地区

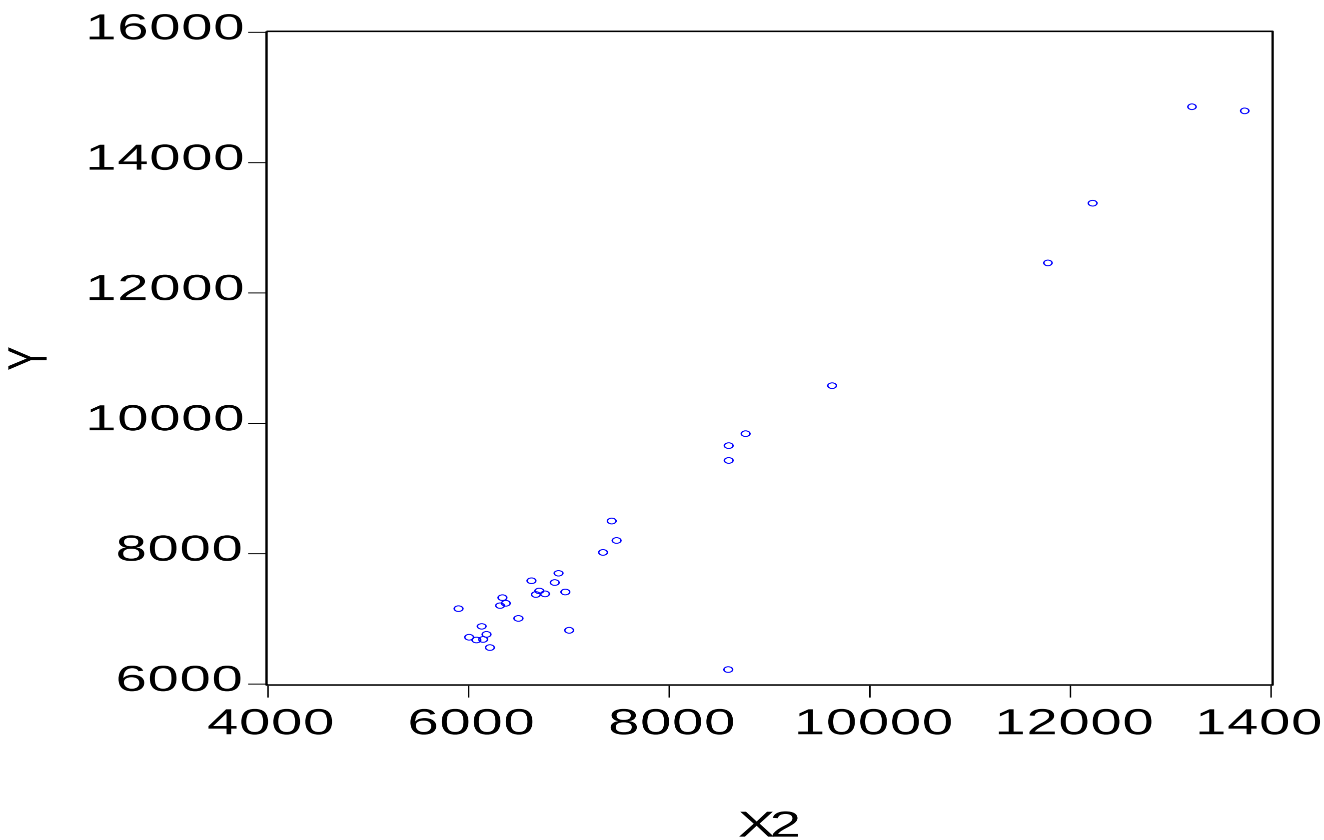
数据

地区	2006年消费支出 Y	2006年可支配收入 X_1	2005年消费支出 X_2	地区	2006年消费支出 Y	2006年可支配收入 X_1	2005年消费支出 X_2
北　京	14825.4	19977.5	13244.2	湖　北	7397.3	9802.7	6736.6
天　津	10548.1	14283.1	9653.3	湖　南	8169.3	10504.7	7505.0
河　北	7343.5	10304.6	6699.7	广　东	12432.2	16015.6	11809.9
山　西	7170.9	10027.7	6342.6	广　西	6792.0	9898.8	7032.8
内蒙古	7666.6	10358.0	6928.6	海　南	7126.8	9395.1	5928.8
辽　宁	7987.5	10369.6	7369.3	重　庆	9398.7	11569.7	8623.3
吉　林	7352.6	9775.1	6794.7	四　川	7524.8	9350.1	6891.3
黑龙江	6655.4	9182.3	6178.0	贵　州	6848.4	9116.6	6159.3
上　海	14761.8	20667.9	13773.4	云　南	7379.8	10069.9	6996.9
江　苏	9628.6	14084.3	8621.8	西　藏	6192.6	8941.1	8617.1
浙　江	13348.5	18265.1	12253.7	陕　西	7553.3	9267.7	6656.5
安　徽	7294.7	9771.1	6367.7	甘　肃	6974.2	8920.6	6529.2
福　建	9807.7	13753.3	8794.4	青　海	6530.1	9000.4	6245.3
江　西	6645.5	9551.1	6109.4	宁　夏	7205.6	9177.3	6404.3
山　东	8468.4	12192.2	7457.3	新　疆	6730.0	8871.3	6207.5
河　南	6685.2	9810.3	6038.0				

变量间关系（Scatter）

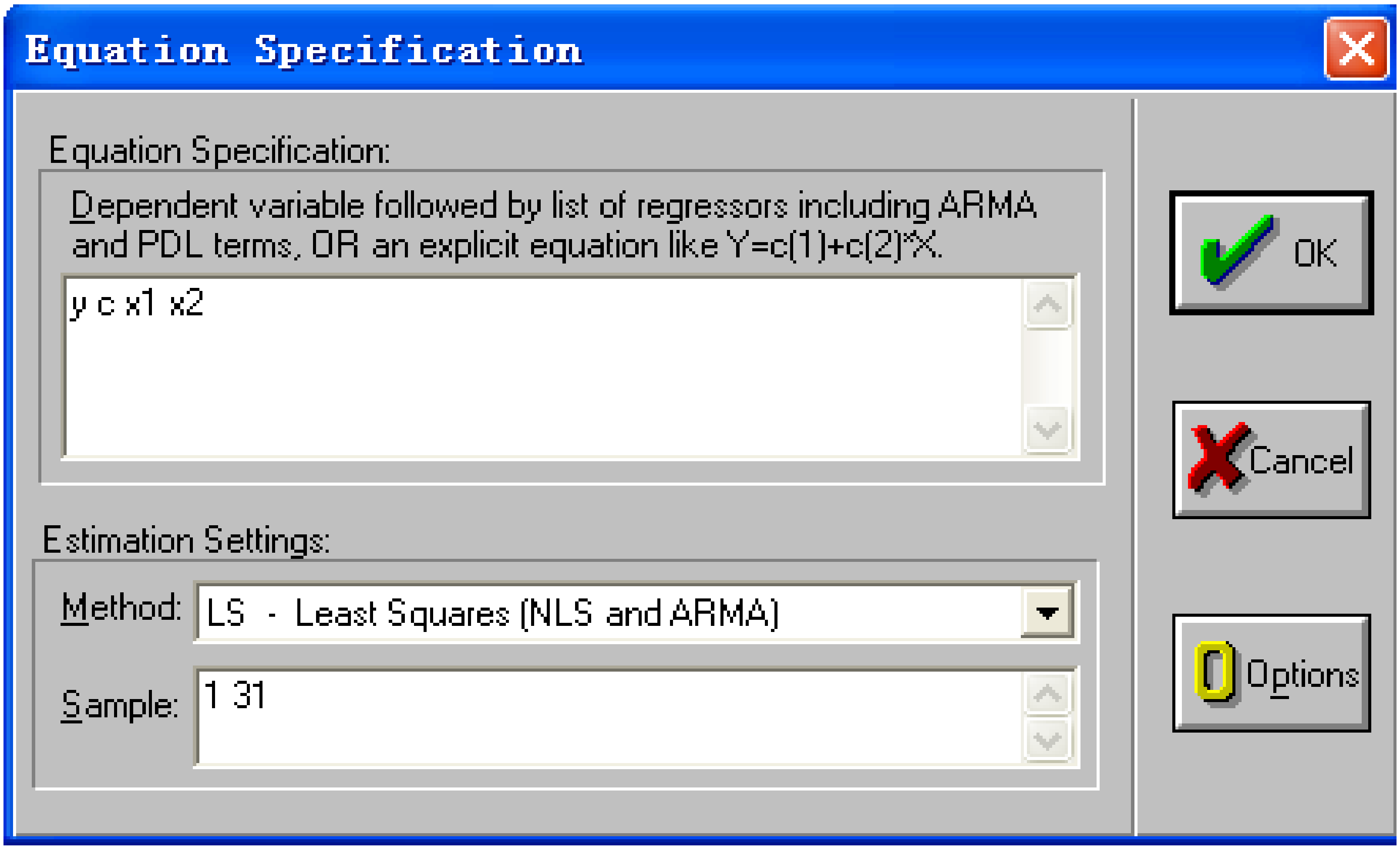


变量间关系



16000
14000
12000
10000
8000
6000
Y
4000
6000
8000
10000
12000
1400
X2

OLS估计


Equation Specification
Equation Specification:
Dependent variable followed by list of regressors including ARMA and PDL terms, OR an explicit equation like Y=c(1)+c(2)*X.
y c x1 x2
Estimation Settings:
Method:
LS - Least Squares (NLS and ARMA)
Sample:
1 31
OK
Cancel
Options

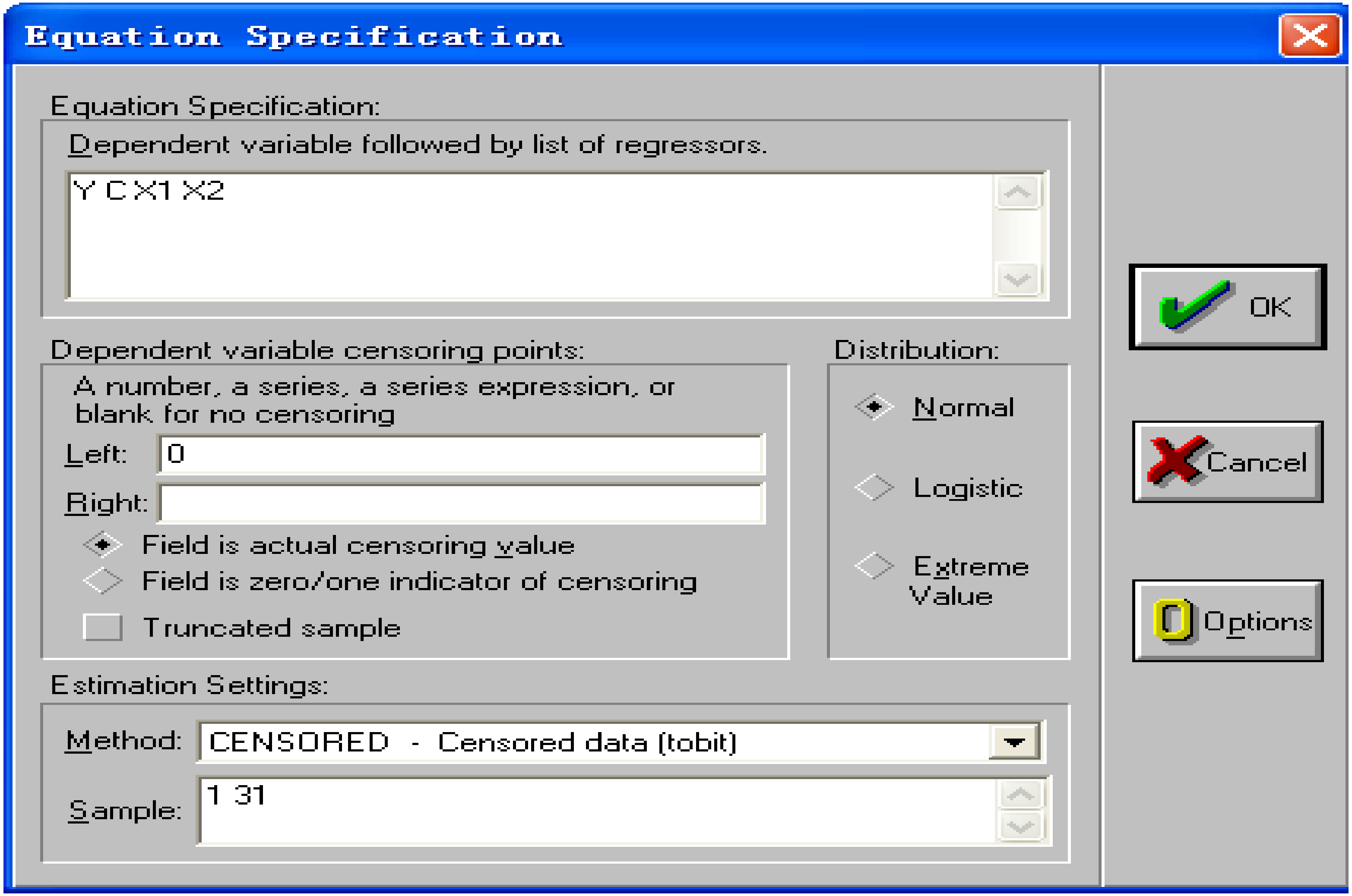
OLS估计结果

Dependent Variable: Y
Method: Least Squares
Date: 09/06/10 Time: 23:06
Sample: 1 31
Included observations: 31

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	143.3265	260.4032	0.550402	0.5864
X1	0.555644	0.075308	7.378320	0.0000
X2	0.250085	0.113634	2.200791	0.0362

R-squared	0.975634	Mean dependent var	8401.468
Adjusted R-squared	0.973893	S.D. dependent var	2388.459
S.E. of regression	385.9169	Akaike info criterion	14.84089
Sum squared resid	4170093.	Schwarz criterion	14.97966
Log likelihood	-227.0337	F-statistic	560.5650
Durbin-Watson stat	1.843488	Prob(F-statistic)	0.000000

ML估计


Equation Specification
Equation Specification:
Dependent variable followed by list of regressors.
Y C X1 X2
Dependent variable censoring points:
A number, a series, a series expression, or blank for no censoring
Left: 0
Right:
Field is actual censoring value
Field is zero/one indicator of censoring
Truncated sample
Distribution:
Normal
Logistic
Extreme Value
OK
Cancel
Options
Estimation Settings:
Method: CENSORED - Censored data (tobit)
Sample: 1 31

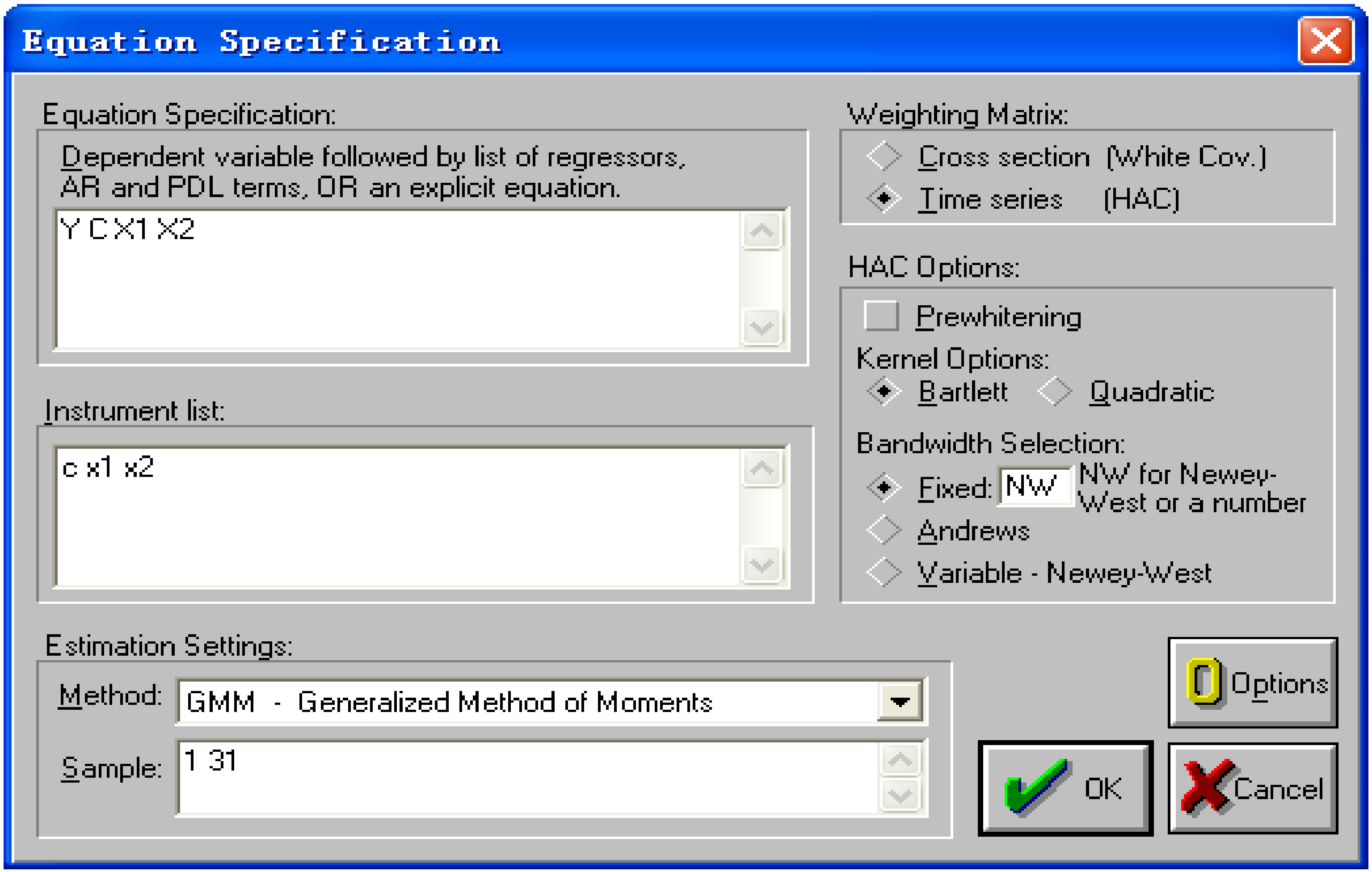
ML估计结果

Dependent Variable: Y
Method: ML - Censored Normal (TOBIT)
Date: 09/06/10 Time: 23:11
Sample: 1 31
Included observations: 31
Left censoring (value) at zero
Convergence achieved after 5 iterations
Covariance matrix computed using second derivatives

	Coefficient	Std. Error	z-Statistic	Prob.
C	143.3265	247.4832	0.579136	0.5625
X1	0.555644	0.071571	7.763508	0.0000
X2	0.250085	0.107996	2.315684	0.0206
		Error Distribution		
SCALE:C(4)	366.7685	46.57999	7.873949	0.0000

R-squared	0.975634	Mean dependent var	8401.468
Adjusted R-squared	0.972926	S.D. dependent var	2388.459
S.E. of regression	392.9986	Akaike info criterion	14.90540
Sum squared resid	4170093.	Schwarz criterion	15.09043
Log likelihood	-227.0337	Hannan-Quinn criter.	14.96572
Avg. log likelihood	-7.323669		

MM估计


Equation Specification
Equation Specification:
Dependent variable followed by list of regressors,
AR and PDL terms, OR an explicit equation.
Y C X1 X2
Instrument list:
c x1 x2
Weighting Matrix:
Cross section (White Cov.)
Time series (HAC)
HAC Options:
Prewhitening
Kernel Options:
Bartlett
Quadratic
Bandwidth Selection:
Fixed:
NW
NW for Newey-West or a number
Andrews
Variable - Newey-West
Estimation Settings:
Method:
GMM - Generalized Method of Moments
Sample:
1 31
Options
OK
Cancel

MM估计结果

Dependent Variable: Y
Method: Generalized Method of Moments
Date: 09/06/10 Time: 23:14
Sample: 1 31
Included observations: 31
No prewhitening
Bandwidth: Fixed (3)
Kernel: Bartlett
Convergence achieved after: 1 weight matrix, 2 total coef iterations
Instrument list: C X1 X2

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	143.3265	199.8715	0.717093	0.4793
X1	0.555644	0.150780	3.685137	0.0010
X2	0.250085	0.229440	1.089983	0.2850

R-squared	0.975634	Mean dependent var	8401.468
Adjusted R-squared	0.973893	S.D. dependent var	2388.459
S.E. of regression	385.9169	Sum squared resid	4170093.
Durbin-Watson stat	1.843488	J-statistic	3.53E-27

§3.3 多元线性回归模型的统计检验
Statistical Test of Multiple Linear Regression Model

一、拟合优度检验*

二、方程的显著性检验(F检验)

三、变量的显著性检验（t检验）

四、参数的置信区间

一、拟合优度检验
Goodness of Fit

1、概念

- **拟合优度检验**：对“样本回归线与样本观测值之间拟合程度”的检验。
- **问题**：采用普通最小二乘估计方法，已经保证了模型最好地拟合了样本观测值，为什么还要检验拟合程度？
- **如何检验**：构造统计量R^2
  - 统计量，亦是随机变量，只能是相对量

2、可决系数与调整的可决系数

- 总离差平方和的分解

记 $TSS=\sum(Y_i-\bar{Y})^2$ 总离差平方和

$ESS=\sum(\hat{Y}_i-\bar{Y})^2$ 回归平方和

$RSS=\sum(Y_i-\hat{Y}_i)^2$ 剩余平方和

$$
\begin{aligned}
TSS &= \Sigma(Y_i-\bar{Y})^2 \\
&= \Sigma((Y_i-\hat{Y}_i)+(\hat{Y}_i-\bar{Y}))^2 \\
&= \Sigma(Y_i-\hat{Y}_i)^2+2\Sigma(Y_i-\hat{Y}_i)(\hat{Y}_i-\bar{Y})+\Sigma(\hat{Y}_i-\bar{Y})^2
\end{aligned}
$$

证明：
该项等于0

$$TSS=\sum(Y_i-\hat{Y}_i)^2+\sum(\hat{Y}_i-\bar{Y})^2=RSS+ESS$$

- **可决系数（ Coefficient of Determination ）**

$$\boxed{R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}}$$

该统计量越接近于1，模型的拟合优度越高。

- 从R^2的表达式中发现，如果在模型中增加解释变量，R^2往往增大。

这就给人一个**错觉**：要使得模型拟合得好，**只要增加解释变量**即可。

但是，由增加解释变量引起的R^2的增大、与拟合好坏无关，所以R^2需调整。

- 调整的可决系数（**adjusted coefficient of determination**）

$$\boxed{\bar{R}^2 = 1 - \frac{RSS/(n-k-1)}{TSS/(n-1)}}$$

其中：n-k-1为残差平方和的自由度，n-1为总体平方和的自由度。

调整的可决系数，多大才是合适的（结合F检验）？

We can also think of R^2 as being equal to the squared correlation coefficient between *the actual y_i and the fitted values $\hat{y}_i$*

$$R^2 = \frac{\left(\sum (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})\right)^2}{\left(\sum (y_i - \bar{y})^2\right)\left(\sum (\hat{y}_i - \bar{\hat{y}})^2\right)} = r^2_{y_i, \hat{y}_i}$$

$$R^2 = \frac{SSE}{SST} = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2} = \frac{\sum(y_i - \hat{u}_i - \bar{y})(\hat{y}_i - \bar{y})}{\sum(y_i - \bar{y})^2}$$

$$= \frac{\sum(y_i - \bar{y})(\hat{y}_i - \bar{y})}{\sum(y_i - \bar{y})^2} = \frac{\sum(y_i - \bar{y})(\hat{y}_i - \bar{y})}{\sum(y_i - \bar{y})^2} \cdot \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(\hat{y}_i - \bar{y})^2}$$

$$= \frac{\sum(y_i - \bar{y})(\hat{y}_i - \bar{y})}{\sum(y_i - \bar{y})^2} \cdot \frac{\sum(y_i - \hat{u}_i - \bar{y})(\hat{y}_i - \bar{y})}{\sum(\hat{y}_i - \bar{y})^2}$$

$$= \frac{\sum(y_i - \bar{y})(\hat{y}_i - \bar{y})}{\sum(y_i - \bar{y})^2} \cdot \frac{\sum(y_i - \bar{y})(\hat{y}_i - \bar{y})}{\sum(\hat{y}_i - \bar{y})^2}$$

$$= \frac{[\sum(y_i - \bar{y})(\hat{y}_i - \bar{y})]^2}{\sum(y_i - \bar{y})^2 \cdot \sum(\hat{y}_i - \bar{y})^2} = \frac{[\sum(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})]^2}{\sum(y_i - \bar{y})^2 \cdot \sum(\hat{y}_i - \bar{\hat{y}})^2}$$

地区城镇居民消费模型（k=2）

Dependent Variable: Y
Method: Least Squares
Date: 09/06/10 Time: 23:06
Sample: 1 31
Included observations: 31

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	143.3265	260.4032	0.550402	0.5864
X1	0.555644	0.075308	7.378320	0.0000
X2	0.250085	0.113634	2.200791	0.0362

R-squared	0.975634	Mean dependent var	8401.468
Adjusted R-squared	0.973893	S.D. dependent var	2388.459
S.E. of regression	385.9169	Akaike info criterion	14.84089
Sum squared resid	4170093.	Schwarz criterion	14.97966
Log likelihood	-227.0337	F-statistic	560.5650
Durbin-Watson stat	1.843488	Prob(F-statistic)	0.000000

地区城镇居民消费模型（k=1）

Dependent Variable: Y
Method: Least Squares
Date: 09/06/10 Time: 23:33
Sample: 1 31
Included observations: 31

与k=2比较，变化不大

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	281.4625	268.9519	1.046516	0.3040
X1	0.714556	0.022760	31.39513	0.0000

R-squared	0.971419	Mean dependent var	8401.468
Adjusted R-squared	0.970433	S.D. dependent var	2388.459
S.E. of regression	410.6950	Akaike info criterion	14.93592
Sum squared resid	4891441.	Schwarz criterion	15.02843
Log likelihood	-229.5068	F-statistic	985.6540
Durbin-Watson stat	1.461481	Prob(F-statistic)	0.000000

二、方程的显著性检验(F检验)

Testing the Overall Significance of a Multiple Regression (the F test)

1、假设检验（Hypothesis Testing）

- 所谓假设检验，就是事先对总体参数或总体分布形式作出一个假设，然后利用样本信息来判断原假设是否合理，即判断样本信息与原假设是否有显著差异，从而决定是否接受或否定原假设。
- 假设检验采用的逻辑推理方法是反证法。先假定原假设正确，然后根据样本信息，观察由此假设而导致的结果是否合理，从而判断是否接受原假设。
- 判断结果合理与否，是基于“小概率事件，在一次抽样中是不会发生的”这一原理的。

2、方程显著性的F检验

- 方程的显著性检验，旨在对模型中被解释变量与解释变量之间的“线性关系在总体上是否显著成立”作出推断。

- 在多元模型中，即检验模型中的参数β_j是否显著不为0。

$$\boxed{Y_i=\beta_0+\beta_1 X_{1i}+\beta_2 X_{2i}+\cdots+\beta_k X_{ki}+\mu_i}$$

$$\boxed{H_0:\beta_1=0,\beta_2=0,\cdots,\beta_k=0}$$

$$\boxed{H_1:\beta_j(j=1,2,\cdots,k)\text{不全为}0}$$

- F检验的思想，仍然来自于总离差平方和的分解式

TSS=ESS+RSS

由于回归平方和 $ESS=\sum \hat{y}_i^2$ 是解释变量X的联合体对被解释变量Y的线性作用的结果，考虑比值

$$ESS/RSS=\sum \hat{y}_i^2 / \sum e_i^2$$

如果这个比值较大，则X的联合体对Y的解释程度高，可认为总体存在线性关系，反之总体上可能不存在线性关系。

因此,可通过该比值的大小对总体线性关系进行推断。

- 在原假设H_0成立的条件下，统计量

$$\boxed{F=\frac{ESS/k}{RSS/(n-k-1)} \sim F(k,n-k-1)}$$

给定显著性水平α，可得到临界值$F_{\alpha}(k,n-k-1)$，由样本求出统计量F的数值，通过：

$$F > F_{\alpha}(k,n-k-1) \quad 或 \quad F \leq F_{\alpha}(k,n-k-1)$$

来拒绝或接受原假设H_0，以判定原方程总体上的线性关系是否显著成立。

地区城镇居民消费模型

Dependent Variable: Y
Method: Least Squares
Date: 09/06/10 Time: 23:06
Sample: 1 31
Included observations: 31

拒绝0假设，犯错误的概率为0

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	143.3265	260.4032	0.550402	0.5864
X1	0.555644	0.075308	7.378320	0.0000
X2	0.250085	0.113634	2.200791	0.0362

R-squared	0.975634	Mean dependent var	8401.468
Adjusted R-squared	0.973893	S.D. dependent var	2388.459
S.E. of regression	385.9169	Akaike info criterion	14.84089
Sum squared resid	4170093.	Schwarz criterion	14.97966
Log likelihood	-227.0337	F-statistic	560.5650
Durbin-Watson stat	1.843488	Prob(F-statistic)	0.000000

3、关于拟合优度检验R^2与方程显著性F检验关系的讨论

$$\bar{R}^2 = 1 - \frac{RSS/(n-k-1)}{TSS/(n-1)}$$

$$F = \frac{ESS/k}{RSS/(n-k-1)}$$

$$\bar{R}^2 = 1 - \frac{n-1}{n-k-1+kF}$$

$$F = \frac{\bar{R}^2/k}{(1-\bar{R}^2)/(n-k-1)}$$

F 与 $\bar{R}^2$ 同向变化：当 $\bar{R}^2$ ＝ 0 时，F ＝ 0；

$\bar{R}^2$ 越大，F 值也越大；

当 $\bar{R}^2$ ＝ 1 时，F 为无穷大。

因此，F检验是所估计回归的总显著性的一个度量，也是$\bar{R}^2$的一个显著性检验。亦即

检验 $H_0: \beta_1 = 0, \beta_2 = 0, \cdots, \beta_k = 0$ 等价于检验 $\bar{R}^2 = 0$

回答前面的问题：$\bar{R}^2$多大才算通过拟合优度检验

对于一般的实际问题：在5%的显著性水平下，F统计量临界值所对应的R^2的水平、是较低的；

所以，不宜过分注重R^2值，应注重模型的经济意义；在进行总体显著性检验时，显著性水平应该控制在5%以内。

三、单变量的显著性检验（t检验）
Testing the Significance of Variables (the t test)

- 方程的总体线性关系显著，不等于每个解释变量对被解释变量的影响都是显著的。

- 必须对每个解释变量进行显著性检验，以决定是否作为解释变量被保留在模型中；
- 这一检验，是由对变量的 t 检验完成的。

1、t统计量

$$Cov(\hat{\beta}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$$

以C_{jj}表示矩阵$(X'X)^{-1}$ 主对角线上的第 j 个元素

$$Var(\hat{\beta}_i) = \sigma^2 c_{ii}$$

$$\hat{\beta}_i \sim N(\beta_i, \sigma^2 c_{ii})$$

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n-k-1} = \frac{\mathbf{e}'\mathbf{e}}{n-k-1}$$

$$t = \frac{\hat{\beta}_j - \beta_j}{s_{\hat{\beta}_j}} = \frac{\hat{\beta}_j - \beta_j}{\sqrt{c_{jj} \frac{\mathbf{e}'\mathbf{e}}{n-k-1}}} \sim t(n-k-1)$$

Given the Gauss-Markov Assumptions

$$Var\left(\hat{\beta}_j\right) = \frac{\sigma^2}{SST_j\left(1 - R_j^2\right)},$$

where $$SST_j = \sum\left(x_{ij} - \bar{x}_j\right)^2$$

which is the Total Sample Variation in x_j,

and R_j^2 is the R^2 from regressing x_j on all other x's .

2、t检验

设计原假设与备择假设：

$$H_0: \beta_i=0 \qquad （i=1,2\ldots k）$$

$$H_1: \beta_i \neq 0$$

给定显著性水平α，可得到临界值$t_{\alpha/2}(n-k-1)$，由样本求出统计量**t**的数值，通过

$$|t| > t_{\alpha/2}(n-k-1) \quad 或 \quad |t| \leq t_{\alpha/2}(n-k-1)$$

判断拒绝或不拒绝原假设H_0，从而**判定对应的解释变量是否应包括在模型中。**

地区城镇居民消费模型

Dependent Variable: Y
Method: Least Squares
Date: 09/06/10 Time: 23:06
Sample: 1 31
Included observations: 31

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	143.3265	260.4032	0.550402	0.5864
X1	0.555644	0.075308	7.378320	0.0000
X2	0.250085	0.113634	2.200791	0.0362

R-squared	0.975634	Mean dependent var	8401.468
Adjusted R-squared	0.973893	S.D. dependent var	2388.459
S.E. of regression	385.9169	Akaike info criterion	14.84089
Sum squared resid	4170093.	Schwarz criterion	14.97966
Log likelihood	-227.0337	F-statistic	560.5650
Durbin-Watson stat	1.843488	Prob(F-statistic)	0.000000

3、关于常数项的显著性检验

- t检验同样可以进行。
- 但是，一般不以t检验决定常数项是否保留在模型中，而是从经济意义方面、分析回归线是否应该通过原点。

四、参数的置信区间

Confidence Interval of Parameter

1、区间估计

- 回归分析希望通过样本得到的参数估计量，能够代替总体参数。
- 假设检验可通过一次抽样的结果、检验总体参数可能的假设值的范围（例如是否为零），但它并没有指出在一次抽样中样本参数值到底离总体参数的真值有多“近”。
- 要判断样本参数的估计值在多大程度上“近似”地替代总体参数的真值，需要通过构造一个以样本参数的估计值为中心的“区间”，来考察它以多大的可能性（概率）包含着真实的参数值。这种方法就是参数检验的置信区间估计。

要判断估计的参数值$\hat{\beta}$离真实的参数值β有多“近”，可预先选择一个概率 α(0<α<1)，并求一个正数 δ，使得随机区间$(\hat{\beta}-\delta,\hat{\beta}+\delta)$包含参数的真值的概率为1-α。即：

$$\boxed{P(\hat{\beta}-\delta\le\beta\le\hat{\beta}+\delta)=1-\alpha}$$

如果存在这样一个区间，称之为置信区间； 1-α称为置信系数（置信度）（confidence coefficient）、置信水平， α称为 <u>显著性水平</u>（Significance Level）；置信区间的端点称为置信限（confidence limit）。

2、参数的置信区间

$$t=\frac{\hat{\beta}_i-\beta_i}{S_{\hat{\beta}_i}}\;\frac{\hat{\beta}_i-\beta_i}{\sqrt{c_{ii}\dfrac{\mathbf{e}'\mathbf{e}}{n-k-1}}}\sim t(n-k-1)$$

在(1−α)的置信水平下

$$(\hat{\beta}_i-t_{\frac{\alpha}{2}}\times s_{\hat{\beta}_i},\hat{\beta}_i+t_{\frac{\alpha}{2}}\times s_{\hat{\beta}_i})$$

3、如何才能缩小置信区间？

- **增大样本容量n**，因为在同样的显著性水平下，n越大、t分布表中的临界值越小；同时，增大样本容量，还可使参数估计量的样本标准误差减小。

- **提高模型的拟合优度**，因为样本参数估计量的标准差与残差平方和呈正比，模型优度越高，残差平方和应越小。

- **提高样本观测值的分散度**，一般情况下，样本观测值越分散，$(X'X)^{-1}$的分母的|X' X|的值越大，致使区间缩小。

§3.4 多元线性回归模型的预测

一、$E(Y_0)$的置信区间

二、Y_0的置信区间

一、E(Y$_0$)的置信区间

$$E(\hat{Y}_0) = E(\mathbf{X_0}\hat{\beta}) = \mathbf{X_0}E(\hat{\beta}) = \mathbf{X_0}\beta = E(Y_0)$$

$$Var(\hat{Y}_0) = E(\mathbf{X_0}\hat{\beta} - \mathbf{X_0}\beta)^2 = E(\mathbf{X_0}(\hat{\beta} - \beta)\mathbf{X_0}(\hat{\beta} - \beta))$$

$$\begin{aligned} Var(\hat{Y}_0) &= E(\mathbf{X_0}(\hat{\beta} - \beta)(\hat{\beta} - \beta)'\mathbf{X}_0') \\ &= \mathbf{X_0}E(\hat{\beta} - \beta)(\hat{\beta} - \beta)'\mathbf{X}_0' \\ &= \sigma^2 \mathbf{X}_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_0' \end{aligned}$$

$$\hat{Y}_0 \sim N(\mathbf{X_0}\beta, \sigma^2\mathbf{X_0}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_0')$$

取随机扰动项的样本估计量$\hat{\sigma}^2$，构造如下t统计量

$$\boxed{\frac{\hat{Y}_0 - E(Y_0)}{\hat{\sigma}\sqrt{\mathbf{X}_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_0'}} \sim t(n-k-1)}$$

于是，得到$(1-\alpha)$的置信水平下$E(Y_0)$的**置信区间**：

$$\boxed{\hat{Y}_0 - t_{\frac{\alpha}{2}} \times \hat{\sigma}\sqrt{\mathbf{X}_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_0'} < E(Y_0) < \hat{Y}_0 + t_{\frac{\alpha}{2}} \times \hat{\sigma}\sqrt{\mathbf{X}_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_0'}}$$

其中，$t_{\alpha/2}$为$(1-\alpha)$的置信水平下的**临界值**。

二、Y_0的置信区间

$$e_0 = Y_0 - \hat{Y}_0$$

$$\begin{aligned} E(e_0) &= E(\mathbf{X}_0\beta + \mu_0 - \mathbf{X}_0\hat{\beta}) \\ &= E(\mu_0 - \mathbf{X}_0(\hat{\beta} - \beta)) \\ &= E(\mu_0 - \mathbf{X}_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mu) \\ &= 0 \end{aligned}$$

$$\begin{aligned} Var(e_0) &= E(e_0^2) \\ &= E(\mu_0 - \mathbf{X}_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mu)^2 \\ &= \sigma^2(1 + \mathbf{X}_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_0') \end{aligned}$$

注：$E(u_0 u_i) = 0$，且 $Euu' = \sigma^2 I$

$(ABC)' = C'B'A'$

$$e_0 \sim N(0, \sigma^2(1+\mathbf{X}_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_0'))$$

$$\hat{\sigma}_{e_0}^2 = \hat{\sigma}^2(1+\mathbf{X}_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_0'))$$

$$t = \frac{\hat{Y}_0 - Y_0}{\hat{\sigma}_{e_0}} \sim t(n-k-1)$$

如何根据置信区间、正确地陈述预测结果？

$$\hat{Y}_0 - t_{\frac{\alpha}{2}} \times \hat{\sigma}\sqrt{1+\mathbf{X}_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_0'} < Y_0 < \hat{Y}_0 + t_{\frac{\alpha}{2}} \times \hat{\sigma}\sqrt{1+\mathbf{X}_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_0'}$$

§3.5　回归模型的其他函数形式

一、模型的类型与变换

二、非线性回归实例

三、非线性最小二乘估计*

说 明

- 在实际经济活动中，经济变量的关系是复杂的，直接表现为线性关系的情况并不多见。
- 如著名的恩格尔曲线(Engle curves)表现为幂函数曲线形式、宏观经济学中的菲利普斯曲线（Pillips cuves）表现为双曲线形式等。
- 但是，大部分非线性关系又可以通过一些简单的数学处理，使之转化为数学上的线性关系，从而可以运用线性回归模型的理论方法。

一、模型的类型与变换

1、变量置换：倒数模型、多项式模型

例如，描述税收与税率关系的**拉弗曲线：**抛物线

$$s = a + b\,r + c\,r^2 \qquad c<0$$

s：税收；　　r：税率

设$X_1 = r$，$X_2 = r^2$，　则原方程变换为

$$s = a + b\,X_1 + c\,X_2 \qquad c<0$$

2、函数变换：幂函数模型、指数函数模型与对数变换法

例如，Cobb-Dauglas生产函数：幂函数

$$Q = AK^{\alpha}L^{\beta}$$

Q：产出量，K：投入的资本；L：投入的劳动

方程两边取对数：

$$\ln Q = \ln A + \alpha \ln K + \beta \ln L$$

3、Taylor/级数展开：复杂函数模型

例如，常替代弹性CES生产函数

$$Q = A(\delta_1 K^{-\rho} + \delta_2 L^{-\rho})^{-\frac{1}{\rho}} e^{\mu}$$

$(\delta_1+\delta_2=1)$

Q:产出量，K：资本投入，L：劳动投入

ρ：替代参数，δ_1、δ_2：分配参数

方程两边取对数后，得到：

$$LnQ = LnA - \frac{1}{\rho} Ln(\delta_1 K^{-\rho} + \delta_2 L^{-\rho}) + \mu$$

将式中$\ln(\delta_1 K^{-\rho} + \delta_2 L^{-\rho})$：“在ρ=0处、展开Taylor级数”，取 关于ρ的线性项 ，即得到一个线性近似式。

如取0阶、1阶、2阶项，令 $m = \delta_1 K^{-\rho} + \delta_2 L^{-\rho}$ 可得：

$$\ln Y = \ln A + \delta_1 m \ln K + \delta_2 m \ln L - \frac{1}{2}\rho\, m \delta_1 \delta_2 \left(\ln\left(\frac{K}{L} \right) \right)^2$$

二、可化为线性的非线性回归实例

例3.5.1 建立中国城镇居民食品消费需求函数模型。

根据需求理论，居民对食品的消费需求函数大致为

$$Q = f(X, P_1, P_0) \qquad (*)$$

Q:居民对食品的需求量，X：消费者的消费支出总额

P_1：食品价格指数，P_0：一般居民消费价格指数。

零阶齐次性，当所有商品价格指数和消费者货币支出总额按同一比例变动时，需求量保持不变：

$$Q = f(X/P_0, P_1/P_0) \qquad (**)$$

为了进行比较，将同时估计（*）式与（）式。**

首先，确定具体的函数形式

根据**恩格尔定律**，居民对食品的消费支出Q与居民的总支出X间呈**幂函数**的变化关系：

$$Q = AX^{\beta_1} P_1^{\beta_2} P_0^{\beta_3}$$

对数变换：

$$\ln(Q) = \beta_0 + \beta_1 \ln X + \beta_2 \ln P_1 + \beta_3 \ln P_0 + \mu \qquad (***)$$

考虑到**零阶齐次性**时

$$\ln(Q) = \beta_0 + \beta_1 \ln(X / P_0) + \beta_2 \ln(P_1 / P_0) + \mu \qquad (****)$$

(****)式也可看成是对（***）式施加如下约束而得

$$\beta_1 + \beta_2 + \beta_3 = 0$$

因此，对（**）式进行“受约束回归”，若约束条件为真，就意味着原需求函数满足零阶齐次性条件。**

表 3.5.1　中国城镇居民消费支出及价格指数（单位：元）

	X (当年价)	X_1 (当年价)	GP (上年=100)	FP (上年=100)	Q (2000年价)	P_0 (2000年=100)	P_1 (2000年=100)
1985	673.2	351.4	111.9	116.5	1315.9	28.1	26.7
1986	799.0	418.9	107.0	107.2	1463.3	30.1	28.6
1987	884.4	472.9	108.8	112.0	1475.0	32.8	32.1
1988	1104.0	567.0	120.7	125.2	1412.5	39.5	40.1
1989	1211.0	660.0	116.3	114.4	1437.2	46.0	45.9
1990	1278.9	693.8	101.3	98.8	1529.2	46.6	45.4
1991	1453.8	782.5	105.1	105.4	1636.3	49.0	47.8
1992	1671.7	884.8	108.6	110.7	1671.4	53.2	52.9
1993	2110.8	1058.2	116.1	116.5	1715.9	61.7	61.7
1994	2851.3	1422.5	125.0	134.2	1718.7	77.2	82.8
1995	3537.6	1771.9	116.8	123.6	1732.1	90.1	102.3
1996	3919.5	1904.7	108.8	107.9	1725.6	98.1	110.4
1997	4185.6	1942.6	103.1	100.1	1758.2	101.1	110.5
1998	4331.6	1926.9	99.4	96.9	1799.8	100.5	107.1
1999	4615.9	1932.1	98.7	95.7	1885.7	99.2	102.5
2000	4998.0	1971.3	100.8	97.6	1971.3	100.0	100.0
2001	5309.0	2027.9	100.7	100.7	2013.8	100.7	100.7
2002	6029.9	2271.8	99.0	99.9	2258.3	99.7	100.6
2003	6510.9	2416.9	100.9	103.4	2323.5	100.6	104.0
2004	7182.1	2709.6	103.3	109.9	2370.2	103.9	114.3
2005	7942.9	2914.4	101.6	103.1	2472.7	105.6	117.9
2006	8696.6	3111.9	101.5	102.6	2573.4	107.2	120.9

原始数据：

X：人均消费

（也应调整为2000年价）

X1：人均食品消费(当年价)

GP：居民消费价格指数

FP：居民食品消费价格指数

调整后数据：

Q：人均食品消费（2000年价）

P0：居民消费价格缩减指数（2000=100）

P1：居民食品消费价格缩减指数（2000=100）

- 按（***）式估计

Dependent Variable: LOG(Q)
Sample: 1985 2006
Included observations: 22

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.531950	0.093107	59.41489	0.0000
LOG(X)	0.539917	0.036530	14.78015	0.0000
LOG(P1)	-0.258012	0.178186	-1.447994	0.1648
LOG(P0)	-0.288561	0.205184	-1.406350	0.1766

R-squared	0.977345	Mean dependent var	7.493909
Adjusted R-squared	0.973569	S.D. dependent var	0.193147
S.E. of regression	0.031401	Akaike info criterion	-3.921001
Sum squared resid	0.017748	Schwarz criterion	-3.722630
Log likelihood	47.13101	F-statistic	258.8448
Durbin-Watson stat	0.696202	Prob(F-statistic)	0.000000

$$\ln(\hat{Q}) = 5.53 + 0.540\ln(X) - 0.258\ln(P_1) - 0.288\ln(P_0)$$

具体解释估计结果及其经济含义。

- 按（****）式估计

Dependent Variable: LOG(Q)
Sample: 1985 2006
Included observations: 22

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.524569	0.083108	66.47481	0.0000
LOG(X/P0)	0.534439	0.023198	23.03776	0.0000
LOG(P1/P0)	-0.275347	0.151143	-1.821763	0.0843

R-squared	0.977296	Mean dependent var	7.493909
Adjusted R-squared	0.974906	S.D. dependent var	0.193147
S.E. of regression	0.030596	Akaike info criterion	-4.009741
Sum squared resid	0.017787	Schwarz criterion	-3.860963
Log likelihood	47.10715	F-statistic	408.9291
Durbin-Watson stat	0.695256	Prob(F-statistic)	0.000000

$$\ln(\hat{Q}) = 5.52 + 0.534\ln(X / P_0) - 0.275\ln(P_1 / P_0)$$

具体解释估计结果及其经济含义。

三、非线性最小二乘估计

1. 普通最小二乘原理

$$y_i = f(x_i, \beta) + \mu_i$$

残差平方和

$$S(\hat{\beta}) = \sum_{i=1}^{n} (y_i - f(x_i, \hat{\beta}))^2$$

取极小值的一阶条件

$$\frac{dS}{d\hat{\beta}} = -2\sum_{i=1}^{n} (y_i - f(x_i, \hat{\beta}))(\frac{-df(x_i, \hat{\beta})}{d\hat{\beta}}) = 0$$

$$\sum_{i=1}^{n} (y_i - f(x_i, \hat{\beta}))(\frac{df(x_i, \hat{\beta})}{d\hat{\beta}}) = 0$$

如何求解非线性方程组？

2. 高斯—牛顿(Gauss-Newton)迭代法

- 高斯—牛顿迭代法的原理

对原始模型：在给定初值处、展开台劳级数，取一阶近似值

$$f(x_i,\hat\beta)\approx f(x_i,\hat\beta_{(0)})+\frac{df(x_i,\hat\beta)}{d\hat\beta}\Big|_{\hat\beta_{(0)}}(\hat\beta-\hat\beta_{(0)})$$

$$z_i(\hat\beta)=\frac{df(x_i,\hat\beta)}{d\hat\beta}$$

$$S(\hat\beta)=\sum_{i=1}^{n}(y_i-f(x_i,\hat\beta_{(0)})-z_i(\hat\beta_{(0)})(\hat\beta-\hat\beta_{(0)}))^2$$

$$=\sum_{i=1}^{n}(y_i-f(x_i,\hat\beta_{(0)})+z_i(\hat\beta_{(0)})\hat\beta_{(0)}-z_i(\hat\beta_{(0)})\hat\beta)^2$$

$$=\sum_{i=1}^{n}(\tilde y_i(\hat\beta_{(0)})-z_i(\hat\beta_{(0)})\hat\beta)^2$$

构造并估计<u>伪线性模型</u>

$$\tilde{y}_i(\hat{\beta}_{(0)}) = z_i(\hat{\beta}_{(0)})\beta + \varepsilon_i$$

构造 <u>伪线性模型</u>

$$S(\hat{\beta}_{(1)}) = \sum_{i=1}^{n}(\tilde{y}_i(\hat{\beta}_{(0)}) - z_i(\hat{\beta}_{(0)})\hat{\beta}_{(1)})^2$$

估计得到参数的第1次迭代值 $\hat{\beta}_{(1)}$

迭代

- **高斯－牛顿迭代法的步骤**

第一步：给出参数估计值$\hat{\beta}$的初值$\hat{\beta}_{(0)}$，将$f(x_i,\hat{\beta})$在$\hat{\beta}_{(0)}$处展开台劳级数，

取一阶近似值；

第二步：计算 $z_i = \left.\dfrac{df(x_i,\hat{\beta})}{d\hat{\beta}}\right|_{\hat{\beta}_{(0)}}$ 和 $\tilde{y}_i = y_i - f(x_i,\hat{\beta}_{(0)}) + z_i \cdot \hat{\beta}_{(0)}$的样本观测值；

第三步：采用普通最小二乘法估计模型 $\tilde{y}_i = z_i\beta + \varepsilon_i$，得到$\beta$的估计值$\hat{\beta}_{(1)}$；

第四步：用$\hat{\beta}_{(1)}$代替第一步中的$\hat{\beta}_{(0)}$，重复这一过程，直至收敛。

3. 牛顿—拉夫森(Newton-Raphson)迭代法

- 自学，掌握以下2个要点
- 牛顿—拉夫森迭代法的原理
  - 对残差平方和展开台劳级数，取二阶近似值；
  - 对残差平方和的近似值求极值；
  - 迭代。
- 与高斯—牛顿迭代法的区别
  - 直接对残差平方和展开台劳级数，而不是对原模型展开；
  - 取二阶近似值，而不是取一阶近似值。

4.应用中的一个困难

- 如何保证迭代所逼近的是总体极小值（即最小值）而不是局部极小值？

- 一般方法是模拟试验：

  随机产生初始值→估计→改变初始值→再估计→反复试验，设定收敛标准（例如100次连续估计结果相同）→直到收敛。

5.非线性普通最小二乘法在软件中的实现

- 给定初值
- 写出模型
- 估计模型
- 改变初值
- 反复估计

6.例题

- 例3.5.1 建立中国城镇居民食品消费需求函数模型。

Dependent Variable: Q
Method: Least Squares
Sample: 1985 2006
Included observations: 22
Convergence achieved after 6 iterations
Q=EXP(C(1))*X^C(2)*P1^C(3)*P0^C(4)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	5.567708	0.083537	66.64931	0.0000
C(2)	0.555715	0.029067	19.11874	0.0000
C(3)	-0.190154	0.143823	-1.322146	0.2027
C(4)	-0.394861	0.159291	-2.478866	0.0233

R-squared	0.983631	Mean dependent var	1830.000
Adjusted R-squared	0.980903	S.D. dependent var	365.1392
S.E. of regression	50.45954	Akaike info criterion	10.84319
Sum squared resid	45830.98	Schwarz criterion	11.04156
Log likelihood	-115.2751	Hannan-Quinn criter.	10.88992
Durbin-Watson stat	0.672163		

$$Q = e^{5.568} X^{0.556} P_1^{-0.190} P_0^{-0.395}$$

$$\ln(\hat{Q}) = 5.53 + 0.540\ln(X) - 0.258\ln(P_1) - 0.288\ln(P_0)$$

线性估计

Dependent Variable: Q
Method: Least Squares
Sample: 1985 2006
Included observations: 22
Convergence achieved after 9 iterations
Q=EXP(C(1))*(X/P0)^C(2)*(P1/P0)^C(3)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	5.525965	0.072685	76.02666	0.0000
C(2)	0.533824	0.019785	26.98163	0.0000
C(3)	-0.242862	0.134014	-1.812218	0.0858

R-squared	0.982669	Mean dependent var	1830.000
Adjusted R-squared	0.980845	S.D. dependent var	365.1392
S.E. of regression	50.53638	Akaike info criterion	10.80939
Sum squared resid	48524.59	Schwarz criterion	10.95817
Log likelihood	-115.9033	Hannan-Quinn criter.	10.84444
Durbin-Watson stat	0.656740		

$$Q = e^{5.526}(X/P_0)^{0.534}(P_1/P_0)^{-0.243}$$

$$\ln(\hat{Q}) = 5.52 + 0.534\ln(X/P_0) - 0.275\ln(P_1/P_0)$$

线性估计

讨论

- 一般情况下，线性化估计和非线性估计结果差异不大；如果差异较大，在确认非线性估计结果为总体最小时，应该怀疑和检验线性模型。

- 非线性估计，确实存在局部极小问题；

  根据参数的经济意义和数值范围，选取迭代初值。

- NLS估计的异方差和序列相关问题。
  - NLS不能直接处理。
  - 应用最大似然估计。

§3.6 受约束回归***

Restricted Regression

一、模型参数的线性约束**

二、对回归模型增加或减少解释变量

三、参数的稳定性*

说　明

- 在建立回归模型时，有时根据经济理论需要对模型中的参数施加一定的约束条件。例如：
  - 需求函数的0阶齐次性条件
  - 生产函数的1阶齐次性条件

- 模型施加约束条件后进行回归，称为受约束回归（**restricted regression**）；
- 未加任何约束的回归称为无约束回归（**unrestricted regression**）。

一、模型参数的线性约束

1、参数的线性约束

$$Y=\beta_0+\beta_1X_1+\beta_2X_2+\cdots+\beta_kX_k+\mu$$

$$\beta_1+\beta_2=1$$

$$\beta_{k-1}=\beta_k$$

$$Y=\beta_0+\beta_1X_1+(1-\beta_1)X_2+\cdots+\beta_{k-1}X_{k-1}+\beta_{k-1}X_k+\mu^*$$

$$Y^*=\beta_0+\beta_1X_1^*+\beta_3X_3+\cdots+\beta_{k-1}X_{k-1}^*+\mu^*$$

$$Y^*=Y-X_2$$

$$X_1^*=X_1-X_2$$

$$X_{k-1}^*=X_{k-1}+X_k$$

$$\hat\beta_0,\hat\beta_1,\hat\beta_3,\cdots,\hat\beta_{k-1}$$

$$\hat\beta_2=1-\hat\beta_1$$

$$\hat\beta_k=\hat\beta_{k-1}$$

2、参数线性约束检验

- 对所考查的具体问题**能否施加约束？**
  需进一步进行相应的检验。
  常用的检验有：**F检验、x^2检验与 t检验。**

- **F检验**
  - 构造统计量；
  - 检验施加约束后模型的解释能力、是否发生显著变化。

$$\boxed{\mathbf{Y} = \mathbf{X}\hat{\beta} + \mathbf{e}} \qquad \boxed{\mathbf{Y} = \mathbf{X}\hat{\beta}_* + \mathbf{e}_*}$$

$$\boxed{\mathbf{e}_* = \mathbf{Y} - \mathbf{X}\hat{\beta}_* = \mathbf{X}\hat{\beta} + \mathbf{e} - \mathbf{X}\hat{\beta}_* = \mathbf{e} - \mathbf{X}(\hat{\beta}_* - \hat{\beta})}$$

$$e^{*\prime} = e' - (\hat{\beta}^* - \hat{\beta})' X'$$

$$X'e = 0$$

$$\boxed{\mathbf{e}_*'\mathbf{e}_* = \mathbf{e}'\mathbf{e} + (\hat{\beta}_* - \hat{\beta})'\mathbf{X}'\mathbf{X}(\hat{\beta}_* - \hat{\beta})}$$

- 受约束样本回归模型的残差平方和$\mathbf{RSS_R}$ 大于 无约束样本回归模型的残差平方和$\mathbf{RSS_U}$。

这意味着：通常情况下，对模型施加约束条件、会降低模型的解释能力。

- 如果约束条件为真，则受约束回归模型与无约束回归模型具有相同的解释能力，RSS_R 与 RSS_U的差异较小。
- 可用“**（RSS_R －RSS_U）<u>的相对大小</u>**”，来检验约束的真实性。

$$RSS_U / \sigma^2 \sim \chi^2(n - k_U - 1)$$

$$RSS_R / \sigma^2 \sim \chi^2(n - k_R - 1)$$

$$(RSS_R - RSS_U) / \sigma^2 \sim \chi^2(k_U - k_R)$$

$$\boxed{F = \frac{(RSS_R - RSS_U)/(k_U - k_R)}{RSS_U/(n - k_U - 1)} \sim F(k_U - k_R, n - k_U - 1)}$$

例3.6.1 中国城镇居民对食品的人均消费需求实例中，对零阶齐次性检验：

无约束回归：$RSS_U=0.017748$，　　$k_U=3$

受约束回归：$RSS_R=0.017787$，　　$K_R=2$

样本容量n=22，　　约束条件个数 $k_U - k_R=3-2=1$

$$\boxed{F=\frac{(0.017787-0.017748)/1}{0.017787/18}=0.0395}$$

取$\alpha=5\%$，查得临界值$F_{0.05}(1,18)=4.41$

结论：<u>不能拒绝</u>中国城镇居民对食品的人均消费需求函数具有零阶齐次特性这一假设。

二、对回归模型增加或减少解释变量

$$\boxed{Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + \mu}$$

$$\boxed{Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + \beta_{k+1} X_{k+1} + \cdots \beta_{k+q} X_{k+q} + \mu}$$

前者可看成是后者“施加q个约束后的**受约束回归**”：**通过约束检验，决定是否增加变量。**

$$\mathrm{H_0}: \quad \beta_{k+1} = \beta_{k+2} = \cdots = \beta_{k+q} = 0$$

$$F = \frac{(RSS_R - RSS_U)/q}{RSS_U/(n-(k+q+1))}$$

$$= \frac{(ESS_U - ESS_R)/q}{RSS_U/(n-(k+q+1))} \sim F(q, n-(k+q+1))$$

三、参数的稳定性

1、邹氏参数稳定性检验

- 为了检验模型在两个连续的时间序列（1,2,…，n_1）与（n_1+1,…，n_1+n_2）中是否稳定，可以将它转变为在合并时间序列(1,2,…，n_1 ，n_1+1,…，n_1+n_2)中模型的约束检验问题。

$$\boxed{Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + \mu}$$

$$\boxed{Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + \mu_1}$$

（n_1+1,…，n_1+n_2）

$$\boxed{Y = \alpha_0 + \alpha_1 X_1 + \cdots + \alpha_k X_k + \mu_2}$$

（1,2,…，n_1）

合并两个时间序列为(1,2,…，n_1，n_1+1,…，n_1+n_2)，则可写出如下无约束回归模型

$$\begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \end{pmatrix} \begin{pmatrix} \beta \\ \alpha \end{pmatrix} + \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$$

如果α=β，表示没有发生结构变化；

因此，可针对如下假设进行检验：H_0: α=β，得到受约束回归模型为：

$$\begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \beta + \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$$

检验的F统计量为：

$$\boxed{F=\frac{(RSS_R-RSS_U)/k}{RSS_U/[n_1+n_2-2(k+1)]}\sim F[k,n_1+n_2-2(k+1)]}$$

$$\boxed{RSS_U=RSS_1+RSS_2}$$

$$\boxed{F=\frac{[RSS_R-(RSS_1+RSS_2)]/k}{(RSS_1+RSS_2)/[n_1+n_2-2(k+1)]}\sim F[k,n_1+n_2-2(k+1)]}$$

- 参数稳定性的检验步骤：
  - 分别以两连续时间序列作为两个样本进行回归，得到相应的残差平方：**RSS_1与RSS_2**
  - 将两序列合并为一个大样本后进行回归，得到大样本下的残差平方和**RSS_R**
  - 计算F统计量的值，与临界值比较。若F值大于临界值，则拒绝原假设，认为发生了结构变化，参数是非稳定的。

- 该检验也被称为邹氏参数稳定性检验（Chow test for parameter stability）。

2、邹氏预测检验

- 如果出现$n_2<k$ ，则往往进行如下的**邹氏预测检验**（Chow test for predictive failure）。

- **邹氏预测检验的基本思想**:
  - 先用前一时间段n_1个样本估计原模型，再用估计出的参数进行后一时间段n_2个样本的预测。
  - 如果2段回归的预测误差相差较大，则说明参数发生了变化，否则说明参数是稳定的。
  - 转变为约束回归问题。

$$
\boxed{F=\frac{(RSS_R-RSS_U)/(k_U-k_R)}{RSS_U/(n-k_U-1)}=\frac{(RSS_R-RSS_1)/n_2}{RSS_1/(n_1-k-1)}}
$$

- 邹氏预测检验步骤：
  - 在两时间段的合并大样本下做OLS回归，得受约束模型的残差平方和RSS_R；
  - 对前一时间段的n_1个子样做OLS回归，得无约束回归残差平方和RSS_1；
  - 计算检验的F统计量，做出判断：
  - 给定显著性水平α，查F分布表，得临界值$F_\alpha(n_2, n_1-k-1)$，如果 $F>F(n_2, n_1-k-1)$，则拒绝原假设，认为预测期发生了结构变化。

例3.6.3 中国城镇居民食品人均消费需求的邹氏检验。

参数稳定性检验

1985~1997:

$$\boxed{\ln\hat{Q} = 4.633 + 0.799\ln(X/P_0) - 1.046\ln(P_1/P_0)}$$

$RSS_1=0.0083$

1998~2006:

$$\boxed{\ln\hat{Q} = 5.223 + 0.606\ln(X/P_0) - 0.224\ln(P_1/P_0)}$$

$RSS_2=0.0008$

1985~2006:

$$\boxed{\ln(\hat{Q}) = 5.52 + 0.534\ln(X/P_0) - 0.275\ln(P_1/P_0)}$$

$RSS_U=0.0178$

$$\boxed{F = \frac{[0.0178-(0.0083+0.0008)]/3}{(0.0083+0.0008)/(22-6)} = 5.10}$$

给定α=5%，查表得临界值$F_{0.05}(3, 16)=3.24$

结论：F值>临界值，拒绝参数稳定的原假设，表明中国城镇居民食品人均消费需求在1998年前后发生了显著变化。

四、非线性约束

说　明

- 非线性约束检验，是建立在<u>最大似然原理</u>基础上的。主要的检验：
  - **最大似然比检验**（likelihood ratio test, LR）
  - **<u>沃尔德检验</u>**（Wald test, Wald）
  - **拉格朗日乘数检验**（Lagrange multiplier test, LM）
- 它们的共同特点是：在大样本下，以共同的检验为基础，而自由度就是<u>约束条件的个数</u>。
- 本节不作为教学内容，供有兴趣的同学自学。