# §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

对于模型

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

给定样本以外的解释变量的观测值 $\mathbf{X}_0=(1,X_{10},X_{20},\ldots,X_{k0})$，可以得到被解释变量的预测值：

$$\hat{Y}_0 = \mathbf{X}_0\hat{\beta}$$

它可以是总体均值 $E(Y_0)$ 或个值 $Y_0$ 的预测。

**为了进行科学预测，还需求出预测值的置信区间，包括 $E(Y_0)$ 和 $Y_0$ 的置信区间。**

# 一、E($Y_0$)的置信区间

易知

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

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

由于$\mathbf{X_0}(\hat{\beta} - \beta)$为标量，因此

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

容易证明

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

取随机扰动项的样本估计量$\hat{\sigma}^2$，构造如下t统计量

$$\frac{\hat{Y}_0 - E(Y_0)}{\hat{\sigma}\sqrt{\mathbf{X_0}(\mathbf{X'X})^{-1}\mathbf{X_0'}}} \sim t(n-k-1)$$

于是，得到(1−α)的置信水平下$E(Y_0)$的置信区间：

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

其中，$t_{\alpha/2}$为(1−α)的置信水平下的临界值。

## 二、$Y_0$的置信区间

如果已经知道实际的预测值$Y_0$，那么预测误差为：

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

容易证明

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

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

$e_0$服从正态分布，即

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

取随机扰动项的样本估计量 $\hat{\sigma}^2$，可得$e_0$的方差的估计量

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

构造t统计量

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

可得给定(1−α)的置信水平下$Y_0$的置信区间：

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

中国居民人均收入-消费支出二元模型例中：

2001年人均GDP：4033.1元，

于是人均居民消费的预测值为

$\hat{Y}_{2001}$=120.7+0.2213×4033.1+0.4515×1690.8=1776.8（元）

实测值（90年价）=1782.2元，相对误差：-0.31%

预测的置信区间：

在95%的置信度下，临界值$t_{\alpha/2}(19)$=2.093，$\hat{\sigma}^2$=705.5，

$$(\mathbf{X}'\mathbf{X})^{-1}=\begin{pmatrix} 1.88952 & 0.00285 & -0.00828 \\ 0.00285 & 0.00001 & -0.00001 \\ -0.00828 & -0.00001 & 0.00004 \end{pmatrix}$$

$$\mathbf{X}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_0 = 0.3938$$

于是$E(\hat{Y}_{2001})$的95%的置信区间为:

$$1776.8 \pm 2.093 \times \sqrt{705.5} \times \sqrt{0.3938}$$

或 （1741.8，1811.7）

同样，易得$\hat{Y}_{2001}$的95%的置信区间为

$$1776.8 \pm 2.093 \times \sqrt{705.5} \times \sqrt{1.3938}$$

或 （1711.1, 1842.4）

# §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

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。

但是，大部分非线性关系又可以通过一些简单的数学处理，使之化为数学上的线性关系，从而可以运用线性回归的方法进行计量经济学方面的处理。

# 一、模型的类型与变换

## 1、倒数模型、多项式模型与变量的直接置换法

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、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、复杂函数模型与级数展开法

例如，常替代弹性CES生产函数

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

Q:产出量，K：资本投入，L：劳动投入

ρ：替代参数，$\delta_1$、$\delta_2$：分配参数

方程两边取对数后，得到：

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

将式中$\ln(\delta_1 K^{-\rho} + \delta_2 L^{-\rho})$在$\rho=0$处展开泰勒级数,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，可得

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

# 并非所有的函数形式都可以线性化

无法线性化模型的一般形式为:

$$Y = f(X_1, X_2, \cdots, X_k) + \mu$$

其中，f($x_1$,$x_2$,…,$X_k$)为非线性函数。如：

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

# §3.6 受约束回归

在建立回归模型时，有时根据经济理论需对模型中变量的参数施加一定的约束条件。

模型施加约束条件后进行回归，称为受约束回归（**restricted regression**）；

不加任何约束的回归称为无约束回归（**unrestricted regression**）。

# 受约束回归

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

# 一、模型参数的线性约束

对模型

$$Y=\beta_0+\beta_1X_1+\beta_2X_2+\cdots+\beta_kX_k+\mu \qquad (*)$$

施加约束

$$\beta_1+\beta_2=1 \qquad \beta_{k-1}=\beta_k$$

得

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

或

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

如果对（**）式回归得出 $\hat{\beta}_0,\hat{\beta}_1,\hat{\beta}_3,\cdots,\hat{\beta}_{k-1}$

则由约束条件可得： $\hat{\beta}_2=1-\hat{\beta}_1 \qquad \hat{\beta}_k=\hat{\beta}_{k-1}$

然而，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，

**主要介绍F检验**

在同一样本下，记无约束样本回归模型为

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

受约束样本回归模型为

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

于是

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

受约束样本回归模型的残差平方和 $\mathbf{RSS_R}$

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

于是

$$\mathbf{e}_*'\mathbf{e}_* \geq \mathbf{e}'\mathbf{e} \qquad (*)$$

**e'e**为无约束样本回归模型的残差平方和 $\mathbf{RSS_U}$

受约束与无约束模型都有相同的**TSS**

由（*）式 $\mathbf{RSS_R} \geq \mathbf{RSS_U}$

从而 $\mathbf{ESS_R} \leq \mathbf{ESS_U}$

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。

但是，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 **$RSS_U$**的差异变小。

可用**$RSS_R$ - $RSS_U$**的大小来检验约束的真实性

根据数理统计学的知识：

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

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

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

于是：

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

**讨论：**

如果约束条件无效， $\mathbf{RSS_R}$ 与 $\mathbf{RSS_U}$的差异较大，计算的F值也较大。

于是，可用计算的F统计量的值与所给定的显著性水平下的临界值作比较，对约束条件的真实性进行检验。

注意，$\boldsymbol{k_U - k_R}$恰为约束条件的个数。

例 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，$P_0$：居民消费价格总指数。

零阶齐次性，当所有商品的价格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与居民的总支出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（****）式进行回归，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（元）及价格指数

| | X (当年价) | X1 (当年价) | GP (上年=100) | FP (上年=100) | XC (1990年价) | Q (1990年价) | P0 (1990=100) | P1 (1990=100) |
|---|---|---|---|---|---|---|---|---|
| 1981 | 456.8 | 420.4 | 102.5 | 102.7 | 646.1 | 318.3 | 70.7 | 132.1 |
| 1982 | 471.0 | 432.1 | 102.0 | 102.1 | 659.1 | 325.0 | 71.5 | 132.9 |
| 1983 | 505.9 | 464.0 | 102.0 | 103.7 | 672.2 | 337.0 | 75.3 | 137.7 |
| 1984 | 559.4 | 514.3 | 102.7 | 104.0 | 690.4 | 350.5 | 81.0 | 146.7 |
| 1985 | 673.2 | 351.4 | 111.9 | 116.5 | 772.6 | 408.4 | 87.1 | 86.1 |
| 1986 | 799.0 | 418.9 | 107.0 | 107.2 | 826.6 | 437.8 | 96.7 | 95.7 |
| 1987 | 884.4 | 472.9 | 108.8 | 112.0 | 899.4 | 490.3 | 98.3 | 96.5 |
| 1988 | 1104.0 | 567.0 | 120.7 | 125.2 | 1085.5 | 613.8 | 101.7 | 92.4 |
| 1989 | 1211.0 | 660.0 | 116.3 | 114.4 | 1262.5 | 702.2 | 95.9 | 94.0 |
| 1990 | 1278.9 | 693.8 | 101.3 | 98.8 | 1278.9 | 693.8 | 100.0 | 100.0 |
| 1991 | 1453.8 | 782.5 | 105.1 | 105.4 | 1344.1 | 731.3 | 108.2 | 107.0 |
| 1992 | 1671.7 | 884.8 | 108.6 | 110.7 | 1459.7 | 809.5 | 114.5 | 109.3 |
| 1993 | 2110.8 | 1058.2 | 116.1 | 116.5 | 1694.7 | 943.1 | 124.6 | 112.2 |
| 1994 | 2851.3 | 1422.5 | 125.0 | 134.2 | 2118.4 | 1265.6 | 134.6 | 112.4 |
| 1995 | 3537.6 | 1766.0 | 116.8 | 123.6 | 2474.3 | 1564.3 | 143.0 | 112.9 |
| 1996 | 3919.5 | 1904.7 | 108.8 | 107.9 | 2692.0 | 1687.9 | 145.6 | 112.8 |
| 1997 | 4185.6 | 1942.6 | 103.1 | 100.1 | 2775.5 | 1689.6 | 150.8 | 115.0 |
| 1998 | 4331.6 | 1926.9 | 99.4 | 96.9 | 2758.9 | 1637.2 | 157.0 | 117.7 |
| 1999 | 4615.9 | 1932.1 | 98.7 | 95.7 | 2723.0 | 1566.8 | 169.5 | 123.3 |
| 2000 | 4998.0 | 1958.3 | 100.8 | 97.6 | 2744.8 | 1529.2 | 182.1 | 128.1 |
| 2001 | 5309.0 | 2014.0 | 100.7 | 100.7 | 2764.0 | 1539.9 | 192.1 | 130.8 |

X：人均消费

X1：人均食品消费

GP：居民消费价格指数

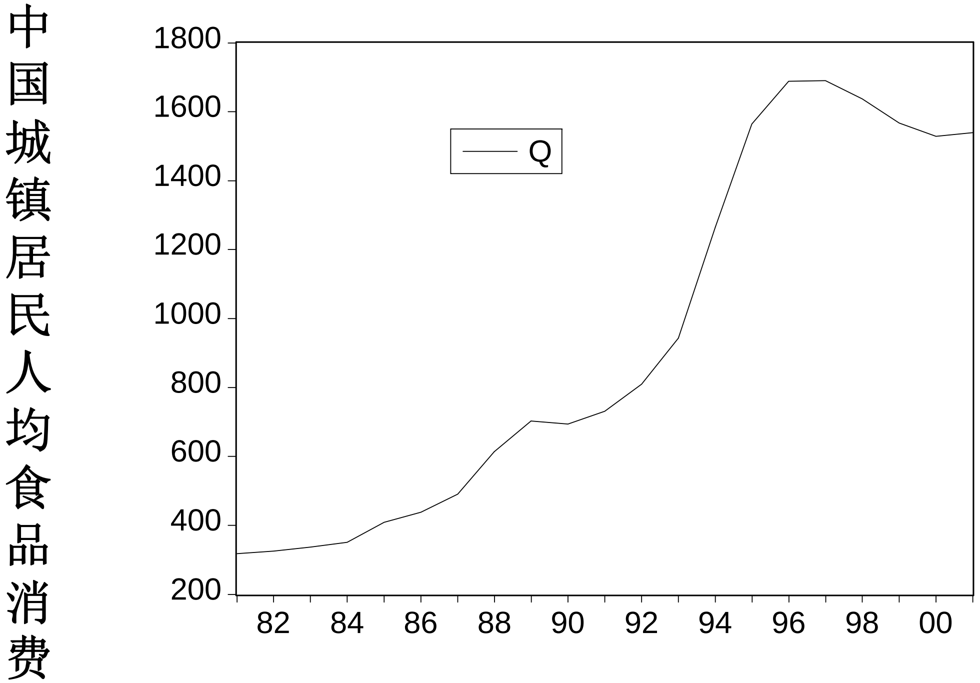
FP：居民食品消费价格指数

XC：人均消费（90年价）

Q：人均食品消费（90年价）

P0：居民消费价格缩减指数（1990=100）

P：居民食品消费价格缩减指数（1990=100




**特征：**

消费行为在1981~1995年间表现出较强的一致性

1995年之后呈现出另外一种变动特征。

**建立1981~1994年中国城镇居民对食品的消费需求模型：**

$$\ln(\hat{Q}) = 3.63 + 1.05\ln(X) - 0.08\ln(P_1) - 0.92\ln(P_0)$$

(9.03) (25.35) (-2.28) (-7.34)

$R^2=0.9987$ $\bar{R}^2=0.9983$ DW=1.50 F=2583.28

各变量的弹性和 $\hat{\beta}_1+\hat{\beta}_2+\hat{\beta}_3$ =0.05，比较接近于零，但不为零。

## 按零阶齐次性表达式回归：

$$\ln(\hat{Q}) = 3.83 + 1.07\ln(X/P_0) - 0.09\ln(P_1/P_0)$$

（75.86）(52.66)　　　(-3.62)

$R^2=0.9986$　　$\bar{R}^2=0.9984$　　DW=1.51　　F=4166.3

为了比较，改写该式为：

$$\begin{aligned}\ln\hat{Q} &= 3.83 + 1.07(\ln X - \ln P_0) - 0.09(\ln P_1 - \ln P_0) \\ &= 3.83 + 1.07\ln X - 0.09\ln P_1 - 0.98\ln P_0\end{aligned}$$

发现与

$$\ln(\hat{Q}) = 3.63 + 1.05\ln(X) - 0.08\ln(P_1) - 0.92\ln(P_0)$$

接近。

意味着：所建立的食品需求函数满足零阶齐次性特征

下面对零阶齐次性检验：

无约束回归：$RSS_U$=0.00324，　　$k_U$=3
受约束回归：$RSS_R$=0.00332，　　$K_R$=2
样本容量n=14，　　约束条件个数$k_U - k_R$=3-2=1

$$F = \frac{(0.003315 - 0.003240)/1}{0.003240/10} = 0.231$$

取α=5%，查得临界值$F_{0.05}(1,10)$=4.96

判断：不能拒绝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

考虑如下两个回归模型

$$Y=\beta_0+\beta_1 X_1+\cdots+\beta_k X_k+\mu \qquad (*)$$

$$Y=\beta_0+\beta_1 X_1+\cdots+\beta_k X_k+\beta_{k+1} X_{k+1}+\cdots\beta_{k+q} X_{k+q}+\mu \qquad (**)$$

(*)式可看成是（**）式的**受约束回归**：

H0： $\beta_{k+1}=\beta_{k+2}=\cdots=\beta_{k+q}=0$

相应的F统计量为：

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

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

## 讨论：

如果约束条件为真，即额外的变量$X_{k+1}$，…，$X_{k+q}$对 Y 没有解释能力，则 F 统计量较小；

否则，约束条件为假，意味着额外的变量对 Y 有较强的解释能力，则 F 统计量较大。

因此，可通过F的计算值与临界值的比较，来判断额外变量是否应包括在模型中。

F 统计量的另一个等价式

$$F=\frac{(R_U^2-R_R^2)/q}{(1-R_U^2)/(n-(k+q+1))}$$

## 三、参数的稳定性

建立模型时往往希望模型的参数是稳定的，即所谓的结构不变，这将提高模型的预测与分析功能。如何检验？

### 1、邹氏参数稳定性检验

假设需要建立的模型为

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

在两个连续的时间序列（1,2,…，$n_1$）与（$n_1$+1,…，$n_1+n_2$）中，相应的模型分别为：

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

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

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

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

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

(*)式施加上述约束后变换为受约束回归模型

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

因此，检验的F统计量为：

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

记$RSS_1$与$RSS_2$为在两时间段上分别回归后所得的残差平方和，容易验证，

$$RSS_U=RSS_1+RSS_2$$

于是

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

## 参数稳定性的检验步骤：

（1）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$RSS_2$

（2）将两序列并为一个大样本后进行回归，得到大样本下的残差平方和$RSS_R$

（3）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**）。

## 例3.6.2 中国城镇居民食品人均消费需求的邹氏检验。

### 1、参数稳定性检验

**1981~1994:**

$$\ln(\hat{Q}) = 3.63 + 1.05\ln(X) - 0.08\ln(P_1) - 0.92\ln(P_0)$$

$RSS_1$=0.003240

**1995~2001:**

$$\ln Q = 13.78 + 0.55\ln X - 3.06\ln P_1 + 0.71\ln P_0$$

(9.96) (7.14) (-5.13) (1.81)

$R^2$=0.9946 $\bar{R}^2$=0.9893 DW=2.80 F=185.37 $RSS_2$ =0.000058

**1981~2001:**

$$\ln Q = 5.00 + 1.21\ln X - 0.14\ln P_1 - 1.39\ln P_0$$

(14.83) (27.26) (-3.24) (-11.17)

$R^2$=0.9982 $\bar{R}^2$ =0.9979 DW=0.93 F=3228.0 $RSS_R$=0.013789

$$F=\frac{[0.013789-(0.003240+0.0000580)]/4}{(0.003240+0.000058)/(21-8)}=10.34$$

给定$\alpha=5\%$，查表得临界值$F_{0.05}(4, 13)=3.18$

**判断：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1994年前后发生了显著变化。**

例1 在经典线性模型基本假定下，对含有三个自变量的多元回归模型：

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \mu$$

你想检验的虚拟假设是 H0：$\beta_1 - 2\beta_2 = 1$。

（1）用 $\hat{\beta}_1, \hat{\beta}_2$ 的方差及其协方差求出 $Var(\hat{\beta}_1 - 2\hat{\beta}_2)$。

（2）写出检验 H0：$\beta_1 - 2\beta_2 = 1$ 的 t 统计量。

（3）如果定义 $\beta_1 - 2\beta_2 = \theta$，写出一个涉及 $\beta_0$、$\theta$、$\beta_2$ 和 $\beta_3$ 的回归方程，以便能直接得到 $\theta$ 估计值 $\hat{\theta}$ 及其标准误。

应用: 有总成本和产出数据如下

表 7—4　总成本（$Y$）与产出（$X$）

| 产出 | 总成本（美元） |
|---|---|
| 1 | 193 |
| 2 | 226 |
| 3 | 240 |
| 4 | 244 |
| 5 | 257 |
| 6 | 260 |
| 7 | 274 |
| 8 | 297 |
| 9 | 350 |
| 10 | 420 |

建立的模型如下:

$$\hat{Y}_i = 141.766\,7 + 63.477\,7X_i - 12.961\,5X_i^2 + 0.939\,6X_i^3$$

$$\text{se} = (6.375\,3) \quad (4.778\,6) \quad (0.985\,7) \quad (0.059\,1)$$

$$\operatorname{cov}(\hat{\beta}_3, \hat{\beta}_4) = -0.057\,6 \qquad R^2 = 0.998\,3 \qquad (7.10.6)$$

其中 $Y$ 为总成本，而 $X$ 为产出。括号中的数字代表估计的标准误。

假如我们要检验假设：立方成本函数中的 $X^2$ 和 $X^3$ 项的系数相等，即 $\beta_3 = \beta_4$ 或 $\beta_3 - \beta_4 = 0$。

例2　教材104页 8

对下列模型：$y_i = \alpha + \beta x_i + 2z_i + u_i$　　　（1）

$y_i = \alpha + \beta x_i - \beta z_i + u_i$　　　（2）

求出β的最小二乘估计值；并将结果与下面的三变量回归方程的最小二乘估计值作比较：

（3）$y_i = \alpha + \beta x_i - \gamma z_i + u_i$　，你认为哪一个估计值更好？

## 例3　教材 104 页 5

考虑下列两个模型：

Ⅰ、$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$

Ⅱ、$(y_i - x_{2i}) = \alpha_1 + \alpha_2 x_{2i} + \alpha_3 x_{3i} + u_i'$

要求：（1）证明：$\hat{\alpha}_2 = \hat{\beta}_2 - 1$ ，$\hat{\alpha}_1 = \hat{\beta}_1$ ，$\hat{\alpha}_3 = \hat{\beta}_3$

（2）证明：残差的最小二乘估计量相同，即：$\hat{u}_i = \hat{u}_i'$

（3）在何种情况下，模型Ⅱ的拟合优度 $R_2^2$ 会小于模型Ⅰ拟合优度 $R_1^2$ 。

例4　考虑柯布-道格拉斯生产函数：

$$Y=\beta_1 L^{\beta_2} K^{\beta_3}$$

其中$Y$=产出，$L$=劳动力投入，$K$=资本投入。把方程（1）的两边同时除以$K$得到：

$$(Y/K)=\beta_1 (L/K)^{\beta_2} K^{\beta_2+\beta_3-1} \qquad (2)$$

取（2）的自然对数得：

$$\ln(Y/K)=\beta_0+\beta_2\ln(L/K)+(\beta_2+\beta_3-1)\ln K+u_i \qquad (3)$$

其中$\beta_0=\ln\beta_1$。

a. 假如你有做回归（3）的数据，你会怎样检验规模报酬不变即$\beta_2+\beta_3=1$这个假设？

b. 如果有规模报酬不变情形，你会怎样解释回归（3）？

c. 用$L$而不用$K$去除方程（1），会有什么不同吗？

例5　纳洛夫（M. Nerlove）曾估计如下的电力产生的成本函数①：

$$Y = AX^{\beta}P_1^{\alpha_1}P_2^{\alpha_2}P_3^{\alpha_3}u$$

其中$Y$=总生产成本；

$X$=千瓦小时产出；

$P_1$=劳动力投入价格；

$P_2$=资本投入价格；

$P_3$=燃料价格；

$u$=干扰项。

理论上，预期价格弹性之和为1，即$\alpha_1+\alpha_2+\alpha_3=1$。引进这一约束，上述成本函数就可写为：

$$(Y/P_3) = AX^{\beta}(P_1/P_3)^{\alpha_1}(P_2/P_3)^{\alpha_2}u \qquad (2$$

换言之，(1) 是无约束成本函数，而 (2) 是受约束成本函数。

根据 29 个中等厂家的一个样本并通过对数变换，纳洛夫得到如下回归结果：

$$\widehat{\ln Y_i} = -4.93 + 0.94 \ln X_i + 0.31 \ln P_1 - 0.26 \ln P_2 + 0.44 \ln P_3$$

$$se = (1.96) \quad (0.11) \quad (0.23) \quad (0.29) \quad (0.07)$$

$$RSS = 0.336 \tag{3}$$

$$\widehat{\ln (Y/P_3)} = -6.55 + 0.91 \ln X + 0.51 \ln(P_1/P_3) + 0.09 \ln (P_2/P_3)$$

$$se = (0.16) \quad (0.11) \quad (0.19) \quad (0.16)$$

$$RSS = 0.364 \tag{4}$$

a. 解释方程（3）和（4）。

b. 你怎样判断约束 $\alpha_1 + \alpha_2 + \alpha_3 = 1$ 是否正确？说明你的计算。

作业 设有模型：$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$，试在下列条件下：

（1）$\beta_1 + \beta_2 = 1$

（2）$\beta_1 = \beta_2$

分别求出 $\beta_1$ 和 $\beta_2$ 的最小二乘估计量。

例2　有如下生产函数：$\ln X = 1.37 + 0.632 \ln K + 0.452 \ln L$

$$\qquad\qquad\qquad (0.257) \qquad\quad (0.219)$$

$$R^2 = 0.98 \qquad\qquad \mathrm{Cov}(b_K, b_L) = 0.055$$

其中括号内数值为参数标准差。请检验以下零假设：

（1）产出量的资本弹性和劳动弹性是等同的；

（2）存在不变规模收益，即 $\alpha + \beta = 1$ 。