线性回归

司继春

上海对外经贸大学统计与信息学院

1 线性回归

1.1 线性回归的估计

以上我们从拟合和几何的角度讨论了线性回归。而在计量经济学中，回归的目的在于解释而非预测，即关心所谓的因果效应（Causal effects）。比如，在计量经济学中，我们可能关心受教育程度对未来收入的影响，其目的不是为了使用受教育程度预测个人的收入，而是关心受教育程度是不是导致了个人可以获得更高的收入。为了得到因果效应，我们就必须在模型中做一些假设。

在上面的讨论中，我们通过最小化：

$$\hat{\beta} = \arg\min_b \sum_{i=1}^{N} e_i^2 = \arg\min_b e'e = \arg\min_b (y - Xb)'(y - Xb)$$

得到了多元线性回归。在这里，最小二乘法强调的是拟合。而如果我们希望得到所谓的因果效应，最关键的假设是所谓的**外生性**（**Exogeneity**）假设，即对于回归方程：

$$y_i = x_i'\beta + u_i$$

外生性假设即：

$$\mathbb{E}(u_i|x_i) = 0$$

即回归方程中的解释变量与误差项无关。

然而现实中很多应用都不满足这一假设。比如在以上受教育程度的例子中，尽管我们可以使用受教育程度预测个人的收入，但是如果我们简单使用线性回归解释个人的收入，现实数据可能并不满足以上的外生性假设。比如，个人的能力（$ability$）可能会影响个人能否上大学，同时也会影响其毕业之后的收入。即，能上大学的个人普遍来说能力更强，可能即使他不上大学，其收入也比较高。我们可以使用如下方程表示：

$$income_i = \gamma \cdot edu_i + x_i'\beta + ability_i + \epsilon_i$$

其中 x_i 为其他外生的控制变量。现实问题是，尽管能力对收入有影响，然而能

力是不可观测的，因而我们实际可以做的回归是：

$$income_i = \gamma \cdot edu_i + x_i'\beta + u_i$$

其中 $u_i = ability_i + \epsilon_i$。由于教育程度与能力相关，因而 $\mathrm{Cov}\left(ability_i, edu_i\right) \neq 0$，从而 $\mathrm{Cov}\left(u_i, edu_i\right) \neq 0$，因而 $\mathbb{E}\left(u_i | x_i\right) \neq 0$，外生性假设无法满足。

当外生性假设无法满足时，我们需要一些其他的计量方法进行处理，比如下面将要介绍的工具变量法。现在，我们假设外生性假设成立，并在此基础上讨论参数的估计和推断问题。

对于回归方程：

$$y_i = x_i'\beta + u_i$$

其中 x_i 为 k 维向量，代表解释变量；β 为解释变量的系数；u_i 为误差项。如果外生性假设成立，即 $\mathbb{E}\left(u_i | x_i\right) = 0$，那么根据条件期望的性质，必然有：

$$\mathbb{E}\left(x_i u_i\right) = \mathbb{E}\left[\mathbb{E}\left(x_i u_i | x_i\right)\right] = \mathbb{E}\left[x_i \mathbb{E}\left(u_i | x_i\right)\right] = \mathbb{E}\left[x_i \cdot 0\right] = 0$$

因而外生性假设保证了误差项 u_i 与 x_i 之间是不相关的。进而，我么可以使用矩估计的方法对 β 进行估计。注意到不可观测的 u_i 可以写为：

$$u_i = y_i - x_i'\beta$$

从而：

$$\mathbb{E}\left[\left(y_i - x_i'\beta\right) x_i\right] = 0$$

注意由于 $u_i = y_i - x_i'\beta$ 为标量，而 x_i 为 $K \times 1$ 的向量，因而以上方程实际上包含着 K 个矩条件。对以上方程进行整理，得到：

$$\mathbb{E}\left(x_i x_i'\right)\beta = \mathbb{E}\left(x_i y_i\right)$$

如果假设 $\mathbb{E}\left(x_i x_i'\right)$ 可逆，那么：

$$\beta = \left[\mathbb{E}\left(x_i x_i'\right)\right]^{-1} \mathbb{E}\left(x_i y_i\right)$$

根据矩估计的思想，如果我们分别使用均值：

$$\frac{1}{N}\sum_{i=1}^{N} x_i x_i' \text{和} \frac{1}{N}\sum_{i=1}^{N} x_i y_i$$

代替总体均值，那么就得到了以上问题的矩估计：

$$\hat{\beta} = \left[\frac{1}{N}\sum_{i=1}^{N} x_i x_i'\right]^{-1} \frac{1}{N}\sum_{i=1}^{N} x_i y_i = \left[\sum_{i=1}^{N} x_i x_i'\right]^{-1} \sum_{i=1}^{N} x_i y_i$$

注意以上估计实际上就是最小二乘估计：$\hat{\beta}=(X'X)^{-1}X'Y$。

以上矩估计是在外生性假设的基础之上得到的。如果在外生性假设上更进一步，不仅假设 $\mathbb{E}(u_i|x_i)=0$，同时假设 $u_i|x_i$ 的分布，比如正态分布，我们还可以得到 β 的极大似然估计。如果假设 $u_i|x_i\sim N(0,\sigma^2)$，那么 $y_i|x_i\sim N(x_i'\beta,\sigma^2)$，因而其条件极大似然函数为：

$$L(y|x)=\sum_{i=1}^{N}\left[-\ln(2\pi)-\ln\sigma-\frac{(y_i-x_i'\beta)^2}{\sigma^2}\right]$$

最大化以上似然函数，实际上等价于最小化 $\sum_{i=1}^{N}(y_i-x_i'\beta)^2$，即最小二乘的目标函数，因而在正态分布的假设下，使用条件极大似然方法得到的 β 的估计同样也是最小二乘估计量。

1.2 回归系数的解释

在以上的回归方程中，由于我们假设了 $\mathbb{E}(u_i|x_i)=0$，那么 y_i 给定 x_i 的条件期望：

$$\mathbb{E}(y_i|x_i)=x_i'\beta$$

其中系数可以理解为条件期望 $\mathbb{E}(y_i|x_i)$ 对 x_i 的偏导，即：

$$\beta_j=\frac{\partial\mathbb{E}(y_i|x_i)}{\partial x_{ij}}$$

即系数 β_j 可以解释为变量 x_{ij} 对条件期望 $\mathbb{E}(y_i|x_i)$ 的偏效应（partial effects）。使用平均偏效应有助于我们解释估计所得到的系数的含义，比如如下几种特殊情形：

1. 若 $y_i=\beta_1\cdot x_{i1}+x_i'\beta+u_i$，那么偏效应：

   $$\frac{\partial\mathbb{E}(y_i|x_{i1},x_i)}{\partial x_{i1}}=\beta_1$$

   即 x_{i1} 每增加一单位，则 y_i 平均增加 β_1 单位。

2. 若 $y_i=\beta_1\cdot x_{i1}+\beta_2x_{i1}^2+x_i'\beta+u_i$，那么偏效应：

   $$\frac{\partial\mathbb{E}(y_i|x_{i1},x_i)}{\partial x_{i1}}=\beta_1+2\beta_2x_{i1}$$

   那么意味着 x_{i1} 对 y_i 的影响是非线性的，如果 $\beta_2>0$，那么 x_{i1} 对 y_i 的影响随着 x_{i1} 的增加而增加。

3. 若 $y_i=\beta_1\cdot x_{i1}+\beta_2x_{i1}x_{i2}+\beta_3x_{i2}+x_i'\beta+u_i$，那么偏效应：

   $$\frac{\partial\mathbb{E}(y_i|x_{i1},x_{i2},x_i)}{\partial x_{i1}}=\beta_1+\beta_2x_{i2}$$

那么意味着 x_{i1} 对 y_i 的影响收到另外一个变量 x_{i2} 的影响，如果 $\beta_2 > 0$，那么 x_{i1} 对 y_i 的影响随着 x_{i2} 的增加而增加。

4. 若 $\ln y_i = \beta_1 \cdot \ln x_{i1} + x_i'\beta + u_i$，那么偏效应：

$$\frac{\partial \mathbb{E}\left(\ln y_i | x_{i1}, x_i\right)}{\partial \ln x_{i1}} = \beta_1$$

注意由于 $\partial \ln y_i = \frac{1}{y_i}\partial y_i$，同理 $\partial \ln x_{i1} = \frac{1}{x_{i1}}\partial x_{i1}$，即百分比变动，因而 β_1 可以解释为当 x_{i1} 增加 1% 时，y_i 增加的百分比变动的期望，即弹性。注意由于

$$\mathbb{E}\left(\ln y_i | x_{i1}, x_i\right) \neq \ln \mathbb{E}\left(y_i | x_{i1}, x_i\right)$$

因而 β_1 不能被解释为当 x_{i1} 增加 1% 时条件期望变动的百分比。

注意以上 3、4 两个例子中，得到的偏效应都是一个变量的函数，即 x_{i1} 对 y_i 的影响不是一个常数。在这种情况下，我们经常会关心所谓的平均偏效应（Average partial effects），即 x_{i1} 对 y_i 的影响的均值：

$$\mathbb{E}\left(\frac{\partial \mathbb{E}\left(y_i | x_i\right)}{\partial x_{ij}}\right)$$

例如在 3 中，平均偏效应为：

$$\mathbb{E}\left[\frac{\partial \mathbb{E}\left(y_i | x_{i1}, x_{i2}, x_i\right)}{\partial x_{i1}}\right] = \beta_1 + \beta_2 \mathbb{E}\left(x_{i2}\right)$$

在实践中，对于有交叉项、平方项存在的回归中，我们经常会对变量去平均。例如在 3 的回归方程中：

$$y_i = \beta_1 \cdot x_{i1} + \beta_2 x_{i1} x_{i2} + \beta_3 x_{i2} + x_i'\beta + u_i$$

令 $x_{i1}^* = x_{i1} - \bar{x}_1$，同时令 $x_{i2}^* = x_{i2} - \bar{x}_2$，我们经常使用回归：

$$y_i = \beta_1 \cdot x_{i1}^* + \beta_2 x_{i1}^* x_{i2}^* + \beta_3 x_{i2}^* + x_i'\beta + u_i$$

如此，我们可以得到平均偏效应为：

$$\mathbb{E}\left[\frac{\partial \mathbb{E}\left(y_i | x_{i1}^*, x_{i2}^*, x_i\right)}{\partial x_{i1}^*}\right] = \beta_1 + \beta_2 \mathbb{E}\left(x_{i2}^*\right) = \beta_1$$

即去平均处理后，得到的回归系数即为平均偏效应。

1.3 线性回归的小样本统计性质

在这里我们将讨论线性回归的小样本统计性质，包括无偏性即参数估计 $\hat{\beta}$ 的小样本分布。

首先对于无偏性，由于：

$$\begin{aligned}\hat{\beta} &= \left(X'X\right)^{-1} X'Y \\ &= \left(X'X\right)^{-1} X'\left(X'\beta + u\right) \\ &= \beta + \left(X'X\right)^{-1} X'u\end{aligned}$$

其中

$$u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix}$$

为误差项向量。从而：

$$\begin{aligned}\mathbb{E}\left(\hat{\beta}\right) &= \beta + \mathbb{E}\left(\left(X'X\right)^{-1} X'u\right) \\ &= \beta + \mathbb{E}\left[\mathbb{E}\left(\left(X'X\right)^{-1} X'u|X\right)\right] \\ &= \beta + \mathbb{E}\left[\left(X'X\right)^{-1} X'\mathbb{E}\left(u|X\right)\right] \\ &= \beta\end{aligned}$$

因而在外生性假设下，最小二乘估计是真实参数 β 的无偏估计。

此外我们还可以计算最小二乘估计 $\hat{\beta}$ 的方差。注意到由于 $\hat{\beta}-\beta = \left(X'X\right)^{-1} X'u$，因而 $\hat{\beta}$ 的条件协方差矩阵为：

$$\begin{aligned}\mathrm{Var}\left(\hat{\beta}|X\right) &= \mathbb{E}\left[\left(\hat{\beta}-\beta\right)\left(\hat{\beta}-\beta\right)'|X\right] \\ &= \mathbb{E}\left[\left(X'X\right)^{-1} X'uu'X\left(X'X\right)^{-1}|X\right] \\ &= \left(X'X\right)^{-1} X'\mathbb{E}\left[uu'|X\right]X\left(X'X\right)^{-1}\end{aligned}$$

在同方差（homoskedasticity）假定下，即假定 $\mathrm{Var}\left(u_i|x_i\right) = \sigma^2$，以及假定不存在自相关，即 $\mathrm{Cov}\left(u_i, u_j\right) = 0$ 的条件下，那么

$$\mathbb{E}\left(uu'|X\right) = \sigma^2 I$$

则上式可以化简为：

$$\begin{aligned}\mathrm{Var}\left(\hat{\beta}|X\right) &= \left(X'X\right)^{-1} X'\mathbb{E}\left[uu'|X\right]X\left(X'X\right)^{-1} \\ &= \left(X'X\right)^{-1} X'\sigma^2 IX\left(X'X\right)^{-1} \\ &= \sigma^2\left(X'X\right)^{-1}\end{aligned}$$

实际上我们可以证明，在 β 的所有线性无偏估计量中，最小二乘估计量是方差最小的估计量。不妨假设存在另外一个 β 的无偏估计量 $\tilde{\beta}=Cy$，由于无偏性，因而：

$$\mathbb{E}\left(\tilde{\beta}\right)=\mathbb{E}\left(Cy\right)=\mathbb{E}\left[C\left(X\beta+u\right)\right]=\mathbb{E}\left(CX\right)\beta+\mathbb{E}\left(Cu\right)$$

上式这意味着 $CX=I$。注意 $\tilde{\beta}$ 的方差为：

$$\mathrm{Var}\left(\tilde{\beta}\right)=\sigma^2CC'$$

因而其两个估计量方差之逆的差为：

$$\begin{aligned}\mathrm{Var}\left(\hat{\beta}\right)^{-1}-\mathrm{Var}\left(\tilde{\beta}\right)^{-1}&=\frac{1}{\sigma^2}\left[X'X-\left(CC'\right)^{-1}\right]\\&=\frac{1}{\sigma^2}\left[X'X-X'C\left(CC'\right)^{-1}CX\right]\\&=\frac{1}{\sigma^2}X'\left[I-C\left(CC'\right)^{-1}C\right]X\end{aligned}$$

注意到由于矩阵 $I-C\left(CC'\right)^{-1}C$ 为幂等矩阵，因而一定是半正定矩阵，从而：

$$\mathrm{Var}\left(\hat{\beta}\right)^{-1}-\mathrm{Var}\left(\tilde{\beta}\right)^{-1}\geq 0$$

从而 $\mathrm{Var}\left(\hat{\beta}\right)\leq\mathrm{Var}\left(\tilde{\beta}\right)$。正是由于以上性质，我们称最小二乘估计量 $\hat{\beta}$ 为**最优线性无偏估计量**（**Best linear unbiased estimator, BLUE**）。

在推断中，我们需要计算 $\hat{\beta}$ 的无条件方差，即：

$$\begin{aligned}\mathrm{Var}\left(\hat{\beta}\right)&=\mathbb{E}\left(\mathrm{Var}\left(\hat{\beta}|X\right)\right)+\mathrm{Var}\left(\mathbb{E}\left(\hat{\beta}|X\right)\right)\\&=\sigma^2\mathbb{E}\left(X'X\right)+\mathrm{Var}\left(\beta\right)\\&=\sigma^2\mathbb{E}\left(X'X\right)\end{aligned}$$

由于 σ^2 未知，因而我们需要对 σ^2 进行估计。记残差：

$$\hat{e}=Y-X\hat{\beta}=MY=M\left(X'\beta+u\right)=Mu$$

从而残差平方和为：

$$\hat{e}'\hat{e}=u'Mu$$

其期望：

$$\begin{aligned}
\mathbb{E}\left(\hat{e}'\hat{e}\right) &= \mathbb{E}\left(u'Mu\right) \\
&= \mathbb{E}\left[\operatorname{tr}\left(u'Mu\right)\right] \\
&= \mathbb{E}\left[\operatorname{tr}\left(Muu'\right)\right] \\
&= \operatorname{tr}\left[\mathbb{E}\left(Muu'\right)\right] \\
&= \operatorname{tr}\left[\mathbb{E}\left(\mathbb{E}\left(Muu'|X\right)\right)\right] \\
&= \operatorname{tr}\left[\mathbb{E}\left(M\mathbb{E}\left(uu'|X\right)\right)\right] \\
&= \operatorname{tr}\left[\mathbb{E}\left(\sigma^2 M\right)\right] \\
&= \sigma^2\mathbb{E}\left(\operatorname{tr}\left(I - X\left(X'X\right)^{-1}X'\right)\right) \\
&= \sigma^2\mathbb{E}\left(\operatorname{tr}\left(I\right) - \operatorname{tr}\left(X\left(X'X\right)^{-1}X'\right)\right) \\
&= \sigma^2\left(N - \operatorname{tr}\left(\left(X'X\right)^{-1}X'X\right)\right) \\
&= \sigma^2\left(N-K\right)
\end{aligned}$$

因而：

$$\sigma^2 = \frac{\mathbb{E}\left(\hat{e}'\hat{e}\right)}{N-K}$$

从而我们可以使用残差平方和对 σ^2 进行估计，即使用：

$$s^2 = \frac{\hat{e}'\hat{e}}{N-K}$$

对 σ^2 进行估计。

以上我们得到了最小二乘估计量的最优无偏线性估计量的性质，进而讨论了误差项方差的估计，然而对于统计推断来说，我们还需要估计量的抽样分布。然而在对误差项 u 不做任何假定的情况下，小样本条件下我们不能得到估计量的精确分布，因而必须对 u 的分布做进一步的假设。一个常用的假设是正态性假设，即假设误差项 u_i 为独立同分布的正态分布，$u_i \sim N\left(0,\sigma^2\right)$，或者 $u \sim N\left(0,\sigma^2 I\right)$。

在正态性假设基础之上，由于 $\hat{\beta}-\beta = \left(X'X\right)^{-1}X'u$，因而：

$$\hat{\beta}|X \sim N\left(\beta, \sigma^2\left(X'X\right)^{-1}\right)$$

即最小二乘估计量 $\hat{\beta}$ 服从联合正态分布。进而，单独某一个系数的分布也是正态分布，即：

$$\hat{\beta}_j|X \sim N\left(\beta, \sigma^2\left(X'X\right)^{-1}_{jj}\right)$$

其中 $\left(X'X\right)^{-1}_{jj}$ 代表矩阵 $\left(X'X\right)^{-1}$ 的第 j 行第 j 列的元素值。进而，对于原假

设：

$$H_0 : \beta_j = b$$

我们可以使用统计量

$$t_j = \frac{\hat{\beta}_j - b}{\sqrt{s^2 \left(X'X\right)^{-1}_{jj}}} \sim t\left(N-K\right)$$

即以上统计量服从自由度为 $N-k$ 的 t 分布。以上统计量可以用来对单个系数进行假设检验。

如果我们希望检验是否所有的 β_j 都等于 0（除常数项），那么可以使用 F 检验。不失一般性，假设截距项为 β_1，那么原假设：

$$H_0 : \beta_2 = \beta_3 = \cdots = \beta_k = 0$$

可以使用统计量：

$$F = \frac{R^2 / \left(K-1\right)}{\left(1-R^2\right) / \left(N-K\right)} \sim F\left(k-1, N-K\right)$$

在原假设条件下，$R^2=0$，因而只需要检验 $F=0$ 即可。实际上，由于：

$$R^2 = \frac{\hat{y}'M_0\hat{y}}{y'M_0y} = 1 - \frac{\hat{e}'\hat{e}}{y'M_0y}$$

因而

$$F = \frac{\hat{y}'M_0\hat{y} / \left(K-1\right)}{\hat{e}'\hat{e} / \left(N-K\right)}$$

其中：

$$M_0\hat{y} = M_0Py = M_0\left(X\beta + Pu\right)$$

在原假设条件下，$X\beta = \beta_1\iota$，而 $M_0\iota = \left(I - \frac{1}{N}\iota\iota'\right)\iota = \iota - \iota = 0$，因而在原假设条件下，$M_0\hat{y} = M_0Pu$。而由于：

$$M_0P = \left(I - \frac{1}{N}\iota\iota'\right)P = P - \frac{1}{N}\iota\iota'$$

为实对称幂等矩阵，且 $\mathrm{tr}\left(M_0P\right) = \mathrm{tr}\left(P\right) - \mathrm{tr}\left(\frac{1}{N}\iota\iota'\right) = K-1$，如果 $u \sim N\left(0, \sigma^2 I\right)$，那么：

$$\frac{1}{\sigma}\hat{y}'M_0\hat{y} = \frac{1}{\sigma}u'M_0Pu \sim \chi^2\left(K-1\right)$$

同理 $\hat{e}=Mu$，$\mathrm{tr}\,(M)=\mathrm{tr}\,(I-P)=N-K$，因而：

$$\frac{1}{\sigma}\hat{e}'\hat{e}=\frac{1}{\sigma}u'Mu\sim\chi^2\,(N-K)$$

最后，注意到 $M_0P\cdot M=M_0\,(PM)=0$，因而 $\frac{1}{\sigma}\hat{y}'M_0\hat{y}$ 与 $\frac{1}{\sigma}\hat{e}'\hat{e}$ 为独立的 χ^2 分布，从而

$$F=\frac{\hat{y}'M_0\hat{y}/\,(K-1)}{\hat{e}'\hat{e}/\,(N-K)}\sim F\,(K-1,N-K)$$

如果我们即不是针对单个系数进行检验，也不是针对所有的系数做联合检验，而是针对一个系数之间的线性关系做检验，即原假设为：

$$H_0:R\beta+q=0$$

其中 R 为 $r\times K$ 维的矩阵，q 为 $r\times 1$ 维向量。例如如果我们需要检验 $\beta_2+\beta_3=1$，同时 $\beta_1=0$，$K=4$，即原假设为：

$$H_0:\beta_2+\beta_3=1,\beta_1=0$$

那么可以取

$$R=\begin{bmatrix}1&0&0&0\\0&1&1&0\end{bmatrix},q=\begin{bmatrix}0\\-1\end{bmatrix}$$

在这种情况下，我们可以使用 Wald 检验。注意到，由于

$$\hat{\beta}|X\sim N\left(\beta,\sigma^2\,(X'X)^{-1}\right)$$

因而

$$R\hat{\beta}+q|X\sim N\left(R\beta+q,\sigma^2R\,(X'X)^{-1}\,R'\right)$$

从而：

$$\left(R\hat{\beta}+q\right)'\left[\sigma^2R\,(X'X)^{-1}\,R'\right]^{-1}\left(R\hat{\beta}+q\right)\sim\chi^2\,(r)$$

然而注意到由于 σ^2 是未知的，因而需要使用 s^2 进行替代。下面我们将会看到，$s^2=\sigma^2+o_p\,(1)$，因而如此替代并不会改变以上统计量的大样本分布，即仍然可以使用

$$\left(R\hat{\beta}+q\right)'\left[s^2R\,(X'X)^{-1}\,R'\right]^{-1}\left(R\hat{\beta}+q\right)\overset{a}{\sim}\chi^2\,(r)$$

进行假设检验。

1.4 方差分析

1.5 线性回归的大样本统计性质

下面我们来讨论最小二乘估计量的大样本性质。首先我们讨论最小二乘估计量的一致性。注意到最小二乘估计量为：

$$\hat{\beta}=\left[\frac{1}{N}\sum_{i=1}^{N}x_ix_i'\right]^{-1}\frac{1}{N}\sum_{i=1}^{N}x_iy_i$$

式中出现了两个平均，我们可以分别对以上两部分使用大数定律。在一定比较宽松的假定下，根据大数定律，有：

$$\frac{1}{N}\sum_{i=1}^{N}x_ix_i'\xrightarrow{p}\mathbb{E}\left(x_ix_i'\right)$$

以及：

$$\frac{1}{N}\sum_{i=1}^{N}x_iy_i\xrightarrow{p}\mathbb{E}\left(x_iy_i\right)=\mathbb{E}\left(x_i\left(x_i'\beta+u\right)\right)=\mathbb{E}\left(x_ix_i'\right)\beta$$

由于矩阵的逆也是连续函数，因而如果 $\left[\mathbb{E}\left(x_ix_i'\right)\right]^{-1}$ 存在，那么：

$$\hat{\beta}=\left[\frac{1}{N}\sum_{i=1}^{N}x_ix_i'\right]^{-1}\frac{1}{N}\sum_{i=1}^{N}x_iy_i\xrightarrow{p}\left[\mathbb{E}\left(x_ix_i'\right)\right]^{-1}\mathbb{E}\left(x_ix_i'\right)\beta=\beta$$

因而一致性成立。

此外，我们还可以建立 $\hat{\beta}$ 的渐进正态性。注意到由于

$$\hat{\beta}=\left[\frac{1}{N}\sum_{i=1}^{N}x_ix_i'\right]^{-1}\frac{1}{N}\sum_{i=1}^{N}x_iy_i=\beta+\left[\frac{1}{N}\sum_{i=1}^{N}x_ix_i'\right]^{-1}\frac{1}{N}\sum_{i=1}^{N}x_iu_i$$

根据大数定律，我们关注 $\sqrt{N}\left(\hat{\beta}-\beta\right)$ 的渐进分布。由于：

$$\sqrt{N}\left(\hat{\beta}-\beta\right)=\left[\frac{1}{N}\sum_{i=1}^{N}x_ix_i'\right]^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}x_iu_i$$

其中第一部分可以根据大数定律，得到 $\left[\frac{1}{N}\sum_{i=1}^{N}x_ix_i'\right]^{-1}\xrightarrow{p}\left[\mathbb{E}\left(x_ix_i'\right)\right]^{-1}$，而第二部分可以使用中心极限定理：

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}x_iu_i\overset{a}{\sim}N\left(\mathbb{E}\left(x_iu_i\right),\mathrm{Var}\left(x_iu_i\right)\right)$$

由于 $\mathbb{E}(x_i u_i)=0$，$\mathrm{Var}(x_i u_i)=\mathbb{E}\left(u_i^2 x_i x_i'\right)$，因而：

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} x_i u_i \overset{a}{\sim} N\left(0,\mathbb{E}\left(u_i^2 x_i x_i'\right)\right)$$

综上，我们可以得到：

$$\sqrt{N}\left(\hat{\beta}-\beta\right) \overset{a}{\sim} N\left(0,\left[\mathbb{E}\left(x_i x_i'\right)\right]^{-1}\mathbb{E}\left(u_i^2 x_i x_i'\right)\left[\mathbb{E}\left(x_i x_i'\right)\right]^{-1}\right)$$

更进一步，如果我们假设 u_i 为独立同分布，即 $\mathrm{Var}(u_i)=\sigma^2, \mathrm{Cov}(u_i,u_j)=0$，那么

$$\mathbb{E}\left(u_i^2 x_i x_i'\right)=\sigma^2\mathbb{E}\left(x_i x_i'\right)$$

从而

$$\sqrt{N}\left(\hat{\beta}-\beta\right) \overset{a}{\sim} N\left(0,\sigma^2\left[\mathbb{E}\left(x_i x_i'\right)\right]^{-1}\right)$$

注意到在以上渐进分布的方差估计中，σ^2 仍然是未知的，因而我们需要对 σ^2 进行估计。在小样本性质中，我们使用

$$s^2=\frac{\hat{e}'\hat{e}}{N-K}$$

对 σ^2 进行估计，现在我们讨论 s^2 的大样本性质，即是否有 $s^2 \xrightarrow{p} \sigma^2$。注意到

$$\hat{e}'\hat{e}=u'Mu=u'u-u'X\left(X'X\right)^{-1}X'u$$

其中 $u'u=\sum_{i=1}^{N} u_i^2$，$X'u=\sum_{i=1}^{N} u_i x_i$，$X'X=\sum_{i=1}^{N} x_i x_i'$，因而：

$$s^2=\frac{N}{N-K}\left[\frac{\sum_{i=1}^{N} u_i^2}{N}-\frac{\sum_{i=1}^{N} u_i x_i'}{N}\left(\frac{\sum_{i=1}^{N} x_i x_i'}{N}\right)^{-1}\frac{\sum_{i=1}^{N} u_i x_i}{N}\right]$$

根据大数定律，$\frac{1}{N}\sum_{i=1}^{N} u_i^2 \xrightarrow{p} \mathbb{E}\left(u_i^2\right)=\sigma^2$，$\frac{1}{N}\sum_{i=1}^{N} u_i x_i \xrightarrow{p} \mathbb{E}(u_i x_i)=0$，因而：

$$s^2=\frac{N}{N-K}\frac{\sum_{i=1}^{N} u_i^2}{N}+o_p(1) \xrightarrow{p} \sigma^2$$

因而 s^2 是 σ^2 的一致估计，进而可以得到 $\hat{\beta}$ 的方差的估计量：

$$\widehat{\mathrm{Var}\left(\hat{\beta}\right)}=s^2\left(X'X\right)^{-1}$$

以上方差的估计量是基于同方差（homoskedasticity）的假定下得到的，然而现实中同方差的假定仍然很强，**异方差**（**heteroskedasticity**）的存在，即 $\mathrm{Var}(u_i)=\sigma_i^2\neq\sigma^2$ 使得以上对于估计量方差的估计变的不再可靠。

需要注意的是，异方差的存在并不影响参数的一致性。一致性只要求外生性假设，即假设 $\mathbb{E}\left(u_i|x_i\right)=0$，或者 $\mathbb{E}\left(x_iu_i\right)=0$ 即可，并不需要同方差假设。然而异方差的存在却会影响推断，即影响估计量的标准误的估计，从而影响假设检验。

由于 $\mathrm{Var}\left(u_i\right)=\sigma_i^2\neq\sigma^2$，因而 $\mathrm{Var}\left(u\right)=\sigma^2 I$ 不再成立，因而我们需要重新估计 $\mathbb{E}\left(u_i^2x_ix_i'\right)$。由于

$$\hat{u}_i=y_i-x_i'\hat{\beta}=y_i-x_i'\beta+x_i'\beta-x_i'\hat{\beta}=u_i+x_i\left(\beta-\hat{\beta}\right)=u_i+o_p\left(1\right)$$

因而我们可以使用

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^N\hat{u}_i^2x_ix_i' &= \frac{1}{N}\sum_{i=1}^N\left(u_i+o_p\left(1\right)\right)^2x_ix_i'\\ &= \frac{1}{N}\sum_{i=1}^N u_i^2x_ix_i'+o_p\left(1\right)\\ &\xrightarrow{p}\mathbb{E}\left(u_i^2x_ix_i'\right)\end{aligned}$$

因而在异方差存在的情况下，我们可以使用 $\frac{1}{N}\sum_{i=1}^N\hat{u}_i^2x_ix_i'$ 对 $\mathbb{E}\left(x_iu_iu_i'x_i'\right)$ 进行估计，因而 $\hat{\beta}$ 的方差的估计量为：

$$\widehat{\mathrm{Var}\left(\hat{\beta}\right)}=\left(X'X\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^N\hat{u}_i^2x_ix_i'\right)\left(X'X\right)^{-1}$$

以上估计量称为怀特异方差稳健标准误（White's heteroskedasticity robust standard error）。