统计量即可发现假定 SLR. 3 是否成立：若 x_i 的样本标准差为零，则 SLR. 3 不成立，否则 SLR. 3 便成立。

最后，为了得到 β_0 和 β_1 的无偏估计量，我们必须加上 2.1 节曾详细讨论过的零条件均值假定。我们现在明确地把它加入我们的假定之中。

假定 SLR. 4 （零条件均值）

给定解释变量的任何值，误差的期望值都为零。换言之，

$$\mathrm{E}(u|x)=0$$

对一个随机样本，这个假定意味着对所有 $i=1, 2, \cdots, n$，都有 $E(u_i|x_i)=0$。

除了限定总体中 u 和 x 的关系，零条件均值假定（和随机抽样假定结合起来）还导致了一个方便的技术简化。具体而言，我们可以以样本中的 x_i 值为条件推导 OLS 估计量的统计性质。技术上讲，以自变量的样本值为条件的统计推导，与把 x_i 视为在重复样本中固定不变的处理方法相同。其过程如下。我们首先选择 n 个样本值 x_1，x_2，$\cdots$，x_n（这些值可以重复）。给定这些值，我们便可以得到 y 的一个样本（可通过获得 u_i 的一个随机样本而实现）。然后利用相同的 x_1，x_2，$\cdots$，x_n，又得到 y 的另一样本。利用同样的 x_1，x_2，$\cdots$，x_n 再得到 y 的一个样本。并依此重复下去。

在重复样本中保持不变的这种构想在非实验背景中不是很现实。例如，在工资—教育一例中抽取个人时，事先选择 *educ* 的值，再在特定受教育程度下抽取个人，这样的想法很难讲得通。在随机抽样时，个人是随机抽取的，工资和受教育程度都同时被记录下来，这才是社会科学领域中进行经验分析时获得多数数据集的典型方法。一旦我们假定$\mathrm{E}(u|x)$，并进行随机抽样，那么，把 x_i 视为非随机的处理办法，在推导中就不会失掉什么。危险之处在于：在重复样本中固定不变的假定总是意味着 u_i 和 x_i 是独立的。在确定简单回归分析何时得到无偏估计量时，用假定 SLR. 4 来思考至关重要。

现在，我们准备证明 OLS 估计量的无偏性。为此，利用 $\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})=\sum_{i=1}^{n}(x_i-\bar{x})y_i$（见附录 A），我们可以把方程（2.19）中的斜率估计量写成

$$\hat{\beta}_1=\frac{\sum_{i=1}^{n}(x_i-\bar{x})y_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \tag{2.49}$$

$\Sigma(x_i-\bar{x})(y_i-\bar{y})=\Sigma(x_i-\bar{x})y_i$

因为现在我们对 $\hat{\beta}_1$ 在所有可能样本中的行为感兴趣，所以 $\hat{\beta}_1$ 应被看成一个随机变量。

把（2.48）的右侧代入（2.49），从而用总体系数和误差表示 $\hat{\beta}_1$，我们便有

$$\hat{\beta}_1=\frac{\sum_{i=1}^{n}(x_i-\bar{x})y_i}{\mathrm{SST}_x}=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(\beta_0+\beta_1x_i+u_i)}{\mathrm{SST}_x} \tag{2.50}$$

其中，为简化起见，我们定义 x_i 的总波动为 $\mathrm{SST}_x=\sum_{i=1}^{n}(x_i-\bar{x})^2$。（这不是 x_i 的样本方差，因为我们没有除以 $n-1$。）利用代数求和运算，我们可以把 $\hat{\beta}_1$ 的分子写为

$$\begin{aligned}&\sum_{i=1}^{n}(x_i-\bar{x})\beta_0+\sum_{i=1}^{n}(x_i-\bar{x})\beta_1x_i+\sum_{i=1}^{n}(x_i-\bar{x})u_i\\&=\beta_0\sum_{i=1}^{n}(x_i-\bar{x})+\beta_1\sum_{i=1}^{n}(x_i-\bar{x})x_i+\sum_{i=1}^{n}(x_i-\bar{x})u_i\end{aligned} \tag{2.51}$$

$\Sigma(x_i-\bar{x})^2=SST_x$

如附录 A 所示，$\sum_{i=1}^{n}(x_i-\bar{x})=0$ 和 $\sum_{i=1}^{n}(x_i-\bar{x})x_i=\sum_{i=1}^{n}(x_i-\bar{x})^2=\mathrm{SST}_x$。所以，我们可以把 $\hat{\beta}_1$ 的分子写为 $\beta_1\mathrm{SST}_x+\sum_{i=1}^{n}(x_i-\bar{x})u_i$。将其置于分母之上，便得到

$d = x_i - \bar{x}$ $\hat{\beta}_1 - \beta_1 = \dfrac{\sum(x_i - \bar{x})u_i}{\sum(x_i - \bar{x})^2}$ 多元矩阵：

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{n}(x_i - \bar{x})u_i}{\mathrm{SST}_x} = \beta_1 + (1/\,\mathrm{SST}_x)\sum_{i=1}^{n} d_i u_i \tag{2.52}$$

$\underset{k\times 1}{E(\hat{\beta} - \beta)} = \underset{k\times 1}{E(X'X)^{-1}X'u}$

$E[u_i|X]$

其中 $d_i = x_i - \bar{x}$。我们现在可以看到，$\hat{\beta}_1$ 的估计量等于总体斜率 β_1 加上误差 $\{u_1, u_2, \cdots, u_n\}$ 的一个线性组合。以 x_i 的值为条件，$\hat{\beta}_1$ 的随机性完全来自样本中的误差。这些误差一般都不为零的事实，正是 $\hat{\beta}_1$ 与 β_1 有差异的原因。

利用方程（2.52）的表述，我们可以证明 OLS 的第一个重要统计性质。

定理 2.1　OLS 的无偏性

利用假定 SLR.1 至 SLR.4，对 β_0 和 β_1 的任何值，我们都有

$$\mathrm{E}(\hat{\beta}_0) = \beta_0 \quad 和 \quad \mathrm{E}(\hat{\beta}_1) = \beta_1 \tag{2.53}$$

$$\underset{N\times K}{X} = \begin{bmatrix} x_1^{(1)} & x_1^{(2)} & \cdots & x_1^{(k)} \\ \vdots & \vdots & & \vdots \\ x_N^{(1)} & x_N^{(2)} & \cdots & x_N^{(k)} \end{bmatrix}$$

$E[u]$

$= E[E[u|X]]$

换言之，$\hat{\beta}_0$ 对 β_0 而言是无偏的，$\hat{\beta}_1$ 对 β_1 而言是无偏的。

证明：在此证明中，期望值都以自变量的样本值为条件。因为 SST_x 和 d_i 都只是 x_i 的函数，所以它们在条件作用下都是非随机的。因此，根据（2.53），并保持暗含的以 $\{x_1, x_2, \cdots, x_n\}$ 为条件，我们有

$$u = \begin{bmatrix} u_1 \\ \vdots \\ u_N \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_K \end{bmatrix}$$

$$\begin{aligned}\mathrm{E}(\hat{\beta}_1) &= \beta_1 + \mathrm{E}\Big[(1/\,\mathrm{SST}_x)\sum_{i=1}^{n} d_i u_i\Big] = \beta_1 + (1/\,\mathrm{SST}_x)\sum_{i=1}^{n}\mathrm{E}(d_i u_i) \\ &= \beta_1 + (1/\,\mathrm{SST}_x)\sum_{i=1}^{n} d_i \boxed{\mathrm{E}(u_i)} = \beta_1 + (1/\,\mathrm{SST}_x)\sum_{i=1}^{n} d_i \cdot 0 = \beta_1\end{aligned}$$

其中我们利用了如下事实，即在假定 SLR.2 和 SLR.4 之下，每个 u_i 的期望值（以 $\{x_1, x_2, \cdots, x_n\}$ 为条件）都是零。

对 $\hat{\beta}_0$ 的证明现在就非常简单了。将方程（2.48）对 i 取均值，便得到 $\bar{y} = \beta_0 + \beta_1\bar{x} + \bar{u}_i$，将其代入 $\hat{\beta}_0$ 的表达式：

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = \beta_0 + \beta_1\bar{x} + \bar{u} - \hat{\beta}_1\bar{x} = \beta_0 + (\beta_1 - \hat{\beta}_1)\bar{x} + \bar{u}$$

那么，根据假定 SLR.2 和 SLR.4，有 $\mathrm{E}(\bar{u}) = 0$，于是以 x_i 的值为条件，我们有

$$\mathrm{E}(\hat{\beta}_0) = \beta_0 + \mathrm{E}[(\beta_1 - \hat{\beta}_1)\bar{x}] + \mathrm{E}(\bar{u}) = \beta_0 + \mathrm{E}[(\beta_1 - \hat{\beta}_1)]\bar{x}$$

但我们知道 $\mathrm{E}(\hat{\beta}_1) = \beta_1$，这就意味着 $\mathrm{E}[(\hat{\beta}_1 - \beta_1)] = 0$。因此，$\mathrm{E}(\hat{\beta}_0) = \beta_0$。以上两个论证对 β_0 和 β_1 的任何值都成立，由此我们就证明了 OLS 的无偏性。

必须记住，无偏性是 $\hat{\beta}_1$ 和 $\hat{\beta}_0$ 的抽样分布性质，并没有告诉我们从特定样本中得到的估计值是什么。我们希望，如果我们得到的样本多少有些"典型"，那么我们的估计值就会"接近于"总体值。不幸的是，我们总有可能得到点估计值远离 β_1 的不幸样本，而且我们永远也不可能确知情况是否如此。你或许想复习一下附录 C 中无偏估计量的内容，特别是表 C.1 中说明无偏性概念的模拟练习。

如果我们的四个假定中有一个不成立，那么无偏性一般也不成立。这就意味着，考虑每个假定的真实性，对一个特定的应用而言是非常重要的。假定 SLR.1 要求 y 和 x 线性相关，还要加上一个干扰项。这有可能不成立。但我们同样知道，可以通过选择 y 和 x 得到我们感兴趣的非线性关系。解决（2.47）不成立的问题要求更高深的方法，而这超出了本书的讨论范围。

今后要做时间序列分析，我们将放松随机抽样假定 SLR.2。但用它来分析横截面会怎样呢？当样本不能代表其背后的总体时，随机抽样在横截面上就会不成立：事实上，有些数据集就是通过有意识地从总体中不同部分过度取样而构造出来的。我们将在第 9 和 17 章中讨论非随机抽样的问题。

我们曾讨论过，SLR.3 在有意义的回归应用中几乎总能成立。否则，我们甚至得不到 OLS 估计值。

$E[\hat{\beta}] = \beta + E[(X'X)^{-1}X'u]$

$= \beta + E\big[E\big(\boxed{(X'X)^{-1}X'}u \,\big|\, X\big)\big]$

$= \beta + E[(X'X)^{-1}X' \underset{\substack{\parallel \\ 0}}{E(u|X)}] = \beta$

我们现在着重讨论假定 SLR. 4。若 SLR. 4 成立，则 OLS 估计量就无偏。同样，若 SLR. 4 不成立，则 OLS 估计量一般都是有偏的。有一些用以确定偏误方向和大小的方法，我们将在第 3 章中研究。

正如我们在 2.1 节中给出的几个例子所说明的，x 与 u 相关的可能性总是我们在使用非实验数据的简单回归分析中所关注的问题。当 u 包含着影响 y 且与 x 也相关的因素时，使用简单回归就会导致伪相关：也就是说，我们发现 y 和 x 的关系实际上源于既影响 y 同时又恰巧与 x 相关的不可观测因素。

例 2.12　学生的数学成绩与学校的午餐项目

令 $math10$ 表示一个高中的十年级学生在一次标准化数学考试中通过的百分比。假设我们想估计联邦资助的学校午餐项目对学生成绩的影响。在其他条件不变的情况下，我们估计午餐项目对学生成绩有正影响，即其他条件不变，若一个学生太贫穷而不能保证正常的饮食，就可以有资格享受学校午餐项目的资助，他或她的成绩应该会提高。用 $lnchprg$ 表示有资格接受午餐计划的学生百分比。那么，一个简单回归模型就是

$$math10 = \beta_0 + \beta_1 lnchprg + u \tag{2.54}$$

其中 u 包含了影响学校整体成绩的学校和学生特征。MEAP93 包含了密歇根州 408 所高中 1992—1993 学年度的数据，利用这些数据，我们得到

$$\widehat{math10} = 32.14 - 0.319 lnchprg$$

$$n=408,\ R^2=0.171$$

这个方程预计，若有资格接受午餐项目的学生增加 10 个百分点，则通过数学考试的学生会减少 3.2 个百分点。我们能相信更高的午餐项目参与率确实导致了更糟糕的成绩吗？答案几乎肯定是否定的。更好的解释是，方程（2.54）中的误差项 u 和 $lnchprg$ 相关。事实上，u 包含着既影响学生成绩又与午餐项目资格高度相关的因素，比如在校学生的贫困率。像学校质量和资源这样的变量也被包含在 u 中，它们都可能与 $lnchprg$ 相关。应该记住，估计值 -0.319 只针对这个特定例子，但它的符号和大小让我们怀疑 u 和 x 可能相关，因此该简单回归是有偏误的。

在简单回归模型中，除遗漏变量外，还有其他原因使 x 与 u 相关。因为同样的问题也会出现在多元回归分析中，所以我们到那时再系统地讨论这个问题。

2-5b　OLS 估计量的方差

除了知道 $\hat{\beta}_1$ 的抽样分布是以 β_1 为中心（$\hat{\beta}_1$ 是无偏的）的之外，能了解我们预期的 $\hat{\beta}_1$ 究竟距离 β_1 大致有多远也非常重要。在其他条件不变的情况下，这就容许我们从所有（至少一大类）无偏估计量中选择一个最佳估计量。度量 $\hat{\beta}_1$（和 $\hat{\beta}_0$）分布的分散程度，最容易操作的一个指标就是方差或其平方根即标准差。（见附录 C 中更详细的讨论。）

结果表明，在假定 SLR. 1 至 SLR. 4 下，OLS 估计量的方差可以计算出来。不过，这些表达式多少有些复杂。有鉴于此，我们增加一个在横截面分析中的传统假定。这个假定要求，以 x 为条件，无法观测变量 u 的方差是一个常数。这就是**同方差**（homoskedasticity）或"常方差"假定。

假定 SLR. 5　（同方差性）

给定解释变量的任何值，误差都具有相同的方差，换言之，

$$\mathrm{Var}(u|x)=\sigma^2$$

工资波动越大。而受教育程度低的人工作机会更少，而且只能得到最低工资；这就降低了受教育程度极低者的工资波动。这种情形如图 2－9 所示。最后，假定 SLR.5 成立与否是一个经验问题，而且我们在第 8 章会说明如何检验假定 SLR.5。

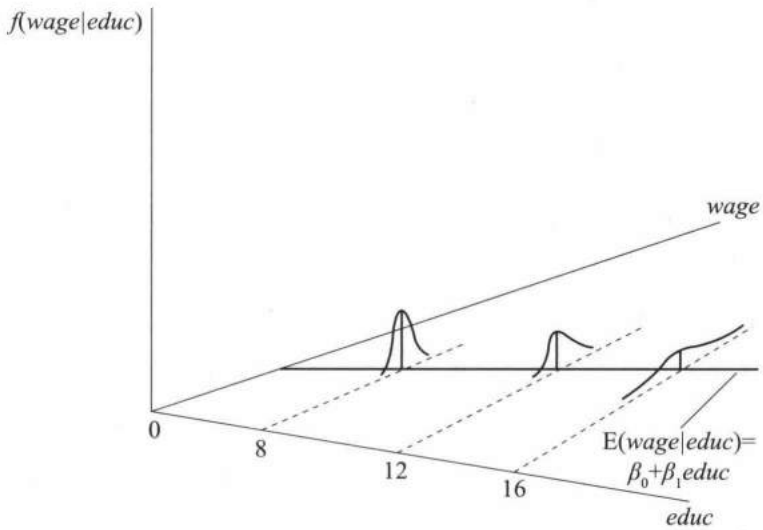


图 2－9　Var(*wage* |*educ*) 随 *educ* 增加

有了同方差假定，我们便可以证明如下定理：

定理 2.2　OLS 估计量的抽样方差

$\sigma^2 \uparrow$, $\mathrm{Var}(\hat{\beta}_1) \uparrow$

在假定 SLR.1 至 SLR.5 下，以样本值 $\{x_1, x_2, \cdots, x_n\}$ 为条件，有

$$\mathrm{Var}(\hat{\beta}_1)=\frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}=\boxed{\sigma^2/\mathrm{SST}_x} \tag{2.57}$$

$SST_x = \sum (x_i - \bar{x})^2$ $\Rightarrow \sum d_i^2 = \sum (x_i-\bar{x})^2 = SST_x$

和

$$\mathrm{Var}(\hat{\beta}_0)=\frac{\sigma^2 n^{-1}\sum_{i=1}^{n}x_i^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \tag{2.58}$$

$d_i = x_i - \bar{x}$

证明：我们只推导 $\mathrm{Var}(\hat{\beta}_1)$ 的公式，把另一个推导留作习题 10。从方程 (2.52)：$\hat{\beta}_1=\beta_1+(1/\mathrm{SST}_x)\sum_{i=1}^{n}d_i u_i$ 开始。因为 β_1 只是一个常数，而且我们以 x_i 为条件，所以 SST_x 和 $d_i=x_i-\bar{x}$ 也是非随机的。而且，因为 u_i 在 i 上（根据随机抽样）是独立的随机变量，故和的方差就是方差的和。利用这些结论，我们有

$$\begin{aligned}\mathrm{Var}(\hat{\beta}_1)&=(1/\mathrm{SST}_x)^2\mathrm{Var}\Big(\sum_{i=1}^{n}d_i u_i\Big)=(1/\mathrm{SST}_x)^2\Big(\sum_{i=1}^{n}d_i^2\,\underbrace{\mathrm{Var}(u_i)}_{\sigma^2}\Big)\\&=(1/\mathrm{SST}_x)^2\Big(\sum_{i=1}^{n}d_i^2\sigma^2\Big)[\text{因为 }\mathrm{Var}(u_i)=\sigma^2,\forall i]\\&=\sigma^2(1/\mathrm{SST}_x)^2\Big(\sum_{i=1}^{n}d_i^2\Big)=\sigma^2(1/\mathrm{SST}_x)^2\mathrm{SST}_x=\boxed{\sigma^2/\mathrm{SST}_x}\end{aligned}$$

这正是我们想要证明的。

方程 (2.57) 和 (2.58) 是简单回归分析的"标准"公式，它们在出现异方差时便失效了。在我们考虑多元回归分析中的置信区间和假设检验问题时，它们的重要性便会表现出来。

多数时候，我们关注的都是 $\mathrm{Var}(\hat{\beta}_1)$。这个方差如何取决于误差方差 σ^2 和 $\{x_1, x_2, \cdots, x_n\}$ 的总波动 SST_x，不难加以概括。1 一方面，误差方差越大，$\mathrm{Var}(\hat{\beta}_1)$ 就越大。因为影响 y 的不可观测因素波动越大，要准确估计 β_1 就越难。2 另一方面，自变量的波动越大越好：随着 x_i 的波动增加，$\hat{\beta}_1$ 的方差就会减小。这一点也符合直觉，因为自变量的样本分布越分散，就越容易找出 $\mathrm{E}(y|x)$ 和 x 之间的关系。也就是说，越容易估计出 β_1。如果 x_i 没有什么波动，就难以查明 $\mathrm{E}(y|x)$ 是如何随着 x 而变化的。当样本容量扩大时，x_i 的总体波动也会增加。因此，较大的样本容量会致使 $\hat{\beta}_1$ 的方差较小。

思考题 2.5

在估计 β_0 的时候，最好是有 $\bar{x}=0$。在这种情况下，$\mathrm{Var}(\hat{\beta}_0)$ 是多少呢？（提示：对样本中的任何数值，都有 $\sum_{i=1}^{n} x_i^2 \geqslant \sum_{i=1}^{n}(x_i-\bar{x})^2$，当且仅当 $\bar{x}=0$ 时取等号。）

这一分析表明，如果我们对 β_1 感兴趣，那么就可以选择使 x_i 尽可能分散。这在实验数据中是可能的，但在社会科学中却极少有这么丰富的样本供我们选择：通常我们要使用通过随机抽样得到的 x_i。但有时候，我们还是有机会得到较大的样本，尽管代价或许会很高。

为了构造置信区间和推导检验统计量，我们需要用到 $\hat{\beta}_1$ 和 $\hat{\beta}_0$ 的标准差 $\mathrm{sd}(\hat{\beta}_1)$ 和 $\mathrm{sd}(\hat{\beta}_0)$。记住通过将方程（2.57）和方程（2.58）中的方差取平方根即可得到。具体而言，就是 $\mathrm{sd}(\hat{\beta}_1)=\sigma/\sqrt{\mathrm{SST}_x}$，其中 σ 是 σ^2 的平方根，$\sqrt{\mathrm{SST}_x}$ 是 SST_x 的平方根。

2-5c　误差方差的估计

（2.57）和（2.58）中的公式使我们能够将影响 $\mathrm{Var}(\hat{\beta}_1)$ 和 $\mathrm{Var}(\hat{\beta}_0)$ 的因素分离出来。但除非在已知 σ^2 这种极端罕见的情况下，否则这些公式都是未知的。不过，我们可以用观测数据去估计 σ^2，从而让我们估计出 $\mathrm{Var}(\hat{\beta}_1)$ 和 $\mathrm{Var}(\hat{\beta}_0)$。

我们现在强调误差（或干扰）与残差的区别再合适不过了，因为这一区别对构造 σ^2 的估计量非常关键。方程（2.48）告诉我们，如何利用随机样本观测把总体模型写成 $y_i=\beta_0+\beta_1 x_i+u_i$，其中 u_i 是第 i 次观测的误差。我们还可以像方程（2.32）那样将 y_i 用其拟合值和残差表示出来：$y_i=\hat{\beta}_0+\hat{\beta}_1 x_i+\hat{u}_i$。比较这两个方程，我们可以看出，一方面，误差出现在包含总体参数 β_0 和 β_1 的方程中。另一方面，残差则出现在使用 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 的估计方程中。误差是无法观测的，但残差却可以从数据中计算出来。

我们可以用方程（2.32）和（2.48）把残差写成误差的函数：

$$\hat{u}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i = (\beta_0 + \beta_1 x_i + u_i) - \hat{\beta}_0 - \hat{\beta}_1 x_i$$

或者

$$\hat{u}_i = u_i - (\hat{\beta}_0 - \beta_0) - (\hat{\beta}_1 - \beta_1)x_i \tag{2.59}$$

尽管 $\hat{\beta}_0$ 的期望值等于 β_0，$\hat{\beta}_1$ 的期望值也等于 β_1，但 $\hat{u}_i$ 却不等于 u_i。但二者之差的期望值确实为零。

既然我们理解了误差和残差的区别，我们就可以回到 σ^2 的估计上来。首先，因为 $\sigma^2=\mathrm{E}(u^2)$，所以 σ^2 的一个无偏"估计量"就是 $n^{-1}\sum_{i=1}^{n}u_i^2$。不幸的是，这不是一个真正的估计量，因为我们观察不到误差 u_i。但是我们有 u_i 的估计值，即 OLS 残差 $\hat{u}_i$。如果我们用 OLS 残差来代替误差，便得到 $n^{-1}\sum_{i=1}^{n}\hat{u}_i^2=\mathrm{SSR}/n$。这是一个真正的估计量，因为对 x 和 y 的任何一个数据样本，它都给出

了一个可以计算的规则。这个估计量的一个小瑕疵是，可以证明它是有偏误的（尽管对很大的 n 来说，这个偏误很小）。因为很容易从它计算出一个无偏估计量，所以我们还是使用后者。

估计量 SST/n 有偏误，本质上是因为它没有考虑 OLS 残差所必须满足的两个约束条件。这些约束由 OLS 的两个一阶条件给出：

$$\sum_{i=1}^{n}\hat{u}_i=0,\ \sum_{i=1}^{n}x_i\hat{u}_i=0 \tag{2.60}$$

考虑这些限制的一个方法是：如果我们知道残差中的 $n-2$ 个，就能够通过（2.60）中的一阶条件所蕴含的约束，得到另外两个残差。因此，OLS 残差只有 $n-2$ 个**自由度**（degrees of freedom），与误差的 n 个自由度相对照。如果我们在方程（2.60）中用 u_i 取代 $\hat{u}_i$，这些约束就不再成立。我们将使用 σ^2 的无偏估计量，对自由度进行调整：

$$\hat{\sigma}^2=\frac{1}{(n-2)}\sum_{i=1}^{n}\hat{u}_i^2=\mathrm{SSR}/(n-2) \tag{2.61}$$

（这个估计量有时记为 s^2，但我们继续使用在估计量上加“帽”的惯常记法。）

定理 2.3　σ^2 的无偏估计

在假定 SLR.1 至 SLR.5 下，我们有

$$\mathrm{E}(\hat{\sigma}^2)=\sigma^2$$

多元版本：$E\left[\frac{\hat{u}'\hat{u}}{n-k}\middle|X\right]=\sigma^2$

证明：如果我们把方程（2.59）对所有 i 进行平均，并利用 OLS 残差均值为零的结论，便得到 $0=\bar{u}-(\hat{\beta}_0-\beta_0)-(\hat{\beta}_1-\beta_1)\bar{x}$；从方程（2.59）中减去它，则得到 $\hat{u}_i=(u_i-\bar{u})-(\hat{\beta}_1-\beta_1)(x_i-\bar{x})$。因此，$\hat{u}_i^2=(u_i-\bar{u})^2+(\hat{\beta}_1-\beta_1)^2(x_i-\bar{x})^2-2(u_i-\bar{u})(\hat{\beta}_1-\beta_1)(x_i-\bar{x})$。对所有 i 求和，又得到 $\sum_{i=1}^{n}\hat{u}_i^2=\sum_{i=1}^{n}(u_i-\bar{u})^2+(\hat{\beta}_1-\beta_1)^2\sum_{i=1}^{n}(x_i-\bar{x})^2-2(\hat{\beta}_1-\beta_1)\sum_{i=1}^{n}u_i(x_i-\bar{x})$。现在如附录 C 所示，等式右边第一项的期望值是 $(n-1)\sigma^2$。第二项的期望值无非就是 σ^2，因为 $\mathrm{E}[(\hat{\beta}_1-\beta_1)^2]=\mathrm{Var}(\hat{\beta}_1)=\sigma^2/\mathrm{SST}_x$。最后，第三项可写为 $-2(\hat{\beta}_1-\beta_1)^2\,\mathrm{SST}_x$，取期望便得到 $-2\sigma^2$。把这三项放在一起，我们得到 $\mathrm{E}(\sum_{i=1}^{n}\hat{u}_i^2)=(n-1)\sigma^2+\sigma^2-2\sigma^2=(n-2)\sigma^2$，因此 $\mathrm{E}[\mathrm{SSR}/(n-2)]=\sigma^2$。

如果把 $\hat{\sigma}^2$ 代入方差公式（2.57）和（2.58），我们就得到 $\mathrm{Var}(\hat{\beta}_1)$ 和 $\mathrm{Var}(\hat{\beta}_0)$ 的无偏估计量。稍后，我们要用到 $\hat{\beta}_1$ 和 $\hat{\beta}_0$ 的标准差估计量，而这就需要估计 σ。σ 的自然估计量就是

$$\hat{\sigma}=\sqrt{\hat{\sigma}^2} \tag{2.62}$$

Jensen's Inequality.

并被称为**回归标准误**（standard error of the regression，SER）。（$\hat{\sigma}$ 的其他名称有估计值标准误和均方根误，但我们不用这些。）尽管 $\hat{\sigma}$ 不是 σ 的无偏估计量，但是我们能够证明它是 σ 的一致估计量（见附录 C），而且它能够很好地服务于我们的目的。

估计值 $\hat{\sigma}$ 值得关注，因为它是影响 y 的不可观测因素的标准差估计值；换言之，它估计了把 x 的影响排除之后 y 的标准差。大多数回归软件包与 R^2、截距、斜率和其他 OLS 统计量一起给出 $\hat{\sigma}$ 值（用上面列出的名字之一）。现在，我们主要的兴趣在于，利用 $\hat{\sigma}$ 来估计 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 的标准差。因为 $\mathrm{sd}(\hat{\beta}_1)=\sigma/\sqrt{\mathrm{SST}_x}$，所以 $\mathrm{sd}(\hat{\beta}_1)$ 的一个自然估计量为

$$\mathrm{se}(\hat{\beta}_1)=\hat{\sigma}/\sqrt{\mathrm{SST}_x}=\hat{\sigma}/\left(\sum_{i=1}^{n}(x_i-\bar{x})^2\right)^{1/2}$$

它被称为 $\hat{\beta}_1$ 的**标准误**（standard error）。注意到，当我们考虑对 y 的不同样本使用 OLS 时，将 $\mathrm{se}(\hat{\beta}_1)$ 看作一个随机变量；因为 $\hat{\sigma}$ 随着样本的不同而变化，所以这是正确的。对于一个给定的样本，$\mathrm{se}(\hat{\beta}_1)$ 就像 $\hat{\beta}_1$ 一样，只是用给定数据计算出来的一个数字而已。

类似地，$\mathrm{se}(\hat{\beta}_0)$ 是通过将 $\mathrm{sd}(\hat{\beta}_0)$ 中的 σ 置换成 $\hat{\sigma}$ 而得到的。任何一个估计值的标准误都能让我们了解这个估计量有多精确。标准误在本书中担当着重要角色；从第 4 章开始，我们将用它们构

定理2.3步骤： $\hat{u}_i = u_i - (\hat{\beta}_0 - \beta_0) - (\hat{\beta}_1 - \beta_1) x_i$ ①

① 求均值： $0 = \bar{u} - (\hat{\beta}_0 - \beta_0) - (\hat{\beta}_1 - \beta_1) \bar{x}$ ②

② 减： $\hat{u}_i = (u_i - \bar{u}) - (\hat{\beta}_1 - \beta_1)(x_i - \bar{x})$

③ 平方，求和： $\sum \hat{u}_i^2 = \boxed{\sum (u_i - \bar{u})^2} \boxed{- 2(\hat{\beta}_1 - \beta_1) \sum (x_i - \bar{x})(u_i - \bar{u})}$

$\frac{1}{n-1} \sum (u_i - \bar{u})^2$ $\quad + (\hat{\beta} - \beta_1)^2 \sum (x_i - \bar{x})^2$

④ $C \cdot \sum (x_i - \bar{x}) = 0$. $\quad -2(\hat{\beta}_1 - \beta_1) \boxed{\sum (x_i - \bar{x}) u_i}$

$$\hat{\beta}_1 - \beta_1 = \frac{\sum (x_i - \bar{x}) u_i}{\sum (x_i - \bar{x})^2} \qquad \sum (x_i - \bar{x}) u_i = (\hat{\beta}_1 - \beta_1) \cdot \sum (x_i - \bar{x})^2$$

$$-2(\hat{\beta}_1 - \beta_1) \sum (x_i - \bar{x}) u_i = -2(\hat{\beta}_1 - \beta_1)^2 \sum (x_i - \bar{x})^2$$

⑤ $E(\hat{\beta}_1 - \beta_1)^2 = E(\hat{\beta}_1^2) - 2E(\beta_1 \hat{\beta}_1) + E(\beta_1^2) = E(\hat{\beta}_1^2) - \beta_1^2$

$$E(\hat{\beta}_1^2) = Var(\hat{\beta}_1) + (E(\hat{\beta}_1))^2 = \frac{\sigma^2}{SST_x} + \beta_1^2.$$

$$E(\hat{\beta}_1 - \beta_1)^2 = \frac{\sigma^2}{SST_x}$$

⑥ 全部期望： $E(\sum \hat{u}_i^2) = (n-1)\sigma^2 - 2\sigma^2 + \sigma^2$

$$= (n-2)\sigma^2$$

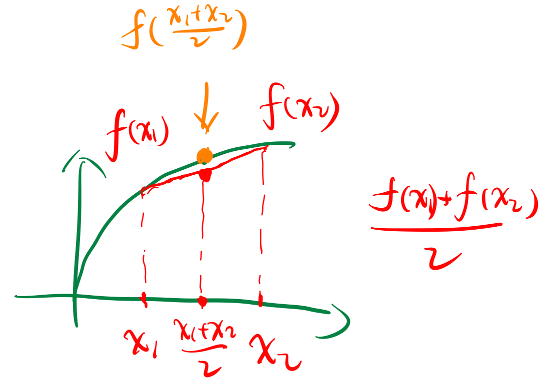
$\frac{1}{n} \sum \hat{u}_i^2$

σ^2 无偏估计： $\frac{1}{(n-2)} \sum \hat{u}_i^2$

$\frac{1}{n-1} \sum (x_i - \bar{x})^2$

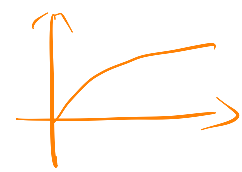
Jensen's Inequality

f 는 concave



$$E(f(u)) \leq f(E(u))$$

$f(x) = \sqrt{x}$



造每一个计量经济程序的检验统计量和置信区间。

消费 y 和 x、x 收入

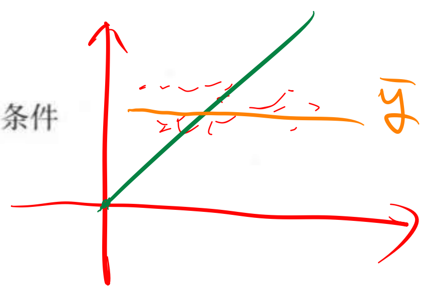
2.6 过原点回归及对常数回归

在少数情形中，我们希望施加如下约束：当 $x=0$ 时，y 的期望值为零。对某些特定关系，这个约束是合理的。例如，若收入（x）为零，那么收入税所得（y）也必须为零。另外，在有些情形中，一个原本有一个非零截距的模型被变换成没有截距的模型。

规范地，我们选择一个斜率估计量（称之为 $\tilde{\beta}_1$）和如下形式的一条线：

$$\tilde{y}=\tilde{\beta}_1 x \tag{2.63}$$

$\tilde{\beta}_1$ 和 $\tilde{y}$ 上面的波浪号用于将这个问题与同时估计截距和斜率的更常见问题区别开来。因为直线（2.63）经过点 $x=0, \tilde{y}=0$，所以得到的方程（2.63）又被称为**过原点回归**（regression through the origin）。为了求方程（2.63）中的斜率估计值，我们仍然使用普通最小二乘，此时最小化的残差平方和为



$$\sum_{i=1}^{n}(y_i-\tilde{\beta}_1 x_i)^2 \tag{2.64}$$

利用一元微积分可以证明，$\tilde{\beta}_1$ 必须满足一阶条件

$$\sum_{i=1}^{n}x_i(y_i-\tilde{\beta}_1 x_i)=0 \tag{2.65}$$

从而解出 $\tilde{\beta}_1$：

$$\tilde{\beta}_1=\frac{\sum_{i=1}^{n}x_i y_i}{\sum_{i=1}^{n}x_i^2} \tag{2.66}$$

前提是并非所有的 x_i 都为零，而这一情形我们已经排除了。

注意，与同时也估计截距（而不是把它设定为零）的斜率估计值相比，$\tilde{\beta}_1$ 有什么不同。当且仅当 $\bar{x}=0$ 时，这两个估计值才是相同的。[见关于 $\hat{\beta}_1$ 的方程（2.49）。] 在应用研究中，用过原点回归求 β_1 估计值的情形并不常见，这也是有道理的：因为如果截距 $\beta_0\neq 0$，那么 $\tilde{\beta}_1$ 就是 β_1 的有偏估计量。习题 8 要求你证明这一点。

在过原点回归被视为适当的情况下，研究者在解释该回归报告的 R^2 时需谨慎。通常在不做其他说明的情况下，R^2 是在计算 SST 时不消除 $\{y_i: i=1, \cdots, n\}$ 的样本均值而得到的，也就是说，R^2 是通过下式计算得到的：

样本外 R^2

横截面 因子模型

$\sum u_i=0$ $\sum x_i u_i=0$

$$1-\frac{\sum_{i=1}^{n}(y_i-\tilde{\beta}_1 x_i)^2}{\sum_{i=1}^{n}y_i^2} \tag{2.67}$$

跟0比

其中，分子部分作为残差平方和是有意义的，而分母部分成立的前提是，我们已知 y 的总体均值为零。之所以通过这种方法计算 R^2，是因为如果我们通过下式的常规方法计算 R^2 时，其结果应为负值：

时间序列

y: 指数收益率

$$1-\frac{\sum_{i=1}^{n}(y_i-\tilde{\beta}_1 x_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2} \tag{2.68}$$

Benchmark.

有 $\bar{y}$