Пропущенные и излишние регрессоры

лекция 15

КОРОТКАЯ ИЛИ ДЛИННАЯ РЕГРЕССИЯ

$$\mathbf{y} = \mathbf{X\beta} + Z\gamma + \boldsymbol{\varepsilon}; \qquad \mathbf{y} = \mathbf{X\beta} + \boldsymbol{\varepsilon}$$

$$R^2_{UR,adj} = 1 - \frac{RSS_{UR}}{TSS}\frac{n-1}{n-k-1}; \qquad R^2_{R,adj} = 1 - \frac{RSS_{R}}{TSS}\frac{n-1}{n-k}$$

$$R^2_{R,adj} - R^2_{UR,adj} = \frac{RSS_{UR}/(n-k-1) - RSS_{R}/(n-k)}{TSS/(n-1)}$$

КОРОТКАЯ ИЛИ ДЛИННАЯ РЕГРЕССИЯ

$$\mathbf{y} = \mathbf{X\beta} + Z\gamma + \boldsymbol{\varepsilon}; \qquad \mathbf{y} = \mathbf{X\beta} + \boldsymbol{\varepsilon}$$

$$R^2_{R,adj} - R^2_{UR,adj} = \frac{RSS_{UR}/(n-k-1) - RSS_R/(n-k)}{TSS/(n-1)}$$

$$F = \frac{(RSS_R - RSS_{UR})/1}{RSS_{UR}/(n-k-1)} \sim F(1; n-k-1) = t^2_{n-k-1} \Rightarrow$$

$$\Rightarrow \frac{1-t_Z^2}{n-k}\frac{RSS_{UR}/(n-k-1)}{TSS/(n-1)} = R^2_{R,adj} - R^2_{UR,adj}$$

$$R^2_{R,adj} < R^2_{UR,adj} \Leftrightarrow |t_Z| > 1$$

$\Rightarrow$ регрессор Z существеннен,

т.к. без него оценки станут "хуже": R^2_{adj} уменьшится

F TESTS OF GOODNESS OF FIT

```
. reg S ASVABC SM SF

  Source |       SS       df       MS              Number of obs =     570
---------+------------------------------           F(  3,   566) =  110.83
   Model |  1278.24153     3  426.080508           Prob > F      =  0.0000
Residual |  2176.00584   566  3.84453329           R-squared     =  0.3700
---------+------------------------------           Adj R-squared =  0.3667
   Total |  3454.24737   569  6.07073351           Root MSE      =  1.9607

------------------------------------------------------------------------------
       S |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  ASVABC |   .1295006   .0099544     13.009   0.000       .1099486    .1490527
      SM |    .069403   .0422974      1.641   0.101       -.013676     .152482
      SF |   .1102684   .0311948      3.535   0.000       .0489967    .1715401
   _cons |   4.914654   .5063527      9.706   0.000       3.920094    5.909214
------------------------------------------------------------------------------
```

Регрессор SM исключать не следует

Выбор переменных множественной регрессии

Оценка значимости включаемой переменной

Значимость включаемой переменной измеряется *t*-статистикой соответствующего коэффициента

Эквивалентный метод - использование *F*-критерия

$$F = \frac{\text{Улучшение качества уравнения}}{\text{Необъясненная сумма квадратов отклонений / оставшееся число степеней свободы}}$$

$$F = \frac{(RSS_k - RSS_{k+1})/1}{RSS_{k+1}/(n-(k+1)-1)} = \frac{(RSS_k - RSS_{k+1})/1}{RSS_{k+1}/(n-k-2)}$$

Эквивалентность предполагает двухстороннюю альтернативу для *t*-критерия

Выбор переменных множественной регрессии

Четыре критерия для включения переменной

1. Роль переменной в уравнении опирается на прочные теоретические основания

2. Высокие значения *t*-статистики

3. Исправленный коэффициент детерминации растет при включении переменной

4. Другие коэффициенты испытывают значительное смещение при включении новой переменной

Выбор переменных множественной регрессии

Последствия невключения в уравнение существенной переменной

Для краткости будем называть переменную существенной, если она должна быть включена в уравнение (согласно правильной теории)

Будем говорить также об исключении переменной из правильно специфицированного уравнения регрессии

Последствия невключения в уравнение существенной переменной

1. Уменьшается возможность правильной оценки и интерпретации уравнения

2. Коэффициенты при оставшихся переменных могут оказаться смещенными

3. Их стандартные ошибки, *t*-статистики и другие показатели качества становятся некорректными и не могут быть использованы для суждения о качестве уравнения

Отсутствие существенной переменной

Механизм разрушения оценок

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$$

Если вторая объясняющая переменная отсутствует, то

$$Y_i = \alpha + \beta_1 X_{1i} + u_i^*$$

где

$$u_i^* = f(\beta_2 X_{2i} + \varepsilon)$$

Если объясняющие переменные коррелированы, то нарушается условие некоррелированности случайного члена и объясняющих переменных

ПРОПУСК СУЩЕСТВЕННЫХ ПЕРЕМЕННЫХ

GeneratedDataProcess $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}; \quad \mathbf{X} = \left(\mathbf{X}_1, \ldots, \mathbf{X}_k\right); \quad \mathbf{Z} = \left(\mathbf{Z}_{k+1}, \ldots, \mathbf{Z}_{k+l}\right)$

$$\hat{\boldsymbol{\delta}}^* = \begin{bmatrix} \hat{\boldsymbol{\beta}}^* \\ \hat{\boldsymbol{\gamma}}^* \end{bmatrix} = \left[\left(\mathbf{XZ}\right)'\left(\mathbf{XZ}\right)\right]^{-1}\left(\mathbf{XZ}\right)'\mathbf{y} = \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}' \\ \mathbf{Z}' \end{bmatrix}\mathbf{y}$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}; \quad \hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$$

[illegible] $E\left(\boldsymbol{\beta} + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Z}\boldsymbol{\gamma} + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\right)$ [illegible]

Оценка МНК смещенная (кроме случая $\boldsymbol{\gamma} = \mathbf{0}$ или $\mathbf{X}'\mathbf{Z} = \mathbf{0}$)

$$\mathbf{X}'\mathbf{Z} = \mathbf{0} \Rightarrow \begin{bmatrix} \hat{\boldsymbol{\beta}}^* \\ \hat{\boldsymbol{\gamma}}^* \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}'\mathbf{Z} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}' \\ \mathbf{Z}' \end{bmatrix}\mathbf{y} = \begin{bmatrix} \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y} \\ \left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{y} \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\gamma}} \end{bmatrix}$$

ПРОПУСК СУЩЕСТВЕННЫХ ПЕРЕМЕННЫХ

GeneratedDataProcess $\quad \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}; \quad \mathbf{X} = \left(\mathbf{X}_1, \ldots, \mathbf{X}_k\right); \quad \mathbf{Z} = \left(\mathbf{Z}_{k+1}, \ldots, \mathbf{Z}_{k+l}\right)$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}; \quad \hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}; \quad \mathbf{M} = \mathbf{I} - \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'$$

$$RSS = \mathbf{e}'\mathbf{e} = \left(\mathbf{My}\right)'\left(\mathbf{My}\right) = \mathbf{y}'\mathbf{My} = \left(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}\right)'\mathbf{M}\left(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}\right) =$$

$$= \boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon} + 2\boldsymbol{\gamma}'\mathbf{Z}'\mathbf{M}\boldsymbol{\varepsilon} + \boldsymbol{\gamma}'\mathbf{Z}'\mathbf{MZ}\boldsymbol{\gamma}$$

$\mathrm{E}\left(\dfrac{RSS}{n-k}\right)$ [illegible] $\mathrm{E}\left(\widehat{\sigma^2}\right) =$ [illegible],

т.к. $\mathbf{Z}'\mathbf{MZ}$ симметрична $\Rightarrow$ положительно определена

(даже в случае $\mathbf{X}'\mathbf{Z} = \mathbf{0} \Rightarrow \mathbf{MZ} = \mathbf{Z} \Rightarrow \mathrm{E}\left(\widehat{\sigma^2}\right) = \sigma^2 + \dfrac{1}{n-k}\boldsymbol{\gamma}'\mathbf{Z}'\mathbf{Z}\boldsymbol{\gamma} > \sigma^2$)

$\widehat{\sigma^2}$ смещена $\Rightarrow$ многие тесты некорректны

ПРОПУСК СУЩЕСТВЕННЫХ ПЕРЕМЕННЫХ

$$\text{GeneratedDataProcess}\quad \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon};\quad \mathbf{X} = \left(\mathbf{X}_1, \ldots, \mathbf{X}_k\right);\quad \mathbf{Z} = \left(\mathbf{Z}_{k+1}, \ldots, \mathbf{Z}_{k+l}\right)$$

$$\hat{\boldsymbol{\delta}}^* = \begin{bmatrix} \hat{\boldsymbol{\beta}}^* \\ \hat{\boldsymbol{\gamma}}^* \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X} \\ \mathbf{Z} \end{bmatrix} \mathbf{y};$$

$$\mathrm{Var}\,\hat{\boldsymbol{\delta}}^* = \sigma^2 \left[\left(\mathbf{XZ}\right)' \left(\mathbf{XZ}\right) \right]^{-1} = \sigma^2 \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} \end{bmatrix}^{-1} = \sigma^2 \begin{bmatrix} \mathbf{Q}_{11} & \mathbf{Q}_{12} \\ \mathbf{Q}_{21} & \mathbf{Q}_{22} \end{bmatrix}$$

$$\mathrm{Var}\,\hat{\boldsymbol{\beta}}^* = \sigma^2 \mathbf{Q}_{11} = \sigma^2 \left(\mathbf{X}'\mathbf{X} - \mathbf{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1} \mathbf{Z}'\mathbf{X} \right)^{-1}$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon};\quad \hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}'\mathbf{y};\quad \mathrm{Var}\,\hat{\boldsymbol{\beta}} = \sigma^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1}$$

$$\mathrm{Var}^{-1}\,\hat{\boldsymbol{\beta}} - \mathrm{Var}^{-1}\,\hat{\boldsymbol{\beta}}^* = \frac{1}{\sigma^2} \mathbf{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1} \mathbf{Z}'\mathbf{X} \geq 0,$$

в смысле положительной определенности $\Rightarrow \mathrm{Var}\,\hat{\boldsymbol{\beta}}^* \geq \mathrm{Var}\,\hat{\boldsymbol{\beta}}$

ПРОПУСК СУЩЕСТВЕННЫХ ПЕРЕМЕННЫХ

$$\text{GeneratedDataProcess} \quad \mathbf{y} = \alpha + \mathbf{x}\beta + \mathbf{z}\gamma + \boldsymbol{\varepsilon};$$

$$\widehat{\beta}^{*} = \frac{\operatorname{Cov}(X,Y)\operatorname{Var}(Z)\text{-}\operatorname{Cov}(Z,Y)\operatorname{Cov}(X,Z)}{\operatorname{Var}(X)\operatorname{Var}(Z) - \left[\operatorname{Cov}(X,Z)\right]^2}$$

$$E(\hat{\beta}^{*}) = \beta; \quad \operatorname{Var}\hat{\beta}^{*} = \frac{\sigma^2}{n\operatorname{Var}(X)} \cdot \frac{1}{1 - r_{X,Z}^2}$$

$$\mathbf{y} = \alpha + \mathbf{x}\beta; \quad \hat{\beta} = \frac{\operatorname{Cov}(X,Y)}{\operatorname{Var}(X)};$$

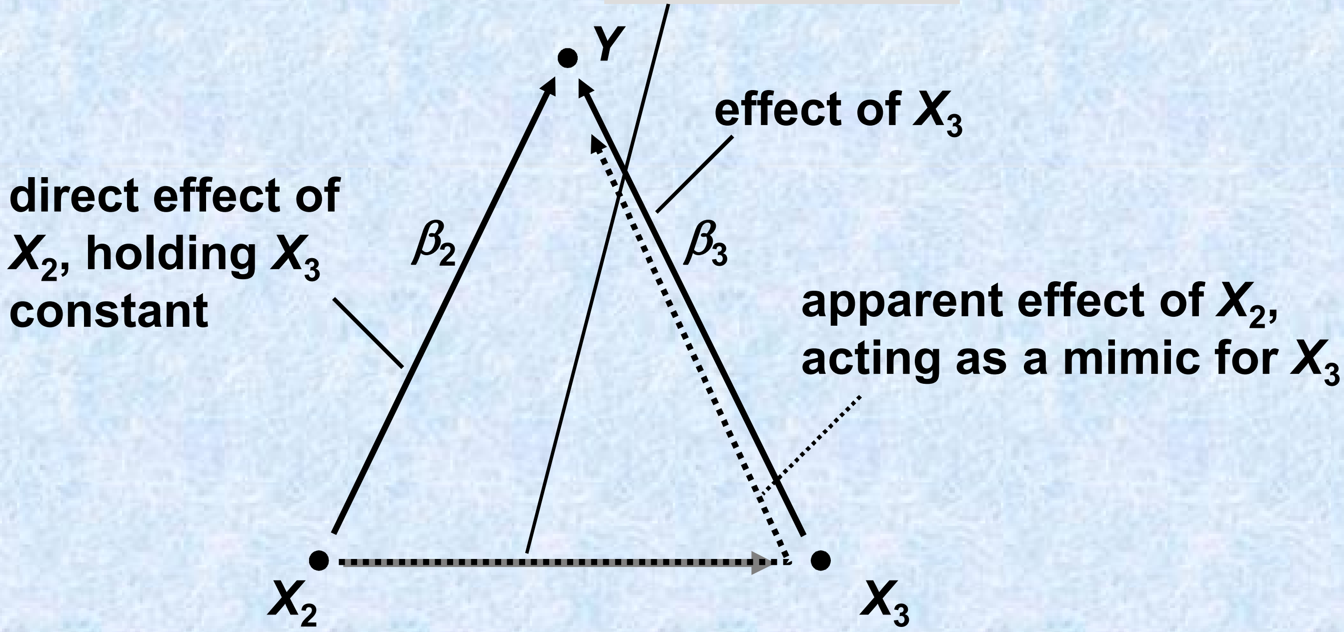
$$E(\hat{\beta}) = \beta + \gamma\frac{\operatorname{Cov}(X,Z)}{\operatorname{Var}(X)}; \quad \operatorname{Var}\hat{\beta} = \frac{\sigma^2}{n\operatorname{Var}(X)}$$

ПРОПУСК СУЩЕСТВЕННЫХ ПЕРЕМЕННЫХ

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$$

$$\hat{Y} = b_1 + b_2 X_2$$

$$E(b_2) = \beta_2 + \beta_3 \boxed{\frac{\text{Cov}(X_2, X_3)}{\text{Var}(X_2)}}$$



The ability of X_2 to mimic X_3 is determined by the slope coefficient obtained when X_3 is regressed on X_2, which of course is Cov(X_2, X_3)/Var(X_2).

ПРОПУСК СУЩЕСТВЕННЫХ ПЕРЕМЕННЫХ

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$$

$$\hat{Y} = b_1 + b_2 X_2$$

$$b_2 = \frac{\mathrm{Cov}(X_2, Y)}{\mathrm{Var}(X_2)} = \frac{\mathrm{Cov}(X_2, [\beta_1 + \beta_2 X_2 + \beta_3 X_3 + u])}{\mathrm{Var}(X_2)}$$

We will now derive the expression for the bias mathematically. Since we are mistakenly fitting a simple regression model, the slope coefficient is calculated as Cov(X_2, Y)/Var(X_2). The first step is to substitute for Y from the true model..

ПРОПУСК СУЩЕСТВЕННЫХ ПЕРЕМЕННЫХ

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$$

$$\hat{Y} = b_1 + b_2 X_2$$

$$b_2 = \frac{\mathrm{Cov}(X_2, Y)}{\mathrm{Var}(X_2)} = \frac{\mathrm{Cov}(X_2, [\beta_1 + \beta_2 X_2 + \beta_3 X_3 + u])}{\mathrm{Var}(X_2)}$$

$$= \frac{\mathrm{Cov}(X_2, \beta_1) + \mathrm{Cov}(X_2, \beta_2 X_2) + \mathrm{Cov}(X_2, \beta_3 X_3) + \mathrm{Cov}(X_2, u)}{\mathrm{Var}(X_2)}$$

We use Covariance Rule 1 to decompose the numerator.

ПРОПУСК СУЩЕСТВЕННЫХ ПЕРЕМЕННЫХ

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$$

$$\hat{Y} = b_1 + b_2 X_2$$

$$b_2 = \frac{\mathrm{Cov}(X_2, Y)}{\mathrm{Var}(X_2)} = \frac{\mathrm{Cov}(X_2, [\beta_1 + \beta_2 X_2 + \beta_3 X_3 + u])}{\mathrm{Var}(X_2)}$$

$$= \frac{\mathrm{Cov}(X_2, \beta_1) + \mathrm{Cov}(X_2, \beta_2 X_2) + \mathrm{Cov}(X_2, \beta_3 X_3) + \mathrm{Cov}(X_2, u)}{\mathrm{Var}(X_2)}$$

$$= \frac{0 + \beta_2 \mathrm{Cov}(X_2, X_2) + \beta_3 \mathrm{Cov}(X_2, X_3) + \mathrm{Cov}(X_2, u)}{\mathrm{Var}(X_2)}$$

The first term is 0 because β_1 is a constant. β_2 and β_3 can be taken out of the second and third terms.

ПРОПУСК СУЩЕСТВЕННЫХ ПЕРЕМЕННЫХ

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$$

$$\hat{Y} = b_1 + b_2 X_2$$

$$b_2 = \frac{\mathrm{Cov}(X_2, Y)}{\mathrm{Var}(X_2)} = \frac{\mathrm{Cov}(X_2, [\beta_1 + \beta_2 X_2 + \beta_3 X_3 + u])}{\mathrm{Var}(X_2)}$$

$$= \frac{\mathrm{Cov}(X_2, \beta_1) + \mathrm{Cov}(X_2, \beta_2 X_2) + \mathrm{Cov}(X_2, \beta_3 X_3) + \mathrm{Cov}(X_2, u)}{\mathrm{Var}(X_2)}$$

$$= \frac{0 + \beta_2 \mathrm{Cov}(X_2, X_2) + \beta_3 \mathrm{Cov}(X_2, X_3) + \mathrm{Cov}(X_2, u)}{\mathrm{Var}(X_2)}$$

$$= \beta_2 + \beta_3 \frac{\mathrm{Cov}(X_2, X_3)}{\mathrm{Var}(X_2)} + \frac{\mathrm{Cov}(X_2, u)}{\mathrm{Var}(X_2)}$$

Hence we have demonstrated that b_2 has three components.

Оценка смещения коэффициента

$$b_1 = \beta_1 + \beta_2 \frac{Cov(x_1, x_2)}{Var(x_1)} + Ошибка выборки$$

$$Смещение = \beta_2 \frac{Cov(x_1, x_2)}{Var(x_1)}$$

Направление смещения зависит от знака истинного значения коэффициента при отсутствующей переменной и выборочной ковариации переменных

ПРОПУСК СУЩЕСТВЕННЫХ ПЕРЕМЕННЫХ

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$$

$$\hat{Y} = b_1 + b_2 X_2$$

$$b_2 = \frac{\mathrm{Cov}(X_2, Y)}{\mathrm{Var}(X_2)} = \beta_2 + \beta_3 \frac{\mathrm{Cov}(X_2, X_3)}{\mathrm{Var}(X_2)} + \frac{\mathrm{Cov}(X_2, u)}{\mathrm{Var}(X_2)}$$

$$E(b_2) = E\left\{\beta_2 + \beta_3 \frac{\mathrm{Cov}(X_2, X_3)}{\mathrm{Var}(X_2)} + \frac{\mathrm{Cov}(X_2, u)}{\mathrm{Var}(X_2)}\right\}$$

ПРОПУСК СУЩЕСТВЕННЫХ ПЕРЕМЕННЫХ

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$$

$$\hat{Y} = b_1 + b_2 X_2$$

$$b_2 = \frac{\mathrm{Cov}(X_2, Y)}{\mathrm{Var}(X_2)} = \beta_2 + \beta_3 \frac{\mathrm{Cov}(X_2, X_3)}{\mathrm{Var}(X_2)} + \frac{\mathrm{Cov}(X_2, u)}{\mathrm{Var}(X_2)}$$

$$E(b_2) = E\left\{\beta_2 + \beta_3 \frac{\mathrm{Cov}(X_2, X_3)}{\mathrm{Var}(X_2)} + \frac{\mathrm{Cov}(X_2, u)}{\mathrm{Var}(X_2)}\right\}$$

$$= E(\beta_2) + E\left\{\beta_3 \frac{\mathrm{Cov}(X_2, X_3)}{\mathrm{Var}(X_2)}\right\} + E\left\{\frac{\mathrm{Cov}(X_2, u)}{\mathrm{Var}(X_2)}\right\}$$

ПРОПУСК СУЩЕСТВЕННЫХ ПЕРЕМЕННЫХ

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$$

$$\hat{Y} = b_1 + b_2 X_2$$

$$b_2 = \frac{\mathrm{Cov}(X_2, Y)}{\mathrm{Var}(X_2)} = \beta_2 + \beta_3 \frac{\mathrm{Cov}(X_2, X_3)}{\mathrm{Var}(X_2)} + \frac{\mathrm{Cov}(X_2, u)}{\mathrm{Var}(X_2)}$$

$$E(b_2) = E\left\{\beta_2 + \beta_3 \frac{\mathrm{Cov}(X_2, X_3)}{\mathrm{Var}(X_2)} + \frac{\mathrm{Cov}(X_2, u)}{\mathrm{Var}(X_2)}\right\}$$

$$= E(\beta_2) + E\left\{\beta_3 \frac{\mathrm{Cov}(X_2, X_3)}{\mathrm{Var}(X_2)}\right\} + E\left\{\frac{\mathrm{Cov}(X_2, u)}{\mathrm{Var}(X_2)}\right\}$$

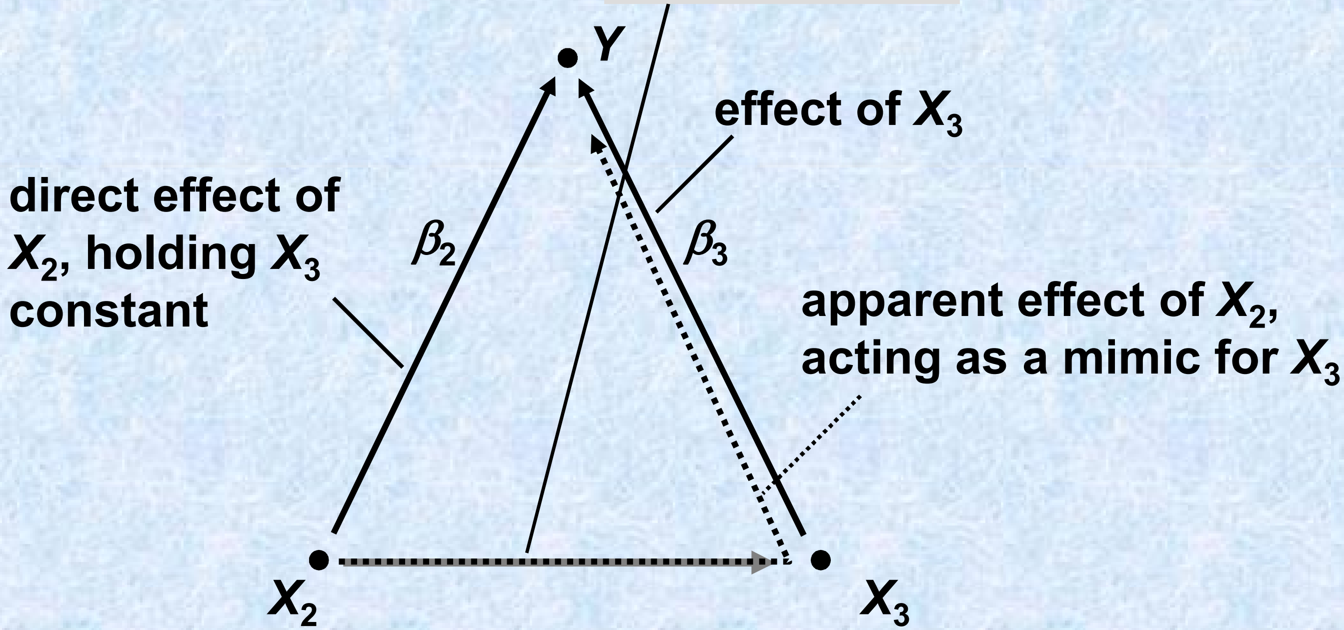
$$= \beta_2 + \beta_3 \frac{\mathrm{Cov}(X_2, X_3)}{\mathrm{Var}(X_2)}$$

ПРОПУСК СУЩЕСТВЕННЫХ ПЕРЕМЕННЫХ

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$$

$$\hat{Y} = b_1 + b_2 X_2$$

$$E(b_2) = \beta_2 + \beta_3 \boxed{\frac{\mathrm{Cov}(X_2, X_3)}{\mathrm{Var}(X_2)}}$$



The ability of X_2 to mimic X_3 is determined by the slope coefficient obtained when X_3 is regressed on X_2, which of course is Cov(X_2, X_3)/Var(X_2).

Последствия невключения в уравнение существенной переменной

1. Уменьшается возможность правильной оценки и интерпретации уравнения

2. Коэффициенты при оставшихся переменных могут оказаться смещенными

3. Их стандартные ошибки, t-статистики и другие показатели качества становятся некорректными и не могут быть использованы для суждения о качестве уравнения

ПРОПУСК СУЩЕСТВЕННЫХ ПЕРЕМЕННЫХ

```
. reg S ASVABC SM

  Source |       SS       df       MS                  Number of obs =     570
---------+------------------------------               F(  2,   567) =  156.81
   Model |  1230.2039      2  615.101949               Prob > F      =  0.0000
Residual | 2224.04347    567  3.92247526               R-squared     =  0.3561
---------+------------------------------               Adj R-squared =  0.3539
   Total | 3454.24737    569  6.07073351               Root MSE      =  1.9805

------------------------------------------------------------------------------
       S |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  ASVABC |   .1381062   .0097494     14.166   0.000       .1189567    .1572556
      SM |    .154783   .0350728      4.413   0.000       .0858946    .2236715
   _cons |   4.791277   .5102431      9.390   0.000        3.78908    5.793475
------------------------------------------------------------------------------
```

ПРОПУСК СУЩЕСТВЕННЫХ ПЕРЕМЕННЫХ

```
. reg S ASVABC SM

  Source |       SS       df       MS                  Number of obs =     570
---------+------------------------------               F(  2,   567) =  156.81
   Model |  1230.2039     2  615.101949               Prob > F      =  0.0000
Residual |  2224.04347   567  3.92247526               R-squared     =  0.3561
---------+------------------------------               Adj R-squared =  0.3539
   Total |  3454.24737   569  6.07073351               Root MSE      =  1.9805

------------------------------------------------------------------------------
       S |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  ASVABC |   .1381062   .0097494     14.166   0.000       .1189567    .1572556
      SM |    .154783   .0350728      4.413   0.000       .0858946    .2236715
   _cons |   4.791277   .5102431      9.390   0.000        3.78908    5.793475
------------------------------------------------------------------------------
```

$$S = \beta_1 + \beta_2 ASVABC + \beta_3 SM + u$$

$$E(b_2) = \beta_2 + \beta_3 \frac{\mathrm{Cov}(ASVABC, SM)}{\mathrm{Var}(ASVABC)}$$

Отсутствие существенной переменной

Направление смещения коэффициента

	$Cov(x_1, x_2)$	
	$+$	$-$
$\beta_2 +$	$+$	$-$
$\beta_2 -$	$-$	$+$

- Ковариация оценивается по выборке
- Знак коэффициента отсутствующей переменной предполагается из теории

ПРОПУСК СУЩЕСТВЕННЫХ ПЕРЕМЕННЫХ

```
. reg S ASVABC SM

  Source |       SS       df       MS                  Number of obs =     570
---------+------------------------------               F(  2,   567) =  156.81
   Model |  1230.2039      2  615.101949               Prob > F      =  0.0000
Residual | 2224.04347    567  3.92247526               R-squared     =  0.3561
---------+------------------------------               Adj R-squared =  0.3539
   Total | 3454.24737    569  6.07073351               Root MSE      =  1.9805

------------------------------------------------------------------------------
       S |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  ASVABC |   .1381062   .0097494     14.166   0.000       .1189567    .1572556
      SM |    .154783   .0350728      4.413   0.000       .0858946    .2236715
   _cons |   4.791277   .5102431      9.390   0.000        3.78908    5.793475
------------------------------------------------------------------------------
```

$$S = \beta_1 + \beta_2 ASVABC + \beta_3 SM + u$$

$$E(b_2) = \beta_2 + \beta_3 \frac{\mathrm{Cov}(ASVABC, SM)}{\mathrm{Var}(ASVABC)}$$

ПРОПУСК СУЩЕСТВЕННЫХ ПЕРЕМЕННЫХ

```
. reg S ASVABC SM

  Source |       SS       df       MS
---------+------------------------------
   Model |  1230.2039      2  615.101949
Residual | 2224.04347    567  3.92247526
---------+------------------------------
   Total | 3454.24737    569  6.07073351

------------------------------------------------------------------------------
       S |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  ASVABC |   .1381062   .0097494     14.166   0.000       .1189567    .1572556
      SM |    .154783   .0350728      4.413   0.000       .0858946    .2236715
   _cons |   4.791277   .5102431      9.390   0.000        3.78908    5.793475
------------------------------------------------------------------------------
```

```
. cor SM ASVABC
(obs=570)

        |       SM   ASVABC
--------+------------------
      SM|   1.0000
  ASVABC|   0.3819   1.0000
```

$$S = \beta_1 + \beta_2 ASVABC + \beta_3 SM + u$$

$$E(b_2) = \beta_2 + \beta_3 \frac{\text{Cov}(ASVABC, SM)}{\text{Var}(ASVABC)}$$

ПРОПУСК СУЩЕСТВЕННЫХ ПЕРЕМЕННЫХ

```
. reg S ASVABC

  Source |       SS       df       MS                  Number of obs =     570
---------+------------------------------               F(  1,   568) =  284.89
   Model |  1153.80864     1  1153.80864               Prob > F      =  0.0000
Residual |  2300.43873   568  4.05006818               R-squared     =  0.3340
---------+------------------------------               Adj R-squared =  0.3329
   Total |  3454.24737   569  6.07073351               Root MSE      =  2.0125

------------------------------------------------------------------------------
       S |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  ASVABC |   .1545378   .0091559     16.879   0.000       .1365543    .1725213
   _cons |   5.770845   .4668473     12.361   0.000       4.853888    6.687803
------------------------------------------------------------------------------
```

ПРОПУСК СУЩЕСТВЕННЫХ ПЕРЕМЕННЫХ

```
. reg S ASVABC SM

------------------------------------------------------------------------------
       S |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  ASVABC |   .1381062   .0097494     14.166   0.000       .1189567    .1572556
      SM |    .154783   .0350728      4.413   0.000       .0858946    .2236715
   _cons |   4.791277   .5102431      9.390   0.000        3.78908    5.793475
------------------------------------------------------------------------------

. reg S ASVABC

------------------------------------------------------------------------------
       S |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  ASVABC |   .1545378   .0091559     16.879   0.000       .1365543    .1725213
   _cons |   5.770845   .4668473     12.361   0.000       4.853888    6.687803
------------------------------------------------------------------------------
```

ПРОПУСК СУЩЕСТВЕННЫХ ПЕРЕМЕННЫХ

```
. reg S SM

  Source |       SS       df       MS                  Number of obs =     570
---------+------------------------------               F(  1,   568) =   83.59
   Model |  443.110436     1  443.110436               Prob > F      =  0.0000
Residual |  3011.13693   568  5.30129742               R-squared     =  0.1283
---------+------------------------------               Adj R-squared =  0.1267
   Total |  3454.24737   569  6.07073351               Root MSE      =  2.3025

------------------------------------------------------------------------------
       S |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
      SM |   .3445198   .0376833      9.142   0.000       .2705041    .4185354
   _cons |   9.506491   .4495754     21.145   0.000       8.623458    10.38952
------------------------------------------------------------------------------
```

$$S = \beta_1 + \beta_2 ASVABC + \beta_3 SM + u$$

$$E(b_3) = \beta_3 + \beta_2 \frac{\mathrm{Cov}(ASVABC, SM)}{\mathrm{Var}(SM)}$$

ПРОПУСК СУЩЕСТВЕННЫХ ПЕРЕМЕННЫХ

```
. reg S ASVABC SM

------------------------------------------------------------------------------
       S |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  ASVABC |   .1381062   .0097494     14.166   0.000       .1189567    .1572556
      SM |    .154783   .0350728      4.413   0.000       .0858946    .2236715
   _cons |   4.791277   .5102431      9.390   0.000        3.78908    5.793475
------------------------------------------------------------------------------

. reg S SM

------------------------------------------------------------------------------
       S |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
      SM |   .3445198   .0376833      9.142   0.000       .2705041    .4185354
   _cons |   9.506491   .4495754     21.145   0.000       8.623458    10.38952
------------------------------------------------------------------------------
```

```
. corr  ASVABC SM,cov
(obs=570)

        |  ASVABC     SM
--------+------------------
ASVABC |   78.654
     SM | 7.48179  6.52508
```

ПРОПУСК СУЩЕСТВЕННЫХ ПЕРЕМЕННЫХ

```
. reg S ASVABC SM
  Source |       SS       df       MS                  Number of obs =     570
---------+------------------------------               F(  2,   567) =  156.81
   Model |   1230.2039     2   615.101949              Prob > F      =  0.0000
Residual |  2224.04347   567   3.92247526              R-squared     =  0.3561
---------+------------------------------               Adj R-squared =  0.3539
   Total |  3454.24737   569   6.07073351              Root MSE      =  1.9805

. reg S ASVABC
  Source |       SS       df       MS                  Number of obs =     570
---------+------------------------------               F(  1,   568) =  284.89
   Model |  1153.80864     1   1153.80864              Prob > F      =  0.0000
Residual |  2300.43873   568   4.05006818              R-squared     =  0.3340
---------+------------------------------               Adj R-squared =  0.3329
   Total |  3454.24737   569   6.07073351              Root MSE      =  2.0125

. reg S SM
  Source |       SS       df       MS                  Number of obs =     570
---------+------------------------------               F(  1,   568) =   83.59
   Model |  443.110436     1   443.110436              Prob > F      =  0.0000
Residual |  3011.13693   568   5.30129742              R-squared     =  0.1283
---------+------------------------------               Adj R-squared =  0.1267
   Total |  3454.24737   569   6.07073351              Root MSE      =  2.3025
```

Does this imply that *ASVABC* explains 33% of the variance in *S* and *SM* 13%? No, because the multiple regression reveals that their joint explanatory power is 0.36, not 0.46.

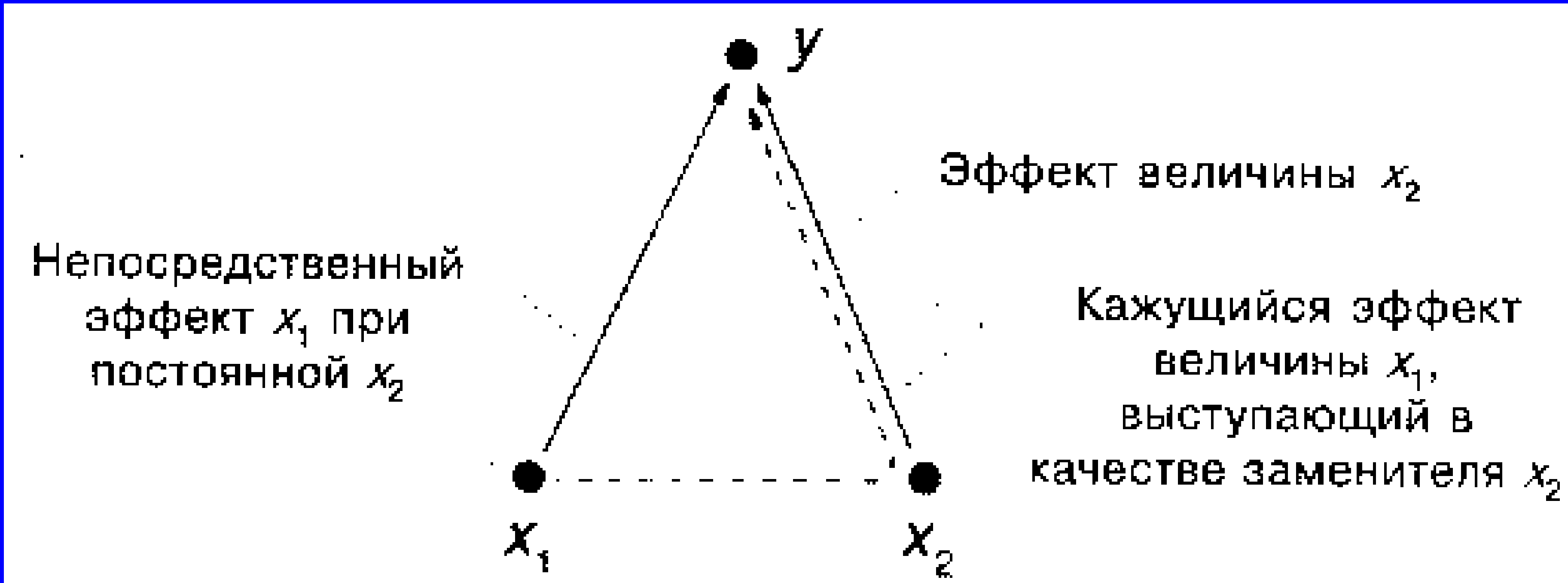
ПРОПУСК СУЩЕСТВЕННЫХ ПЕРЕМЕННЫХ

```
. reg S ASVABC SM
  Source |       SS       df       MS              Number of obs =     570
---------+------------------------------           F(  2,   567) =  156.81
   Model |   1230.2039     2  615.101949           Prob > F      =  0.0000
Residual |  2224.04347   567  3.92247526           R-squared     =  0.3561
---------+------------------------------           Adj R-squared =  0.3539
   Total |  3454.24737   569  6.07073351           Root MSE      =  1.9805

. reg S ASVABC
  Source |       SS       df       MS              Number of obs =     570
---------+------------------------------           F(  1,   568) =  284.89
   Model |  1153.80864     1  1153.80864           Prob > F      =  0.0000
Residual |  2300.43873   568  4.05006818           R-squared     =  0.3340
---------+------------------------------           Adj R-squared =  0.3329
   Total |  3454.24737   569  6.07073351           Root MSE      =  2.0125

. reg S SM
  Source |       SS       df       MS              Number of obs =     570
---------+------------------------------           F(  1,   568) =   83.59
   Model |  443.110436     1  443.110436           Prob > F      =  0.0000
Residual |  3011.13693   568  5.30129742           R-squared     =  0.1283
---------+------------------------------           Adj R-squared =  0.1267
   Total |  3454.24737   569  6.07073351           Root MSE      =  2.3025
```

In the second regression, *ASVABC* is partly acting as a proxy for *SM*, and this inflates its apparent explanatory power. Similarly, in the third regression, *SM* is partly acting as a proxy for *ASVABC*, again inflating its apparent explanatory power.

Отсутствие существенной переменной

Направление смещения коэффициента



Коэффициент детерминации может остаться большим за счет кажущегося эффекта замещающей переменной

ПРОПУСК СУЩЕСТВЕННЫХ ПЕРЕМЕННЫХ

```
. reg LGEARN S MALE

  Source |       SS       df       MS
---------+------------------------------
   Model |   28.951332     2   14.475666
Residual |  124.850561   567  .220194992
---------+------------------------------
   Total |  153.801893   569  .270302096

------------------------------------------------------------------------------
  LGEARN |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
       S |   .0818944   .0079976     10.240   0.000       .0661858     .097603
    MALE |   .2285156   .0397695      5.746   0.000       .1504021    .3066291
   _cons |    1.19254   .1134845     10.508   0.000       .9696386    1.415441
------------------------------------------------------------------------------
```

```
. cor S MALE
(obs=570)

        |        S     MALE
--------+------------------
       S|   1.0000
    MALE|  -0.0581   1.0000
```

$$LGEARN = \beta_1 + \beta_2 S + \beta_3 MALE + u$$

$$E(b_2) = \beta_2 + \beta_3 \frac{\mathrm{Cov}(S, MALE)}{\mathrm{Var}(S)}$$

If we omit *MALE* from the regression, the coefficient of *S* should be subject to a downward bias. β_3 is likely to be positive, and we know that Var(*S*) is positive, but Cov(*S*, *MALE*) is negative because *S* and *MALE* are negatively correlated.

ПРОПУСК СУЩЕСТВЕННЫХ ПЕРЕМЕННЫХ

```
. reg LGEARN S MALE

  Source |       SS       df       MS
---------+------------------------------
   Model |   28.951332     2   14.475666
Residual |  124.850561   567  .220194992
---------+------------------------------
   Total |  153.801893   569  .270302096
```

```
. cor S MALE
(obs=570)

        |        S     MALE
--------+------------------
       S|   1.0000
    MALE|  -0.0581   1.0000
```

```
------------------------------------------------------------------------------
  LGEARN |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
       S |   .0818944   .0079976     10.240   0.000       .0661858     .097603
    MALE |   .2285156   .0397695      5.746   0.000       .1504021    .3066291
   _cons |    1.19254   .1134845     10.508   0.000       .9696386    1.415441
------------------------------------------------------------------------------
```

$$LGEARN = \beta_1 + \beta_2 S + \beta_3 MALE + u$$

$$E(b_3) = \beta_3 + \beta_2 \frac{\mathrm{Cov}(S, MALE)}{\mathrm{Var}(MALE)}$$

ПРОПУСК СУЩЕСТВЕННЫХ ПЕРЕМЕННЫХ

```
. reg LGEARN S MALE
```

LGEARN	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
S	.0818944	.0079976	10.240	0.000	.0661858	.097603
MALE	.2285156	.0397695	5.746	0.000	.1504021	.3066291
_cons	1.19254	.1134845	10.508	0.000	.9696386	1.415441

```
. reg LGEARN S
```

LGEARN	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
S	.0792256	.0082061	9.655	0.000	.0631077	.0953435
_cons	1.358919	.1127785	12.049	0.000	1.137406	1.580433

```
. reg LGEARN MALE
```

LGEARN	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
MALE	.2048652	.0431797	4.744	0.000	.1200538	.2896767
_cons	2.313324	.032605	70.950	0.000	2.249282	2.377365

ПРОПУСК СУЩЕСТВЕННЫХ ПЕРЕМЕННЫХ

```
. reg LGEARN S MALE
  Source |       SS       df       MS              Number of obs =     570
---------+------------------------------           F(  2,   567) =   65.74
   Model |   28.951332     2   14.475666           Prob > F      =  0.0000
Residual |  124.850561   567  .220194992           R-squared     =  0.1882
---------+------------------------------           Adj R-squared =  0.1854
   Total |  153.801893   569  .270302096           Root MSE      =  .46925

. reg LGEARN S
  Source |       SS       df       MS              Number of obs =     570
---------+------------------------------           F(  1,   568) =   93.21
   Model |   21.681253     1   21.681253           Prob > F      =  0.0000
Residual |   132.12064   568   .23260676           R-squared     =  0.1410
---------+------------------------------           Adj R-squared =  0.1395
   Total |  153.801893   569  .270302096           Root MSE      =  .48229

. reg LGEARN MALE
  Source |       SS       df       MS              Number of obs =     570
---------+------------------------------           F(  1,   568) =   22.51
   Model |  5.86288165     1  5.86288165           Prob > F      =  0.0000
Residual |  147.939011   568  .260456005           R-squared     =  0.0381
---------+------------------------------           Adj R-squared =  0.0364
   Total |  153.801893   569  .270302096           Root MSE      =  .51035
```

A comparison of R^2 for the three regressions shows that the sum of R^2 in the simple regressions is actually less than R^2 in the multiple regression. However, in this case the difference is small because the negative correlation between *S* and *MALE* is small (-0.05)

Последствия включения в уравнение несущественной переменной

Для краткости будем называть переменную <u>несущественной</u>, если она не должна быть включена в уравнение (согласно правильной теории)

Будем говорить также о включении лишней переменной в правильное уравнение регрессии

Последствия включения в уравнение несущественной переменной

1. Не теряется возможность правильной оценки и интерпретации уравнения

2. Коэффициенты при прочих переменных остаются несмещенными

3. Стандартные ошибки растут, *t*-статистики уменьшаются, эффективность оценок падает

4. Несущественная переменная может быть значимой, уравнение с ней - давать лучшую оценку

5. Увеличивается риск мультиколлинеарности

ВКЛЮЧЕНИЕ НЕСУЩЕСТВЕННЫХ ПЕРМЕННЫХ

$$\text{GeneratedDataProcess}\quad \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon};\quad \mathbf{X} = \left(\mathbf{X}_1, \ldots, \mathbf{X}_k\right);$$

$$\hat{\boldsymbol{\beta}}^* = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y};\quad \mathrm{Var}\,\hat{\boldsymbol{\beta}}^* = \sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon};\quad \mathbf{Z} = \left(\mathbf{Z}_{k+1}, \ldots, \mathbf{Z}_{k+l}\right);$$

$$\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\gamma}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}' \\ \mathbf{Z}' \end{bmatrix}\mathbf{y} = \begin{bmatrix} \mathbf{Q}_{11} & \mathbf{Q}_{12} \\ \mathbf{Q}_{21} & \mathbf{Q}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{X}' \\ \mathbf{Z}' \end{bmatrix}\mathbf{y} \Rightarrow$$

$$\Rightarrow \hat{\boldsymbol{\beta}} = \left(\mathbf{Q}_{11}\mathbf{X}' + \mathbf{Q}_{12}\mathbf{Z}'\right)\mathbf{y}$$

ВКЛЮЧЕНИЕ НЕСУЩЕСТВЕННЫХ ПЕРМЕННЫХ

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}; \quad \mathbf{Z} = \left(\mathbf{Z}_{k+1}, \ldots, \mathbf{Z}_{k+l}\right); \quad \mathbf{M}_{\mathbf{Z}} = \mathbf{I} - \mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'$$

$$\mathbf{M}_{\mathbf{Z}}\mathbf{y} = \mathbf{M}_{\mathbf{Z}}\mathbf{X}\boldsymbol{\beta} + \mathbf{M}_{\mathbf{Z}}\boldsymbol{\varepsilon}; \quad \hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{M}_{\mathbf{Z}}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{M}_{\mathbf{Z}}\mathbf{y} = \mathbf{R}\mathbf{y}$$

$$\mathrm{E}\,\hat{\boldsymbol{\beta}} = \mathrm{E}\left[\left(\mathbf{X}'\mathbf{M}_{\mathbf{Z}}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{M}_{\mathbf{Z}}\mathbf{X}\boldsymbol{\beta} + \mathbf{e}_{\mathbf{Z}}\right] = \boldsymbol{\beta}$$

$$\mathrm{Var}\,\hat{\boldsymbol{\beta}} = \mathrm{Var}\left(\mathbf{R}\mathbf{y}\right) = \mathbf{R}\left(\mathrm{Var}\,\mathbf{y}\right)\mathbf{R}' = \mathbf{R}\left(\mathrm{Var}\,\boldsymbol{\varepsilon}\right)\mathbf{R}' = \mathbf{R}\left(\sigma^2\mathbf{I}\right)\mathbf{R}' =$$

$$\sigma^2\left(\mathbf{X}'\mathbf{M}_{\mathbf{Z}}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{M}_{\mathbf{Z}}\mathbf{M}'_{\mathbf{Z}}\mathbf{X}\left(\mathbf{X}'\mathbf{M}_{\mathbf{Z}}\mathbf{X}\right)^{-1} = \sigma^2\left(\mathbf{X}'\mathbf{M}_{\mathbf{Z}}\mathbf{X}\right)^{-1} = \sigma^2\left[\mathbf{X}'\mathbf{X} - \mathbf{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{X}\right]^{-1}$$

$$\mathrm{Var}\,\hat{\boldsymbol{\beta}}^* = \sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}$$

$$\mathrm{Var}^{-1}\,\hat{\boldsymbol{\beta}}^* - \mathrm{Var}^{-1}\,\hat{\boldsymbol{\beta}} = \frac{1}{\sigma^2}\mathbf{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{X} \geq 0,$$

т.к. $\mathbf{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{X}$ симметрична $\Rightarrow$ положительно определена $\Rightarrow$

$\Rightarrow \mathrm{Var}\,\hat{\boldsymbol{\beta}} \geq \mathrm{Var}\,\hat{\boldsymbol{\beta}}^*$

ВКЛЮЧЕНИЕ НЕСУЩЕСТВЕННЫХ ПЕРМЕННЫХ

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Z\gamma} + \boldsymbol{\varepsilon}; \quad \mathbf{X} = \left(\mathbf{X}_1, \ldots, \mathbf{X}_k\right); \;\; \mathbf{Z} = \left(\mathbf{Z}_{k+1}, \ldots, \mathbf{Z}_{k+l}\right)$$

$$\mathbf{M} = \mathbf{I} - \mathbf{XZ}\left(\left(\mathbf{XZ}\right)' \mathbf{X}'\mathbf{Z}\right)^{-1} \left(\mathbf{XZ}\right)';$$

$$\mathbf{M}\left[\mathbf{XZ}\right] = \left[\left(\mathbf{MX}\right)\left(\mathbf{MZ}\right)\right] = \mathbf{0} \Rightarrow \mathbf{MX} = \mathbf{0}$$

$$RSS = \mathbf{e}'^{*}\mathbf{e}^{*} = \mathbf{e}'^{*}\mathbf{M}\mathbf{e}^{*} = \left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}^{*} - \mathbf{Z}\hat{\boldsymbol{\gamma}}^{*} + \boldsymbol{\varepsilon}\right)' \mathbf{M}\left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}^{*} - \mathbf{Z}\hat{\boldsymbol{\gamma}}^{*} + \boldsymbol{\varepsilon}\right) =$$

$$= \mathbf{y}'\mathbf{M}\mathbf{y} = \left(\mathbf{X\beta} + \boldsymbol{\varepsilon}\right)' \mathbf{M}\left(\mathbf{X\beta} + \boldsymbol{\varepsilon}\right) = \boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon}$$

$$\mathrm{E}\left(\frac{RSS}{n-k-l}\right) = \mathrm{E}\left(s^2\right) = \mathrm{E}\left(\widehat{\sigma^2}\right) = \sigma^2 \text{ несмещена}$$

Последствия включения в уравнение несущественной переменной

1. Не теряется возможность правильной оценки и интерпретации уравнения

2. Коэффициенты при прочих переменных остаются несмещенными

3. Стандартные ошибки растут, *t*-статистики уменьшаются, эффективность оценок падает

4. Несущественная переменная может быть значимой, уравнение с ней - давать лучшую оценку

5. Увеличивается риск мультиколлинеарности

ВКЛЮЧЕНИЕ НЕСУЩЕСТВЕННЫХ ПЕРМЕННЫХ

```
. reg LGEARN S ASVABC

  Source |       SS       df       MS                  Number of obs =     570
---------+------------------------------               F(  2,   567) =   57.45
   Model |  25.9166749     2  12.9583374               Prob > F      =  0.0000
Residual |  127.885218   567  .225547121               R-squared     =  0.1685
---------+------------------------------               Adj R-squared =  0.1656
   Total |  153.801893   569  .270302096               Root MSE      =  .47492

------------------------------------------------------------------------------
  LGEARN |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
       S |   .0544266   .0099018      5.497   0.000        .034978    .0738753
  ASVABC |   .0114733   .0026476      4.333   0.000       .0062729    .0166736
   _cons |   1.118832    .124107      9.015   0.000       .8750665    1.362598
------------------------------------------------------------------------------
```

ВКЛЮЧЕНИЕ НЕСУЩЕСТВЕННЫХ ПЕРМЕННЫХ

```
. reg LGEARN S ASVABC SM SF

  Source |       SS       df       MS                  Number of obs =     570
---------+------------------------------               F(  4,   565) =   29.22
   Model |  26.3617806     4  6.59044515               Prob > F      =  0.0000
Residual |  127.440112   565  .22555772                R-squared     =  0.1714
---------+------------------------------               Adj R-squared =  0.1655
   Total |  153.801893   569  .270302096               Root MSE      =  .47493

------------------------------------------------------------------------------
  LGEARN |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
       S |   .0511811   .0101812      5.027   0.000       .0311835    .0711788
  ASVABC |    .010444   .0027481      3.800   0.000       .0050463    .0158417
      SM |   .0071835   .0102695      0.699   0.485      -.0129876    .0273547
      SF |    .004794   .0076389      0.628   0.531      -.0102101    .0197981
   _cons |   1.073972   .1324621      8.108   0.000       .8137933     1.33415
------------------------------------------------------------------------------
```

ВКЛЮЧЕНИЕ НЕСУЩЕСТВЕННЫХ ПЕРМЕННЫХ

```
. reg LGEARN S ASVABC

------------------------------------------------------------------------------
  LGEARN |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
       S |   .0544266   .0099018      5.497   0.000        .034978    .0738753
  ASVABC |   .0114733   .0026476      4.333   0.000       .0062729    .0166736
   _cons |   1.118832    .124107      9.015   0.000       .8750665    1.362598
------------------------------------------------------------------------------

. reg LGEARN S ASVABC SM SF

------------------------------------------------------------------------------
  LGEARN |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
       S |   .0511811   .0101812      5.027   0.000       .0311835    .0711788
  ASVABC |    .010444   .0027481      3.800   0.000       .0050463    .0158417
      SM |   .0071835   .0102695      0.699   0.485      -.0129876    .0273547
      SF |    .004794   .0076389      0.628   0.531      -.0102101    .0197981
   _cons |   1.073972   .1324621      8.108   0.000       .8137933     1.33415
------------------------------------------------------------------------------
```

ВКЛЮЧЕНИЕ НЕСУЩЕСТВЕННЫХ ПЕРМЕННЫХ

```
. reg LGEARN S ASVABC

------------------------------------------------------------------------------
  LGEARN |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
       S |   .0544266   .0099018      5.497   0.000        .034978    .0738753
  ASVABC |   .0114733   .0026476      4.333   0.000       .0062729    .0166736
   _cons |   1.118832    .124107      9.015   0.000       .8750665    1.362598
------------------------------------------------------------------------------
```

```
. reg LGEARN S ASVABC SM SF

------------------------------------------------------------------------------
  LGEARN |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
       S |   .0511811   .0101812      5.027   0.000       .0311835    .0711788
  ASVABC |    .010444   .0027481      3.800   0.000       .0050463    .0158417
      SM |   .0071835   .0102695      0.699   0.485      -.0129876    .0273547
      SF |    .004794   .0076389      0.628   0.531      -.0102101    .0197981
   _cons |   1.073972   .1324621      8.108   0.000       .8137933     1.33415
------------------------------------------------------------------------------
```

The standard errors are larger in the misspecified model, reflecting the loss of efficiency.

ВКЛЮЧЕНИЕ НЕСУЩЕСТВЕННЫХ ПЕРМЕННЫХ

```
. reg LGEARN S ASVABC

------------------------------------
  LGEARN |      Coef.   Std. Err.
---------+--------------------------
       S |   .0544266   .0099018
  ASVABC |   .0114733   .0026476
   _cons |   1.118832    .124107
------------------------------------
```

```
. cor S ASVABC SM SF
(obs=570)

        |        S   ASVABC       SM       SF
--------+------------------------------------
       S|   1.0000
  ASVABC|   0.5779   1.0000
      SM|   0.3582   0.3819   1.0000
      SF|   0.4066   0.4179   0.6391   1.0000
```

```
. reg LGEARN S ASVABC SM SF

------------------------------------
  LGEARN |      Coef.   Std. Err.
---------+--------------------------
       S |   .0511811   .0101812
  ASVABC |    .010444   .0027481
      SM |   .0071835   .0102695
      SF |    .004794   .0076389
   _cons |   1.073972   .1324621
------------------------------------
```

. pcorr SM S ASVABC SF
(obs=570)

Partial correlation of SM with

Variable	Corr.	Sig.
S	0.1277	0.002
ASVABC	0.0805	0.055
SF	0.4926	0.000

. pcorr SF S ASVABC SM
(obs=570)

Partial correlation of SFwith

Variable	Corr.	Sig.
S	0.1454	0.001
ASVABC	0.1423	0.001
SM	0.4926	0.000

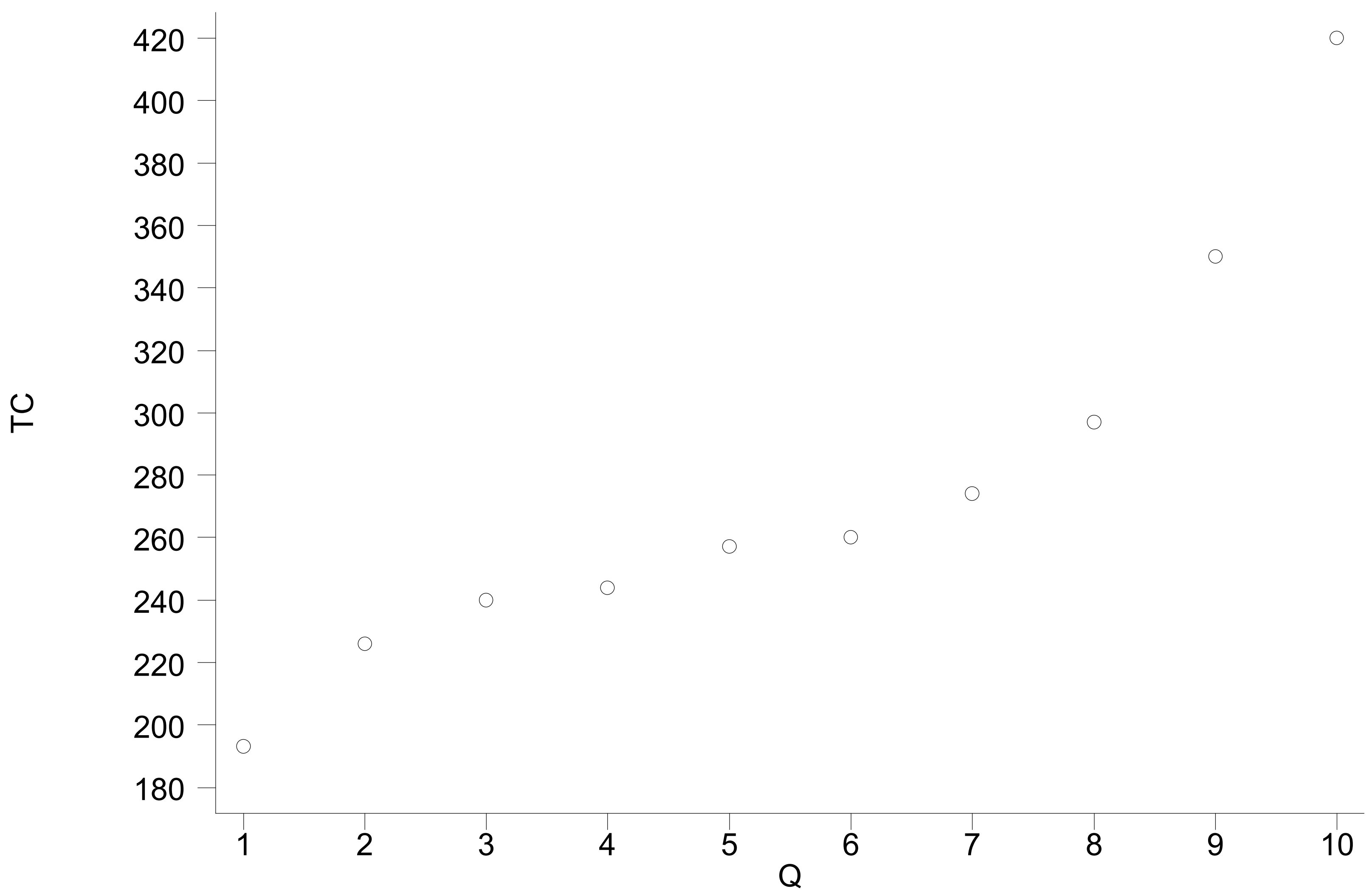
However, the loss of efficiency is not very great. The parental education variables are correlated with both *S* and *ASVABC* but, with a sample as large as the present one, the correlation has to be greater for the loss of efficiency to become a serious problem.

ОШИБОЧНАЯ СПЕЦИФИКАЦИЯ МОДЕЛИ

Consequences of Variable Misspecification

		True Model	
		$Y = \beta_1 + \beta_2 X_2 + u$	$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$
Fitted Model	$\hat{Y} = b_1 + b_2 X_2$	**Correct specification, no problems**	**Coefficients are biased (in general). Standard errors are invalid.**
	$\hat{Y} = b_1 + b_2 X_2 + b_3 X_3$	**Coefficients are unbiased (in general), but inefficient. Standard errors are valid (in general)**	**Correct specification, no problems**

ОГРАНИЧЕНИЯ В ОБОБЩЕННОЙ МОДЕЛИ

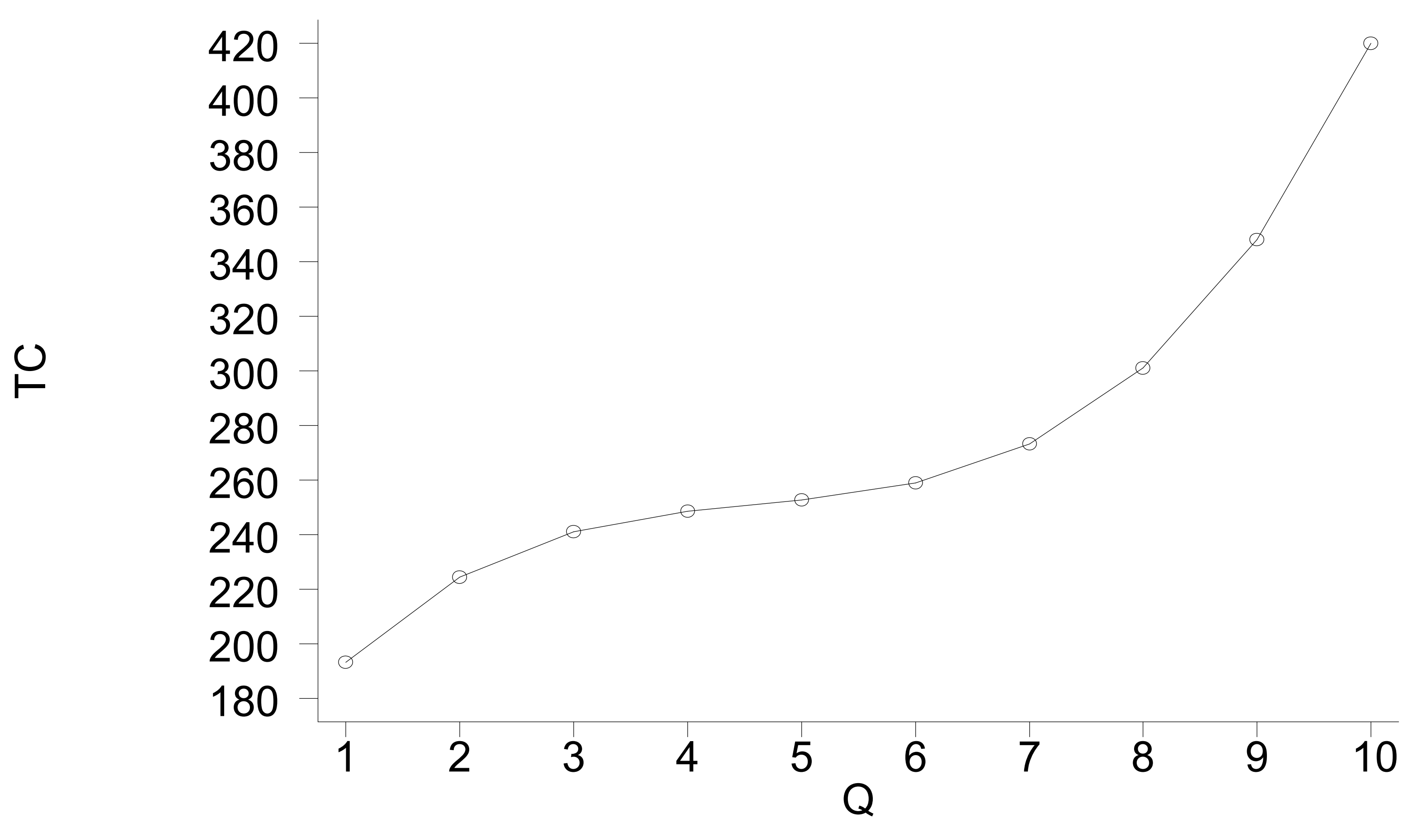


ОГРАНИЧЕНИЯ В ОБОБЩЕННОЙ МОДЕЛИ

```
. reg  TC Q
  Source |       SS            df       MS              Number of obs =      10
-------------+------------------------------           F(  1,     8) =   42.28
     Model |  32780.3667     1  32780.3667              Prob > F      =  0.0002
  Residual |  6202.53333     8  775.316667              R-squared     =  0.8409
-------------+------------------------------           Adj R-squared =  0.8210
     Total |    38982.90     9  4331.43333              Root MSE      =  27.845

------------------------------------------------------------------------------
        TC |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         Q |   19.93333    3.06558     6.50   0.000     12.86409    27.00257
     _cons |   166.4667   19.02142     8.75   0.000     122.6032    210.3301
------------------------------------------------------------------------------

. ovtest
Ramsey RESET test using powers of the fitted values of TC
       Ho:  model has no omitted variables
                  F(3, 5) =    165.83
                  Prob > F =      0.0000
```

ОГРАНИЧЕНИЯ В ОБОБЩЕННОЙ МОДЕЛИ

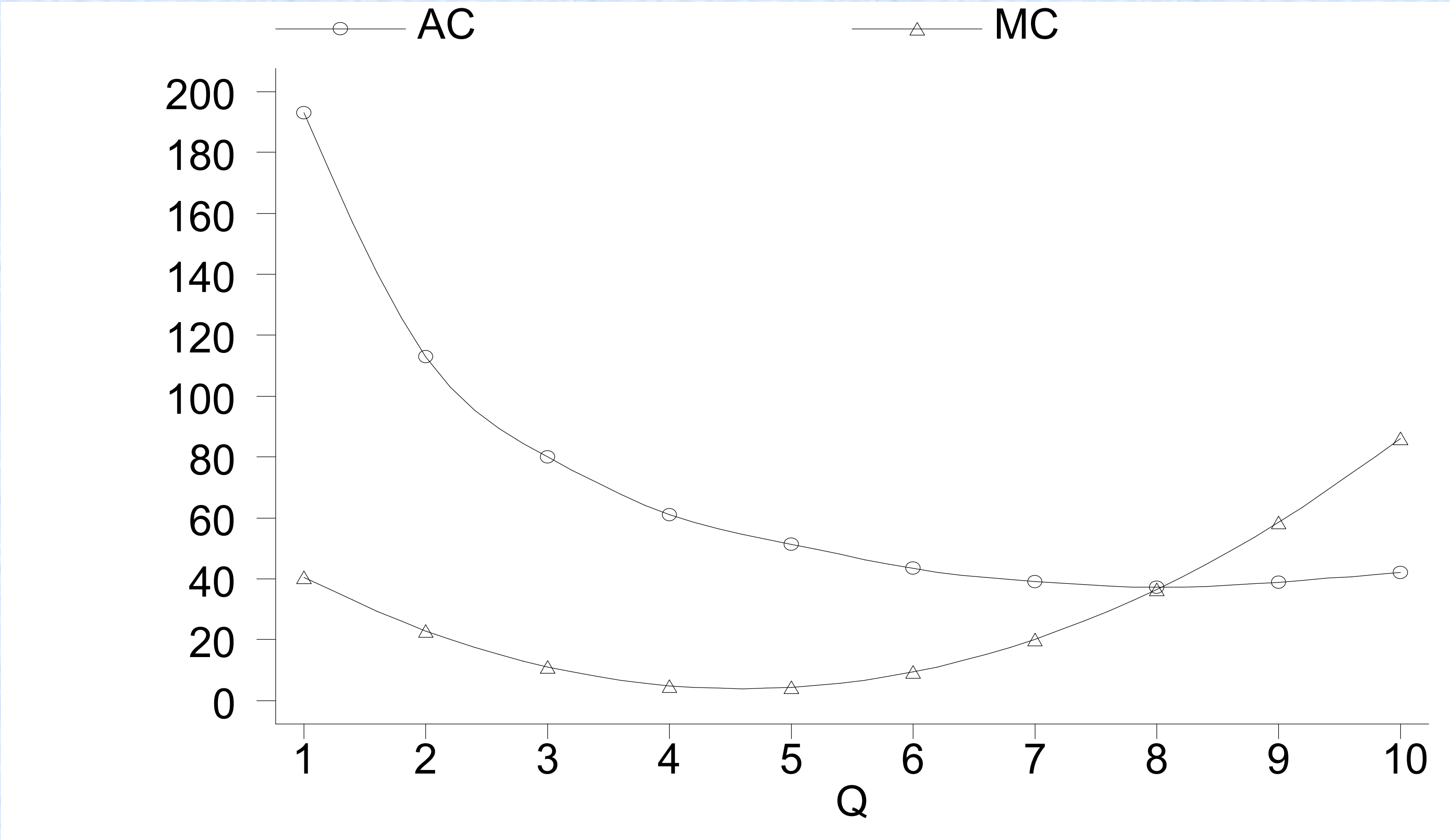
$$TC(Q) = b_0 + b_1 Q + b_2 Q^2 + b_3 Q^3$$



$$0 \le TC(0) = b_0; \qquad b_3 Q^3 \underset{Q \to +\infty}{\sim} TC(Q) \approx +\infty \Rightarrow b_3 \ge 0$$

ОГРАНИЧЕНИЯ В ОБОБЩЕННОЙ МОДЕЛИ

$$MC\left(Q\right)=b_1+2b_2Q+3b_3Q^2 \qquad AC\left(Q\right)=\frac{b_0}{Q}+b_1+b_2Q+b_3Q^2$$



$$0\le MC\left(0\right)=b_1; \qquad MC\left(Q\right)\ge 0 \Rightarrow b_2^2-3b_1b_3\le 0$$

ОГРАНИЧЕНИЯ В ОБОБЩЕННОЙ МОДЕЛИ

$$TC\left(Q\right) = b_0 + b_1 Q + b_2 Q^2 + b_3 Q^3 + \varepsilon$$

```
. reg  TC  Q3 Q2 Q

      Source |       SS           df       MS              Number of obs =      10
-------------+------------------------------              F(  3,     6) = 1202.22
       Model |  38918.1562     3  12972.7187           Prob > F      =  0.0000
    Residual |  64.7438228     6  10.7906371          R-squared     =  0.9983
-------------+------------------------------              Adj R-squared =  0.9975
       Total |    38982.90     9  4331.43333           Root MSE      =  3.2849

------------------------------------------------------------------------------
          TC |      Coef.       Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          Q3 |   .9395882   .0591056    15.90   0.000      .794962    1.084214
          Q2 |  -12.96154   .9856646   -13.15   0.000    -15.37337    -10.5497
           Q |   63.47766   4.778607    13.28   0.000     51.78483    75.17049
       _cons |   141.7667   6.375322    22.24   0.000     126.1668    157.3665
------------------------------------------------------------------------------
```

```
. di _b[ Q2]^2-3*_b[Q]*_b[Q3]
-10.927103

. testnl _b[ Q2]^2=3*_b[Q]*_b[Q3]

(1)  _b[ Q2]^2 = 3*_b[Q]*_b[Q3]

               F(1, 6) =       35.31
           Prob > F =        0.0010
```

```
. ovtest
Ramsey RESET test using powers of the
fitted values of TC
      Ho:  model has no omitted variables
              F(3, 3) =      1.76
           Prob > F =      0.3270
```

Конец лекции