meet_01_21_march_2016

26 марта 2016 г.

1 Вариант 2015

1.1 Задача 1

Линеаризуем (делим на x_i и переворачиваем):

$$\frac{x_i}{y_i} = \theta_0 + \theta_1 x_i + u_i$$

Боремся с гетероскедастичностью:

$$\frac{1}{y_i} = \theta_1 + \theta_0 \frac{1}{x_i} + \nu_i$$

Применяем обычные формулы для θ_0 и θ_1:

$$\hat{\theta}_0 = 1,5$$

$$\hat{\theta}_1 = -0,5$$

1.2 Задача 2

1.2.1 1

Занумеруем Ивана Петрова первым номером.

Замечаем, что $\widehat{\theta}_1$ входит только в первое слагаемое при минимизации RSS:

$$RSS = (y_1 - \hat{\theta}_0 - \hat{\theta}_1 - \hat{\theta}_2 w_1)^2 + (y_2 - \hat{\theta}_0 - \hat{\theta}_2 w_2)^2 + (y_3 - \hat{\theta}_0 - \hat{\theta}_2 w_3)^2 + \ldots$$

Отсюда $\hat{u}_1 = 0$, и

$$\hat{\theta}_1 = y_1 - \hat{\theta}_0 - \hat{\theta}_2 w_1$$

Находим $\hat{\theta}_0$ и $\hat{\theta}_2$ для парной регрессии по всем наблюдениям кроме Ивана.

В новой регрессии $\sum\limits_{i=2}^{100} y_i = 481.2$, $\sum\limits_{i=2}^{100} w_i = 98$, $\sum\limits_{i=2}^{100} w_i^2 = 196$, $\sum\limits_{i=2}^{100} w_i y_i = 862.4$.

$$\hat{\theta}_2 = \frac{\sum\limits_{i=2}^{100} w_i y_i - \bar{y}\bar{w}n}{\sum\limits_{i=2}^{100} w_i^2 - \bar{w}^2 n} = 3.9 \hat{\theta}_0 = 0.9994 = 1 \hat{\theta}_1 = 9.9994 = 10$$

Тестирование гипотезы (три способа).

1. Способ через разложение $\hat{\theta}_1$:

$$Var(\hat{\theta}_1) = Var(y_1 - \hat{\theta}_0 - \hat{\theta}_2 w_1) = Var(y_1) + Var(\hat{\theta}_0 + \hat{\theta}_2 w_1),$$

т.к. оценки $\hat{\theta}_0$ и $\hat{\theta}_2$ получены без Ивана Петрова.

$$\hat{\mathbb{V}}\mathrm{ar}(y_1) = se^2(u_i) = \frac{RSS_{UR}}{n-k} = \frac{380.24}{97} = 3.92$$

Заметим, что в RSS_{UR} слагаемое $(y_1 - \hat{\theta}_0 - \hat{\theta}_1 - \hat{\theta}_2 w_1)^2$ равно нулю, значит, в модели без Ивана Петрова RSS будет таким же. Оценки коэффициентов в двух моделях будут одинаковыми, а вот дисперсии — разными.

Матрица $W'W$ для 99 наблюдений:

$$W'W = \begin{pmatrix} 99 & 98 \\ 98 & 196 \end{pmatrix}$$

$$(W'W)^{-1} = \begin{pmatrix} 0.02 & -0.01 \\ -0.01 & 0.0101 \end{pmatrix}$$

$$\widehat{\mathbb{V}\mathrm{ar}}(\widehat{\theta}_0 + \widehat{\theta}_2 w_1) = 0.02 \cdot se^2(u_i) + 0.0101 \cdot 4 \cdot se^2(u_i) + 2 \cdot (-0.01) \cdot 2 \cdot se^2(u_i) = 0.08$$

$$se^2(u_i) = 4$$

$\frac{10}{2} = 5 \Rightarrow$ гипотеза о незначимости отвергается на любом разумном уровне значимости

2. Способ через VIF:

$$se^2(\hat{\theta}_1) = VIF_x \cdot \frac{RSS/(n-k)}{TSS_x} = \frac{\hat{\sigma}^2}{RSS_x}$$

$$RSS_x = \hat{\varepsilon}'\hat{\varepsilon} = x'(I - H_W)x = x'x - x'H_W x$$

Здесь другая матрица $W'W$, потому что по 100 наблюдениям:

$$W'W = \begin{pmatrix} 100 & 100 \\ 100 & 200 \end{pmatrix}$$

$$(W'W)^{-1} = \begin{pmatrix} 0.02 & -0.01 \\ -0.01 & 0.01 \end{pmatrix}$$

$$W'x = \begin{pmatrix} \sum x_i \\ \sum w_i x_i \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

$$x'x - x'H_W x = 1 - \frac{2}{100} = \frac{98}{100}$$

$$se^2(\hat{\theta}_1) = \frac{380.24/97}{98/100} = 4$$

$\frac{10}{2} = 5 \Rightarrow$ гипотеза о незначимости отвергается на любом разумном уровне значимости

3. Способ через RSS_R и RSS_{UR}:

$$RSS_R = \sum_{i=1}^{100}(y_i - \widehat{\theta}_0 - \widehat{\theta}_2 w_i)^2 = y'(I-H)y = y'y - y'Hy$$

$$H = W(W'W)^{-1}W', \text{ где в матрице W 100 наблюдений}$$

$$W'y = \begin{pmatrix} \sum y_i \\ \sum w_i y_i \end{pmatrix} = \begin{pmatrix} 500 \\ 900 \end{pmatrix}$$

$$W'W = \begin{pmatrix} 100 & 100 \\ 100 & 200 \end{pmatrix}$$

$$(W'W)^{-1} = \begin{pmatrix} 0.02 & -0.01 \\ -0.01 & 0.01 \end{pmatrix}$$

$$y'Hy = 4100$$

Найдём $\sum y_i^2$, раскрыв скобки в RSS_{UR}:

$$RSS_{UR} = 0 + \sum_{i=2}^{100}(y_i - \hat{\theta}_0 - \hat{\theta}_2 w_i)^2 = \sum_{i=2}^{100} y_i^2 - 2\hat{\theta}_0 \sum_{i=2}^{100} y_i - 2\hat{\theta}_2 \sum_{i=2}^{100} w_i y_i + 99\hat{\theta}_0^2 + \hat{\theta}_2^2 \sum_{i=2}^{100} w_i^2 + 2\hat{\theta}_0\hat{\theta}_2 \sum_{i=2}^{100} w_i,$$

откуда

$$\sum_{i=2}^{100} y_i^2 = 4224,8$$

Получаем:

$$RSS_R = 4224,8 + 18,8^2 - 4100 = 478,24$$

F-статистика:

$$\frac{(RSS_R - RSS_{UR})/1}{RSS_{UR}/(n-k)} = 25 \Rightarrow \text{гипотеза о незначимости отвергается на любом разумном уровне значимости}$$

Заодно мы на практике убеждаемся в том, что при проверке гипотезе об одном линейном ограничении значения F-статистики и t-статистики связано соотношением $F = t^2$.

1.2.2 2

- Если на первом шаге 2МНК использовать только z, то получается коэффициент 0 и второй шаг 2МНК невозможен. (похоже авторы задания имели ввиду этот вариант)
- Если на первом шаге 2МНК использовать 1 или w или (1 и w), то оценки первого шага мы можем получить, но второй шаг невозможен (из-за строгой коллинеарности)
- Если на первом шаге 2МНК использовать 1 и z, то недостаточно информации, чтобы получить оценки первого шага, но если бы она была, то всё было бы ок.

Остаётся в качестве инструментов использовать только константу, z_i и w_i вместе.

Для случая, когда общее число инструментальных переменных равно общему числу эндогенных переменных

$$\hat{\beta}_{IV} = (Z'X)^{-1}Z'y$$

Доказательство формулы:

$$\hat{X} = Z(Z'Z)^{-1}Z'X$$

и

$$H_Z = Z(Z'Z)^{-1}Z'$$

$$\hat{\beta}_{IV} = (\hat{X}'\hat{X})^{-1}\hat{X}'y = (X'H_ZX)^{-1}X'$$

Пусть $H_ZX = ZA$, тогда

$$\hat{\beta}_{IV} = (X'H_ZX)^{-1}(H_ZX)'y = ((ZA)'X)^{-1}(ZA)'y = (Z'X)^{-1}Z'y$$

$$A = \begin{pmatrix} 1 & 0 & a_{13} \\ 0 & 1 & a_{23} \\ 0 & 0 & a_{33} \end{pmatrix}$$

$$X = \begin{pmatrix} 1 & w_1 & x_1 \\ . & . & . \\ . & . & . \\ 1 & w_n & x_n \end{pmatrix},\ Z = \begin{pmatrix} 1 & w_1 & z_1 \\ . & . & . \\ . & . & . \\ 1 & w_n & z_n \end{pmatrix}$$

$$Z'X = \begin{pmatrix} n & \sum\limits_{i=1}^{100} w_i & \sum\limits_{i=1}^{100} x_i \\ \sum\limits_{i=1}^{100} w_i & \sum\limits_{i=1}^{100} w_i^2 & \sum\limits_{i=1}^{100} x_i w_i \\ \sum\limits_{i=1}^{100} z_i & \sum\limits_{i=1}^{100} w_i z_i & \sum\limits_{i=1}^{100} x_i z_i \end{pmatrix} = \begin{pmatrix} 100 & 100 & 1 \\ 100 & 200 & 2 \\ 100 & 100 & 0 \end{pmatrix}$$

$$(Z'X)^{-1} = \begin{pmatrix} \frac{2}{100} & -\frac{1}{100} & 0 \\ -\frac{2}{100} & \frac{1}{100} & \frac{1}{100} \\ 1 & 0 & -1 \end{pmatrix}$$

$$Z'y = \begin{pmatrix} 500 \\ 900 \\ 500 \end{pmatrix}$$

$$\widehat{\beta}_{IV} = \begin{pmatrix} 1 \\ 4 \\ 0 \end{pmatrix}$$

Получаем, что гипотеза о незначимости не отвергается.

1.2.3 3

Если опечатка и подразумевалось: $y_i = \theta_0 + \theta_1 x_i + \theta_2 w_i + \theta_3 x_i w_i + u_i$, то жесткая мультиколлинеарность так как $w_i x_i = 2x_i$.

Если нет опечатки:

$$RSS = (y_1 - \hat{\theta}_0 - \hat{\theta}_1 - \hat{\theta}_2 w_1 - \hat{\theta}_2 w_1)^2 + (y_2 - \hat{\theta}_0 - \hat{\theta}_2 w_2)^2 + (y_3 - \hat{\theta}_0 - \hat{\theta}_2 w_3)^2 + \dots$$

Аналогично, первое слагаемое всегда 0, оценки $\hat{\theta}_0$ и $\hat{\theta}_2$ такие же,

$$\hat{\theta}_1 = 2.2$$

RSS_{UR} будет таким же, потому что первое слагаемое опять будет равно нулю, а все остальные не изменятся.

Значимость переменной проверим через разложение:

$$\mathbb{V}\mathrm{ar}(\hat{\theta}_1) = \mathbb{V}\mathrm{ar}(y_1 - \hat{\theta}_0 - 2 \cdot \hat{\theta}_2 w_1) = \mathbb{V}\mathrm{ar}(y_1) + \mathbb{V}\mathrm{ar}(\widehat{\theta}_0 + 2 \cdot \hat{\theta}_2 w_1)$$

$$se^2(u_i) = \frac{RSS_{UR}}{n-k} = \frac{380.24}{100-4} = 3.96$$

По аналогии с п.1 находим дисперсии оценок $\widehat{\theta}_0$ и $\widehat{\theta}_2$ и получаем:

$$\widehat{\mathbb{V}\mathrm{ar}}(\hat{\theta}_1) = se^2(u_i) + 0.02 \cdot se^2(u_i) + 4 \cdot 4 \cdot 0.0101 \cdot se^2(u_i) + 2 \cdot 2 \cdot 2 \cdot (-0.01) \cdot se^2(u_i) = 4.3633$$

$$\frac{2.2}{2.09} = 1.053 \Rightarrow \text{гипотеза о незначимости не отвергается}$$