Решение варианта 1-го тура III Студенческой универсиады по эконометрике МГУ им. М.В.Ломоносова 2014 года

Уракова Екатерина

Задача 2.

а) $\varepsilon = \frac{\partial PRICE}{\partial DIST} * \frac{DIST}{PRICE}$ – искомая эластичность.

Нам неизвестно $\frac{\partial PRICE}{\partial DIST}$, поэтому постараемся его вычислить. С одной стороны,

$$\frac{\partial \,{}^{PRICE}/_{TOTSP}}{\partial DIST} = \frac{\frac{\partial PRICE}{\partial DIST} * TOTSP - \frac{\partial TOTSP}{\partial DIST} * PRICE}{TOTSP^2}$$
$$= \frac{\partial PRICE}{\partial DIST} * \frac{1}{TOTSP} - \frac{\partial TOTSP}{\partial DIST} * \frac{PRICE}{TOTSP^2} = \left\{\frac{\partial TOTSP}{\partial DIST} = 0\right\}$$
$$= \frac{\partial PRICE}{\partial DIST} * \frac{1}{TOTSP}.$$

С другой стороны, из таблицы 2.2 видим, что $\frac{\partial \,{}^{PRICE}/_{TOTSP}}{\partial DIST} = -1{,}3527 * \frac{1}{DIST}$.

Следовательно, $\frac{\partial PRICE}{\partial DIST} = -1{,}3527 * \frac{TOTSP}{DIST}$.

Тогда искомая эластичность равна $\varepsilon = \frac{\partial PRICE}{\partial DIST} * \frac{DIST}{PRICE} = -1{,}3527 * \frac{TOTSP}{DIST} * \frac{DIST}{PRICE} =$ $\{$используем средние значения из таблицы 2.1$\} = -0{,}247$.

б) $\varepsilon = \frac{\partial PRICE}{\partial LIVESP} * \frac{LIVESP}{PRICE}$ – искомая эластичность. С одной стороны,

$$\frac{\partial \,{}^{PRICE}/_{TOTSP}}{\partial LIVESP} = \frac{\frac{\partial PRICE}{\partial LIVESP} * TOTSP - \frac{\partial TOTSP}{\partial LIVESP} * PRICE}{TOTSP^2} = \left\{\frac{\partial TOTSP}{\partial LIVESP} = 1\right\}$$
$$= \frac{\partial PRICE}{\partial LIVESP} * \frac{1}{TOTSP} - \frac{PRICE}{TOTSP^2}.$$

С другой стороны, $\frac{\partial \,{}^{PRICE}/_{TOTSP}}{\partial LIVESP} = \frac{-0{,}903}{TOTSP} * \frac{\partial TOTSP}{\partial LIVESP} = \frac{-0{,}903}{TOTSP}$. Тогда

$$\frac{\partial PRICE}{\partial LIVESP} * \frac{1}{TOTSP} - \frac{PRICE}{TOTSP^2} = \frac{-0{,}903}{TOTSP}$$
$$\frac{\partial PRICE}{\partial LIVESP} = \frac{PRICE}{TOTSP} - 0{,}903.$$
$$\varepsilon = \frac{\partial PRICE}{\partial LIVESP} * \frac{LIVESP}{PRICE} = \frac{PRICE}{TOTSP} * \frac{LIVESP}{PRICE} - 0{,}903 * \frac{LIVESP}{PRICE}$$
$$= \{\text{используем средние значения из таблицы 2.1}\} = 0{,}421.$$

в) **(1-WALK)*METRDIST**: Если до метро на транспорте, то при прочих равных условиях с ростом расстояния до метро на 1 минуту стоимость 1 м^2 падает на 0,058 тыс.долл.

WALK*METRDIST: Если до метро пешком, то при прочих равных условиях с ростом расстояния до метро на 1 минуту стоимость 1 м^2 падает на 0,038 тыс.долл.

LOG(DIST): при прочих равных условиях, при увеличении расстояния до центра на 1% стоимость 1 м^2 падает на 0,0135 тыс.долл.

FLOOR1: при прочих равных условиях стоимость 1 м^2 квартиры, расположенной на 1м этаже, ниже стоимости 1 м^2 квартиры, расположенной не на 1м этаже, на 0,412 тыс.долл.

LOG(TOTSP): при прочих равных устовиях при увеличении общей площади квартиры на 1%, стоимость 1 м^2 уменьшаетсыя на 0,009 тыс.долл.

г) $\frac{PRICE}{TOTSP} = 9{,}77 + 0{,}37 * KITSP - 0{,}01KITSP^2 \to \max(KITSP)$

$$\frac{\partial\, {}^{PRICE}/_{TOTSP}}{\partial KITSP} = 0{,}37 - 0{,}013295 * 2 * KITSP = 0$$

$$\widehat{KITSP}_{opt} = 14{,}0166.$$

Вообще, $\widehat{KITSP}_{opt} = \hat{\beta} = -0{,}5 * \frac{0{,}013295}{0{,}37} = -0{,}5 * \frac{\widehat{\beta_2}}{\widehat{\beta_3}}$.

$\hat{\beta}$ асимптотически нормальна, следовательно, $g(\hat{\beta}) = -0{,}5 * \frac{\widehat{\beta_2}}{\widehat{\beta_3}}$ также асимптотически нормальна. Тогда, чтобы оценить дисперсию $\hat{\beta}$, используем *дельта-метод*.

$$\widehat{V(\hat{\beta})} = (\frac{\partial g}{\partial \widehat{\beta_2}})^2 \widehat{V(\widehat{\beta_2})} + (\frac{\partial g}{\partial \widehat{\beta_3}})^2 \widehat{V(\widehat{\beta_3})} + 2 * \left(\frac{\partial g}{\partial \widehat{\beta_2}} * \frac{\partial g}{\partial \widehat{\beta_3}}\right) * cov(\widehat{\beta_2}, \widehat{\beta_3})$$

$$= \left(-0{,}5 * \frac{1}{\widehat{\beta_3}}\right)^2 * V(\widehat{\beta_2}) + \left(0{,}5 * \frac{\widehat{\beta_2}}{\widehat{\beta_3}^2}\right)^2 + 2 * \left(-0{,}25 * \frac{1}{\widehat{\beta_3}} * \frac{\widehat{\beta_2}}{\widehat{\beta_3}^2}\right)$$

$$* cov(\widehat{\beta_2}, \widehat{\beta_3}) = 2{,}08^2.$$

Следовательно, истинное значение оптимальной площади кухни лежит в пределах:

$$KITSP_{opt} \in (14{,}0166 - 2{,}08 * 1{,}96;\ 14{,}0166 + 2{,}08 * 1{,}96)$$

$$KITSP_{opt} \in (9{,}9;\ 18{,}1).$$

Задача 3. Инструментальные переменные

1) Регрессия экзаменационного балла по посещаемости скорее всего не позволит состоятельно оценить интересующий исследователя эффект, так как в регрессии будет пропущена существенная переменная (например, талант студента, который, безусловно, влияет на балл, полученный на экзамене). Пусть истинная регрессия имеет вид:

$$exam_i = \theta_1 + \theta_2 * d_i + \theta_3 * talent_i + \varepsilon_i,$$

а исследователь хочет оценить модель

$$exam_i = \theta_1 + \theta_2 * d_i + u_i,$$

где $u_i = \theta_3 * talent_i + \varepsilon_i$.

Тогда $\widehat{\theta_2} = \frac{\widehat{cov(d_i, exam_i)}}{\widehat{var(d_i)}} \xrightarrow{P} \frac{cov(d_i, exam_i)}{var(d_i)} = \frac{cov(\theta_1 + \theta_2 * d_i + \theta_3 * talent_i + \varepsilon_i, d_i)}{var(d_i)} = \theta_2 + \theta_3 * \frac{cov(talent_i, d_i)}{var(d_i)}$.

Если талант студента и его посещаемость коррелированы ($cov(talent_i, d_i) \neq 0$), то МНК-оценка $\widehat{\theta_2}$ в модели $exam_i = \theta_1 + \theta_2 * d_i + u_i$ не будет состоятельной.

Замечание: напомним, что оценка является состоятельной, когда $\widehat{\theta_2} \xrightarrow{P} \theta_2$.

2) Новая фиктивная переменная - удаленность от университета – описывается следующим образом: $far_i = \begin{cases} 1, \text{если живет далеко от университета} \\ 0, \text{если живет близко} \end{cases}$. Фиктивная переменная удаленности от университета может быть хорошей инструментальной переменной для посещения лекций, поскольку удовлетворяет требованиям экзогенности ($cov(far_i, u_i) = 0$) и релевантности ($cov(far_i, d_i) \neq 0$). Действительно, из приведенной таблицы видим, что среди тех студентов, кто живет далеко, больше непосещавших, а среди тех студентов, кто живет близко, больше посещавших, а это значит, что переменные far_i и d_i коррелированы.

3) Опишем сначала процедуру состоятельного оценивания и покажем, чему равна 2МНК-оценка коэффициента θ_2.

<u>1 шаг</u>: оцениваем регрессию переменной посещаемости на удаленность проживания от университета, чтобы с помощью регрессии на экзогенную переменную отделить те d_i, которые коррелированы со случайными ошибками модели u_i:

$d_i = \alpha_1 + \alpha_2 * far_i + v_i$.

Получаем прогнозные значения $\widehat{d_i}$: $cov(\widehat{d_i}, u_i) = 0$. Поэтому на втором шаге можно оценивать регрессию экзаменационного балла на «скорректированную» посещаемость $\widehat{d_i}$, не боясь получить несостоятельные оценки коэффициентов.

2 шаг: оцениваем модель $exam_i = \theta_1 + \theta_2 * \widehat{d_\iota} + u_i$ и получаем состоятельные оценки коэффициентов.

$$\widehat{\theta_2}^{TSLS} = \frac{\widehat{cov(exam_\iota, \widehat{d_\iota})}}{\widehat{var(\widehat{d_\iota})}} = \frac{\widehat{cov(exam_\iota, \widehat{\alpha_1} + \widehat{\alpha_2} * far_i)}}{var(\widehat{\alpha_1} + \widehat{\alpha_2} * far_\iota)} = \frac{\widehat{\alpha_2} * \widehat{cov(exam_\iota, far_i)}}{(\widehat{\alpha_2})^2 * \widehat{var(far_\iota)}}$$

$$= \frac{\widehat{cov(exam_\iota, far_i)}}{\widehat{\alpha_2} * \widehat{var(far_\iota)}} = \frac{\widehat{var(far_\iota)}}{\widehat{cov(d_\iota, far_i)}} * \frac{\widehat{cov(exam_\iota, far_i)}}{\widehat{var(far_\iota)}}$$

$$= \frac{\widehat{cov(exam_\iota, far_i)}}{\widehat{cov(d_\iota, far_i)}}.$$

Для удобства дальнейших вычислений перепишем данные в таблицу следующим образом:

far=1 d=1 n=60 exam=90	far=0 d=1 n=90 exam=80
far=1 d=0 n=100 exam=40	far=0 d=0 n=70 exam=30

1 шаг:

$$ESS = \sum_{i=1}^{320} (d_i - \widehat{d_\iota})^2 = \sum_{i=1}^{320} (d_i - \widehat{\alpha_1} + \widehat{\alpha_2} * far_i)^2 \to \min(\widehat{\alpha_1}, \widehat{\alpha_2})$$

$$\begin{cases} \dfrac{\partial ESS}{\partial \widehat{\alpha_1}} = -2 \sum (d_i - \widehat{\alpha_1} + \widehat{\alpha_2} * far_i) = 0 \\ \dfrac{\partial ESS}{\partial \widehat{\alpha_2}} = -2 \sum far_i * (d_i - \widehat{\alpha_1} + \widehat{\alpha_2} * far_i) = 0 \end{cases}$$

$$\begin{cases} \sum (d_i - \widehat{\alpha_1} + \widehat{\alpha_2} * far_i) = 0 \\ \sum far_i * (d_i - \widehat{\alpha_1} + \widehat{\alpha_2} * far_i) = 0 \end{cases}$$

$$\begin{cases} \sum d_i - 320 * \widehat{\alpha_1} - \widehat{\alpha_2} \sum far_i = 0 \\ \sum far_i * d_i - \widehat{\alpha_1} \sum far_i - \widehat{\alpha_2} \sum {far_i}^2 = 0 \end{cases}$$

Подставляя значения сумм для фиктивных переменных из таблицы, имеем:

$$\begin{cases} 150 - 320 * \widehat{\alpha_1} - 160 * \widehat{\alpha_2} = 0 \\ 60 - 160 * \widehat{\alpha_1} - 160 * \widehat{\alpha_2} = 0 \end{cases}.$$

Решая систему линейных уравнений, получаем, что $\widehat{\alpha_1} = 0{,}5625$, а $\widehat{\alpha_2} = -0{,}1875$.

Так, оцененное уравнение регрессии 1-го шага имеет вид:

$$\widehat{d_\iota} = 0{,}5625 - 0{,}1875 * far_i.$$

2 шаг:

$$ESS = \sum_{i=1}^{320} (exam_i - \widehat{exam_\iota})^2 = \sum_{i=1}^{320} (exam_i - \widehat{\theta_1} + \widehat{\theta_2} * \widehat{d_\iota})^2 \to \min(\widehat{\theta_1}, \widehat{\theta_2})$$

$$\begin{cases} \dfrac{\partial ESS}{\partial \widehat{\theta_1}} = -2\sum\left(exam_i - \widehat{\theta_1} + \widehat{\theta_2} * \widehat{d_\iota}\right) = 0 \\ \dfrac{\partial ESS}{\partial \widehat{\theta_2}} = -2\sum \widehat{d_\iota} * \left(exam_i - \widehat{\theta_1} + \widehat{\theta_2} * \widehat{d_\iota}\right) = 0 \end{cases}$$

$$\begin{cases} \sum exam_i - 320 * \widehat{\theta_1} - \widehat{\theta_2} * 0{,}5625 * 320 + 0{,}1875 * \widehat{\theta_2} \sum far_i = 0 \\ 0{,}5625 \sum exam_i - 0{,}1875 \sum far_i * exam_i - \widehat{\theta_1} \sum \widehat{d_\iota} - \widehat{\theta_2} \sum \widehat{d_\iota}^{\,2} = 0 \end{cases}$$

$$\sum exam_i = 60 * 90 + 100 * 40 + 90 * 80 + 70 * 30 = 16900$$

$$\sum far_i = 160$$

$$\sum far_i * exam_i = 60 * 90 + 100 * 40 = 7600$$

$$\sum \widehat{d_\iota} = \sum (0{,}5625 - 0{,}1875 * far_i) = 0{,}5625 * 320 - 0{,}1875 \sum far_i = 180 - 30 = 150$$

$$\sum \widehat{d_\iota}^{\,2} = \sum (0{,}5625 - 0{,}1875 * far_i)^2 = 73{,}125$$

Подставляя вычисленные значения, получим систему линейных уравнений:

$$\begin{cases} 16900 - 320 * \widehat{\theta_1} - 150 * \widehat{\theta_2} = 0 \\ 8081{,}25 - 150 * \widehat{\theta_1} - 73{,}125 * \widehat{\theta_2} = 0 \end{cases}.$$

$$\widehat{\theta_2} = 56{,}67,$$

$$\widehat{\theta_1} = 26{,}25.$$

Ответ: $\widehat{\theta_2} = 56{,}67$.

Задача 4. Временные ряды

Дан AR(2) процесс: $x_t = \delta + \theta_1 * x_{t-1} + \theta_2 * x_{t-2} + \varepsilon_t$.

а) Необходимо восстановить уравнение, описывающее динамику этого временного ряда, то есть найти соответствующие коэффициенты δ, θ_1 и θ_2. Для этого, во-первых, выпишем математическое ожидание для данного процесса:

$$E(x_t) = E(\delta + \theta_1 * x_{t-1} + \theta_2 * x_{t-2} + \varepsilon_t) = \delta + \theta_1 * E(x_{t-1}) + \theta_2 * E(x_{t-2}).$$

Поскольку процесс стационарен (по условию), то $E(x_t) = E(x_{t-1}) = E(x_{t-2}) = 10$. Следовательно, мы можем переписать математическое ожидание для исследуемого процесса следующим образом: $10 = \delta + \theta_1 * 10 + \theta_2 * 10$. Получено уравнение [1] для конечной системы уравнений.

Во-вторых, так как нам известны значения коэффициентов автокорреляции первого и второго порядка, то необходимо вывести для них формулы.

$\gamma_1 = cov(x_t, x_{t-1}) = cov(\delta + \theta_1 * x_{t-1} + \theta_2 * x_{t-2} + \varepsilon_t, x_{t-1}) = \begin{Bmatrix} cov(\delta, x_{t-1}) = 0 \\ cov(\varepsilon_t, x_{t-1}) = 0 \end{Bmatrix} = \theta_1 * cov(x_{t-1}, x_{t-1}) + \theta_2 * cov(x_{t-2}, x_{t-1}) = \theta_1 * \gamma_0 + \theta_2 * \gamma_1$. Отсюда выразим $\gamma_1 = \frac{\theta_1}{1-\theta_2} * \gamma_0$.

Зная, что $\rho_k = \frac{\gamma_k}{\gamma_0}$, очевидно:

$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{\theta_1}{1-\theta_2} = 0{,}5$ и $\theta_1 + 0{,}5 * \theta_2 - 0{,}5 = 0$. [2]

Аналогично получим и для γ_2: $\gamma_2 = cov(x_t, x_{t-2}) = cov(\delta + \theta_1 * x_{t-1} + \theta_2 * x_{t-2} + \varepsilon_t, x_{t-2}) = \begin{Bmatrix} cov(\delta, x_{t-2}) = 0 \\ cov(\varepsilon_t, x_{t-2}) = 0 \end{Bmatrix} = \theta_1 * cov(x_{t-1}, x_{t-2}) + \theta_2 * cov(x_{t-2}, x_{t-2}) = \theta_1 * \gamma_1 + \theta_2 * \gamma_0$. Очевидно, что

$\rho_2 = \frac{\gamma_2}{\gamma_0} = \frac{\theta_1 * \gamma_1 + \theta_2 * \gamma_0}{\gamma_0} = \theta_1 * \rho_1 + \theta_2 = 0{,}4$. [3]

Получим следующую систему линейных уравнений относительно θ_1, θ_2 и δ:

$$\begin{cases} \theta_1 * 10 + \theta_2 * 10 + (\delta - 10) = 0 \\ \theta_1 + 0{,}5 * \theta_2 - 0{,}5 = 0 \\ \theta_1 * \rho_1 + \theta_2 - 0{,}4 = 0 \end{cases}$$

Решая ее, получим $\theta_1 = 0{,}4$; $\theta_2 = 0{,}2$; $\delta = 4$ и восстановленное уравнение для описания авторегрессионного процесса 2-го порядка:

$$x_t = 4 + 0{,}4 * x_{t-1} + 0{,}2 * x_{t-2} + \varepsilon_t.$$

б) Поскольку у исследуемого процесса математическое ожидание не равно нулю, то для построения наилучшего прогноза необходимо перейти к процессу с нулевым математическим ожиданием.

Пусть $E(x_t) = E(x_{t-1}) = E(x_{t-2}) = \mu$, тогда $\mu = \delta + \theta_1 * \mu + \theta_2 * \mu$ и $\mu = \frac{\delta}{1-\theta_1-\theta_2}$.

Вычтем $\mu = \frac{\delta}{1-\theta_1-\theta_2}$ из обеих частей уравнения:

$$x_t - \frac{\delta}{1-\theta_1-\theta_2} = \delta - \frac{\delta}{1-\theta_1-\theta_2} + \theta_1 * x_{t-1} + \theta_2 * x_{t-2} + \varepsilon_t,$$

$$x_t - \frac{\delta}{1-\theta_1-\theta_2} = -\frac{\delta * \theta_1}{1-\theta_1-\theta_2} - \frac{\delta * \theta_2}{1-\theta_1-\theta_2} + \theta_1 * x_{t-1} + \theta_2 * x_{t-2} + \varepsilon_t,$$

$$(x_t - \frac{\delta}{1-\theta_1-\theta_2}) = \theta_1 * (x_{t-1} - \frac{\delta}{1-\theta_1-\theta_2}) + \theta_2 * (x_{t-2} - \frac{\delta}{1-\theta_1-\theta_2}) + \varepsilon_t.$$

Произведем соответствующую замену переменных: $z_t = x_t - \frac{\delta}{1-\theta_1-\theta_2}$, $z_{t-1} = x_{t-1} - \frac{\delta}{1-\theta_1-\theta_2}$ и $z_{t-2} = x_{t-2} - \frac{\delta}{1-\theta_1-\theta_2}$. В результате получим новый авторегрессионный процесс второго порядка с нулевым математическим ожиданием:

$z_t = \theta_1 * z_{t-1} + \theta_2 * z_{t-2} + \varepsilon_t$, или, зная коэффициенты, $z_t = 0{,}4 * z_{t-1} + 0{,}2 * z_{t-2} + \varepsilon_t$.

$$z_1 = x_1 - \frac{\delta}{1-\theta_1-\theta_2} = 8 - \frac{4}{1-0{,}4-0{,}2} = -2,$$

$$z_2 = x_2 - \frac{\delta}{1-\theta_1-\theta_2} = 6 - \frac{4}{1-0{,}4-0{,}2} = -4.$$

Следовательно, наилучший прогноз: $\widehat{z_3} = E\left(z_3|_{z_1,z_2}\right) = 0{,}4 * z_2 + 0{,}2 * z_1 = 0{,}4 * (-4) + 0{,}2 * (-2) = -2.$

$$\widehat{z_4} = E\left(z_4|_{z_1,z_2}\right) = E(0{,}4 * z_3 + 0{,}2 * z_2) = 0{,}4 * E\left(z_3|_{z_1,z_2}\right) + 0{,}2 * z_2 = -0{,}4 * 2 - 0{,}2 * 4 = -1{,}6.$$

Таким образом, $\widehat{x_4} = \widehat{z_4} + \frac{\delta}{1-\theta_1-\theta_2} = 10 - 1{,}6 = 8{,}4.$

Ответ: а) $x_t = 4 + 0{,}4 * x_{t-1} + 0{,}2 * x_{t-2} + \varepsilon_t$;   б) $\widehat{x_4} = 8{,}4.$