

Универсиада по эконометрике
экономического факультета МГУ имени М. В. Ломоносова
при поддержке Комплекса экономической политики города Москвы
19 апреля 2025 г.
100 баллов, каждая задача по 25 баллов

У Д А Ч И !

Задача 1.

Аналитик Дмитрий из Департамента экономической политики и развития города Москвы изучает влияние открытия новых станций метрополитена на время, которое москвичи тратят на дорогу. Он делит весь город на небольшие участки и отбирает только те, в пределах которых в 2023 году станция метро отсутствовала. Обозначим число таких участков n .

Для каждого из этих участков Дмитрий вычисляет время, затрачиваемое типичным жителем участка на дорогу на работу в 2023 и в 2024 г. В 2024 году на 50 % этих участков открылась новая станция метро ($W_i = 1$). (Т. е. рядом с домом типичного жителя участка i появилось метро, что сократило время, которое он тратит на дорогу на работу.)

Дмитрий также понимает, что помимо прямых эффектов могут существовать и косвенные положительные внешние эффекты для жителей других участков. Даже если станция метро открылась на других участках, она может быть удобной для перемещений по городу до другой необходимой локации или для пересадки, и сократить время на дорогу до работы. И таким образом, можно записать следующую модель:

$$Y_{it} = \alpha_i + \sum_{j=1}^n \gamma_{ijt} W_j + \varepsilon_{it} \quad (1)$$

где Y_{it} — это время, которое i -тый респондент тратит на дорогу в неделю в среднем за год t ($t = 2023$ или 2024), W_j — бинарная переменная, принимающая значение 1, если в 2024 г. открылась новая станция метро на каком-то из участков j ($j=1, \dots, n$), ε_{it} — случайный шок. Дополнительно предполагаем, что случайный шок не коррелирует с W_j .

Вопрос 1.

Допустим в модели (1), $\gamma_{ijt}=0$ для любых $i \neq j$. Т. е. на время, затрачиваемое на дорогу жителем участка i , не влияет открытие «чужих» станций метро. Покажите, что оценка

$$\hat{\tau} = \frac{2}{n} \sum_{i: W_i=1} (Y_{i,2024} - Y_{i,2023}) - \frac{2}{n} \sum_{i: W_i=0} (Y_{i,2024} - Y_{i,2023}) \quad (2)$$

является несмещённой для агрегированного коэффициента $\tau_{\text{простое}} = \frac{1}{n} \sum_{i=1}^n \gamma_{ii,2024}$. Дайте интерпретацию $\tau_{\text{простое}}$.

Вопрос 2

Теперь ограничения $\gamma_{ijt}=0$ для любых $i \neq j$ из предыдущего вопроса снимаются, т. е. допускается влияние открытия «чужой» станции метро на время поездки. Определим ещё один агрегированный коэффициент $\tau_{\text{новое}} = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \gamma_{ij,2024}$. Дайте интерпретацию $\tau_{\text{новое}}$. Является ли оценка $\hat{\tau}$, заданная выражением (2), несмещённой для коэффициента $\tau_{\text{новое}}$?



**ЭКОНОМИКА
МОСКВЫ**



Вопрос 3 Аналитик рассматривает третью («хитрую») оценку вида:

$$\hat{\tau}_{\text{хитрое}} = \frac{k}{n} \sum_{i=1}^n (Y_{i,2024} - Y_{i,2023}) \quad (3),$$

где k – некоторая константа. Существует ли такая константа k , при которой эта оценка является несмещённой оценкой для агрегированного коэффициента $\tau_{\text{новое}}$? Покажите это.

Задача 2

Вы стажёр в Аналитическом Центре при Правительстве Москвы. Вам необходимо оценить влияние пилотной программы благоустройства на удовлетворённость жизнью в городе. Вы решили использовать данные опроса 16200 респондентов, из них 8100 проживают в местах, где было проведено новое благоустройство, и 8100, где нет. Их ответы закодированы в 4 градациях: 0.01 (для наименьшего уровня), 0.1, 1 и 10 (полностью удовлетворён). Данные приведены в таблице.

	Нового благоустройства нет					Проведено новое благоустройство					
Балл удовлетворённости респондента жизнью в городе (<i>y</i>)	0.01	0.1	1	10		0.01	0.1	1	10		
Округ Москвы	Количество респондентов					А	Количество респондентов				Б
СЗАО	40	26	7	0	10	20	18	8	0	10	
САО	30	17	8	0	10	20	28	7	0	10	
СВАО	10	69	3	0	10	20	48	5	0	10	
ВАО	50	35	6	0	10	30	27	7	0	10	
ЮВАО	10	39	6	0	10	10	49	5	0	10	
ЮАО	50	55	4	0	10	0	10	99	0	100	
ЮЗАО	40	36	76	92	1000	0	150	85	0	100	
ЗАО	50	455	2334	762	10000	0	40	2866	713	10000	
ЦАО	20	238	2816	716	10000	0	20	3128	687	10000	
Общее количество респондентов	300	970	5260	1570		100	390	6210	1400		

В таблице указано количество респондентов, которые попали в выборку в разбивке по административному округу, наличию благоустройства и баллу удовлетворённости жизнью.

В колонках А и Б рассчитаны суммарные баллы удовлетворённости респондентов по округам города (СЗАО, САО и т. д.). Например, в СЗАО $0.01 \cdot 40 + 0.1 \cdot 26 + 1 \cdot 7 + 10 \cdot 0 = 10$.

Вопрос 1: Получив данные, вы оцениваете следующую регрессию по всем респондентам:

$$\log_{10}(p_i) = \alpha + \beta W_i + \varepsilon_i \quad (1)$$

где i – номер респондента ($i = 1, \dots, 16200$); p_i – балл удовлетворённости жизнью i -го респондента; W_i – индикатор, равный 1, если новое благоустройство проведено, и 0, если нет; ε_i – случайный шок. Вычислите оценку коэффициента β .



ЭКОНОМИКА
МОСКВЫ



Вопрос 2: Проинтерпретируйте коэффициент β в формате: «Программа благоустройства <увеличивает/уменьшает> удовлетворённость жизнью <каких объектов> в среднем на <сколько>.»

Вопрос 3: Затем, используя колонки А и Б в таблице, Вы делаете оценку по данным, агрегированным на уровне округов Москвы:

$$\log_{10}(p_j) = \alpha + \beta W_j + \varepsilon_j \quad (2)$$

где j – номер округа с учётом наличия или отсутствия нового благоустройства (т. е. $j = 1, \dots, 18$); p_j – в отличие от p_i из модели (1), это суммарный балл удовлетворённости жизнью респондентов из округа j ; W_j – индикатор, равный единице, если в j проведено новое благоустройство, и 0, если нет.

Вычислите оценку коэффициента β .

Вопрос 4: Проинтерпретируйте коэффициент β в формате: «Программа благоустройства <увеличивает/уменьшает> удовлетворённость <каких объектов> в среднем на <сколько>.»

Вопрос 5: Вы заметили, что оценки из вопросов 1 и 3 сильно различаются, и вам пришла в голову другая идея – построить модель следующего вида:

$$p_i = 10^{\alpha + \beta W_i + \varepsilon_i} \quad (3)$$

где i – номер респондента, p_i – балл удовлетворённости жизнью i -го респондента (берётся логарифм по основанию 10), W_i – индикатор, равный 1, если новое благоустройство проведено, и 0, если нет, ε_i – случайный шок. Вычислите оценку коэффициента $10^{\beta} - 1$. Как изменится ваш ответ, если, как и в вопросе 2, мы будем пользоваться агрегированными на уровне округа?

Вопрос 6: Объясните, какие преимущества может иметь регрессия из вопроса 5 над логарифмически-линейной регрессией из вопросов 1 и 3. Вы можете привести в качестве аргумента релевантную целям анализа интерпретацию коэффициента, но можете и на полезные свойства оценки, которые вы в процессе анализа выяснили.



ЭКОНОМИКА
МОСКВЫ



Задача 3.

Василий изучает влияние развития финансового сектора на темпы роста ВВП в странах мира за несколько лет. Он оценивает две регрессии вида:

$$y_{it} = \alpha + \alpha_1 FIN_{it-1} + \alpha_2 FIN_{it-1}^2 + \alpha_5 OPEN_{it-1} + \alpha_6 GCF_{it-1} + \varepsilon_{it} \quad (1)$$

$$y_{it} = \alpha + \beta_1 FIN_{it-1} + \beta_2 FIN_{it-1}^2 + \beta_3 FIN_{it-1} * d_{it-1} + \beta_4 FIN_{it-1}^2 * d_{it-1} + \beta_5 OPEN_{it-1} + \beta_6 GCF_{it-1} + \varepsilon_{it} \quad (2)$$

где y_{it} — темпы роста ВВП; FIN_{it} — показатель развития финансового сектора (Василий выбрал кредит частному сектору в долях от ВВП); d_{it} — бинарная переменная, отвечающая за наличие (=1) или отсутствие (=0) действующих санкций в стране i в году t ; $OPEN_{it}$ — открытость экономики (контрольная переменная), рассчитанная как сумма экспорта и импорта в % от ВВП; GCF_{it} — валовое накопление основного капитала, % ВВП (контрольная переменная); ε_{it} — случайный шок.

Т.е. в моделях (1) и (2) предполагается нелинейная (параболическая) зависимость темпов роста ВВП от развития финансового сектора.

Василий оценивает модели (1) и (2) по МНК и получает следующие результаты (см. таблицу). В правом столбце таблицы 1 указаны коэффициенты модели (2).

Переменная	Модель 1	Модель 2
Константа	Василий не предоставил данные	0,15
FIN_{it-1}	Василий не предоставил данные	1,16
FIN_{it-1}^2	Василий не предоставил данные	-0,82
$FIN_{it-1} * d_{it-1}$	Василий не предоставил данные	0,715
$FIN_{it-1}^2 * d_{it-1}$	Василий не предоставил данные	-0,6
$OPEN_{it-1}$	Василий не предоставил данные	0,215
GCF_{it-1}	Василий не предоставил данные	0,43
Число наблюдений	1807	1807
Коэффициент детерминации (R2)	0,2	0,4

1. Можно ли на основе полученных из модели (2) оценок утверждать на уровне значимости 5%, что открытость экономики значимо влияет на темпы роста ВВП? Сформулируйте гипотезу в терминах коэффициентов модели и проверьте её соответствующим тестом.
2. Найдите оценку оптимального (максимизирующего темпы роста ВВП) значения развития финансового сектора в условиях отсутствия санкций, используя модель (2).



**ЭКОНОМИКА
МОСКВЫ**



3. Можно ли утверждать на основании модели (2), что введение санкций не влияет на величину оптимального (максимизирующего темпы роста ВВП) развития финансового сектора? Сформулируйте гипотезу в терминах коэффициентов модели и проверьте её соответствующим тестом на уровне значимости 5%.
4. Верно ли, что в условиях санкций оптимальное (максимизирующее темпы роста ВВП) значение развития финансового сектора равно 0,5? Сформулируйте гипотезу в терминах коэффициентов модели (2) и проверьте её соответствующим тестом.
5. Поясните, по каким причинам переменные включаются в модели (1) и (2) с лагом?

Справочно:

Вы можете пользоваться таблицами распределения, прикрепленными к Вашим листам решения.

Таблица 2. Оценка ковариационной матрицы оценок коэффициентов модели:

0,01	0,15	0,22	0,11	0,17	0,55	0,661
0,15	0,25	0,33	0,14655	0,72	0,316	0,112
0,22	0,33	0,16	0,23	0,9	0,32	0,101
0,11	0,14655	0,23	0,04	0,32	0,532	0,514
0,17	0,72	0,9	0,32	0,0169	0,134	0,1421
0,55	0,316	0,32	0,532	0,134	0,01	0,43342
0,661	0,112	0,101	0,514	0,1421	0,43342	0,04

Задача 4.

Существует так называемая проблема смещения из-за пропуска существенных переменных (omitted variable bias). Рассмотрим ситуацию, когда истинная связь между показателями y , x_1 и x_2 описывается моделью:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i \quad (1)$$

Однако была оценена парная регрессия без x_2 :

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} \quad (2)$$

Вопрос 1. Докажите, что величина смещения оценки коэффициента $\hat{\beta}_1$ из регрессии (2) равна $\beta_2 \frac{\text{cov}(x_1, x_2)}{\text{var}(x_1)}$

Вопрос 2. Приведите содержательный экономический пример, когда это смещение равно нулю. Либо объясните, почему такая ситуация невозможна.

Вопрос 3. Пусть x_1 и x_2 – бинарные переменные, N – общее число наблюдений, N_1 – число единиц среди значений переменной x_1 , N_2 – число единиц среди значений переменной x_2 . Причём содержательный смысл x_1



**ЭКОНОМИКА
МОСКВЫ**



и x_2 таков, что они не могут одновременно принимать значение единица. Докажите, что в таких условиях смещение равно $-\beta_2 \frac{N_2}{N-N_1}$. Вы можете привести доказательство в скалярной или в матричной форме.

Вопрос 4. Приведите содержательный пример для ситуации, описанной в вопросе 3.

Вопрос 5. Рассмотрим случай множественной регрессии. Истинная связь между переменными записана в матричном виде:

$$Y = X_1\beta_1 + X_2\beta_2 + \varepsilon \quad (3)$$

где Y – вектор зависимой переменной размерности $N \times 1$; X_1 – матрица размерности $N \times (1+k_1)$, то есть состоящая из столбца единиц (т.к. модель (3) включает константу) и k_1 столбцов регрессоров; X_2 – матрица размерности $N \times k_2$, то есть состоящая из k_2 столбцов регрессоров; X_1 и X_2 могут содержать не только бинарные переменные; β_1 – вектор коэффициентов размерности $(1+k_1) \times 1$; β_2 – вектор коэффициентов размерности $k_2 \times 1$; ε – вектор случайных шоков размерности $N \times 1$.

Оценена регрессия:
$$\hat{Y} = X_1\hat{\beta}_1 \quad (4)$$

Докажите, что вектор смещений для оценок $\hat{\beta}_1$ из (4) равен $(X_1^T X_1)^{-1} (X_1^T X_2) \beta_2$.

Вопрос 6. Рассмотрим случай, когда матрицы X_1 и X_2 состоят только из бинарных переменных, причём нет наблюдений, для которых любые две бинарные переменные одновременно принимали бы значение 1.

Утверждение 1. Докажите, что квадратная матрица $X_1^T X_1$ является «стреловидной» (arrow matrix), а именно

$$X_1^T X_1 = \begin{pmatrix} N & N_1 & N_2 & \dots & \dots & N_{k_1} \\ N_1 & N_1 & 0 & \dots & \dots & 0 \\ N_2 & 0 & N_2 & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \dots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & 0 \\ N_{k_1} & 0 & 0 & \dots & 0 & N_{k_1} \end{pmatrix}, \text{ где } N - \text{общее число наблюдений, } N_1, N_2, \dots, N_{k_1} - \text{объёмы наблюдений,}$$

для которых соответствующие бинарные переменные из матрицы X_1 принимают значение 1. А все элементы, кроме главной диагонали, первой строки и первого столбца, нулевые.

(В дальнейших пунктах Вы можете пользоваться результатом Утверждения 1, даже если не доказали его.)

Для доказательства Утверждения 2, сформулированного ниже, Вы можете воспользоваться следующим готовым результатом:

Лемма¹: Для всякой «стреловидной» матрицы $A = \begin{pmatrix} \alpha & z^T \\ z & D \end{pmatrix}$ размерности $m \times m$, где α – скаляр, z – вектор-столбец размерности $(m-1) \times 1$, z^T – вектор-строка размерности $1 \times (m-1)$, D – диагональная матрица размерности

¹ Clarke D., A convenient omitted variable bias formula for treatment effect models. Economics Letters (2018), <https://doi.org/10.1016/j.econlet.2018.10.035>.



ЭКОНОМИКА
МОСКВЫ



$(m-1)*(m-1)$ обратная матрица имеет вид $A^{-1} = \begin{pmatrix} 0 & \vec{0}^T \\ 0 & D^{-1} \end{pmatrix} + \rho u u^T$, где 0 – нулевой скаляр, $\vec{0}$ – нулевой вектор-столбец, $\vec{0}^T$ – нулевая вектор-строка, $\rho = \frac{1}{\alpha - Z^T D^{-1} Z}$ – скаляр, $u = [-1 \quad D^{-1} Z]^T$ – вектор.

Утверждение 2. Обратная матрица $(X_1^T X_1)^{-1}$ имеет вид

$$(X_1^T X_1)^{-1} = \begin{pmatrix} \frac{1}{\lambda} & -\frac{1}{\lambda} & -\frac{1}{\lambda} & \dots & \dots & -\frac{1}{\lambda} \\ -\frac{1}{\lambda} & (\frac{1}{N_1} + \frac{1}{\lambda}) & \frac{1}{\lambda} & \dots & \dots & \frac{1}{\lambda} \\ -\frac{1}{\lambda} & \frac{1}{\lambda} & (\frac{1}{N_2} + \frac{1}{\lambda}) & \dots & \dots & \frac{1}{\lambda} \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ -\frac{1}{\lambda} & \frac{1}{\lambda} & \dots & \dots & \frac{1}{\lambda} & (\frac{1}{N_{k1}} + \frac{1}{\lambda}) \end{pmatrix}, \text{ где } \lambda = N - N_1 - N_2 - \dots - N_{k1}.$$

(В дальнейших пунктах Вы можете пользоваться результатом Утверждения 2, даже если не доказали его.)

*Утверждение 3. Докажите, что $(1+k1)*k2$ матрица $X_1^T X_2$ равна*
$$\begin{pmatrix} M_1 & M_2 & M_3 & \dots & \dots & M_{k2} \\ 0 & 0 & 0 & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 0 \end{pmatrix},$$

где N – общее число наблюдений, $M_1, M_2, \dots, M_{k2}$ – число наблюдений, для которых бинарные переменные из матрицы X_2 принимают значение 1

(В дальнейших пунктах Вы можете пользоваться результатом Утверждения 3, даже если не доказали его.)

Утверждение 4. Докажите, что вектор $(X_1^T X_1)^{-1} (X_1^T X_2) \beta_2$ (из вопроса 5) можно представить в виде

$$\begin{pmatrix} \frac{M_1}{\lambda} & -\frac{M_2}{\lambda} & -\frac{M_3}{\lambda} & \dots & \dots & -\frac{M_{k2}}{\lambda} \\ -\frac{M_1}{\lambda} & -\frac{M_2}{\lambda} & -\frac{M_3}{\lambda} & \dots & \dots & -\frac{M_{k2}}{\lambda} \\ -\frac{M_1}{\lambda} & -\frac{M_2}{\lambda} & -\frac{M_3}{\lambda} & \dots & \dots & -\frac{M_{k2}}{\lambda} \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ -\frac{M_1}{\lambda} & -\frac{M_2}{\lambda} & \dots & \dots & -\frac{M_{k2}}{\lambda} \end{pmatrix} \beta_2$$

Иными словами, смещение оценки j-того коэффициента $\hat{\beta}_1^j$ из модели (4) для j от 1 до k1 составляет величину $-\beta_2^1 \frac{M_1}{\lambda} - \beta_2^2 \frac{M_2}{\lambda} - \dots - \beta_2^{k2} \frac{M_{k2}}{\lambda}$, где $\beta_2^1, \beta_2^2, \dots, \beta_2^{k2}$ – компоненты вектора β_2 из модели (3).



**ЭКОНОМИКА
МОСКВЫ**

