

## Задание 1 (25 баллов)

В некотором исследовании анализируется влияние шоков нефтяных цен на инфляцию, а также воздействие выбора режима денежно-кредитной политики на силу этого влияния. Оценены модели на панельных данных по странам мира за период с 2000 по 2017 годы.

Авторы опираются на следующую спецификацию модели:

$$\Delta \ln P_{it} = \phi \Delta \ln P_{it-1} + \lambda \Delta \ln e_{it} + \beta \Delta \ln e_{it} IT_{it} + \theta \Delta \ln OP_{it} + \tau \Delta \ln OP_{it} IT_{it} + \delta X_{it} + \mu_i + \varepsilon_{it}$$

Индекс  $i$  обозначает страну, а  $t$  - год.

$P$  — индекс потребительских цен. Соответственно,  $\Delta \ln P$  — логарифмический темп прироста цен, то есть инфляция. Включение в правую часть уравнения лагированного значения зависимой переменной отражает инерционный характер инфляции, объясняющийся адаптивным характером инфляционных ожиданий части экономических агентов, который подтверждается рядом современных работ;

$e$  — номинальный курс национальной валюты данной страны (увеличение этой переменной означает ослабление национальной валюты);

$OP$  — уровень цен на нефть;

$IT$  — бинарная переменная, равная единице в те периоды, в которые данная страна таргетирует инфляцию, и равная нулю в остальные периоды;

$X$  — вектор контрольных переменных;

$\mu$  — фиксированные страновые эффекты.

Полученные в ходе моделирования уравнения регрессии для двух промежутков времени представлены в таблице 1.

(а) Опираясь на эти результаты, поясните, какие из пяти перечисленных ниже утверждений являются верными для периода 2000-2017? Затем, поясните, какие из них являются верными для периода 2000-2008?

1. Ослабление национальной валюты ускоряет инфляцию;
2. В странах, которые таргетируют инфляцию, ослабление национальной валюты разгоняет инфляцию слабее, чем во всех остальных странах;
3. Рост цен на нефть разгоняет инфляцию;
4. В странах, которые таргетируют инфляцию, рост цен на нефть разгоняет инфляцию слабее, чем во всех остальных странах;
5. В странах, не таргетирующих инфляцию, ослабление национальной валюты на один процент в тот же самый год увеличивает уровень инфляции меньше, чем на один процент.

Обоснуйте свои ответы. Используйте пятипроцентный уровень значимости для тестирования необходимых гипотез.

(b) Какие из ваших ответов по поводу корректности утверждений в предыдущем задании изменятся, если вместо пятипроцентного уровня значимости вы будете использовать десятипроцентный уровень?

*Примечания: В круглых скобках под оценками коэффициентов указаны состоятельные в условиях гетероскедастичности и автокорреляции стандартные ошибки. \* и \*\*\* обозначают значимость на десяти- и однопроцентном уровнях, соответственно.*

Таблица 1: Результаты оценивания базовой спецификации модели

Оцениваемые параметры	Модель 1 2000-2017	Модель 2 2000-2008
$\phi$	0,453 *** (0,082)	0,367 *** (0,089)
$\lambda$	0,270 *** (0,049)	0,397 *** (0,048)
$\beta$	-0,148 *** (0,053)	-0,148 *** (0,053)
$\theta$	0,004 (0,011)	0,043 * (0,026)
$\tau$	0,015 (0,015)	-0,035 (0,021)
Контрольные переменные	Да	Да
Число наблюдений	831	303

## Задание 2 (25 баллов)

Исследователь Иван планирует эксперимент по оценке эффективности новых тренировочных курсов для работников компании ABC. Производительность труда  $i$ -го работника ( $y_i$ ) описывается следующей моделью:

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$$

Здесь  $x_i$  — это бинарная переменная, которая равна единице для работников, которые прошли тренировочные курсы, и равна нулю для всех остальных.

$\varepsilon_i$  — независимые нормально распределенные случайные величины с нулевым математическим ожиданием, причем для всех, кто не посещал тренировочные курсы, дисперсия  $\varepsilon_i$  одинакова и равна  $\sigma^2$ . Для тех, кто посещал тренировочные курсы она одинакова и равна  $9 * \sigma^2$ . Будем считать, что Иван знает об этом соотношении дисперсий, но не знает значения  $\sigma^2$ .

Иван может собрать данные о производительности труда 100 работников. Он сам решает, сколько работников в его выборке будет иметь опыт прохождения тренировочных курсов (обозначим это число буквой  $m$ ), а сколько работников не будут иметь такого опыта.

- (a) (20) Предположим, что Иван использует для оценки интересующего его эффекта обычный МНК. Какое значение  $m$  ему следует выбрать для обеспечения максимальной точности результатов оценивания?
- (b) (5) Уже после того, как Иван выбрал значение  $m$  и собрал данные, его коллега отметил, что точность можно улучшить, если вместо обычного МНК применить другой метод, который обеспечит получение эффективных оценок. Какой метод имел в виду коллега?

### Задание 3 (25 баллов)

В некоей экономике кривая Филлипса имеет следующий вид:

$$\pi_t = \pi_t^e - 2 * (u_t - 5)$$

$\pi_t$  — фактический уровень инфляции в году  $t$ ,  $\pi_t^e$  — ожидаемый уровень инфляции в году  $t$ ,  $u_t$  — фактический уровень безработицы в году  $t$ . Все переменные измерены в процентах.

Известно, что ожидаемый уровень инфляции описывается процессом авторегрессии второго порядка:  $\pi_t^e = 0,6\pi_{t-1}^e + 0,3\pi_{t-2}^e + v_t$ , где  $v_t$  — белый шум с единичной дисперсией.

Фактический уровень безработицы описывается процессом скользящего среднего первого порядка:  $u_t = 5 + \varepsilon_t - 0,25\varepsilon_{t-1}$ , где  $\varepsilon_t$  — белый шум с единичной дисперсией.  $\varepsilon_t$  и  $v_t$  не коррелированы друг с другом.

- (a) Является ли фактическая инфляция стационарным процессом?
- (b) Известно, что в 2019 и 2020 годах ожидаемый уровень инфляции составлял 20 процентов. Постройте прогноз фактического уровня инфляции на 2022 год и определите 95-процентный доверительный интервал для этого прогноза.

## Задание 4 (25 баллов)

В данном задании вам предстоит рассмотреть простой пример проблемы причинного вывода при наличии эффекта окружения (имеется в виду peer effects, но, к сожалению, для этого термина нет хорошего перевода на русский язык). Представьте, что вы хотите определить, как воздействие на индивидуальном уровне влияет на результаты обучения в старшей школе, например, на баллы за единый государственный экзамен (ЕГЭ). Если быть более точными, то рассматриваемое вами – назначение тьютора, который работает на экономическом факультете МГУ и помогает старшекласснику подготовиться к экзаменам.

Для оценивания данного эффекта воздействия, вы собираете информацию о различных школах ( $s$  означает школу), а в каждой школе вы наблюдаете  $n_s$  единиц (число учеников одинаковое в каждой школе). Вы понимаете, что помимо прямых эффектов (тьютор способствует улучшению результатов назначенного ему ученика), воздействие также может иметь косвенный эффект. Чтобы отследить данный эффект, вы формулируете следующую модель потенциального исхода (для  $\beta \neq 1$ ):

$$Y_i(\{\omega_j\}_{j \in s(i)}) = \alpha_{s(i)} + \beta \sum_{j \in s(i)} \frac{E[Y_j(\{\omega_l\}_{l \in s(i)})]}{n_s} + \tau\omega_i + u_i \quad (1)$$

$$E[u_i] = 0$$

В данной модели  $s(i)$  – школа, к которой относится индивид  $i$ , а  $\omega_i \in \{0, 1\}$  – потенциальное воздействие (назначен тьютор или нет). Данная модель генерирует эффект окружения: влияют ли (ожидаемые) потенциальные результаты других учеников в одной и той же школе на потенциальные результаты  $i$ -го ученика.

1. Коэффициент  $\beta$  измеряет *эндогенный* эффекты окружения – как результаты других учеников влияют на результаты  $i$ -го ученика. Приведите примеры из реальной жизни, когда справедливо ожидать, что коэффициент  $\beta$  отрицательный, а когда – что коэффициент  $\beta$  положительный.
2. Покажите, что приведенная выше модель эквивалента (генерирует те же результаты) следующей модели:

$$Y_i(\{\omega_j\}_{j \in s(i)}) = \lambda_{s(i)} + \gamma \sum_{j \in s(i)} \frac{\omega_j}{n_s} + \tau\omega_i + u_i \quad (2)$$

$$E[u_i] = 0$$

а также выразите  $\lambda_{s(i)}$  и  $\gamma$  через  $\alpha_{s(i)}, \beta, \tau$ . Коэффициент  $\gamma$  отражает *экзогенный* эффект окружения – как на результаты  $i$ -го ученика влияет назначение тьютора его одноклассникам.

3. Предположим, что  $n_s$  одинаково, и в каждом классе мы случайным образом назначаем тьюторов ровно половине учеников. Предложите оценку  $\tau$ . Можем ли мы оценить  $\beta$  (или  $\gamma$ ) в данном случае?
4. Преположим, что  $n_s$  кратно 4, и вы рассматриваете долю учеников  $x$  в каждой школе, где  $x \in \{\frac{1}{4}, \frac{1}{2}, \frac{3}{4}\}$ . Для каждой школы вы случайным образом выбираете  $x$  (одинаковым образом), а затем случайным образом выбираете учеников так, что ровно на  $x$  из класса оказывается воздействие. Покажите, что в данном случае  $\tau, \gamma, \beta$  могут быть определены и предложите для них оценки (подумайте о сравнениях внутри школ и между школами).
5. Наконец предположим, что вы собираетесь провести эксперимент, описанный в пункте 4. К сожалению, что-то пошло не так, и вместо того, чтобы случайным образом определять долю учеников, на которых будет оказано воздействие, данное решение принимается директорами каждой школы. Некоторые из них выбирают оказывать воздействие (назначать тьютора) на  $\frac{3}{4}$  класса, в то время как другие выбрали оказывать воздействие лишь на четверть. Тьюторы по-прежнему назначаются ученикам случайным образом. Как данный факт влияет на ваш анализ, проведенный в пункте 4?

## Задание 5 (25 баллов)

Предположим, что вы хотите определить разрыв в заработной плате мужчин и женщин. Для этого вы собираете случайную выборку наблюдений, для каждого из которых у вас есть данные про пол  $X_i$  (мужской/женский), индикатор  $D_i$ , отражающий работает ли  $i$ -ый индивид или нет, а для наблюдений с  $D_i = 1$  — логарифм заработной платы  $Y_i$ .

Пусть  $Y_i^*$  является логарифмом потенциальной заработной платы  $i$ -го индивида. Для работающих индивидов ( $D_i = 1$ ) мы видим, что  $Y_i = Y_i^*$ , но для остальных ( $D_i = 0$ )  $Y_i^*$  ненаблюдаема. Наша цель — изучить функцию  $h$  (или по крайней мере что-то про нее):

$$\begin{aligned} Y_i^* &= h(X_i) + \epsilon_i \\ E[\epsilon_i | X_i] &= 0 \end{aligned} \quad (1)$$

Решение о выходе на работу зависит от двух факторов: характеристик индивида и некоторого внешнего стимула (инструмента), который мы будем обозначать  $z$ . Формально каждый индивид  $i$  принимает решение о работе, используя следующее правило при заданном уровне стимула  $z$ :

$$D_i(z) = \{u_i < p(z, X_i)\} \quad (2)$$

где  $p(z, x) \in [0, 1]$  для всех значений  $x, z$ . Пусть  $Z_i$  — подлинное значение стимула, который получает индивид  $i$ , так что  $D_i = D_i(Z_i)$ . В данном задании мы предполагаем, что  $Z_i$  распределено случайным образом, что формально означает:

$$Z_i \text{ не зависит от } (X_i, \epsilon_i, u_i). \quad (3)$$

Также предположим, что  $Z_i$  непрерывно распределено на интервале  $[z_0, z_1]$ .

1. Объясните, почему в большинстве случаев мы получаем следующее:

$$h(X_i) \neq E[Y_i | X_i, D_i = 1] \quad (4)$$

или, другими словами, среднее значение логарифма заработной платы для работающих мужчин и женщин не равно их средней потенциальной заработной плате.

2. Предположим, что  $u_i$  не зависит от  $X_i$  и распределена одинаково на  $[0, 1]$ . Покажите, что следующее выражение верно:

$$E[D_i | Z_i, X_i] = p(Z_i, X_i) \quad (5)$$

3. При условии соблюдения предпосылок из предыдущего вопроса, предположим, что выполняется следующее ограничение:

$$(Z_i, X_i) \text{ не зависят от } (\epsilon_i, u_i). \quad (6)$$

Покажите, что в данном случае выполняется следующее:

$$E[Y_i | X_i, D_i = 1] = h(X_i) + \lambda(p(Z_i, X_i)) \quad (7)$$

где  $\lambda(\cdot)$  — некая неизвестная функция, такая что  $\lambda(1) = 0$ . Объясните, почему предпосылка о независимости может не выполняться, в частности, почему ее выполнение не гарантируется фактом того, что  $Z_i$  случайно распределена, и  $u_i$  не зависит от  $X_i$ .

4. В условиях соблюдения предпосылок из прошлого вопроса предположим, что существуют  $z_m, z_w \in [z_0, z_1]$ , такие что  $p(z_m, 0) = p(z_w, 1)$ . Покажите, что в данном случае мы можем определить следующую разность:

$$\Delta h := h(1) - h(0) \quad (8)$$

Докажите, что до тех пор, пока не существует такой  $z \in [z_0, z_1]$ , что  $\max\{p(z, 0), p(z, 1)\} = 1$ , невозможно выявить, исходя из данных, ни  $h(0)$ , ни  $h(1)$ . Что это означает для нашей первоначальной задачи, т.е. можем ли мы определить гендерный разрыв в заработной плате, исходя из данных?

5. Теперь предположим, что вы хотите применить данную методологию к другому вопросу. В частности, вы хотите определить долю населения, заразившегося определенной болезнью. У вас есть данные исследования, в которых вы наблюдаете характеристики единицы наблюдения  $X_i$ , решение о том, сдавать тест или нет  $D_i$  и результаты теста  $Y_i$ . Кроме того, для каждой единицы вы наблюдаете уровень стимула  $Z_i$ , который был случайным образом назначен данному индивиду для того, чтобы мотивировать его пройти тестирование. Предположим, что тест совершенен, то есть он положителен тогда и только тогда, когда человек заражен. Что говорит анализ, проведенный вами в предыдущих вопросах, в контексте данной проблемы? Когда/При каких условиях мы сможем узнать долю инфицированного населения на основе этих данных?