

Задачи по эконометрике вводного уровня

Задачник

Ф. С. Картаев



Экономический
факультет
МГУ
имени
М.В. Ломоносова

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М. В. Ломоносова
Экономический факультет



Ф. С. Картаев

Задачи по эконометрике вводного уровня

Задачник

Москва
2026

УДК 330.43
ББК 65в6
К27

Каргаев Ф. С.

К27 **Задачи по эконометрике вводного уровня:** Задачник. — М.: Экономический факультет МГУ имени М. В. Ломоносова, 2026. — 55 с. — URL: <https://www.econ.msu.ru/elibrary/is/bef/#top>

ISBN 978-5-907909-12-0

УДК 330.43
ББК 65в6

ISBN 978-5-907909-12-0

© Экономический факультет
МГУ имени М. В. Ломоносова, 2026

Содержание

Предисловие	4
Семинар 1. Введение. Повторение важных определений, а также первое знакомство с МНК	5
Семинары 2–3. Парная регрессия	9
Семинары 4–5. Классическая линейная модель множественной регрессии	16
Семинары 6–8. Множественная регрессия: мультиколлинеарность, фиктивные переменные, нелинейные модели	24
Семинар 9. Гетероскедастичность	36
Семинар 10. Модель со стохастическими регрессорами и асимптотический подход в эконометрике	40
Семинары 11–12. Инструментальные переменные	43

Предисловие

Задачник обобщает опыт ведения семинарских занятий по курсу эконометрики на экономическом факультете МГУ имени М. В. Ломоносова и содержит как расчетные, так и теоретические задания по ключевым темам, связанным с анализом пространственных данных: моделям парной и множественной регрессии, методу наименьших квадратов (МНК), проблемам спецификации уравнений регрессии, асимптотическим методам в эконометрике и инструментальным переменным. Задачник будет полезен преподавателям эконометрики, а также студентам, изучающим вводный курс этой дисциплины.

Задачник хорошо сочетается с цифровым учебником: Картаев Ф. С. Дружелюбная эконометрика (<https://books.econ.msu.ru/Introduction-to-Econometrics/>), в котором можно ознакомиться со всеми необходимыми для решения представленных здесь задач методами. Там же доступны и массивы данных для выполнения эмпирических упражнений.

Для удобства работы задачи разделены по темам на семинары, что позволяет использовать представленный материал для проведения занятий по базовому курсу эконометрики.

Семинар 1

Введение.

Повторение важных определений, а также первое знакомство с МНК

Задание 1

В этом задании вам предлагается вспомнить свойства математического ожидания, дисперсии и ковариации.

Предположим, что вы располагаете определенной суммой денег для инвестирования. Для простоты будем считать, что эта сумма равна одному рублю. Вы планируете инвестировать долю w в ценную бумагу A , а оставшуюся долю $(1 - w)$ в ценную бумагу B .

Годовая доходность от вложения в ценную бумагу A составляет R_A (это означает, например, что вложение одного рубля в эту ценную бумагу приносит через год $1 + R_A$ рублей). Известно, что R_A — случайная величина с математическим ожиданием $0,3$ и стандартным отклонением $0,1$.

Годовая доходность от вложения в ценную бумагу B составляет R_B . Известно, что R_B — случайная величина с математическим ожиданием $0,5$ и стандартным отклонением $0,2$.

Корреляция между случайными величинами R_A и R_B составляет $0,25$.

Если вы вложите долю w в ценную бумагу A , долю $(1 - w)$ в ценную бумагу B , то доходность вашего портфеля ценных бумаг составит:

$$R = wR_A + (1 - w)R_B.$$

- (а) Пусть $w = 0,5$. Используя необходимые свойства ковариации и дисперсии, вычислите математическое ожидание и стандартное отклонение R .

- (б) Пусть $w = 0,5$. Вычислите ковариацию доходности портфеля и доходности ценной бумаги A .
- (в) Определите значение w , при котором стандартное отклонение доходности портфеля будет минимальным.

Задание 2

Пусть y_1, y_2, \dots, y_{10} — случайная выборка, математическое ожидание $E(y_i) = \mu$, а дисперсия $V(y_i) = \sigma^2$. Исследователя интересует математическое ожидание μ , и он рассматривает три возможные оценки этого параметра, которые можно вычислить на основе имеющихся данных:

$$(1) \hat{\mu} = \frac{y_1 + \dots + y_{10}}{10};$$

$$(2) \hat{\mu} = \frac{y_1 + \dots + y_{10}}{9};$$

$$\hat{\mu} = \frac{1}{15}(y_1 + \dots + y_5) + \frac{2}{15}(y_6 + \dots + y_{10}).$$

- (а) В чем состоит разница между μ и $\hat{\mu}$?
- (б) Какие из указанных в пунктах (1)—(3) оценок являются несмещенными?
- (в) Какая из несмещенных оценок в этом примере более эффективна?

Задание 3

Исследователь планирует изучить эффективность нового лекарства от горной болезни, с которой сталкиваются люди, оказавшись на большой высоте. У него будет возможность собрать данные о 200 альпинистах, часть из которых, находясь на высоте, принимала новое лекарство, а часть — нет. Впоследствии в результате комплексного обследования уровень здоровья каждого из альпинистов будет оценен по специальной 10-балльной шкале (1 — очень плохо, 10 — очень хорошо). После этого исследователь

планирует сравнить средний уровень здоровья альпинистов из группы, принимавшей лекарство, со средним уровнем здоровья альпинистов, которые обходились без него.

Рассматриваются три варианта тестирования лекарства:

- Вариант 1. Каждый альпинист, участвующий в эксперименте, самостоятельно решает, в каких количествах ему принимать лекарство.
- Вариант 2. Все альпинисты-любители принимают лекарство, а все профессиональные альпинисты — нет.
- Вариант 3. Альпинисты участвуют в лотерее, в ходе которой случайным образом определяется, кто из них будет принимать лекарство, а кто не будет.

Какой из трех вариантов предпочтителен, если цель исследователя состоит в получении корректной оценки эффективности лекарства? Аргументируйте свой ответ.

Задание 4

Имеются следующие данные о переменных x и y :

y	1	3	1	3	7		
x	2	2	6	4	6		

- (а) Найдите МНК-оценки параметров в линейной регрессии y на x :

$$\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i.$$

- (б) Вычислите сумму квадратов остатков и коэффициент R^2 .

Задание 5

Докажите, что в парной регрессии y на x (с константой) коэффициент детерминации R^2 равен квадрату выборочного коэффициента корреляции между переменными x и y .

Задание 6

В этой задаче рассматривается влияние преобразования объясняющей переменной на МНК-оценки. Формально обоснуйте свой ответ на каждый из вопросов:

- (а) Что произойдет с МНК-оценками коэффициентов в парной регрессии y на x , если добавить константу c к каждому наблюдению x ? Как изменится (если изменится) коэффициент детерминации R^2 ?
- (б) Что произойдет с МНК-оценками коэффициентов в парной регрессии y на x , если домножить каждое наблюдение x на константу $c \neq 0$? Как изменится (если изменится) коэффициент детерминации R^2 ?

Задание 7

Анализируется модель парной регрессии y на x (с константой). В ходе МНК-оценивания модели на основе данных о 150 наблюдениях исследователь получил следующие результаты:

$$\hat{y}_i = 10,4 + 2,0 \cdot x_i, \quad R^2 = 0,8.$$

Если теперь, используя те же самые данные, оценить параметры модели $\hat{x}_i = \hat{\alpha}_1 + \hat{\alpha}_2 y_i$, то чему будет равна МНК-оценка коэффициента при переменной y ?

Семинары 2–3

Парная регрессия

Задание 1

Исследователь анализирует зависимость потребления риса от уровня дохода для группы из 20 потребителей. (Вспомнив микроэкономику, вы можете догадаться, что если считать всех потребителей однородными, то такую зависимость можно назвать кривой Энгеля.) Все потребители из этой группы сталкиваются с одинаковыми ценами на рис и другие товары, и только уровни дохода у них различны, поэтому исследователь использует модель парной регрессии.

Обозначим x_i — ежемесячный располагаемый доход i -го потребителя (в тысячах денежных единиц), y_i — ежемесячное потребление риса i -м потребителем (в килограммах).

Имеются следующие данные о переменных x и y :

$$\sum_{i=1}^{20} x_i = 20; \quad \sum_{i=1}^{20} x_i^2 = 40; \quad \sum_{i=1}^{20} y_i = 42; \quad \sum_{i=1}^{20} y_i^2 = 108;$$

$$\sum_{i=1}^{20} x_i \cdot y_i = 60.$$

- (а) Вычислите МНК-оценки коэффициентов в регрессии:

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

Запишите полученное уравнение регрессии и коэффициент R^2 .

- (б) При уровне значимости 5% проверьте значимость переменной x .

- (в) Дайте содержательную интерпретацию коэффициента при переменной x .
- (г) Вспомнив соответствующие определения из курса микроэкономики и вычислив необходимую эластичность, определите, является ли рис для этой группы потребителей низкокачественным товаром, товаром первой необходимости или предметом роскоши?
- (д) При уровне значимости 5% проверьте гипотезу о том, что коэффициент β_2 равен единице.
- (е) Постройте 95%-ный доверительный интервал для коэффициента β_2 .

Задание 2

В этом задании вам предлагается проанализировать взаимосвязь между площадью квартиры и ее ценой. Вам доступны следующие данные о московском рынке недвижимости в 2012 г. (файл Price1): *Price* — рыночная цена однокомнатной квартиры в Москве (руб.), выкуп которой был осуществлен с 10.01.2012 по 28.09.2012; *TotalArea* — общая площадь квартиры (m^2).

- (а) Оцените регрессию переменной *Price* на переменную *TotalArea*. Запишите оцененное уравнение регрессии, указав коэффициент детерминации и (в скобках под соответствующими коэффициентами) стандартные ошибки. Постройте диаграмму рассеяния с линией регрессии.
- (б) Является ли коэффициент при переменной *TotalArea* статистически значимым? Дайте содержательную интерпретацию для этого коэффициента.
- (в) Постройте 95%-ный доверительный интервал для коэффициента при переменной *TotalArea*.
- (г) Риелтора интересует ожидаемая цена квартиры площадью 35 m^2 . Постройте прогноз рыночной цены для данной квартиры и 95%-ный доверительный интервал для этого прогноза.

Задание 3

В задании вам предлагается поработать с данными из статьи: Hamermesh, D. S., Parker, A. Beauty in the Classroom: Professors' Pulchritude and Putative Pedagogical Productivity (*Economics of Education Review*, 2005, 24(4):369–376. DOI: 10.1016/j.econedurev.2004.07.013.).

Вам потребуется файл TeachingRatings.

Ниже представлена табл. 1 с описанием переменных, но в этом задании вам потребуются только некоторые из них. Остальные переменные пригодятся на следующих семинарах.

Таблица 1

Описание переменных

Переменная	Описание
<i>Course_eval</i>	Оценка, которую выставили студенты преподавателю по итогам прочитанного им курса. Измеряется по шкале от 1 (очень плохо) до 5 (отлично)
<i>Beauty</i>	Рейтинг физической привлекательности преподавателя (каждого преподавателя оценивало несколько человек, после чего рейтинг вычислялся как средняя из их оценок). Рейтинг нормирован таким образом, чтобы его среднее значение равнялось нулю
<i>Female</i>	Переменная, которая равна единице, если преподаватель женщина, и равна нулю, если он мужчина
<i>Minority</i>	Переменная, которая равна единице, если преподаватель не является белым, и равна нулю, если он белый
<i>NNenglish</i>	Переменная, которая равна нулю, если для преподавателя английский язык является родным, и равна единице в противном случае
<i>Intro</i>	Переменная, которая равна единице, если курс является вводным (обычно такие курсы читаются перво- или второкурсникам), и равна нулю в противном случае
<i>Onecredit</i>	Переменная, которая равна единице, если курс является коротким факультативным курсом (например, йога, танцы или аэробика), и равна нулю в противном случае
<i>Age</i>	Возраст преподавателя

- (а) Оцените регрессию переменной *Course_eval* на переменную *Beauty*.
 (а.1) Запишите оцененное уравнение регрессии, указав коэффициент детерминации и (в скобках под соответствующими

- коэффициентами) стандартные ошибки. Постройте диаграмму рассеяния с линией регрессии.
- (а.2) Является ли коэффициент при переменной *Beauty* статистически значимым? Дайте содержательную интерпретацию для этого коэффициента.
 - (а.3) Постройте 95%-ный доверительный интервал для коэффициента при переменной *Beauty*.
 - (а.4) Велика ли доля дисперсии оценок за курс, которая объясняется красотой профессора? Объясните свой ответ.
- (б) Оцените регрессию переменной *Course_eval* на переменную *Beauty*, используя только данные о преподавателях-женщинах, и выполните для этого случая задания (а.1)–(а.3).
- (в) Оцените регрессию переменной *Course_eval* на переменную *Beauty*, используя только данные о преподавателях-мужчинах, и выполните для этого случая задания (а.1)–(а.3).

Задание 4

Рассмотрим модель регрессии без константы:

$$y_i = \theta x_i + \varepsilon_i, \quad i = 1, 2, 3, \dots, n.$$

При решении предполагайте выполнение предпосылок классической линейной модели регрессии.

- (а) Найдите МНК-оценку для коэффициента θ . Покажите, что оценка является несмещенной.
- (б) Вычислите дисперсию оценки $\hat{\theta}$. Как меняется точность оценки с ростом числа наблюдений?
- (в) На лекции мы обсуждали несколько формул для вычисления R^2 . Для регрессии с константой все они были эквивалентны. Будут ли они эквивалентны для регрессии без константы? Обоснуйте свой ответ. Приведите пример данных, когда в регрессии без константы коэффициент R^2 , рассчитанный по какой-нибудь из этих формул, меньше нуля или больше единицы.

Задание 5

Рассмотрим модель регрессии на константу:

$$y_i = \theta + \varepsilon_i, \quad i = 1, 2, 3, \dots, n.$$

При решении предполагайте выполнение предпосылок классической линейной модели регрессии.

- (а) Найдите МНК-оценку для коэффициента θ . Покажите, что оценка является несмещенной.
- (б) Вычислите дисперсию оценки $\hat{\theta}$. Как меняется точность оценки с ростом числа наблюдений?
- (в) Вычислите R^2 .

Задание 6

Это задание следует решать при помощи электронных таблиц MS Excel или чего-нибудь аналогичного.

Рассматривается модель парной регрессии:

$$y_i = 20 + 3 \cdot x_i + \varepsilon_i.$$

Объясняющая переменная x — детерминированная (неслучайная), принимает значения 1, 2, 3, ..., 10 по 20 раз каждое, таким образом, число наблюдений равно 200.

Ошибки ε_i — независимые, нормально распределенные случайные величины с нулевым математическим ожиданием и дисперсией $\sigma^2 = 1$.

- (а) Осуществите генерацию двухмерной выборки:
 1. Введите значения переменной x в столбец A (ячейки $A2$ – $A201$).
 2. Сгенерируйте ошибки наблюдений: в меню «Анализ данных» выберите пункт «Генерация случайных чисел». В открывшемся меню выберите нормальное распределение с математическим ожиданием 0 и дисперсией 1. В пункте «Случайное рассеивание» (*Random seed*) укажите любое число. В качестве области вывода (*Output range*) укажите интервал $\$B\2 : $\$B\201 .

3. В столбце C (ячейки $C2$ – $C201$) сгенерируйте наблюдаемые значения переменной y по формуле $y_i = 20 + 3 \cdot x_i + \varepsilon_i$.
- (б) Проведите оценку коэффициентов модели парной регрессии. Теперь будем считать, что истинные значения коэффициентов $\beta_1 = 20$ и $\beta_2 = 3$ нам неизвестны. Попробуем оценить их, используя сгенерированную нами выборку и соответствующие формулы МНК-оценок коэффициентов. Вычислите оценки коэффициентов $\hat{\beta}_1$ и $\hat{\beta}_2$, коэффициент детерминации R^2 . Близки ли полученные оценки параметров к их истинным значениям?
- (в) Повторите пункты (а) и (б), увеличив стандартное отклонение случайной ошибки в четыре раза. Сравните полученные модели. Интерпретируйте результат.

Задание 7

Рассматривается классическая линейная модель парной регрессии $y_i = \beta_1 + \beta_2 \cdot x_i + \varepsilon_i$. Имеется следующая информация о 10 наблюдениях анализируемых переменных:

$$\sum_{i=1}^{10} x_i = 20; \quad \sum_{i=1}^{10} x_i^2 = 50; \quad \sum_{i=1}^{10} y_i = 8; \quad \sum_{i=1}^{10} y_i^2 = 26;$$

$$\sum_{i=1}^{10} x_i \cdot y_i = 10.$$

Для одиннадцатого наблюдения дано $x_{11} = 5$. Предполагая, что это наблюдение удовлетворяет исходной модели, вычислите наилучший линейный несмещенный прогноз y_{11} и оцените его точность, построив для него 95%-ный доверительный интервал.

Задание 8

Имеются некоторые данные о переменных x и y :

y	16	9	7	5	3
x	12	9	6	3	0

- (а) Вычислите МНК-оценки коэффициентов в регрессии:

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

- (б) Вычислите сумму квадратов остатков и R^2 .
- (в) Постройте 95%-ный доверительный интервал для коэффициента β_2 .

Семинары 4–5

Классическая линейная модель множественной регрессии

Задание 1

На основе 20 наблюдений была оценена следующая модель регрессии (в скобках указаны стандартные ошибки оценок коэффициентов):

$$\hat{y}_i = 2,4 + \underset{(0,6)}{6,9}x_i + \underset{(0,3)}{5,1}w_i.$$

Кроме того, известно, что $TSS = 2000$, а сумма квадратов остатков равна 200.

- (а) Вычислите значение коэффициента R^2 , значение скорректированного коэффициента R^2_{adj} и стандартную ошибку регрессии.
- (б) Проверьте значимость уравнения в целом: сформулируйте и проверьте гипотезу о том, что все коэффициенты при переменных уравнения одновременно равны нулю.
- (в) Значим ли коэффициент при переменной x ? Сформулируйте и проверьте соответствующую гипотезу.
- (г) Проверьте гипотезу о том, что коэффициент при переменной x равен 7.
- (д) Постройте 99%-ный доверительный интервал для коэффициента при переменной x .
- (е) После того как исследователь добавил в модель еще две переменные (p и s), R^2 в этой модели увеличился до 0,95. Осуществив соответствующий тест, определите, стоило ли добавлять в модель эти переменные?

Примечание: все гипотезы в этой задаче проверяйте при уровне значимости 1%. Приводите расчетные значения используемых t - и F -статистик, а также критические значения из соответствующих таблиц с указанием числа степеней свободы.

Задание 2. Моделирование отдачи от образования

Исходный файл с данными EARNINGS.xls доступен на сайте on.econ.

В задаче используются данные из учебника К. Доугерти¹. Первоначальный источник: *National Longitudinal of Youth*. Эта база данных представляет собой результаты обследования общенациональной американской репрезентативной выборки мужчин и женщин.

Цель нашего исследования состоит в том, чтобы определить, влияет ли образование на уровень заработной платы типичного работника в США.

На этом семинаре будут предприняты только первые шаги в достижении данной цели. Позже мы вернемся к обсуждению отдачи от образования.

В вашем распоряжении имеются следующие данные о 540 работниках (270 мужчин и 270 женщин):

Earnings — текущий часовой заработок, долл. США;

S — продолжительность обучения (число полных лет обучения);

Exp — общий стаж работы после окончания учебы, лет;

Female — пол респондента (0 — для мужчин, 1 — для женщин).

Проведите предварительный анализ данных:

- (а) Импортируйте данные в эконометрический пакет. Вычислите и проанализируйте описательные статистики для переменных *Earnings*, *S*, *Exp*, *Female*. Скопируйте полученную таблицу в MS Word².

¹ Доугерти К. Введение в эконометрику: учеб. 3-е изд. / пер. с англ. М.: ИНФРА-М, 2009. Используется сет данных EAEF21. Подробное описание данных можно найти в приложении этого учебника (с. 444).

² Если на вашем компьютере установлен TeX, то можно взаимодействовать и с ним тоже. Если вам удобнее этот вариант, то вы можете использовать его.

- (б) Вычислите матрицу парных коэффициентов корреляции между переменными. Интерпретируйте полученные результаты: соответствуют ли знаки коэффициентов вашим ожиданиям?
- (в) Постройте диаграмму рассеяния, характеризующую зависимость *Earnings* от *S*. Постройте диаграмму рассеяния, характеризующую зависимость *Earnings* от *Exp*. Интерпретируйте результаты.
- (г) Оцените параметры модели парной регрессии:

$$Earnings_i = \beta_1 + \beta_2 S_i + \varepsilon_i.$$

Запишите полученное уравнение регрессии, указав коэффициент R^2 и стандартные ошибки оценок коэффициентов.

Дайте ответы на следующие вопросы:

1. Значимо ли уровень образования влияет на заработок?
2. Чему равна стандартная ошибка регрессии и как можно интерпретировать этот результат?

Охарактеризуйте общее качество уравнения регрессии.

- (д) Оцените параметры модели множественной регрессии:

$$Earnings_i = \beta_1 + \beta_2 S_i + \beta_3 Exp_i + \varepsilon_i.$$

Запишите полученное уравнение регрессии, указав коэффициент R^2 и стандартные ошибки оценок коэффициентов.

Дайте ответы на следующие вопросы:

1. Значимо ли уравнение в целом?
2. Используя 1%-ный уровень значимости, укажите, какие из факторов значимо влияют на заработок. Соответствуют ли знаки коэффициентов вашим ожиданиям?
3. Дайте содержательную интерпретацию коэффициента при переменной *Exp*.
4. Дайте содержательную интерпретацию коэффициента при переменной *S*.

Коэффициенты в уравнении оказались значимыми, однако R^2 довольно низкий. Это может свидетельствовать о том, что есть и другие важные факторы, влияющие на нашу зависимую переменную, помимо уже включенных в модель. Оцените, улучшается ли качество модели при добавлении некоторых новых переменных.

- (е) Оцените регрессию заново, добавив в модель переменную *Female*. В чем особенность этой переменной по сравнению с остальными? Оправданно ли включение данной переменной в модель с содержательной точки зрения? Оправданно ли включение переменной с точки зрения теста на значимость? Как можно интерпретировать полученный результат?
- (ж) Представьте три оцененные модели в виде единой сводной таблицы. Сохраните ее в MS Word.
- (з) Используя тест для сравнения «короткой» и «длинной» регрессии, сравните модель из пункта (е) и модель парной регрессии из пункта (г). Интерпретируйте полученный результат.
- (и) Оцените модель из пункта (д) отдельно для мужчин и женщин. Сравните полученные результаты.

Дополнительные замечания:

- Нам пока не удалось получить полностью удовлетворительную модель, поэтому мы еще вернемся к этому примеру, когда будем анализировать более продвинутые методы анализа пространственных выборок. Возможно, дело в том, что мы учли не все важные контрольные переменные или неправильно выбрали форму зависимости.
- С учетом проблемы эндогенности следует осторожно относиться к полученным оценкам влияния образования на заработную плату. Дело в том, что образование обычно коррелировано с ненаблюдаемыми характеристиками (например, уровнями интеллекта и мотивации), которые тоже влияют на заработную плату индивида. Таким образом, может оказаться, что зарплата более образованных людей больше *не потому, что уровень их образования выше, а потому, что они способнее других*. Однако, возможно, что от образования все-таки есть прок³. Для того чтобы отделить эффект влияния ненаблюдаемых характеристик от эффекта самого образования, используют, например, метод инструментальных переменных, который будет обсуждаться в дальнейшем.

³ Преподавателям хотелось бы в это верить ☺

Задание 3

Исследуется зависимость среднедушевого потребления алкоголя по странам мира от различных факторов.

$$\text{Модель 1: } Alco = \beta_1 + \beta_2 \cdot GDP_i + \beta_3 \cdot Musl_i + \beta_4 \cdot Budd_i + \beta_5 \cdot Hindu_i + \varepsilon_i,$$

где $Alco_i$ — среднедушевое потребление чистого спирта на человека, л; GDP_i — ВВП на душу населения, долл. США; $Musl_i$, $Budd_i$, $Hindu_i$ — доли населения, исповедующего соответственно мусульманство, буддизм и индуизм, в % от общей численности населения. В ходе МНК-оценивания модели на основе данных о 180 странах получены следующие результаты: сумма квадратов остатков⁴ $ESS = 200$, объясненная сумма квадратов $RSS = 300$.

Также для проверки гипотезы о том, что религия не оказывает существенного влияния на потребление алкоголя, были оценены параметры модели 2.

$$\text{Модель 2: } Alco_i = \beta_1 + \beta_2 \cdot GDP_i + \varepsilon_i.$$

В модели 2 по сравнению с моделью 1 значение объясненной суммы квадратов RSS изменилось на 100.

- (а) Вычислите R^2 в модели 1.
- (б) В модели 2 по сравнению с моделью 1 значение RSS увеличилось или уменьшилось? Почему? Вычислите R^2 в модели 2.
- (в) Влияет ли религия на потребление алкоголя? Сделайте вывод на основе соответствующего статистического теста.

⁴ Напомним, что в разных учебниках сумма квадратов остатков обозначается по-разному: где-то ESS (*error sum of squares*), где-то RSS (*residuals sum of squares*), а где-то другими способами. Мы здесь используем обозначения учебника: Картаев Ф. С. Введение в эконометрику. М.: Экономический факультет МГУ имени М. В. Ломоносова, 2019. 472 с. Однако призываем вас быть внимательными при чтении разных источников.

Задание 4

Рассматривается классическая линейная модель множественной регрессии $y_i = \beta_1 + \beta_2 x_i^{(2)} + \beta_3 x_i^{(3)} + \varepsilon_i$. Имеются следующие данные о тысяче наблюдений соответствующих переменных:

$$\sum_{i=1}^{1000} x_i^{(2)} = 1000; \sum_{i=1}^{1000} x_i^{(3)} = 1000; \sum_{i=1}^{1000} x_i^{(2)} x_i^{(3)} = 1000; \sum_{i=1}^{1000} (x_i^{(2)})^2 = 3000;$$

$$\sum_{i=1}^{1000} (x_i^{(3)})^2 = 2000;$$

$$\sum_{i=1}^{1000} x_i^{(2)} y_i = 1000; \sum_{i=1}^{1000} x_i^{(3)} y_i = 2000; \sum_{i=1}^{1000} y_i = 0.$$

- Вычислите МНК-оценки коэффициентов модели.
- Пусть также известно, что сумма квадратов остатков в оцененной регрессии оказалась равна 997 000. Вычислите оценку ковариационной матрицы вектора оценок коэффициентов.
- Запишите оцененное уравнение регрессии в стандартной форме, указав в скобках под оценками коэффициентов соответствующие стандартные ошибки. При уровне значимости 5% проверьте значимость переменной $x^{(3)}$.
- При уровне значимости 5% проверьте гипотезу $\beta_2 + 2\beta_3 = 5$.

Задание 5

Оценивание модели производственной функции типичной фирмы некоторой отрасли $\ln y_i = \beta_1 + \beta_2 \cdot \ln x_i^{(2)} + \beta_3 \cdot \ln x_i^{(3)} + \varepsilon_i$ с помощью МНК по 27 наблюдениям дало следующие результаты: $\ln \hat{y}_i = 5 + 0,5 \cdot \ln x_i^{(2)} + 0,4 \cdot \ln x_i^{(3)}$, $\sum e_i^2 = 40$.

Также известно, что $(X^T X)^{-1} = 10^{-3} \begin{pmatrix} 4890 & -100 & -25 \\ -100 & 2,25 & 0,25 \\ -25 & 0,25 & 0,5 \end{pmatrix}$,

где X — матрица регрессоров (результаты приведены с округлением).

- (а) Проверьте значимость каждого из коэффициентов.
- (б) Можно ли утверждать, что технология производства типичной фирмы данной отрасли характеризуется постоянной отдачей от масштаба? Сформулируйте соответствующую гипотезу в терминах модели. Протестируйте ее, используя соответствующий F -тест на линейное ограничение.
- (в) Проверьте ту же самую гипотезу, используя t -тест. Можно ли использовать этот тест, если вы тестируете одновременное выполнение двух линейных ограничений?

Задание 6

В городе N цена жилого дома (обозначим эту переменную $Price$) линейно зависит от его жилой площади ($Area$), количества этажей ($Height$) и удаленности от центра города ($Distance$). Жители города любят жить в домах побольше, пониже и поближе к центру. Однако из-за того, что в центре города мало свободной земли, в среднем чем ближе к центру расположен дом, тем у него меньше площадь. Высокие дома в среднем характеризуются большей жилой площадью, чем низкие. Исследователь хочет оценить, на сколько рублей увеличивается цена квартиры в результате увеличения ее площади на единицу. Не включая в уравнение прочие упущенные факторы, он оценивает параметры модели парной регрессии:

$$\widehat{Price}_i = \hat{\beta}_1 + \hat{\beta}_2 \cdot Area_i.$$

Будет ли оценка коэффициента при переменной $Area$ завышенной, заниженной или несмещенной? Формально обоснуйте свой ответ.

Задание 7

Имеются данные о 100 наблюдениях переменных x и y :

$$\sum_{i=1}^{100} x_i = 0; \quad \sum_{i=1}^{100} x_i^2 = 100; \quad \sum_{i=1}^{100} y_i = 1000;$$

$$\sum_{i=1}^{100} y_i^2 = 10\,200; \quad \sum_{i=1}^{100} x_i y_i = 100.$$

- (а) Рассматривается классическая линейная модель регрессии:

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

Найдите МНК-оценки параметров модели.

- (б) Вычислите сумму квадратов остатков и значение коэффициента R^2 .
- (в) Пусть теперь имеется информация о переменной z :

$$\sum_{i=1}^{100} z_i = 0; \quad \sum_{i=1}^{100} z_i^2 = 300; \quad \sum_{i=1}^{100} x_i z_i = 100; \quad \sum_{i=1}^{100} y_i z_i = 200.$$

Рассматривается классическая линейная модель регрессии:

$$y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \varepsilon_i.$$

Вычислите МНК-оценки коэффициентов модели.

- (г) Для модели из пункта (в) вычислите сумму квадратов остатков и значение коэффициента R^2 .
- (д) Для модели из пункта (в) постройте 99%-ный доверительный интервал для коэффициента β_2 .
- (е) При уровне значимости 1% проверьте гипотезу $\beta_2 = \beta_3 = 0$.

Семинары 6–8

Множественная регрессия: мультиколлинеарность, фиктивные переменные, нелинейные модели

Задание 1

По результатам 1000 наблюдений было оценено следующее уравнение регрессии (в скобках указаны стандартные отклонения оценок коэффициентов):

$$\hat{y}_i = -117,0 + 20,0 \cdot \ln x_i + 20,0 \cdot z_i, \quad R^2 = 0,95.$$

(19,1) (4,2) (5,1)

- (а) Дайте интерпретацию коэффициента при переменной z : выберите **единственную** нужную формулировку из предложенного списка и впишите соответствующее число. При прочих равных условиях: при увеличении переменной z на 1% переменная y увеличивается на ____ %;
- при увеличении переменной z на 1 ед. переменная y увеличивается на ____ %;
- при увеличении переменной z на 1% переменная y увеличивается на ____ ед.;
- при увеличении переменной z на 1 ед. переменная y увеличивается на ____ ед.
- (б) Дайте интерпретацию коэффициента при переменной $\ln x$: выберите **единственную** нужную формулировку из предложенного

списка и впишите соответствующее число. При прочих равных условиях:

при увеличении переменной x на 1% переменная y увеличивается на ____ %;

при увеличении переменной x на 1 ед. переменная y увеличивается на ____ %;

при увеличении переменной x на 1% переменная y увеличивается на ____ ед.;

при увеличении переменной x на 1 ед. переменная y увеличивается на ____ ед.

Задание 2

По результатам 1000 наблюдений было оценено следующее уравнение регрессии (в скобках указаны стандартные отклонения оценок коэффициентов):

$$\ln \hat{y}_i = -10,0 + 0,07 \cdot \ln x_i + 0,07 \cdot z_i + 0,90 \cdot d_i, \quad R^2 = 0,95.$$

(2,0)
(0,01)
(0,01)
(0,1)

- (а) Дайте интерпретацию коэффициента при переменной $\ln x$: выберите **единственную** нужную формулировку из предложенного списка и впишите соответствующее число. При прочих равных условиях:
- при увеличении переменной x на 1% переменная y увеличивается на ____ %;
- при увеличении переменной x на 1 ед. переменная y увеличивается на ____ %;
- при увеличении переменной x на 1% переменная y увеличивается на ____ ед.;
- при увеличении переменной x на 1 ед. переменная y увеличивается на ____ ед.
- (б) Дайте интерпретацию коэффициента при переменной z : выберите **единственную** нужную формулировку из предложенного списка и впишите соответствующее число. При прочих равных условиях:

при увеличении переменной z на 1% переменная y увеличивается на ____ %;

при увеличении переменной z на 1 ед. переменная y увеличивается на ____ %;

при увеличении переменной z на 1% переменная y увеличивается на ____ ед.;

при увеличении переменной z на 1 ед. переменная y увеличивается на ____ ед.

- (в) Дайте интерпретацию коэффициента при переменной d : выберите **единственную** нужную формулировку из предложенного списка и впишите соответствующее число. При прочих равных условиях:
- при увеличении переменной d на 1% переменная y увеличивается на ____ %;
- при увеличении переменной d на 1 ед. переменная y увеличивается на ____ %;
- при увеличении переменной d на 1% переменная y увеличивается на ____ ед.;
- при увеличении переменной d на 1 ед. переменная y увеличивается на ____ ед.

Задание 3

Исследуется зависимость потребления индивида от его располагаемого дохода:

$$c_i = \beta_1 + \beta_2 \cdot \text{income}_i + \varepsilon_i.$$

- (а) Пусть в выборке есть представители двух регионов: А и Б. Как при помощи фиктивных переменных учесть потенциальное различие функций потребления в двух регионах, если вы считаете, что предельная склонность к потреблению в них одинакова, в то время как автономное потребление может быть разным?
- (б) Как при помощи фиктивных переменных проверить гипотезу о том, что предельные склонности к потреблению индивидов с доходом выше и ниже уровня 100 тыс. руб. совпадают?

Задание 4

Оценивается производственная функция типичной фирмы некоторой отрасли:

$$\ln Y_i = \beta_1 + \beta_2 \cdot \ln K_i + \beta_3 \cdot \ln L_i + \varepsilon_i.$$

Выборка состоит из 100 отечественных и 100 иностранных фирм. Исследователь предполагает, что у двух этих групп фирм различаются эластичности выпуска по труду, в то время как все остальные параметры производственной функции идентичны. Как можно учесть данное различие при помощи фиктивной переменной? Опишите переменную, которую следует добавить в модель, и напишите уравнение, которое следует оценить.

Задание 5

На основе данных о 200 квартирах города Готэм было оценено следующее уравнение регрессии (все переменные оказались значимыми, $R^2 = 0,94$):

$$\ln \hat{P}_i = 1,00 + 0,90 \cdot \ln S_i + 0,20 \cdot Center_i \cdot \ln S_i + 0,03 \cdot Center_i + 0,04 \cdot Metro_i + 0,05 \cdot Metro_i \cdot Center_i,$$

где P_i — цена i -й квартиры, тыс. долл.; S_i — площадь i -й квартиры, m^2 ; $Center_i$ — фиктивная переменная, равная единице, если i -я квартира расположена в центре города, и равная нулю в противном случае; $Metro_i$ — фиктивная переменная, равная единице, если i -я квартира расположена в пешей доступности от метро, и равная нулю, если от квартиры до метро надо добираться на общественном транспорте.

- (а) На сколько процентов при прочих равных условиях увеличивается цена квартиры в центре города при увеличении ее площади на 1%?
- (б) Для квартир, расположенных в центре города, на сколько процентов дороже квартира рядом с метро по сравнению с такой же квартирой, расположенной не рядом с метро?

- (в) Для квартир, расположенных не в центре города, на сколько процентов дороже квартира рядом с метро по сравнению с такой же квартирой, расположенной не рядом с метро?

Задание 6

На рынке некоторого города продаются телевизоры только трех фирм: «Альфа», «Бета» и «Гамма». Исследователь анализирует зависимость цены телевизора от диагонали экрана и марки производителя. В его распоряжении имеется информация о 100 моделях телевизоров. Для каждого наблюдения ему известна цена телевизора в долларах (обозначим ее P), длина диагонали экрана в дюймах (обозначим ее $Diag$) и марка производителя. Исследователь ввел следующие фиктивные переменные:

$Alfa_i$ — фиктивная переменная, равная единице, если i -я модель телевизора произведена фирмой «Альфа», и равная нулю во всех остальных случаях;

$Beta_i$ — фиктивная переменная, равная единице, если i -я модель телевизора произведена фирмой «Бета», и равная нулю во всех остальных случаях.

На основе доступных данных было оценено следующее уравнение регрессии (все переменные оказались значимыми):

$$\ln \hat{P}_i = 2,00 + 0,05 \cdot Diag_i + 0,07 \cdot Alfa_i + 0,06 \cdot Beta_i + 0,03 \cdot Diag_i \cdot Beta_i, \\ R^2 = 0,95.$$

- (а) На сколько процентов увеличивается цена телевизора фирмы «Бета» при увеличении диагонали его экрана на один дюйм?
- (б) На сколько процентов увеличивается цена телевизора фирмы «Гамма» при увеличении диагонали его экрана на один дюйм?
- (в) Как можно интерпретировать коэффициент при переменной $Alfa_i$? Заполните пропуски в формулировке: «При прочих равных условиях телевизоры фирмы «Альфа» на _____ % дороже, чем _____».

Задание 7

Исходные данные для этого задания содержатся в файле *Training*.

Руководство крупной торговой сети планирует выяснить, помогает ли тренинг по продажам увеличить эффективность работы менеджеров по продажам.

Для решения этой задачи вы располагаете следующими данными:

Sales — объем продаж данного менеджера, тыс. руб. за период;

Training — фиктивная переменная, равная единице, если в начале данного периода менеджер прошел тренинг по продажам (работники, которые направлялись на курсы, выбирались из общей совокупности работников компании при помощи специальной лотереи);

Female — фиктивная переменная, равная единице для менеджеро-женщин и нулю для мужчин;

Experience — опыт работы менеджера в годах;

Capital — фиктивная переменная, равная единице, если менеджер работает в столичном отделении компании, и равная нулю в противном случае;

IQ — результаты *IQ*-теста менеджера (все менеджеры при приеме на работу в данную компанию проходят этот тест).

- (а) Оцените регрессию переменной *sales* на переменные *Training*, *Female*, *Experience*, *Capital* и *IQ*. В этом и последующих пунктах не забудьте использовать состоятельные в условиях гетероскедастичности стандартные ошибки.

Значимо ли уравнение в целом (при уровне значимости 5%)? Какие из переменных являются значимыми (при уровне значимости 5%)?

Дайте содержательную интерпретацию коэффициента при переменной *Training*.

- (б) Оцените модель из пункта (а) заново, исключив из нее переменную *Capital* и добавив переменную *Training*Capital*. Дайте содержательную интерпретацию коэффициента при добавленной переменной, а также коэффициента при переменной *Training* в новой модели.
- (в) Решите пункты (а) и (б) заново, используя в качестве зависимой переменной **логарифм** переменной *Sales*.

Как поменяется интерпретация коэффициентов при переменных *Training* и *Training*Capital* ?

- (г) Сделайте тест Рамсея для последней из оцененных моделей, интерпретируйте полученный результат.

Задание 8

Исходный файл с данными: ef.xls.

Имеются следующие данные о 150 абитуриентах, сдававших вступительный экзамен в магистратуру экономического факультета:

Y — количество баллов за вступительный экзамен по экономической теории;

D — фиктивная переменная, равная единице, если соответствующий абитуриент посещал подготовительные курсы для поступающих, и равная нулю в противном случае;

EF — фиктивная переменная, равная единице, если соответствующий абитуриент является выпускником бакалавриата данного экономического факультета, и равная нулю в противном случае.

Вас интересует ответ на следующий вопрос: помогают ли курсы подготовиться к экзамену?

- (а) Оцените регрессию переменной Y на константу и переменную D . Интерпретируйте полученный результат.
- (б) Оцените регрессию переменной Y на константу, переменную D и переменную EF . Интерпретируйте полученные результаты. Что можно сказать об уравнении из предыдущего пункта в свете полученных вами результатов — была ли смещена оценка коэффициента при переменной D из-за пропуска существенной переменной (*Omitted variable bias*)?
- (в) Оцените регрессию переменной Y на константу и переменные D , EF и $(D*EF)$. Интерпретируйте полученные результаты: **на сколько баллов** увеличится ожидаемый результат экзамена для выпускника экономического факультета, если он посещал подготовительные курсы? А для выпускника другого вуза?

- (г) Оцените модель из пункта (г), используя теперь в качестве зависимой переменной логарифм количества баллов по экономической теории. Интерпретируйте полученные результаты: **на сколько процентов** увеличится ожидаемый результат экзамена для выпускника экономического факультета, если он посещал подготовительные курсы? А для выпускника другого вуза?

Задание 9

Рассматривается следующая модель:

$$y_i = \beta_1 + \beta_2 \ln x_i + \beta_3 d_i \ln x_i + \varepsilon_i,$$

где y_i — величина спроса i -го индивида на товар Φ , кг; x_i — доход индивида, тыс. руб.; d_i — бинарная переменная, равная единице для мужчин и нулю для женщин.

Оценка параметров модели при помощи МНК на основе данных о 1000 индивидов дала следующие результаты:

$$\hat{y}_i = \underset{(2,12)}{2,34} + \underset{(2,00)}{8,00} \ln x_i - \underset{(4,00)}{6,00} d_i \ln x_i.$$

Также известно, что $\widehat{cov}(\hat{\beta}_2, \hat{\beta}_3) = 2,50$ и $\widehat{corr}(\ln x_i, d_i \ln x_i) = 0,7$.

- (а) Оказывает ли изменение дохода влияние на потребление товара Φ женщинами (сформулируйте и проверьте соответствующую гипотезу при уровне значимости 1%)? Если да, то укажите, на сколько килограммов (или процентов) и в каком направлении меняется спрос на товар Φ в результате увеличения дохода женщины на 1%?
- (б) Оказывает ли изменение дохода влияние на потребление товара Φ мужчинами (сформулируйте и проверьте соответствующую гипотезу при уровне значимости 1%)? Если да, то укажите, на сколько килограммов (или процентов) и в каком направлении меняется спрос на товар Φ в результате увеличения дохода мужчины на 1%?
- (в) Есть ли в рассматриваемой модели существенная мультиколлинеарность? Обоснуйте свой ответ, вычислив соответствующий коэффициент VIF .

Задание 10

Вопросы этого задания связаны с анализом результатов, представленных в табл. 2:

AHE — средний доход, долл./ч, в ценах 1998 г.;

College — фиктивная переменная, равная единице, если работник окончил колледж, и равная нулю в противном случае;

Female — фиктивная переменная, равная единице для женщин и нулю для мужчин;

Age — возраст в годах;

Northeast — фиктивная переменная, равная единице, если работник из Северо-Восточного региона, и равная нулю в противном случае;

Midwest — фиктивная переменная, равная единице, если работник со Среднего Запада, и равная нулю в противном случае;

South — фиктивная переменная, равная единице, если работник из Южного региона, и равная нулю в противном случае;

West — фиктивная переменная, равная единице, если работник из Западного региона, и равная нулю в противном случае.

- (a) Добавьте обозначения «***» (уровень значимости 1%), «**» (5%) и «*» (10%) для того, чтобы охарактеризовать статистическую значимость каждого из коэффициентов.
- (б) В выборке присутствуют данные о работниках только из четырех регионов: *Northeast*, *Midwest*, *South*, *West*. Почему в этом случае переменная *West* не могла быть добавлена в модель (3)?
- (в) Существуют ли статистически значимые региональные особенности на рассматриваемом рынке труда? Обоснуйте свой ответ, используя необходимый статистический тест.
- (г) Дайте содержательную интерпретацию коэффициента при переменной *Northeast* модели (3). Постройте 95%-ный доверительный интервал для этой переменной.

Таблица 2

**Результаты оценки регрессии среднего дохода в час
на фиктивные переменные пола и образования,
а также на другие характеристики работников
на основе данных обследования населения за 1998 г.**

Зависимая переменная: средний доход (<i>AHE</i> , долл./ч)			
Регрессор	(1)	(2)	(3)
<i>College</i>	5,46 (0,21)	5,48 (0,21)	5,44 (0,21)
<i>Female</i>	-2,64 (0,20)	-2,62 (0,20)	-2,62 (0,20)
<i>Age</i>		0,29 (0,04)	0,29 (0,04)
<i>Northeast</i>			0,69 (0,30)
<i>Midwest</i>			0,60 (0,28)
<i>South</i>			-0,27 (0,26)
<i>Intercept</i>	12,69 (0,14)	4,40 (1,05)	4,40 (1,05)
Стандартная ошибка регрессии	6,27	6,22	6,21
R^2	0,176	0,190	0,194
Число наблюдений	4000	4000	4000

Источник: Stock, J., Watson, M. *Introduction to econometrics*. 3-d ed. Pearson, Addison Wesley, 2010.

- (д) Сопоставьте качество трех анализируемых моделей, сделайте выбор в пользу одной из них. Подробно аргументируйте свой ответ.
- (е) На основе выбранной вами в предыдущем пункте модели оцените ожидаемую зарплату для Мэри, которая является 25-летней работницей со Среднего Запада, не учившейся в колледже.

Задание 11

В некотором аналитическом центре работает n аналитиков. Их производительность труда описывается следующей функцией:

$$y_i = \beta_1 + \beta_2 \cdot MSU_i + \beta_3 \cdot HSE_i + \varepsilon_i; \quad i = 1, 2, 3, \dots, n,$$

где y_i — производительность труда i -го аналитика; MSU_i — переменная, равная единице, если i -й аналитик является выпускником магистратуры МГУ, и равная нулю в противном случае; HSE_i — переменная, равная единице, если i -й аналитик является выпускником магистратуры НИУ ВШЭ, и равная нулю в противном случае (все аналитики в компании оканчивали либо только магистратуру МГУ, либо только магистратуру НИУ ВШЭ, либо не оканчивали магистратуру вовсе); ε_i — случайные ошибки, которые удовлетворяют всем предпосылкам классической линейной модели множественной регрессии; β_2 и β_3 — положительные параметры.

Прочие факторы, кроме указанных в уравнении, не влияют на производительность труда аналитиков.

Экономист ААА готовит исследование «Магистерские программы МГУ и производительность труда». На основе данных про всех аналитиков компании при помощи МНК он оценивает параметры модели парной регрессии $\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 \cdot MSU_i$, игнорируя вторую бинарную переменную.

- (а) Пусть производительность аналитиков из МГУ и ВШЭ различна: $\beta_2 > \beta_3$. Будет ли $\hat{\beta}_2$ несмещенной оценкой параметра β_2 ? Если да, то почему? Если нет, то будет ли она завышена или занижена?

Формально обоснуйте свой ответ (вы можете использовать готовые формулы для МНК-оценок коэффициентов, но все остальное следует подробно вывести).

- (б) Изменится ли ваш ответ на вопросы предыдущего пункта, если в действительности производительность аналитиков из МГУ и ВШЭ совпадает ($\beta_2 = \beta_3$)?

Задание 12

В файле Cottage.xlsx имеются следующие данные о двух сотнях коттеджей:

Living_area — жилая площадь коттеджа, м²;

Total_area — общая площадь коттеджа, м²;

Land — площадь участка, на котором расположен коттедж, сотки;

Dist — расстояние от города до участка с коттеджем, км;

Lake — фиктивная переменная, равная единице для коттеджей, расположенных на берегу естественного водоема (реки или озера);

Price — цена коттеджа, млн руб.

Ваша цель исследовать, как цена коттеджа зависит от остальных переменных здесь переменных.

- (а) Проведите предварительный анализ данных. Ожидаете ли вы встретить проблему мультиколлинеарности? Почему? Оцените линейную модель, в которой цена коттеджа зависит от всех переменных, кроме фиктивной (к ней мы вернемся позже). Проанализируйте качество модели и интерпретируйте результаты.
- (б) Подтвердились ли ваши опасения по поводу мультиколлинеарности? Попробуйте устранить проблему, исключив из регрессии переменную *Living_area*. Оцените новую модель и проанализируйте результаты.
- (в) Теперь оцените логарифмическую версию модели из предыдущего пункта. Проанализируйте качество модели и интерпретируйте результаты.
- (г) Для модели из предыдущего пункта осуществите тест Чоу, чтобы выяснить, совпадают ли модели цены коттеджа для домов, расположенных рядом с водоемом и вдали от него.
- (д) Оцените модель, в которой логарифм цены зависит от фиктивной переменной *Lake*, а также от логарифмов переменных *Total_area*, *Land* и *Dist*. Проанализируйте качество модели и интерпретируйте результаты.
- (е) Создайте переменную наклона: $LakC \cdot \ln Total_area$. Оцените модель, в которой логарифм цены зависит от этой новой переменной, а также от логарифмов переменных *Total_area*, *Land* и *Dist*. Проанализируйте качество модели и интерпретируйте результаты.
- (ж) Для последней оцененной модели осуществите тест Рамсея и интерпретируйте его результаты.

Семинар 9

Гетероскедастичность

Задание 1

Исследователь при помощи МНК оценил коэффициенты в следующем уравнении регрессии:

$$y_i = \beta_1 + \beta_2 \cdot x_i^{(1)} + \beta_3 \cdot x_i^{(2)} + \beta_4 \cdot x_i^{(3)} + \varepsilon_i.$$

Число наблюдений равно 180. После этого он решил провести тест Уайта (с перекрестными эффектами) на гетероскедастичность.

- (а) Выпишите в явном виде уравнение регрессии, которое должен оценить исследователь.
- (б) Пусть в уравнении, которое вы записали в предыдущем пункте, коэффициент детерминации оказался равен 0,45. Закончите проведение теста. Сделайте соответствующий вывод.

Задание 2

Рассматривается модель $y_i = \beta_1 + \beta_2 \cdot x_i + \varepsilon_i$, для которой выполнены все предположения классической линейной модели множественной регрессии, за одним исключением: в модели присутствует гетероскедастичность. Известна ее функциональная форма: $\sigma_i^2 = c^2 \cdot x_i^2$.

Y	1,00	1,00	1,50	1,50	1,00
X	1,00	1,00	0,50	0,50	0,25

- (а) Покажите, что если от исходной модели перейти к взвешенной модели:

$$\frac{y_i}{x_i} = \beta_1 \cdot \frac{1}{x_i} + \beta_2 + u_i,$$

то в этой новой модели гетероскедастичность будет отсутствовать.

- (б) Используя данные таблицы, вычислите оценки обобщенного МНК параметров β_1 и β_2 .

Задание 3

Рассматривается модель регрессии $y_i = \beta x_i + \varepsilon_i$, где ε_i — независимые случайные величины с нулевым математическим ожиданием и дисперсией $V(\varepsilon_i) = \sigma_0^2 \cdot z_i$. В вашем распоряжении имеется выборка из n наблюдений (x_i, y_i, z_i) , $z_i > 0$, $i = 1, 2, 3, \dots, n$.

Используя взвешенный МНК, найдите эффективную оценку параметра β . (Задайте оценку $\hat{\beta}$ как функцию от исходных данных (x_i, y_i, z_i) , $i = 1, 2, 3, \dots, n$.)

Задание 4. Анализ факторов, влияющих на качество и количество продукции растениеводства

Исходные данные: файл AGRICULTURE доступен на on.econ.

Используются данные 2010 г. об урожайности яровой и озимой пшеницы в Спасском районе Пензенской области. Данные взяты из учетных реестров компаний Спасского района Пензенской области.

Цель исследования: определить, какие факторы оказывают влияние на урожайность пшеницы. Предполагается, что в первую очередь это трудозатраты на обработку полей, а также расход различных видов удобрений. Ниже представлен полный список переменных, о которых имеется информация:

Prodp — урожайность в денежном выражении, тыс. руб./га;

Size — размер пахотного поля, га;

Labour – трудозатраты, руб./га;

Fung1 – фунгициды протравители семян, руб./га;

Fung2 – фунгициды во время роста, руб./га;

Girb – гербициды, руб./га;

Insec – инсектициды, руб./га;

Ydob1 – аммофос во время сева, руб./га;

Ydob2 – аммиачная селитра во время роста, руб./га.

- (а) Оцените зависимость урожайности в денежном выражении от константы и переменных *Labour*, *Fung1*, *Fung2*, *Ydob1*, *Ydob2*, *Girb*, *Insec*. Запишите уравнение регрессии в стандартной форме, указав коэффициент детерминации и (в скобках под соответствующими коэффициентами) стандартные ошибки. Какие из переменных значимы на 5%-ном уровне значимости?
- (б) Проведите тест Уайта при уровне значимости 5%. Запишите расчетное значение используемой статистики и соответствующее *P*-значение. Сделайте соответствующий вывод.
- (в) Оцените зависимость из пункта (1), используя состоятельные в условиях гетероскедастичности стандартные ошибки. Какие из переменных значимы на 5%-ном уровне значимости? Сравните выводы о значимости переменных: в пункте (1) переменная *Fung2* была значима, а пункте (3) — нет. Выводам какого из пунктов следует доверять? Почему?
- (г) Теперь оцените модель с коррекцией на гетероскедастичность (в программе *Gret!*: *Модель – Другие линейные модели – С коррекцией гетероскедастичности*). Запишите уравнение регрессии в стандартной форме, указав коэффициент детерминации и (в скобках под соответствующими коэффициентами) стандартные ошибки. Сравните полученные результаты с моделями из предыдущих пунктов.
- (д) Какое уравнение для остатков оценивается эконометрическим пакетом для применения взвешенного МНК в предыдущем пункте?

Задание 5

Рассмотрим уравнение регрессии $y_i = \beta + \varepsilon_i$, $i = 1, 2, 3, \dots, n$. Пусть ошибки регрессии удовлетворяют следующим условиям:

$$E(\varepsilon_i) = 0, E(\varepsilon_i \varepsilon_j) = 0 \text{ при } i \neq j, E(\varepsilon_i^2) = \sigma^2 \cdot x_i, x_i > 0.$$

- (а) Вспомните оценку обычного МНК $\hat{\beta}_{\text{МНК}}$ для этой модели. Вычислите ее дисперсию (обратите внимание, что раньше вы вычисляли эту дисперсию в условиях гомоскедастичности, а теперь вам нужно сделать это в условиях гетероскедастичности).
- (б) Найдите оценку взвешенного МНК. Покажите, что она является несмещенной. Вычислите ее дисперсию.
- (в) Сравните дисперсии оценок, полученные в пунктах (а) и (б). Интерпретируйте результат.

Задание 6

Рассмотрим уравнение регрессии $y_i = \beta \cdot x_i + \varepsilon_i$, $i = 1, 2, 3, \dots, n$. Пусть ошибки регрессии удовлетворяют следующим условиям: $E(\varepsilon_i) = 0$, $E(\varepsilon_i \varepsilon_j) = 0$ при

$$i \neq j, E(\varepsilon_i^2) = a \cdot x_i^2, \sum_{i=1}^n x_i^2 = n.$$

- (а) Вспомните оценку обычного МНК $\hat{\beta}_{\text{МНК}}$ для этой модели. Вычислите ее дисперсию.
- (б) Найдите оценку взвешенного МНК. Покажите, что она является несмещенной. Вычислите ее дисперсию.
- (в) Сравните дисперсии оценок, полученные в пунктах (а) и (б). Интерпретируйте результат.

Семинар 10

Модель со стохастическими регрессорами и асимптотический подход в эконометрике

Задание 1

Рассматривается модель:

$$y_i = 1 + 5 \cdot x_i - 3 \cdot x_i^2 + \varepsilon_i.$$

Случайные ошибки имеют нулевое условное математическое ожидание при заданном x_i : $E(\varepsilon_i | x_i) = 0$. Также известно, что $Ex_i = 2$, $Ex_i^2 = 8$.

Вычислите:

- (а) Ey_i ;
- (б) $E(y_i | x_i = 4)$;
- (в) $E(y_i | x_i)$.

Задание 2

Исследователь анализирует влияние опыта работы на доход индивида при помощи следующей модели:

$$\ln Earnings_i = \beta_1 + \beta_2 \cdot Exp_i + \beta_3 \cdot Exp_i^2 + \beta_4 \cdot Female_i + \varepsilon_i,$$

где $Earnings_i$ — доходы i -го индивида, долл./ч; Exp_i — опыт работы i -го индивида, лет; $Female_i$ — бинарная переменная, равная единице, если i -й индивид женщина, и равная нулю, если мужчина.

На основе оценки уравнения по 500 наблюдениям он получил следующие результаты:

$$\widehat{\ln Earnings}_i = 2,263 + 0,0818 Exp_i - 0,0023 Exp_i^2 - 0,3218 Female_i,$$

$$R^2 = 0,103, \quad SEE = 0,565.$$

(В скобках под оценками коэффициентов указаны робастные стандартные ошибки.)

Используя асимптотический подход:

- (а) Вычислите P -значение (P -value) для теста на незначимость переменной $Female$. Вы отвергаете нулевую гипотезу при уровне значимости 5%? А при уровне значимости 1%?
- (б) Вычислите P -значение (P -value) для проверки гипотезы о том, что коэффициент при переменной $Female$ равен $-0,2$. Вы отвергаете нулевую гипотезу при уровне значимости 5%? А при уровне значимости 1%?
- (в) Постройте 99%-ный доверительный интервал для коэффициента при переменной $Female$.

Задание 3

В условиях предыдущего задания обратите внимание, что по мере увеличения опыта работы доход индивида сначала растет, а затем снижается.

- (а) Вычислите оценку опыта работы, при котором доход индивида максимален.
- (б) Пусть также известно, что $\widehat{cov}(\widehat{\beta}_2, \widehat{\beta}_3) = -3 \cdot 10^{-5}$. При 5%-ном уровне значимости проверьте гипотезу о том, что доход индивида максимален при опыте работы, равном 20 годам.

Задание 4

Рассмотрим модель парной регрессии на фиктивную переменную: $y_i = \alpha + \beta x_i + \varepsilon_i$. В вашем распоряжении есть данные о переменных x , y . Известно, что случайная ошибка ε_i может принимать только два значения: $+1$ или -1 . Более того, известно совместное распределение рассматриваемых переменных:

вероятность того, что $x_i = 0$ и $\varepsilon_i = -1$, равна $1/8$;

вероятность того, что $x_i = 0$ и $\varepsilon_i = 1$, равна $3/8$;

вероятность того, что $x_i = 1$ и $\varepsilon_i = -1$, равна $3/8$;

вероятность того, что $x_i = 1$ и $\varepsilon_i = 1$, равна $1/8$.

- (а) Выполняется ли в данном случае предпосылка $E(\varepsilon_i) = 0$? А предпосылка о том, что для любого x_i выполняется равенство $E(\varepsilon_i | x_i) = 0$?
- (б) Вычислите предел по вероятности для МНК-оценки параметра β . Интерпретируйте полученный результат.
- (в) Предложите какую-нибудь состоятельную оценку параметра β .

Задание 5

Рассмотрим регрессионную модель:

$$y_i = \beta_1 \cdot x_i + \beta_2 \cdot w_i + \varepsilon_i.$$

Известно, что безусловные математические ожидания регрессоров x_i , w_i равны нулю и x_i распределены независимо от (w_i, ε_i) .

Для представленной модели выполнены все предпосылки линейной модели со стохастическими регрессорами (которые сформулированы на лекции) за одним исключением: w_i положительно коррелирована с ε_i .

- (а) Будет ли МНК-оценка $\hat{\beta}_1$ состоятельной?
- (б) Будет ли МНК-оценка $\hat{\beta}_2$ состоятельной?

Семинары 11–12

Инструментальные переменные

Задание 1

Анализируется модель $y_i = \beta_1 + \beta_2 \cdot x_i + \varepsilon_i$, $i = 1, 2, 3, \dots, n$. Известно, что случайные ошибки коррелированы с объясняющей переменной x , следовательно, обычный МНК даст несостоятельные оценки. Поэтому исследователь использует следующую процедуру оценивания (называемую двухшаговым МНК):

Шаг 1. Оценивает регрессию переменной x на переменную z :

$$\widehat{x}_i = \widehat{\alpha}_1 + \widehat{\alpha}_2 \cdot z_i.$$

Шаг 2. Оценивает регрессию переменной y на прогнозные значения переменной, полученные на предыдущем шаге: $\widehat{y}_i = \widetilde{\beta}_1 + \widetilde{\beta}_2 \cdot \widehat{x}_i$.

Выведите формулы оценок коэффициентов $\widetilde{\beta}_1$ и $\widetilde{\beta}_2$, получаемых в результате реализации этой процедуры.

Задание 2

На рынке сигарет в некоторой стране функция спроса в i -м регионе имеет вид:

$$\ln(Q_i) = \beta_1 + \beta_2 \cdot \ln(P_i) + \varepsilon_i.$$

Функция предложения описывается соотношением:

$$\ln(Q_i) = \gamma_1 + \gamma_2 \cdot \ln(P_i) + \gamma_3 \cdot \ln(T_i) + u_i,$$

где Q_i — количество сигарет в i -м регионе; P_i — цена сигарет в i -м регионе; T_i — налог с продаж в i -м регионе; ε_i — независимые и одинаково распределенные случайные величины, характеризующие шоки спроса (не коррелированы с налогами); u_i — независимые и одинаково распределенные случайные величины, характеризующие шоки предложения.

- (а) Объясните, почему МНК-оценка эластичности спроса по цене в рассматриваемой модели будет несостоятельной. Для этого вычислите предел по вероятности для МНК-оценки $\hat{\beta}_2^{OLS}$. Определите, если это возможно, будет ли МНК давать завышенную или заниженную оценку эластичности спроса?
- (б) Предложите процедуру состоятельного оценивания эластичности спроса по цене. Формально обоснуйте свой ответ, вычислив соответствующий предел по вероятности.

Задание 3

Исследователь анализирует спрос на сигареты в 48 американских штатах. Он исследует зависимость величины спроса (Q) от цены (P), используя в качестве инструмента для цены ставку налога, взимаемого с производителей в соответствующем штате (T). Ниже представлены результаты применения двухшагового МНК.

Регрессия первого шага:

$$\widehat{\ln P}_i = 4,63 + 0,03 \cdot \ln T_i, \quad R^2 = 0,47.$$

(0,03) (0,005)

Регрессия второго шага (в скобках указаны робастные к гетероскедастичности стандартные ошибки для двухшагового МНК):

$$\widehat{\ln Q}_i = 9,72 - 1,08 \cdot \widehat{\ln P}_i.$$

(1,53) (0,32)

- (а) Является ли инструмент рассматриваемой модели слабым?
- (б) Постройте 95%-ный доверительный интервал для эластичности спроса по цене. Можно ли на основе полученного интервала утверждать, что цена значимо влияет на потребление сигарет? Можно ли на основе полученного интервала утверждать, что спрос на сигареты является эластичным?

Задание 4

Пусть в рамках предыдущей задачи исследователь теперь считает, что на спрос влияет не только цена сигарет, но и уровень доходов потребителей. Таким образом, он анализирует модель:

$$\ln(Q_i) = \beta_0 + \beta_1 \cdot \ln(P_i) + \beta_2 \cdot \ln(Income_i) + \varepsilon_i,$$

где $\ln(P_i)$ — логарифм цены сигарет в i -м регионе, эндогенная переменная; $\ln(Income_i)$ — логарифм дохода на душу населения в i -м регионе, экзогенная переменная.

Исследователь использует два инструмента для эндогенной переменной: $\ln(rtaxso_i)$ — логарифм налога с продаж, общего для всех товаров в i -м регионе; $\ln(rtax_i)$ — логарифм акциза на продажу сигарет в i -м регионе.

- (а) Запишите уравнение, которое следует оценить исследователю в качестве первого шага двухшагового МНК в рассматриваемой модели.
- (б) F -статистику для тестирования какой гипотезы следует вычислить исследователю, чтобы выяснить, являются ли инструменты слабыми? Запишите в явном виде формулу для данной F -статистики, расшифровав все обозначения. Пусть эта F -статистика оказалась равна 900. Интерпретируйте полученный результат.
- (в) При оценивании регрессии второго шага исследователь получил следующую модель (в скобках указаны робастные к гетероскедастичности стандартные ошибки для двухшагового МНК):

$$\widehat{\ln(Q_i)} = 8,40 - 1,20 \cdot \widehat{\ln P_i} + 0,46 \cdot \ln(Income_i),$$

(1,20)
(0,20)
(0,31)

Какие из переменных являются значимыми? Дайте содержательную интерпретацию полученных результатов (не забудьте, что следует интерпретировать только коэффициенты при значимых переменных).

Задание 5

Исследователя интересует ответ на вопрос: *как переменная X влияет на переменную Y ?* Исследователь предполагает, что переменная X является эндогенной. Он располагает данными о переменной X , экзогенной переменной W , а также о нескольких переменных, которые могут использоваться как инструменты для переменной X . Обозначим эти потенциальные инструменты $Z^{(1)}$, $Z^{(2)}$, $Z^{(3)}$ и $Z^{(4)}$. Исследователь оценил пять моделей, используя двухшаговый МНК и различные комбинации инструментов. Результаты оценивания моделей представлены в табл. 3. Какая из них позволяет наиболее адекватно ответить на вопрос исследователя? Аргументируйте свой выбор.

Почему в левой нижней ячейке таблицы стоит прочерк?

Таблица 3

Модели переменной Y

Dependent variable: Y	Number of observations: 964				
Regressor	(1)	(2)	(3)	(4)	(5)
X	7,1 (0,8)	7,0 (1,2)	3,4 (0,1)	3,2 (0,2)	9,9 (2,1)
W	4,2 (0,9)	4,0 (2,3)	3,3 (0,8)	4,1 (1,3)	2,9 (2,1)
<i>Intercept</i>	1,4 (0,2)	1,5 (0,2)	1,3 (0,3)	1,2 (0,3)	1,8 (0,4)
Instrumental variable(s)	$Z^{(1)}$	$Z^{(1)}, Z^{(2)}$	$Z^{(2)}, Z^{(3)}$	$Z^{(1)}, Z^{(3)}$	$Z^{(1)}, Z^{(4)}$
First-stage F-statistic	2,1	34,2	96,7	46,8	3,2
P-value for overidentifying restrictions J-test	—	0,127	0,001	0,009	0,181

Задание 6

В некоторой отрасли промышленности фирмы определяют запасы готовой продукции в зависимости от ожидаемых годовых объемов продаж (модель 1):

$$y_i = \beta_1 + \beta_2 \cdot x_i^e,$$

где y_i — запасы готовой продукции i -й фирмы в 2012 г.; x_i^e — ожидаемый i -й фирмой годовой объем продаж в 2012 г.

Фактический объем продаж i -й фирмы в 2012 г. (x_i) отличается от ожидаемого на случайную величину u_i :

$$x_i = x_i^e + u_i, E(u_i) = 0, V(u_i) = \sigma_u^2 = \text{const.}$$

При этом распределение u_i независимо от x_i^e : $\text{Cov}(x_i^e, u_i) = 0$.

Информация об ожиданиях фирм недоступна исследователю. Конечно, он мог бы провести опрос, но у него нет уверенности в том, что фирмы предоставят ему честные ответы о своих ожиданиях. Зато в распоряжении исследователя есть данные о фактических объемах продаж.

Используя обычный МНК, исследователь оценивает параметры для модели 2:

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

- (а) Покажите, что $\varepsilon_i = -\beta_2 \cdot u_i$.
- (б) Вычислите предел по вероятности для МНК-оценки $\widehat{\beta_2}^{MНК}$. Интерпретируйте результат.
- (в) Пусть исследователь применяет двухшаговый МНК для получения оценок коэффициентов модели 2, используя переменную z как инструмент для переменной x .

Какими свойствами должна обладать переменная z , для того чтобы полученная исследователем оценка коэффициента была состоятельной? Обоснуйте свой ответ формально, вычислив соответствующий предел по вероятности.

- (г) Предложите пример переменной, которую можно было бы использовать как инструмент для фактического объема продаж в 2012 г. Аргументируйте свой выбор.

Задание 7

Оценивание отдачи от образования.

В задаче используются данные из: Вербик М. Путеводитель по современной эконометрике / пер. с англ. В. А. Банникова. Науч. ред. и предисл. С. А. Айвазяна. М.: Научная книга, 2008.

Файл с исходными данными: *schooling*.

Имеются следующие данные примерно о 3 тыс. граждан США за 1976 г.:

Ed76 — количество лет обучения;

Age76 — возраст, лет;

Exp76 — опыт работы, лет;

Wage76 — зарплата (центов/ч);

Lwage76 — логарифм зарплат;

Black — фиктивная переменная, которая равна 1, если рассматриваемый индивид афроамериканец;

Smsa76 — фиктивная переменная, которая равна 1, если индивид жил в крупном городе с пригородами (метрополией);

Nearc4 — фиктивная переменная, которая равна 1, если индивид проживал рядом с колледжем;

South76 — фиктивная переменная, которая равна 1, если индивид жил на юге;

Momed — число лет образования матери;

Daded — число лет образования отца.

Исследовательский вопрос, на который мы попытаемся ответить в этом задании: влияет ли образование на доходы индивидов в США?

- (а) Используя обычный МНК, оцените зависимость логарифма зарплат от количества лет обучения, опыта, опыта в квадрате и фиктивных переменных *Black*, *Smsa76* и *South76*. Не забудьте применить состоятельные в условиях гетероскедастичности стандартные ошибки.
- (б) Какова отдача от образования в представленной модели: на сколько процентов при прочих равных условиях увеличивается зарплата при увеличении числа лет обучения на один год?

Как справедливо замечает в своем исследовании *Card* (1995), число лет обучения может быть коррелировано со случайными ошибками модели. Возможное объяснение следующее: существует ненаблюдаемая характеристика индивида (скажем, его уровень способностей). Так как переменная *способности* не наблюдаема, то она учитывается только в случайных ошибках. Однако если более способные индивиды склонны много учиться, то окажется, что переменная *Ed76* и переменная *способности* коррелированы. А следовательно, переменная *Ed76* коррелирована

со случайными ошибками (возникает проблема эндогенности). Это приведет к смещению оценки коэффициента при *Ed76*. Для решения данной возможной проблемы воспользуемся методом инструментальных переменных.

- (в) Обсудите причины, по которым переменную *Nearc4* можно использовать в качестве инструмента для переменной *Ed76*. Какой недостаток есть у этой переменной в качестве инструмента?

Используя фиктивную переменную наличия вблизи колледжа в качестве инструментальной переменной для продолжительности обучения и считая все контрольные переменные экзогенными, оцените уравнение для логарифма заработной платы.

Интерпретируйте результаты: как изменилась отдача от образования? Осталось ли влияние образования значимым?

- (г) Интерпретируйте результаты теста на слабые инструменты и результаты теста Хаусмана для модели из предыдущего пункта.
- (д) Заново оцените модель из пункта (в), добавив в качестве инструментов для образования переменные *Momed* и *Daded*. Объясните, почему они могут являться подходящими инструментами. В оцененной модели интерпретируйте результаты теста на слабые инструменты, теста Хаусмана и теста Саргана. Почему в этом пункте возможно проведение теста Саргана, а в пункте (в) — нет?
- (е) Представьте результаты моделирования в виде сводной таблицы. Приведите там, помимо оценок коэффициентов, стандартных ошибок и *R*-квадрата, результаты сделанных вами тестов. Сопоставив результаты оценки разных моделей, дайте ответ на исследовательский вопрос, сформулированный в начале задания.

Задание 8. Институты и экономический рост

Задание основано на статье: Acemoglu, D., Johnson, S., Robinson, J.A. The Colonial Origins of Comparative Development: An Empirical Investigation (*American Economic Review*, 2001, 91(5). DOI: 10.1257/aer.91.5.1369). На момент написания этих строк статья имела около 2 тысяч цитирований и занимала 8-е место в списке самых цити-

руемых статей из топ-5 научных экономических журналов в мире (речь о работах, опубликованных с 1990 г. по настоящее время).

Авторы исследования проверяют гипотезу о том, что качество институтов имеет значение для обеспечения высоких темпов экономического роста. Точнее, о том, что страны, в которых права собственности защищены лучше, имеют преимущество в накоплении капитала, а это в конечном счете обеспечивает более высокий долгосрочный ВВП на душу населения.

Исходный файл с данными: *Acemoglu*.

В файле содержатся данные о следующих переменных:

Countryn — название страны;

Shortnam — краткий код страны;

$\lg DP$ — логарифм ВВП на душу населения в 1995 г.;

$\log Mort$ — логарифм уровня смертности колонистов (см. детали в тексте статьи);

Latitude — широта, на которой расположена столица данной страны (измерена как расстояние от экватора и нормирована таким образом, чтобы изменяться в пределах от нуля до единицы);

Prot — мера защиты прав собственности в данной стране (мера определена так, что более высокие значения соответствуют более хорошей защите прав собственности);

Euro — доля населения европейского происхождения в данной стране по состоянию на 1975 г.

В задании вам предлагается воспроизвести основные результаты авторов статьи, следуя их логике. Прежде чем переходить к расчетам, прочитайте статью.

- (а) Постройте диаграмму рассеяния, характеризующую зависимость переменной $\lg dp$ от переменной *Prot*. При помощи обычного МНК оцените параметры соответствующей парной регрессии. Вы должны реплицировать результаты столбца 2 табл. 2 из статьи⁵. Почему полученная зависимость не может быть интерпретирована как причинно-следственная связь? (Назовите не менее

⁵ Здесь и далее возможны незначительные отклонения получаемых вами численных значений от аналогичных значений, представленных в статье. Это объясняется уточнениями в статистических данных, а также использованием различных форм состоятельных в условиях гетероскедастичности стандартных ошибок.

двух различных причин несостоятельности оценки коэффициента при переменной *Prot.*)

- (б) Оцените новую регрессию, включив в исходное уравнение переменную *Latitude*. Ваши результаты должны быть аналогичны результатам столбца 5 табл. 2 из статьи. Опираясь на логику авторов статьи, поясните, почему *Latitude* может быть важной контрольной переменной в данном случае. Интерпретируйте полученные результаты. Подтверждают ли они, что *Latitude* была важной пропущенной переменной для исходной модели?
- (в) Прежде чем переходить к применению двухшагового МНК, оцените уравнение приведенной формы: регрессию логарифма ВВП по смертности поселенцев. Постройте диаграмму рассеяния, характеризующую эту связь. Как можно объяснить полученные результаты?
- (г) Теперь перейдем к оценке влияния качества институтов на логарифм ВВП при помощи двухшаговой МНК. В качестве инструмента используйте переменную *Logmort*.

Поясните содержательно, почему авторы считают данный инструмент валидным, а такая эмпирическая стратегия позволяет корректно тестировать наличие причинно-следственной связи между качеством институтов и ВВП.

- (д) Оцените регрессию первого шага двухшагового МНК. Проиллюстрируйте результат диаграммой рассеяния. Вычислив соответствующую *F*-статистику, поясните, является ли используемый инструмент сильным (релевантным).
- (е) Оцените регрессию второго шага двухшагового МНК. Интерпретируйте полученные результаты.

Как изменилась оценка влияния качества институтов на ВВП по сравнению с оценкой, полученной на основе обычного МНК? Что такое изменение говорит о причине эндогенности регрессора в первоначальной модели (какая из причин эндогенности преобладала в ней)?

- (ж) Оцените альтернативную спецификацию двухшаговой МНК-регрессии, добавив туда контрольную переменную аналогично тому, как это сделано в столбце 2 табл. 4 из статьи. Устойчив ли вывод о положительном влиянии качества институтов на ВВП?

Задание 9. Эмпирический пример

- (а) Используя функцию генерации случайных чисел в Excel, создайте по 2000 наблюдений для трех независимых одинаково (нормально) распределенных случайных величин:

$$\varepsilon_i \sim N(0,1);$$

$$u_i \sim N(0,1);$$

$$v_i \sim N(0,1).$$

- (б) Вычислите по 2000 значений переменных x , y , z , используя следующие расчетные формулы:

$$x_i = \varepsilon_i + u_i;$$

$$z_i = v_i + u_i;$$

$$y_i = 2 + 0,4 \cdot x_i + \varepsilon_i.$$

- (в) Будем считать, что мы не наблюдаем случайные величины ε , u и v , а располагаем только данными о переменных x , y , z . Оцените регрессию y по x ($y_i = \alpha + \beta \cdot x_i + \varepsilon_i$), используя обычный МНК. Проверьте гипотезу $\beta = 0,4$.
- (г) Можно ли доверять полученным МНК-оценкам? Почему?
- (д) Оцените регрессию y на x ($y_i = \alpha + \beta \cdot x_i + \varepsilon_i$), используя двухшаговый МНК (z — инструмент для x). Сопоставьте $\widehat{\beta}_{OLS}$, $\widehat{\beta}_{TSL}$ и истинное значение коэффициента. Интерпретируйте результаты.
- (е) Осуществите тест Хаусмана для сравнения регрессий, полученных в пунктах (б) и (д). Интерпретируйте результаты.

Задание 10

Страна Кейнсиания является закрытой экономикой без непосредственного государственного вмешательства, поэтому основное макроэкономическое тождество в ней имеет следующий вид:

$$Y_t = C_t + I_t,$$

где Y_t — ВВП в месяце t ; C_t — совокупное потребление в месяце t ; I_t — инвестиции в месяце t .

Потребление следующим образом зависит от дохода:

$$C_t = C_a + mpc \cdot Y_t + \varepsilon_t,$$

где ε_t — случайные шоки потребления, которые представляют собой независимые одинаково распределенные случайные величины с нулевым математическим ожиданием и постоянной дисперсией, случайные шоки потребления не коррелированы с инвестициями. Структурных сдвигов в Кейнсиании не происходит, поэтому автономное потребление и предельная склонность к потреблению неизменны.

Вы знаете указанные выше факты об экономике Кейнсиании, а также располагаете ежемесячными данными о динамике потребления, инвестиций и ВВП в этой стране за последние 100 лет. Вас интересует, чему равна предельная склонность к потреблению в Кейнсиании.

Ваш ассистент осуществил предварительные расчеты и вычислил **выборочные** ковариации и дисперсии для имеющихся в вашем распоряжении переменных:

$$\widehat{Cov}(C, Y) = 150, \quad \widehat{var}(C) = 100, \quad \widehat{var}(Y) = 300.$$

- (а) Найдите состоятельную оценку предельной склонности к потреблению в Кейнсиании. Обоснуйте свой ответ.
- (б) Эконометрист Джон Инконсистент, располагая той же самой информацией, что и вы, решил оценить предельную склонность к потреблению. Используя обычный МНК, он оценивает регрессию потребления на доход. Вычислите оценку, которую получит Джон. Объясните, почему результату, который он получит, нельзя доверять?

Картаев Ф. С.

Задачи по эконометрике вводного уровня

Задачник

ISBN 978-5-907909-12-0



9 785907 909120 >

Электронное издание сетевого распространения.
3,5 печ. л. Опубликовано 16.05.2026.
Издательство «ЭФ МГУ имени М.В. Ломоносова»;
www.econ.msu.ru; +7 (495) 939-17-15