

Задачи по Эконометрике

Артамонов Н.В.

18 января 2021 г.

Оглавление

1	Метод наименьших квадратов	5
2	Линейная модель регрессии	9
2.1	Стандартная линейная модель	9
2.2	Стандартная линейная модель. Практика	14
2.3	Проблемы спецификации	19
2.4	Проблемы спецификации. Практика	23
2.5	Линейная модель с гетероскедастичностью	24
2.6	Линейная модель с автокорреляцией	27
3	Системы одновременных уравнений	31
4	Метод инструментальных переменных	35
5	Временные ряды	37
5.1	Модели стационарных временных рядов	37
5.2	Модели стационарных временных рядов. Практика	39
5.3	Модели распределенных лагов	40
5.4	Модели распределенных лагов. Практика	42
5.5	TS – ряды	43
5.6	DS – ряды	45
5.7	DS-ряды. Практика	46
5.8	Многомерные модели временных рядов	48
5.8.1	VAR и коинтеграция	48
5.9	Из пособия	49
6	Бинарный выбор	53
6.1	LPM-модель	53

6.2	LPM-модель. Практика	55
6.3	probit/logit-модель	57
6.4	probit/logit-модель. Практика	61
7	Панельные данные	67
7.1	Модели панельных данных	67
7.2	Модели панельных данных. Практика	71
A	Описание наборов данных	73
B	Результаты оценивания регрессий	83

Глава 1

Метод наименьших квадратов

№1. Пусть задано n наблюдений (точек на плоскости) $\{x_i, y_i\}_{i=1}^n$. Для линейной функции $y = \beta_0 + \beta_1 x$

1. применив метод наименьших квадратов выведите систему нормальных уравнений для нахождения параметров (оптимальной) прямой, наименее уклоняющейся от заданных наблюдений (точек на плоскости);
2. выведите формулы для оценок $\hat{\beta}_0$ и $\hat{\beta}_1$ коэффициентов оптимальной прямой;
3. покажите, что для оценок коэффициентов верно

$$\hat{\beta}_1 = \frac{\text{cov}(x, y)}{\text{Var}(x)}, \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \cdot \bar{x}.$$

№2. Пусть задано n наблюдений (точек на плоскости) $\{x_i, y_i\}_{i=1}^n$. Для линейной функции $y = \beta x$

1. применив метод наименьших квадратов выведите систему нормальных уравнений для нахождения параметров (оптимальной) прямой, наименее уклоняющейся от заданных наблюдений (точек на плоскости);
2. выведите формулы для оценки $\hat{\beta}$ коэффициента оптимальной прямой.

№3. Выведите систему нормальных уравнений для функций

$$y = \beta_0 + \beta_1 x + \beta_2 z \qquad y = \beta_0 + \beta_1 x + \beta_2 x^2$$

№4. Пусть $\hat{\beta}$ есть OLS-оценка коэффициента наклона линейной функции y на x без константы, а $\hat{\gamma}$ – OLS-оценка коэффициента наклона линейной функции x на y без константы. Верно ли для этих оценок равенство

$$\hat{\gamma} = \frac{1}{\hat{\beta}}?$$

№5. Пусть $\hat{\beta}_1$ есть OLS-оценка коэффициента наклона линейной функции y на x с константой, а $\hat{\gamma}_1$ – OLS-оценка коэффициента наклона линейной функции x на y с константой. Верно ли равенство

$$\hat{\gamma}_1 = \frac{1}{\hat{\beta}_1}?$$

Ответ поясните.

№6. Пусть $\hat{\beta}_1$ есть OLS-оценка коэффициента наклона линейной функции y на x с константой, а $\hat{\gamma}_1$ – OLS-оценка коэффициента наклона линейной функции x на y с константой. Покажите, что

$$\hat{\gamma}_1 = \frac{1}{\hat{\beta}_1} \iff \text{corr}(x, y) = \pm 1.$$

№7. Пусть $\hat{\beta}_0, \hat{\beta}_1$ – OLS-оценки коэффициентов линейной функции y на x , а $\hat{\gamma}_0, \hat{\gamma}_1$ – OLS-оценки коэффициентов линейной функции $(c_1 y)$ на $(c_2 x)$ ($c_1, c_2 \neq 0$). Покажите, что

$$\hat{\gamma}_1 = \frac{c_1}{c_2} \cdot \hat{\beta}_1, \quad \hat{\gamma}_0 = c_1 \hat{\beta}_0.$$

№8. Пусть $\hat{\beta}_0, \hat{\beta}_1$ – OLS-оценки коэффициентов линейной функции y на x , а $\hat{\gamma}_0, \hat{\gamma}_1$ – OLS-оценки коэффициентов линейной функции $(y + c_1)$ на $(x + c_2)$. Покажите, что

$$\hat{\gamma}_1 = \hat{\beta}_1, \quad \hat{\gamma}_0 = \hat{\beta}_0 + c_1 - c_2 \hat{\beta}_1.$$

№9. Пусть $\hat{\beta}_0, \hat{\beta}_1$ – OLS-оценки коэффициентов линейной функции y на x , а $\hat{\gamma}_0, \hat{\gamma}_1$ – OLS-оценки коэффициентов линейной функции $(y + cx)$ на x ($c \neq 0$). Как связаны $\hat{\beta}_0, \hat{\beta}_1$ и $\hat{\gamma}_0, \hat{\gamma}_1$

№10. Пусть $\hat{\beta}_0, \hat{\beta}_1$ – OLS-оценки коэффициентов линейной функции y на x , а $\hat{\gamma}_0, \hat{\gamma}_1$ – OLS-оценки коэффициентов линейной функции $y - \bar{y}$ на $x - \bar{x}$. Покажите, что $\hat{\gamma}_0 = 0$ и $\hat{\gamma}_1 = \hat{\beta}_1$.

№11. Рассмотрим линейную функцию $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$. Покажите, что систему нормальных уравнений можно записать в виде

$$\begin{aligned}\beta_0 + \beta_1 \cdot \bar{x}_1 + \beta_2 \cdot \bar{x}_2 &= \bar{y} \\ \text{Var}(x_1)\beta_1 + \text{cov}(x_1, x_2)\beta_2 &= \text{cov}(x_1, y) \\ \text{cov}(x_1, x_2)\beta_1 + \text{Var}(x_2)\beta_2 &= \text{cov}(x_2, y)\end{aligned}$$

Найдите формулы для OLS-оценок $\hat{\beta}_0, \hat{\beta}_1$ и $\hat{\beta}_2$.

№12. Рассмотрим линейную функцию $y = \beta_1 x_1 + \beta_2 x_2$.

1. Выведите систему нормальных уравнений для нахождения OLS-оценок коэффициентов.
2. Найдите явные формулы для OLS-оценок коэффициентов.

Глава 2

Линейная модель регрессии

Для вычисления P-значений используйте следующие функции MS Excel

t-тест (двустор. альтернатива)	СТЫЮДРАСП(t,df,2)
t-тест (одностор. альтернатива)	СТЫЮДРАСП(t,df,1)
F-тест	ФРАСП(F, df1, df2)

2.1 Стандартная линейная модель

№1. Для выборки

y	2	-1	1	3	-2	0
x_1	1	-1	2	0	2	1
x_2	1	2	-1	2	1	2

1. Напишите матрицы \mathbf{X} и \mathbf{y} для модели регрессии

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \text{error}.$$

2. Запишите в матричном виде систему нормальных уравнений.

№2. Рассмотрим парную модель линейной регрессии без константы

$$y_i = \beta x_i + u_i.$$

Покажите, что для этой модели выполнено равенство

$$\sum_{i=1}^n y_i^2 = \sum_{i=1}^n \hat{y}_i^2 + \sum_{i=1}^n e_i^2,$$

где $\hat{y}_i = \hat{\beta} x_i$ – предсказанные значения и $e_i = y_i - \hat{y}_i$ – остатки.

№3. Была оценена производственная функция Кобба – Дугласа:

Зависимая переменная: $\ln(Q)$; объём выборки: 142

	const	$\ln(K)$	$\ln(L)$	$\ln(H)$
coefficient	3.45	0.47	0.12	0.34
Std Error	1.45	0.23	0.56	0.13
t-Statistic	–	–	–	–
Prob.	0.0210	0.0461	0.8312	0.0116

- Дайте интерпретацию коэффициентов модели.
- Заполните прочерки.
- Проверьте значимость коэффициентов при уровне значимости 5%.
- Запишите оцененное уравнение регрессии.
- Дайте экономическую интерпретацию гипотезе

$$H_0 : \beta_1 + \beta_2 + \beta_3 = 1.$$

- Пусть F -статистика для гипотезы из предыдущего пункта равна 1.76. Тестируйте гипотезу. Вычислите P -значение.

№4. По предприятиям энергетической отрасли была оценена производственная функция Кобба–Дугласа (K – капитал, L – труд) по данным за февраль 2006 года.

Зависимая переменная: $\ln(Q)$; объём выборки 30

	const	$\ln(K)$	$\ln(L)$	$R^2 = 0.45$
coefficient	3.45	0.47	0.32	Prob(F -stat)=0.00031
Prob.	0.0210	0.461	0.528	DW = 3.43

- Проверьте значимость влияния капитала, значимость влияния труда, совместную значимость влияния труда и капитала.
- Как можно объяснить результаты предыдущего пункта? Дайте не менее двух возможных объяснений.

№5. Для мужчин возрастной группы 28–38 лет была оценена модель регрессии зависимости почасовой оплаты *wage* от возраста *age*, стажа на текущем месте работы *tenure*, семейного статуса *married*, места жительства *urban*, количества братьев и сестер *sibs*, и результатов IQ-теста.

Зависимая переменная: $\ln(wage)$; объем выборки 935

	const	<i>age</i>	<i>age</i> ²	<i>tenure</i>	<i>married</i>	<i>urban</i>	<i>sibs</i>	<i>iq</i>
coeff.	3.3868	0.1166	-0.0015	0.0113	0.1901	0.1879	-0.0092	0.0082
Prob.	0.0408	0.42423	0.3153	0.0000	0.0000	0.0000	0.0987	0.0000

$R^2 = 0.2085$ F-statistic=34.8769 Prob(F-stat)=0.0000

1. Дайте интерпретацию коэффициентов модели регрессии.
2. Дайте интерпретацию коэффициента R^2 и проверьте значимость регрессии «в целом».
3. Проверьте значимость влияния фактора возраст и фактора квадрат возраста.
4. F-статистика для проверки совместной значимости *age* и *age*² равна 8.4986. Значимо ли совместное влияние факторов возраст и квадрат возраста?
5. Как можно объяснить результаты, полученные в п.3 и п.4? Дайте развернутый ответ.

№6. Для набора данных *sleep75*, Таблица А.3, были оценены несколько моделей регрессии. В Таблице В.1 приведены результаты оценивая, в скобках указаны стандартные ошибки коэффициентов.

1. Запишите спецификации каждой из регрессий
2. Проверьте значимость коэффициентов регрессии №2 и №5
3. Проверьте значимость «в целом» регрессии №5. Обязательно сформулируйте тестируемую гипотезу
4. Значимо ли влияние возраста на продолжительность сна? Дайте развернутый ответ, указав, в частности, тестируемую гипотезу.

5. Значимо ли совместное влияние на продолжительность сна семейного статуса и членства в профсоюзе? Дайте развёрнутый ответ, указав, в частности, тестируемую гипотезу.
6. Предположим, что 31-летний человек живёт в мегаполисе на севере и работает 40 часов в неделю. Используя регрессию №2 оцените продолжительность его сна.

№7. Для набора данных `sleep75`, Таблица A.3, были оценены несколько моделей регрессии. В Таблице B.2 приведены результаты оценивая, в скобках указаны стандартные ошибки коэффициентов.

1. Проверьте значимость коэффициентов регрессии №3
2. Проверьте значимость «в целом» регрессии №5. Обязательно сформулируйте тестируемую гипотезу
3. Значимо ли влияние возраста на продолжительность сна? Дайте развёрнутый ответ, указав, в частности, тестируемую гипотезу.
4. Значимо ли влияют на продолжительность сна семейный статус и членство в профсоюзе? Дайте развёрнутый ответ, указав, в частности, тестируемую гипотезу.
5. Пусть 40-летний человек работает 35 часов в неделю. Используя данные Таблицы B.2 оцените его продолжительность сна (час/нед)

№8 (F-тест/тест Вальда). Для набора данных `Labour`, Таблица A.1, рассмотрим регрессию логарифма выпуска от логарифма капитала, логарифма числа сотрудников и зарплаты.

В таблице 2.1 приведены результаты тестирования некоторых гипотез. Дайте их интерпретацию. Подтверждаются ли эти гипотезы?

Таблица 2.1: Результаты тестирования		
	Гипотеза	Статистика
1	$H_0 : \beta_{capital} + \beta_{labour} = 1$	77.14
2	$H_0 : \beta_{capital} = \beta_{labour}$	310.67

№9 (Тест Чоу). Для набора данных `sleep75`, Таблица A.3, рассмотрим регрессию продолжительности сна (**час/нед!**) на занятость, возраст и квадрат возраста. В Таблице B.3 приведены результаты оценивания регрессии по всей выборке, только по мужчинам и только по женщинам соответственно.

Проведите тест Чоу на структурные изменения и сформулируйте тестируемую гипотезу.

№10 (Тест Чоу). Для набора данных `wage2`, Таблица A.4, рассмотрим регрессию логарифма зарплаты на возраст, `south`, `urban`, IQ. В Таблице B.4 приведены результаты оценивания регрессии по всей выборке, только по женатым и только по неженатым соответственно.

Проведите тест Чоу на структурные изменения и сформулируйте тестируемую гипотезу¹.

Мультиколлинеарность

№11. Для набора данных `sleep75`, Таблица A.3, рассмотрим несколько регрессий. В Таблице B.5 приведены результаты оценивания, в Таблице B.6 приведены показатели VIF для этих регрессий. В Таблице B.7 приведена корреляционная матрица для переменных модели.

1. для первой регрессии тестируйте значимость занятости, квадрата занятости и совместную значимость. Какие можно сделать выводы?
2. для первой регрессии тестируйте значимость возраста, квадрата возраста и совместную значимость. Какие можно сделать выводы?

№12. Для набора данных `sleep75`, Таблица A.3, рассмотрим несколько регрессий. В Таблице B.8 приведены результаты оценивания, в Таблице B.9 приведены показатели VIF для этих регрессий. В Таблице B.10 приведена корреляционная матрица для переменных модели

1. для первой регрессии тестируйте значимость регрессоров с гендерной дамми.
2. Для первой регрессии тестируйте значимость влияния гендера. Какие можно сделать выводы?

¹Указание: используйте соотношение $s^2 = \text{RSS} / (n - k - 1)$

Бинарные переменные

№13. Для набора данных `wage2`, Таблица A.4, были оценены несколько регрессий, Результаты оценивания приведены в Таблице B.11.

Дайте интерпретацию коэффициентов к каждой регрессии.

№14. Для набора данных `sleep75`, Таблица A.3, рассмотрим несколько регрессий. Результаты оценивания приведены в Таблице B.12.

Дайте интерпретацию коэффициентов к каждой регрессии.

№15. Для набора данных `Diamond`, Таблица A.11, рассмотрим несколько регрессий. Результаты оценивания приведены в Таблице B.13.

1. Дайте интерпретацию коэффициентам первой регрессии
2. Тестируйте значимость влияния цвета. Сформулируйте тестируемую гипотезу
3. Тестируйте значимость влияния выбора органа сертификации. Сформулируйте тестируемую гипотезу

2.2 Стандартная линейная модель. Практика

Основы линейной регрессии

Для решения задачи понадобятся следующие пакеты:

```
car, lmtest, wooldridge, Ecdat
```

№1. Для набора данных `sleep75`, Таблица A.3, оцените регрессию продолжительности сна на занятость, возраст, гендерный фактор, место жительства (живет ли в мегаполисе), географический фактор и членство в профсоюзе.

1. Напишите спецификацию модели регрессии
2. Дайте интерпретацию коэффициентам
3. Вычислите TSS, RSS и ESS. Проверьте равенство

$$TSS = ESS + RSS$$

4. Вычислите R^2 и дайте его интерпретацию
5. Проведите t-тест для коэффициентов регрессии (тестируйте значимость коэффициентов). Укажите тестируемые гипотезы
6. Проведите F-тест для регрессии (тестируйте значимость регрессии). Укажите тестируемую гипотезу.

№2. Для набора данных `wage2`, Таблица A.4 оцените линейную регрессию логарифма зарплаты на возраст, семейный статуса, географический фактор `south`, место жительства `urban` и результаты IQ-теста

1. Напишите спецификацию модели регрессии
2. Дайте интерпретацию коэффициентам
3. Вычислите TSS, RSS и ESS. Проверьте равенство

$$TSS = ESS + RSS$$

4. Вычислите R^2 и дайте его интерпретацию
5. Проведите t-тест для коэффициентов регрессии (тестируйте значимость коэффициентов). Укажите тестируемые гипотезы
6. Проведите F-тест для регрессии (тестируйте значимость регрессии). Укажите тестируемую гипотезу.

№3. Для набора данных `Labour`, Таблица A.1 оцените линейную регрессию логарифма выпуска на логарифм капитала и логарифм числа сотрудников.

1. Напишите спецификацию модели регрессии
2. Дайте интерпретацию коэффициентам
3. Вычислите TSS, RSS и ESS. Проверьте равенство

$$TSS = ESS + RSS$$

4. Вычислите R^2 и дайте его интерпретацию

5. Проведите t-тест для коэффициентов регрессии (тестируйте значимость коэффициентов). Укажите тестируемые гипотезы
6. Проведите F-тест для регрессии (тестируйте значимость регрессии). Укажите тестируемую гипотезу.

№4 (F-тест/тест Вальда). Для набора данных `sleep75`, Таблица [A.3](#), рассмотрим регрессию продолжительности сна от занятости, возраста, квадрата возраста, гендерного фактора, семейного статуса, места жительства (живет ли в мегаполисе), географического фактора, членства в профсоюзе.

1. тестируйте значимость влияния возраста и сформулируйте тестируемую гипотезу.
2. тестируйте (совместную) значимость влияния семейного статуса и членства в профсоюзе и сформулируйте тестируемую гипотезу

№5 (F-тест/тест Вальда). Для набора данных `wage2`, Таблица [A.4](#) рассмотрим линейную регрессию логарифма зарплаты на возраст, семейный статус, географический фактор `south`, место жительства `urban`, результаты IQ-теста и уровень образования родителей.

1. тестируйте значимость влияния уровня образования родителей и сформулируйте тестируемую гипотезу.
2. тестируйте гипотезу

$$H_0 : \beta_{meduc} = \beta_{educ}$$

и дайте её интерпретацию.

№6 (F-тест/тест Вальда). Для набора данных `Labour`, Таблица [A.1](#) рассмотрим линейную регрессию логарифма выпуска на логарифм капитала и логарифм числа сотрудников.

1. Тестируйте гипотезу

$$H_0 : \beta_{labour} = \beta_{capital}$$

и дайте её интерпретацию

2. Тестируйте гипотезу

$$H_0 : \beta_{labour} + \beta_{capital} = 1$$

и дайте её интерпретацию

№7 (Тест Чоу). Для набора данных `sleep75`, Таблица [A.3](#), рассмотрим регрессию продолжительности сна от занятости, возраста, квадрата возраста, места жительства (живет ли в мегаполисе) и географического фактора.

1. Тестируйте эту модель на структурные изменения относительно гендерного фактора.
2. Тестируйте эту модель на структурные изменения относительно семейного статуса.

№8 (Тест Чоу). Для набора данных `wage2`, Таблица [A.4](#) рассмотрим линейную регрессию логарифма зарплаты на возраст, результаты IQ-теста и переменную `KWW`

1. Тестируйте эту модель на структурные изменения относительно географического фактора.
2. Тестируйте эту модель на структурные изменения относительно семейного статуса.
3. Тестируйте эту модель на структурные изменения относительно относительно жительства в городе .

Мультиколлинеарность

Для решения задачи понадобятся следующие пакеты:

`car`, `lmtest`, `wooldridge`, `Ecdat`

№9. Для набора данных `wage2`, Таблица [A.4](#) оцените линейную регрессию логарифма зарплаты на возраст, квадрат возраста, семейный статус, географический фактор `south`, место жительства `urban` и результаты IQ-теста.

1. Тестируйте значимость коэффициентов при возрасте и его квадрате
2. Тестируйте значимость влияния возраста на зарплату.
3. Вычислите показатели VIF для факторов. Какие можно сделать выводы?
4. Вычислите описательные статистики для возраста.
5. Нарисуйте график зависимости логарифма зарплаты от возраста
6. Оцените регрессию квадрата возраста на остальные факторы. Какие можно сделать выводы?
7. Нарисуйте графики квадрата возраста против возраста и квадрата возраста против предсказанного квадрата из предыдущей регрессии. Какие можно сделать выводы?
8. Можно ли первоначальную регрессию использовать для прогнозирования зарплаты? Ответ поясните.

№10. Для набора данных `sleep75`, Таблица A.3 оцените регрессию `sleep` на `totwrk`, `male`, `age`, `male*totwrk`, `male*age`.

Как можно прокомментировать полученный результат?

Бинарные переменные

Для решения задачи понадобятся следующие пакеты:

```
car, lmtest, wooldridge
```

№11. Для набора данных `diamonds`, Таблица A.10, оцените следующие регрессии:

1. логарифм цены на логарифм веса и тип огранки
2. логарифм цены на логарифм веса и цвет
3. логарифм цены на логарифм веса и прозрачность

и дайте интерпретацию их коэффициентам.

№12. Для набора данных *Diamond*, Таблица A.11, оцените регрессию логарифма цена на вес, квадрат веса и все факторные переменные

1. дайте интерпретацию значимым коэффициентам
2. тестируйте значимость влияния каждой факторной переменной

и дайте интерпретацию их коэффициентам.

№13. Для набора данных *sleep75*, Таблица A.3 оцените линейную регрессию продолжительности сна от занятости, гендерного фактора, произведения гендерного фактора на занятость, возраста, произведения гендерного фактора на возраст, семейного статуса, членства в профсоюзе.

1. Напишите спецификацию оцениваемой регрессии
2. Дайте интерпретацию значимым коэффициентам регрессии
3. Тестируйте значимость влияния гендерного фактора на продолжительность сна.

2.3 Проблемы спецификации

№1. Исследуется влияние образования *Educ*, пола *Male* и возраста *Age* на почасовую оплату труда *Wage* для возрастной группы 25 – 75 лет. Рассматривается модель регрессии

$$\ln Wage = \beta_0 + \beta_1 Educ + \beta_2 Male + \beta_3 Age + \beta_4 Age^2 + \text{error}$$

С какой целью в модель регрессии был включен квадрат возраста? Какой ожидаемый знак у оценки $\hat{\beta}_4$? Как можно объяснить включение именно *квадрата* возраста (а не, например, третьей степени)?

Как можно объяснить использования в качестве зависимой переменной логарифма зарплаты?

№2. Предложите и экономически обоснуйте модель регрессии для оценки влияния на почасовую оплату *wage* возраста *age*, гендерного признака *female* и недельной занятости *hours*

1. для возрастной группы 30–37 лет
2. для возрастной группы 30–80 лет

Дайте интерпретацию параметров модели

№3. Рассмотрим модель регрессии зависимости цены автомобиля Price от его возраста Age. Какая модель предпочтительней?

Эконометрист полагает, что скорость падения цены автомобиля с возрастом уменьшается. Как учесть это предположение в модели регрессии?

№4. Предложите и экономически обоснуйте модель регрессии для оценки влияния на зарплату CEO (фактор Salary) объема продаж фирмы sales, ее рыночной стоимости mktval, продолжительности работы в фирме ceotenure, возраста age. Дайте интерпретацию параметров модели.

№5. Предложите и экономически обоснуйте модель регрессии для оценки влияния на зарплату CEO (фактор salary) объема продаж фирмы sales, доходности на собственный капитал roe (в %) и доходности акций ros (также в %) если

1. модель регрессии рассматривается только для фирм с объемом продаж \$200 млн– \$250 млн в год
2. модель регрессии рассматривается для фирм с объемами продаж \$10 млн – \$1 млрд в год

Дайте интерпретацию параметров модели.

№6. Рассматриваются следующие регрессионные модели зависимости зарплаты CEO (S) от годового уровня продаж фирмы ($Sales$), дохода на собственный капитал roe и доходности акций ros ($n = 7340$):

$$\widehat{\ln(S)} = \beta_0 + \beta_1 \ln(Sales) + \beta_2 roe + \beta_3 ros \quad R^2 = 0.2129$$

$$\widehat{S} = \beta_0 + \beta_1 Sales + \beta_2 Sales^2 + \beta_3 roe + \beta_4 roe^2 + \beta_5 ros \quad R^2 = 0.2139$$

Можно ли сравнить эти модели по критерию R^2 или \bar{R}^2 ? Ответ обоснуйте. Если можно, то в пользу какой модели можно сделать выбор?

№7. Для мужчин возрастной группы 28–38 лет оценили модели регрессии, описывающие зависимость почасовой оплаты $wage$ от возраста age , стажа на текущем месте работы $tenure$, семейного статуса $married$ (бинарная переменная), места жительства $urban$ (бинарная переменная), количества братьев и сестер $sibs$, результатов IQ-теста (в скобках указаны t -статистики).

Зависимая переменная: $\ln(wage)$; объем выборки 935

	const	<i>age</i>	<i>age</i> ²	<i>tenure</i>	<i>married</i>	<i>urban</i>	<i>sibs</i>	<i>iq</i>	<i>R</i> ²
№1	3.3868 (2.0480)	0.1166 (1.1701)	-0.0015 (-1.0048)	0.0113 (4.4614)	0.1901 (4.7400)	0.1879 (6.8530)	-0.0092 (-1.6526)	0.0082 (9.6327)	0.2085
№2	5.5812 (53.7947)			0.0140 (0.0025)	0.2041 (5.0699)	0.1902 (6.8872)	-0.0106 (-1.8818)	0.0080 (9.2794)	0.1939
№3	5.0396 (29.6252)	0.0166 (3.9984)		0.0113 (4.4663)	0.1897 (4.7295)	0.1893 (6.9086)	-0.0095 (-1.6983)	0.0082 (9.6038)	0.2076
№4	2.7556 (1.6302)	0.1577 (1.5464)	-0.0021 (-1.3791)	0.0107 (4.1183)	0.1801 (4.3819)			0.0089 (10.5712)	0.1656

Какая модель регрессии предпочтительней?

№8. Для набора данных `sleep75`, Таблица A.3, были оценены несколько моделей регрессии. В Таблице B.1 приведены результаты оценивая, в скобках указаны стандартные ошибки коэффициентов. Какая регрессия предпочтительней?

№9. Для набора данных `sleep75`, Таблица A.3, были оценены несколько моделей регрессии. В Таблице B.2 приведены результаты оценивая, в скобках указаны стандартные ошибки коэффициентов. Какая регрессия предпочтительней?

№10. Рассмотрим линейную регрессионную модель зависимости зарплаты *Wage* от уровня школьного образования *Educ* (в годах), возраста *Age*, наличия высшего образования *Heduc* (фиктивная переменная), школьного образования родителей *Feduc*, *Meduc* и пола *Male*:

$$\ln(\widehat{Wage}) = \beta_0 + \beta_1 Educ + \beta_2 Age + \beta_3 Heduc + \beta_4 Feduc + \beta_5 Meduc + \beta_6 Male \quad n = 500 \quad R^2 = 0.302$$

- Пусть $\hat{\beta}_3 = 0.34$. Дайте интерпретацию полученной оценке.
- Пусть $\hat{\beta}_6 = 0.02$ и $s_6 = 0.002$. Дайте интерпретацию полученной оценке. Значима ли дискриминация в оплате труда между мужчинами и женщинами на 1%-м уровне значимости?
- Для тестирования справедливости линейной спецификации обычно используют RESET-тест. Напишите спецификацию вспомогательной регрессии для применения RESET-теста (с включением вспомогательных переменных до 4-й степени). Сформулируйте проверяемую статистическую гипотезу.

- d) Пусть во вспомогательной регрессии $R_1^2 = 0.361$. Тестируйте нулевую гипотезу из пункта c) при уровне значимости 1%. Какой можно сделать вывод?
- e) Можно ли доверять выводам, сделанным в пп. a) и b)?

№11. Рассмотрим регрессионную модель зависимости зарплаты CEO (S) от годового уровня продаж фирмы ($sales$) и дохода на собственный капитал roe :

$$\ln(S) = \beta_0 + \beta_1 \ln(Sales) + \beta_2 roe + \text{error} \quad n = 22 \quad R^2 = 0.213$$

Напишите спецификацию вспомогательной регрессии для применения RESET-теста на справедливость данной спецификации (с включением вспомогательных переменных до 4-й степени). Сформулируйте проверяемую статистическую гипотезу. Пусть во вспомогательной регрессии $R_1^2 = 0.421$. Тестируйте нулевую гипотезу при уровне значимости 5%. Какой можно сделать вывод?

№12. Рассмотрим линейную регрессионную модель зависимости продолжительности сна $sleep$ (мин в неделю) от недельной занятости $totwrk$ (в мин), пола $male$, количества детей $kids$ и уровня образования $educ$.

- a) Опишите процедуру RESET-теста и укажите проверяемую статистическую гипотезу
- b) Если в RESET-тесте нулевая гипотеза отвергается, то какие возможны корректировки модели (приведите не менее двух)?

№13. Для количественного описания зависимости зарплаты CEO (фактор $salary$) от объема продаж фирмы $sales$, ее рыночной стоимости $mktval$ и стажа работы в должности в фирме $ceoten$ рассматривается модель регрессии

$$\ln(salary) = \beta_0 + \beta_1 \ln(sales) + \beta_2 \ln(mktval) + \beta_3 ceoten + \text{error}$$

Дайте описание теста на согласованность выборочных данных с этой спецификацией модели регрессии.

№14. В декабре 2007 были оценены (на одних и тех же данных) две модели регрессии зависимости зарплаты CEO S от годового уровня продаж фирмы $sales$, дохода на собственный капитал roe и доходности акций ros

Модель №1: зависимая переменная $\ln(S)$, объем выборки 27

	const	$\ln(sales)$	roe	ros
coefficient	2.4	0.14	0.008	0.0013
Std. Error	1.02	0.013	0.0023	0.0003

$$R^2 = 0.101 \quad DW = 0.23$$

Модель №2: зависимая переменная $\ln(S)$, объем выборки 27

	const	$\ln(sales)$	roe	ros	roe^2	ros^2
coefficient	2.31	0.18	0.006	0.0009	-0.00003	-0.0000078
Std. Error	1.02	0.026	0.0024	0.00034	0.00004	0.0000063

$$R^2 = 0.127 \quad DW = 2.03$$

- a) По каким критериям можно сравнить эти модели? Укажите не менее двух и дайте их описание.
- b) Какая модель предпочтительней? Ответ обоснуйте.

№15. Рассматривается регрессионная модель зависимости недельной занятости $hours$ от почасовой оплаты $wage$ и уровня образования $educ$

$$hours = \beta_0 + \beta_1 wage + \beta_2 wage^2 + \beta_3 educ + error$$

- a) Как можно объяснить включение в модель регрессии квадрата (и именно квадрата) почасовой оплаты? Какой ожидаемый знак у оценки коэффициента $\hat{\beta}_2$? Ответ обоснуйте.
- b) Предположим, что при оценке модели регрессии знак оценки коэффициента при квадрате получился противоположным ожидаемому. Как это можно объяснить? Дайте несколько возможных объяснений (не менее двух).

2.4 Проблемы спецификации. Практика

№1. Для данных из набора [A.3](#)

1. постройте графики $sleep$ против $totwrk$ и age .
2. предложите спецификацию регрессии зависимости продолжительности сна от занятости и возраста.

3. нужно включать в число объясняющих переменных квадраты регрессоров?

№2. Для данных из набора [A.1](#)

1. постройте графики output против capital и labour.
2. постройте графики $\log(\text{output})$ против $\log(\text{capital})$ и $\log(\text{labour})$
3. предложите спецификацию регрессии зависимости выпуска от занятости и возраста.
4. нужно включать в число объясняющих переменных квадраты регрессоров?

2.5 Линейная модель с гетероскедастичностью

№1. Рассмотрим модель регрессии

$$y_i = \beta x_i + u_i$$

для ошибки которой выполнены условия

$$E(u_i|\mathbf{X}) = 0, \quad \text{cov}(u_i, u_j|\mathbf{X}) = 0 \quad (i \neq j), \quad \text{Var}(u_i|\mathbf{X}) = \sigma^2 x_i^2.$$

1. Будет ли OLS-оценка коэффициента β несмещенной? Состоятельной? Оптимальной?
2. Предложите несмещенную оценку коэффициента β с меньшей дисперсией. Как это соотносится с теоремой Гаусса–Маркова?

№2. Рассмотрим модель регрессии $y_i = \beta_0 + \beta_1 x_i + u_i$ для ошибок которой выполнено

$$E(u_i|\mathbf{X}) = 0, \quad \text{Var}(u_i|\mathbf{X}) = \sigma^2 \exp(10x_i), \quad \text{cov}(u_i, u_j|\mathbf{X}) = 0 \quad (i \neq j).$$

Пусть $\hat{\beta}_1$ есть OLS-оценка коэффициента β_1 .

1. Будет ли $\hat{\beta}_1$ несмещенной оценкой β_1 ? Состоятельной оценкой?
2. Будет ли $\hat{\beta}_1$ наилучшей оценкой коэффициента β_1 (с минимальной дисперсией)? Если нет, то предложите оценку коэффициента с меньшей дисперсией (более точную оценку).

№3. Рассмотрим модель регрессии $y_i = \beta_0 + \beta_1 x_i + u_i$ ($i = 1, \dots, n$) для ошибок которой выполнены условия (δ_0, δ_1 неизвестны и $\delta_1 \neq 0$)

$$E(u_i|\mathbf{X}) = 0 \quad \text{cov}(u_i, u_j|\mathbf{X}) = 0 \quad (i \neq j) \quad \text{Var}(u_i|\mathbf{X}) = \exp(\delta_0 + \delta_1 x_i)$$

- a) Будут ли оценки наименьших квадратов коэффициентов в этой модели регрессии эффективными? Ответ поясните.
- b) Если оценки наименьших квадратов не будут эффективными, то как получить более эффективные оценки коэффициентов (при больших выборках)?

№4. Исследуется зависимость цена коттеджа Price от удаленности от МКАД Dist, площади Hsize и площади участка Lsize. Эконометрист полагает, что дисперсия ошибки в линейной модели регрессии зависит от площади дома и площади участка.

1. Опишите процедуру теста Breusch–Pagan’a и укажите проверяемую статистическую гипотезу.
2. Можно ли применять этот тест при небольших выборках?
3. Если тест отвергает нулевую гипотезу, то какие могут быть причины?

№5. По методу наименьших квадратов была оценена регрессионная модель зависимости недельной занятости hours от почасовой оплаты труда wage (в \$100) и уровня образования educ

Зависимая переменная: hours; объем выборки 1321

	const	wage	wage ²	educ
coefficient	23.4	0.84	-0.008	-0.04
Std. Error	1.02	0.013	0.0027	0.023

Статистика теста White’a равна 6.34.

1. Можно ли для прогнозирования занятости использовать полученные оценки коэффициентов? Ответ обоснуйте.
2. Можно ли проверять значимость влияния уровня образования на занятость с использованием OLS t-статистики? Ответ поясните.

№6. По методу наименьших квадратов была оценена регрессионная модель зависимости недельной занятости $hours$ от почасовой оплаты труда $wage$ (в \$100) и уровня образования $educ$

Зависимая переменная: $hours$; объем выборки 5321

	const	$wage$	$wage^2$	$educ$	
coefficient	23.4	0.84	-0.008	-0.04	ESS = 380.12
Std. Error	1.02	0.013	0.0027	0.023	RSS = 130.23

Статистика теста Breusch–Pagan’a на зависимость дисперсии ошибки модели регрессии от почасовой оплаты и пола человека равна 8.75.

1. Можно ли для прогнозирования занятости использовать полученные оценки коэффициентов? Ответ обоснуйте.
2. Можно ли проверять значимость влияния уровня образования на занятость с помощью OLS t -статистики? Ответ поясните.

№7. Для количественного описания зависимости зарплаты CEO ($salary$) от объема продаж фирмы $sales$, её рыночной стоимости $mktval$ и стажа работы в должности в фирме $ceoten$ была оценена модель регрессии

Зависимая переменная: $\ln(salary)$; объем выборки 177

	const	$\ln(sales)$	$\ln(mktval)$	$ceoten$
coefficient	4.5038	0.1629	0.1092	0.0117
OLS s.e.	0.2572	0.0392	0.0496	0.0053
HC s.e.	0.2801	0.0382	0.0499	0.0073

1. Как можно объяснить выбор такой спецификации модели?
2. Проверьте значимость коэффициентов модели с использованием стандартных ошибок коэффициентов метода наименьших квадратов и стандартных ошибок, устойчивых к гетероскедастичности (НС-стандартных ошибок). Какой можно сделать вывод?

№8. Для модели регрессии из предыдущей задачи была оценена вспомогательная регрессия

Зависимая переменная: \hat{u}^2 ; объем выборки 177

	const	$\ln(\text{sales})$	$\ln(\text{mktval})$	ceoten
coefficient	-0.0521	-0.0805	0.0992	0.0188
s.e.	0.3052	0.0466	0.0588	0.0063
Prob.	0.8647	0.0856	0.0938	0.0033

R-squared=0.0657 F-statistic=4.0537 Prob(F-statistic)=0.0082

1. Какие выводы можно сделать на основе этой регрессии?
2. Можно ли (экономически) интерпретировать эту модель регрессии?

№9. Для количественного описания зависимости зарплаты CEO (salary) от объема продаж фирмы sales , ее рыночной стоимости mktval и стажа работы в должности в фирме ceoten рассматривается модель регрессии

$$\ln(\text{salary}) = \beta_0 + \beta_1 \ln(\text{sales}) + \beta_2 \ln(\text{mktval}) + \beta_3 \text{ceoten} + \text{error}$$

Графический анализ остатков показал, что разброс остатков уменьшается с увеличением фактора ceoten и нет явной зависимости разброса остатков от $\ln(\text{sales})$ и $\ln(\text{mktval})$.

Дайте описание теста на гетероскедастичность ошибки модели регрессии, который можно применить в данной ситуации.

№10. Для пространственных данных рассмотрим модель регрессии

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 z_i + u_i$$

$$E(u_i | \mathbf{X}) = 0 \quad \text{Var}(u_i | \mathbf{X}) = \sigma^2 x_i^2 z_i^2.$$

Как получить эффективную оценку коэффициентов β_1 , β_2 и β_3 ? Дайте развернутый ответ.

2.6 Линейная модель с автокорреляцией

№1. Рассмотрим модель регрессии

$$y_t = \beta_0 + \beta_1 x_t + u_t$$

для ошибок которой выполнено

$$E u_t = 0, \quad \text{Var}(u_t) = \sigma^2, \quad u_t = 0.15 u_{t-1} + w_t$$

и w_t – белый шум. Пусть $\hat{\beta}_1$ есть OLS-оценка коэффициента β_1 .

1. Сформулируйте условия Гаусса–Маркова для w_t .
2. Будет ли $\hat{\beta}_1$ несмещенной оценкой β_1 ? Состоятельной оценкой?
3. Будет ли $\hat{\beta}_1$ наилучшей оценкой коэффициента β_1 (с минимальной дисперсией)? Если нет, то предложите оценку коэффициента с меньшей дисперсией (более точную оценку).

№2. Оценивается кривая Филлипса с ожиданиями, связывающая инфляцию с безработицей, безработицей с единичным лагом и с инфляцией с лагом один.

1. Опишите тест на автокорреляцию первого порядка ошибок этой модели регрессии и укажите проверяемую статистическую гипотезу.
2. Можно ли применять этот тест при небольших выборках?
3. Если тест отвергает нулевую гипотезу, то какие могут быть причины?

№3. По квартальным данным за пять с половиной лет по методу наименьших квадратов была оценена функция спроса с учетом сезонности (факторы *Spr*, *Summ*, *Fall* – весна, лето, осень)

Зависимая переменная: $\ln(Q)$

	const	$\ln(P)$	<i>Spr</i>	<i>Summ</i>	<i>Fall</i>	
coefficient	4.4	0.64	-0.08	-0.13	-0.07	$R^2 = 0.13$
s.e.	1.02	0.013	0.013	0.023	0.03	DW = 0.78

- a) Можно ли для прогнозирования объема продаж использовать полученные оценки коэффициентов модели регрессии? Ответ обоснуйте.
- b) Можно ли для проверки значимости регрессии «в целом» использовать стандартную F -статистику? Ответ обоснуйте.

№4. По четырехнедельным историческим данным была оценена регрессионная модель зависимости спроса на мороженное *consumption* (в пинтах на человека) от среднего недельного дохода домохозяйства *income* (в \$), цены за пинту *price* (в \$) и средней температуры *temp* (по Фаренгейту)

Зависимая переменная: *consumption*; объем выборки 30

	const	<i>income</i>	<i>price</i>	<i>temp</i>
coeff	0.1973149	0.0033078	-1.044413	0.0034584
Std. Err	0.2702161	0.0011714	0.834357	0.0004455

$$R^2 = 0.7890 \quad R_{adj}^2 = 0.6866 \quad DW = 1.021169$$

1. Тестируйте ошибку модели регрессии на автокорреляцию первого порядка.
2. Какие можно сделать выводы?

№5. Для временных рядов рассмотрим модель регрессии (ϵ_t – белый шум)

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 z_t + u_t \quad u_t = -0.25u_{t-1} + \epsilon_t.$$

Как получить эффективную оценку коэффициентов β_1 и β_2 ? Дайте развернутый ответ.

№6. Рассмотрим макроэкономическую кейнсианскую функцию потребления без автономного потребления

$$C_t = \beta_1 Y_t + \beta_2 T_t + \beta_3 r_t + u_t,$$

где C_t – потребление, Y_t – национальный доход, T_t – налоговые поступления, r_t – ставка рефинансирования. Модель оценивается по историческим данным и предполагается, что объясняющие переменные экзогенны.

Дайте описание теста на автокоррелированность первого порядка ошибок этой модели регрессии.

Глава 3

Системы одновременных уравнений

№1. Рассмотрим кейнсианскую модель потребления

$$\begin{aligned}C_t &= \beta_0 + \beta_1 Y_t + u_t \\ Y_t &= C_t + I_t,\end{aligned}\tag{3.1}$$

где C_t – агрегированное потребление, Y_t – национальный доход, I_t – инвестиции. Коэффициент β_1 имеет смысл предельной склонности к потреблению. Предполагается, что u_t (шок потребления) удовлетворяет условиям Гаусса–Маркова и $\text{cov}(I_t, u_t) = 0$.

1. Дайте интерпретацию уравнений модели.
2. Напишите приведенную форму системы нормальных уравнений
3. Укажите алгоритм оценки коэффициентов модели.

№2. Рассмотрим модель спроса–предложения

$$Q_{dt} = \beta_0 + \beta_1 P_t + \mathbf{x}'_t \boldsymbol{\beta} + u_t\tag{3.2}$$

$$Q_{st} = \gamma_0 + \gamma_1 P_t + \mathbf{z}'_t \boldsymbol{\gamma} + w_t\tag{3.3}$$

где $\mathbf{x}'_t = (x_{1t}, \dots, x_{kt})$ – факторы, влияющие на спрос (например, цвет), $\mathbf{z}'_t = (z_{1t}, \dots, z_{lt})$ – факторы, влияющие на предложение, коэффициенты $\boldsymbol{\beta}' = (\beta_2, \dots, \beta_{k+1})$ и $\boldsymbol{\gamma}' = (\gamma_2, \dots, \gamma_{l+1})$, u_t – шоки спроса, w_t – шоки предложения. Предполагаем, что $\text{cov}(u_t, w_t) = 0$ и факторы \mathbf{x} и \mathbf{z} различаются. Рынок находится в равновесии.

1. Выразите Q и P через \mathbf{x} , \mathbf{z} и шоки u и w .
2. Оценивается функция спроса (3.2). Можно ли применить метод наименьших квадратов к первому уравнению модели? Ответ обоснуйте.
3. Предложите метод оценивания функции спроса.
4. В чем отличие модели спроса–предложения (3.2), (3.3) от модели только одной функции спроса

$$Q_{dt} = \beta_0 + \beta_1 P_t + \mathbf{x}'_t \boldsymbol{\beta} + u_t?$$

Дайте экономическую интерпретацию каждой модели.

№3. Рассмотрим эконометрическую модель, связывающую недельную занятость $hours$ и месячную зарплату $wage$ замужних женщин

$$\begin{aligned} hours &= \beta_0 + \beta_1 \log(wage) + \beta_2 educ + \beta_3 age + \beta_4 kidslt6 + \beta_5 nwifeinc + error_1 \\ \log(wage) &= \gamma_0 + \gamma_1 hours + \gamma_2 exper + \gamma_3 exper^2 + \gamma_4 educ + error_2 \end{aligned}$$

где $educ$ – уровень образования, age – возраст, $exper$ – опыт работы, $kidslt6$ – число детей до 6 лет, $nwifeinc$ – прочий доход (например, доход мужа).

1. Экономиста интересует оценка коэффициента γ_1 . Дайте интерпретацию этого коэффициента.
2. Дайте экономические обоснования того, что занятость и зарплата в одном уравнении являются зависимой переменной, а в другом объясняющей переменной.
3. Выполнено ли условие идентифицируемости? Ответ обоснуйте.
4. Какой метод следует использовать для оценки коэффициента γ_1 ? Обоснуйте ответ и дайте описание метода оценки коэффициентов.

№4 (Практика). Рассмотрим систему одновременных уравнений из предыдущей задачи:

$$\begin{aligned} hours &= \beta_0 + \beta_1 \log(wage) + \beta_2 educ + \beta_3 age + \beta_4 kidslt6 + \beta_5 nwifeinc + error_1 \\ \log(wage) &= \gamma_0 + \gamma_1 hours + \gamma_2 exper + \gamma_3 exper^2 + \gamma_4 educ + error_2 \end{aligned}$$

Для данных из файла `mroz.gdt`

- Оцените (раздельно) уравнения системы методом наименьших квадратов (OLS). Тестируйте значимость β_1 и γ_1 .
- Оцените каждое уравнение системы используя двухшаговый метод наименьших квадратов (2SLS). Для этого используйте метод инструментальных переменных, выбрав в качестве инструментов **все** экзогенные факторы модели. Сравните полученные оценки коэффициентов с OLS-оценками. Тестируйте значимость β_1 и γ_1 .

Глава 4

Метод инструментальных переменных

№1. Рассмотрим кейнсианскую модель потребления

$$\begin{aligned}C &= \beta_0 + \beta_1 Y + u \\ Y &= C + I,\end{aligned}\tag{4.1}$$

где C – агрегированное потребление, Y – национальный доход, I – инвестиции. Коэффициент β_1 имеет смысл предельной склонности к потреблению. Предполагается, что u (шок потребления) удовлетворяет условиям Гаусса–Маркова и $\text{cov}(I, u) = 0$.

- a) Выразите C и Y через I и u .
- b) Оценивается уравнение (4.1). Имеет ли место проблема эндогенности? Обоснуйте ответ
- c) Если на предыдущий пункт ответ положительный, то будет ли фактор I инструментом для Y ? Ответ обоснуйте.
- d) Если на предыдущие пункты ответы положительны, то найдите IV-оценку коэффициента β_1 (напишите формулу или укажите алгоритм нахождения).

№2. Модель регрессии $\ln Wage = \beta_0 + \beta_1 Educ + \beta_2 Age + \beta_3 Age^2 + u$ описывает зависимость почасовой оплаты от уровня образования и возраста человека. Но ошибка u в этой модели содержит (ненаблюдаемый) фактор

«индивидуальные способности человека», который положительно коррелирует с уровнем образования.

- а) Можно ли для оценки отдачи от образования использовать метод наименьших квадратов? Ответ обоснуйте.
- б) Для учета фактора «индивидуальные способности человека» в модель регрессии была включена объясняющая переменная *iqscores* (результат IQ-теста), которая, как предполагает эконометрист, приближенно оценивает индивидуальные способности человека (т.е. *iqscores* есть прокси-переменная). Можно ли для оценки отдачи от образования использовать метод наименьших квадратов для новой модели регрессии (с включенным фактором *iqscores*)? Ответ поясните.

№3. Рассмотрим модель регрессии

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u, \quad \text{cov}(x_1, u) = 0 \quad \text{cov}(x_2, u) \neq 0.$$

Будет ли оценка наименьших квадратов $\hat{\beta}_1$ коэффициента β_1 состоятельной? Ответ кратко поясните. Пусть z – инструмент для фактора x_2 . Каким условиям должен удовлетворять инструмент? Опишите двухшаговый метод оценки коэффициентов регрессии методом инструментальных переменных.

№4. Рассмотрим модель ADL с автокоррелированными ошибками

$$y_t = \beta_0 + \beta_1 y_{t-1} + \gamma_1 x_t + u_t \quad u_t = \rho u_{t-1} + v_t \quad (|\rho| < 1, \rho \neq 0)$$

где v_t удовлетворяет условиям Гаусса–Маркова. Как автокоррелированность ошибок влияет на свойства оценок наименьших квадратов коэффициентов регрессии?

Глава 5

Временные ряды

5.1 Модели стационарных временных рядов

№1. Постройте автоковариационную и автокорреляционную функцию для модели $MA(1)$, $MA(2)$, $MA(3)$.

№2. Для модели $ARMA(1, 1)$

$$y_t = \beta_0 + \beta_1 y_{t-1} + u_t + \theta_1 u_{t-1} \quad u_t \sim WN$$

1. Вычислите Ey_t и $\text{Var}(y_t)$.
2. Найдите представление временного ряда в виде $MA(\infty)$.
3. Вычислите автоковариационную и автокорреляционную функцию.

№3. Для модели $ARMA(1, 2)$

$$y_t = \beta_0 + \beta_1 y_{t-1} + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} \quad u_t \sim WN$$

1. Вычислите Ey_t и $\text{Var}(y_t)$.
2. Найдите представление временного ряда в виде $MA(\infty)$.
3. Вычислите автоковариационную и автокорреляционную функцию.

№4. Задаёт ли модель

$$x_t = 10 + 1.2x_{t-1} - 0.2x_{t-2} + u_t \quad u_t \sim WN$$

стационарный ряд? Ответ поясните.

№5. По временному ряду длины $n = 110$ были оценены следующие модели ARMA:

1. $y_t = 16 + 0.67y_{t-1} + u_t - 0.3u_{t-1}$, $s^2 = 1.63$;
2. $y_t = 15.1 + 0.6y_{t-1} - 0.1y_{t-2} + u_t$, $s^2 = 1.8$;
3. $y_t = 17.1 + 0.71y_{t-1} + u_t - 0.34u_{t-1} + 0.12u_{t-2}$, $s^2 = 1.86$.

Какую модель вы выберете?

№6. По 100 наблюдениям была оценена модель

$$x_t = 2 + 0.5x_{t-1} - 0.06x_{t-2} + u_t.$$

и проверена её адекватность. Известно, что $x_{100} = 3.1$, $x_{99} = 2.5$, $x_{98} = 2.8$, $x_{97} = 3.6$. Найдите \hat{x}_{102} , \hat{x}_{103} и \hat{x}_{104} .

№7. Временной ряд x_t – первая разность логарифма реального ВВП США с 1984:2 по 2013:2. Для этого ряда были оценены следующие модели (в таблице приведены значения информационных критериев и Q_{LB} -статистики с лагом 4 для остатков):

модель	MA(1)	AR(1)	ARMA(1,1)	AR(2)	MA(2)	AR(3)	ARMA(2,1)
AIC	-870.85	-880.17	-883.55	-885.54	-882.83	-884.10	-884.54
BIC	-862.57	-871.88	-872.50	-874.49	-871.78	-870.29	-870.73
Q_{LB}	21.5127	11.0688	3.68299	1.34523	4.21655	0.963561	0.494841

Какая модель предпочтительней? Ответ обоснуйте.

№8. Временной ряд x_t – первая разность логарифма реального ВВП США с 1984:2 по 2013:2 (квартальные данные). Была оценена модель ARMA(2,1)

	const	x_{t-1}	x_{t-2}	u_{t-1}
коэфф.	0.007	-0.026	0.397	0.378
P-значение	$1.02 \cdot 10^{-10}$	0.9172	0.0006	0.1534

Известны значения временного ряда и остатки модели

t	2012:2	2012:3	2012:4	2013:1	2013:2
x_t	0.003	0.007	0.0004	0.003	0.006
\hat{u}_t	-0.007	0.001	-0.006	-0.002	0.003

Постройте прогноз на четвёртый квартал 2013 года (2013:4).

5.2 Модели стационарных временных рядов. Практика

Для практических задач данные нужно взять с сайта Федерального Банка г. С.-Луиса www.stlouisfed.org

№1. Рассмотрим месячные данные по денежной массе M2 для US с 1990 г. по н.в. Постройте ряд x_t – первая разность логарифма M2.

1. Постройте график x_t . Какие можно сделать выводы?
2. Постройте ACF до лага 20. Какие можно сделать выводы? Подтверждается ли гипотеза что x_t – белый шум?
3. Какую модель ARMA выбрать для ряда x_t ? Проверьте “адекватность” выбранной модели.
4. Сделайте прогноз на 7 шагов по выбранной и оценённой модели.

№2. Рассмотрим квартальные данные по ВВП US с 1985 г. по н.в. Постройте ряд x_t – первая разность логарифма ВВП.

1. Постройте график x_t . Какие можно сделать выводы?
2. Постройте ACF до лага 20. Какие можно сделать выводы? Подтверждается ли гипотеза что x_t – белый шум?
3. Какую модель ARMA выбрать для ряда x_t ? Проверьте “адекватность” выбранной модели.
4. Сделайте прогноз на 4 шага по выбранной и оценённой модели.

№3. Рассмотрим месячные данные по краткосрочной ставке 3-х месячных Treasury Bills с 1990 г. по н.в. Постройте ряд x_t – первая разность ставки.

1. Постройте график x_t . Какие можно сделать выводы?
2. Постройте ACF до лага 20. Какие можно сделать выводы? Подтверждается ли гипотеза что x_t – белый шум?
3. Какую модель ARMA выбрать для ряда x_t ? Проверьте “адекватность” выбранной модели.

4. Сделайте прогноз на 7 шагов по выбранной и оценённой модели.

№4. Рассмотрим месячные данные по долгосрочной ставке 10-ти летних бондов с 1990 г. пл н.в. Постройте ряд x_t – первая разность ставки.

1. Постройте график x_t . Какие можно сделать выводы?
2. Постройте АСФ до лага 20. Какие можно сделать выводы? Подтверждается ли гипотеза что x_t – белый шум?
3. Какую модель ARMA выбрать для ряда x_t ? Проверьте “адекватность” выбранной модели.
4. Сделайте прогноз на 7 шагов по выбранной и оценённой модели.

5.3 Модели распределенных лагов

№1. Рассмотрим модель FDL(2)

$$y_t = \mu + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \text{error}$$

1. Напишите функцию импульсного отклика (impulse response function) для краткосрочной зависимости.
2. Напишите функцию импульсного отклика (impulse response function) для долгосрочной зависимости.
3. Напишите уравнение долгосрочной зависимости и дайте его интерпретацию.

№2. Рассмотрим модель FDL(4)

$$y_t = \mu + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \beta_3 x_{t-3} + \gamma_0 z_t + \gamma_1 z_{t-1} + \gamma_2 z_{t-2} + u_t.$$

Сформулируйте условия Гаусса–Маркова для ошибок этой модели регрессии. Найдите долгосрочные мультипликаторы y по x и по z . Напишите уравнение «долгосрочной зависимости» y от x и z .

№3. Рассмотрим модель ADL

$$y_t = \mu + \phi y_{t-1} + \beta x_t + \text{error}$$

1. Когда для этой модели выполнено условие стационарности?
2. Напишите функцию импульсного отклика для краткосрочной зависимости.
3. Напишите функцию импульсного отклика для долгосрочной зависимости.
4. Напишите уравнение долгосрочной зависимости и дайте его интерпретацию.

№4. Рассмотрим модель ADL

$$y_t = \mu + \phi y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \text{error}$$

1. Когда для этой модели выполнено условие стационарности?
2. Напишите функцию импульсного отклика для краткосрочной зависимости.
3. Напишите функцию импульсного отклика для долгосрочной зависимости.
4. Напишите уравнение долгосрочной зависимости и дайте его интерпретацию.

№5. Рассмотрим модель ADL

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \beta x_t + \text{error}$$

1. Когда для этой модели выполнено условие стационарности?
2. Напишите функцию импульсного отклика для краткосрочной зависимости.
3. Напишите функцию импульсного отклика для долгосрочной зависимости.
4. Напишите уравнение долгосрочной зависимости и дайте его интерпретацию.

№6. Рассмотрим модель ADL(2,2)

$$y_t = 5 + \frac{5}{6}y_{t-1} - \frac{1}{6}y_{t-2} + 2.1x_t + 1.2x_{t-1} + 0.2x_{t-2} + 4.3z_t + 2.3z_{t-1} + u_t$$

для ошибок выполнены условия Гаусса–Маркова.

1. Выполнено ли для этой модели условие стационарности?
2. Если да, то найдите долгосрочные мультипликаторы y по x и z и напишите уравнение «долгосрочной зависимости» y от x и z .

№7. Рассмотрим модель ADL(2,1)

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \beta_0 x_t + \beta_1 x_{t-1} + u_t,$$

где x_t – стационарный временной ряд и $u_t \sim \text{WN}$.

1. Когда для этой модели выполнено условие стационарности?
2. Напишите уравнение долгосрочной зависимости и дайте его интерпретацию.

5.4 Модели распределенных лагов. Практика

Для практических задач данные нужно взять с сайта Федерального Банка г. С.-Луиса www.stlouisfed.org

№1. Рассмотрим недельные исторические данные для US по M2, краткосрочной 3-х месячной ставке и долгосрочной 10-ти летней ставке с 1990 года по н.в. Пусть $dM2$ – первая разность логарифма M2, $dTB3$ – первая разность краткосрочной ставки, $dTCM10$ – первая разность долгосрочной ставки.

1. Постройте модель FDL зависимости $dM2$ от $dTB3$ и $dTCM10$. Какой порядок шагов выбрать?
2. Дайте интерпретацию коэффициентов оценённой модели.
3. Напишите уравнение долгосрочной зависимости и дайте его интерпретацию.

№2. Рассмотрим квартальные исторические данные для US по ВВП, M2, краткосрочной 3-х месячной ставке и долгосрочной 10-ти летней ставке с 1990 года по н.в. Пусть IGDP – первая разность логарифма ВВП, dM2 – первая разность логарифма M2, dTB3 – первая разность краткосрочной ставки, dTSM10 – первая разность долгосрочной ставки.

1. Постройте модель FDL зависимости IGDP от dM2 , dTB3 и dTSM10 . Какой порядок шагов выбрать?
2. Дайте интерпретацию коэффициентов оценённой модели.
3. Напишите уравнение долгосрочной зависимости и дайте его интерпретацию.

5.5 TS – ряды

№1. Рассмотрим модель временного ряда

$$x_t = \gamma_0 + \gamma_1 t + \phi x_{t-1} + u_t \quad u_t \sim \text{WN}$$

1. Сформулируйте условия стационарности.
2. Напишите уравнение тренда для временного ряда.

№2. Рассмотрим модель временного ряда

$$x_t = \gamma_0 + \gamma_1 t + \gamma_2 t^2 + \phi x_{t-1} + u_t \quad u_t \sim \text{WN}$$

1. Сформулируйте условия стационарности.
2. Напишите уравнение тренда для временного ряда.

№3. Рассмотрим модель временного ряда

$$x_t = \gamma_0 + \gamma_1 t + \phi_1 x_{t-1} + \phi_2 x_{t-2} + u_t \quad u_t \sim \text{WN}$$

1. Сформулируйте условия стационарности.
2. Напишите уравнение тренда для временного ряда.

№4. Для TS-рядов (v_t – стационарный ряд)

1. $x_t = \beta_0 + \beta_1 t + v_t$;
2. $x_t = \beta_0 + \beta_1 t + \beta_2 t^2 + v_t$

Найдите Δx_t и $\Delta^2 x_t$.

№5. Для TS-рядов ($u_t \sim \text{WN}$)

1. $x_t = \gamma_0 + \gamma_1 t + \phi x_{t-1} + u_t$;
2. $x_t = \gamma_0 + \gamma_1 t + \gamma_2 t^2 + \phi x_{t-1} + u_t$
3. $x_t = \gamma_0 + \gamma_1 t + \phi_1 x_{t-1} + \phi_2 x_{t-2} + u_t$

найдите уравнение для Δx_t и $\Delta^2 x_t$.

№6. Рассмотрим модель квадратичного тренда с автокорреляцией первого порядка

$$\begin{aligned} x_t &= \beta_0 + \beta_1 t + \beta_2 t^2 + v_t \\ v_t &= \rho v_{t-1} + u_t \quad u_t \sim \text{WN} \quad |\rho| < 1 \end{aligned}$$

1. Проведите авторегрессионное преобразование для получения эффективной оценки тренда.
2. Как связаны коэффициенты преобразованного уравнения с параметрами тренда?

№7. Для TS-ряда

$$x_t = \gamma_0 + \gamma_1 t + \gamma_2 t^2 + \phi x_{t-1} + u_t + \theta u_{t-1} \quad u_t \sim \text{WN} \quad |\phi| < 1$$

1. Найдите уравнение для первой разности x_t .
2. Найдите уравнение для второй разности x_t .

№8. Рассмотрим модель временного ряда для ВВП

$$\ln \text{GDP}_t = \gamma_0 + \gamma_1 t + \phi_1 \ln \text{GDP}_{t-1} + \phi_2 \ln \text{GDP}_{t-2} + u_t, \quad u_t \sim \text{WN}$$

1. При каких условиях эта модель временного ряда задаёт TS-ряд?
2. Найдите тренд ВВП (для случая TS).

№9. Рассмотрим модель тренда для ВВП (в млн. \$)

$$\widehat{GDP}_t = \underset{(23.4)}{230.4} + \underset{(0.038)}{0.32} t \quad (t = 0, \dots, 20)$$

Дайте интерпретацию коэффициентов модели.

Проверьте значимость тренда при альтернативе о тенденции к росту ВВП. Уровень значимости 5%

№10. Рассмотрим модель тренда для численности населения (в млн. чел.)

$$\widehat{\ln POP}_t = \underset{(0.4)}{2.4} + \underset{(0.01)}{0.023} t \quad (t = 0, \dots, 23)$$

Дайте интерпретацию коэффициентов модели.

Проверьте значимость тренда при альтернативе о тенденции к росту населения. Уровень значимости 1%.

№11. Исследует регрессионная модель влияния численности населения (POP) на ВВП (GDP) на основе временных рядов

$$\ln GDP_t = \beta_0 + \beta_1 \ln POP_t + \text{error}.$$

Однако временные ряды для ВВП и численности населения как правило имеют тенденцию к росту. Как это повлияет на статистические выводы для регрессионной модели?

5.6 DS – ряды

№1. Для временного ряда

$$x_t = x_{t-1} + u_t \quad u_t \sim \text{WN}(0, \sigma^2) \quad x_0 = 0$$

найдите $E x_t$ и $\text{Var } x_t$.

№2. Для временного ряда ($\mu \neq 0$)

$$x_t = \mu + x_{t-1} + u_t \quad u_t \sim \text{WN}(0, \sigma^2) \quad x_0 = 0$$

найдите $E x_t$ и $\text{Var } x_t$.

№3. Для временного ряда ($\mu \neq 0$)

$$x_t = \mu + \beta t + x_{t-1} + u_t \quad u_t \sim \text{WN}(0, \sigma^2) \quad x_0 = 0$$

найдите $E x_t$ и $\text{Var } x_t$.

№4. Пусть $x_t \sim \text{ARIMA}(1, 1, 0)$. Запишите представление ряда в виде (нестационарной) модели $\text{ARMA}(2, 0)$. Покажите, что авторегрессионный многочлен модели ARMA имеет единичный корень кратности 1.

№5. Пусть $x_t \sim \text{ARIMA}(1, 2, 0)$. Запишите представление ряда в виде (нестационарной) модели $\text{ARMA}(3, 0)$. Покажите, что авторегрессионный многочлен модели ARMA имеет единичный корень кратности 2.

№6. Пусть $x_t \sim \text{ARIMA}(2, 2, 3)$. Запишите представление ряда в виде (нестационарной) модели ARMA . Покажите, что авторегрессионный многочлен модели ARMA имеет единичный корень и найдите кратность этого корня.

№7. Пусть $x_t \sim \text{ARIMA}(3, 1, 1)$. Запишите представление ряда в виде (нестационарной) модели ARMA . Покажите, что авторегрессионный многочлен модели ARMA имеет единичный корень и найдите кратность этого корня.

№8. Пусть $x_t \sim \text{ARIMA}(1, 2, 2)$.

1. Что означает $x_t \sim \text{ARIMA}(1, 2, 2)$?
2. Запишите представление ряда в виде (нестационарной) модели ARMA .

№9. Рассмотрим модель временно ряда

$$x_t = 2x_{t-1} - x_{t-2} + u_t + 0.5u_{t-1} - 0.1u_{t-2} \quad u_t \sim \text{WN}$$

1. Покажите, что $x_t \sim I(2)$.
2. Найдите порядок модели $\text{ARIMA}(p, k, q)$, соответствующей ряду x_t .

5.7 DS-ряды. Практика

Для практических задач данные нужно взять с сайта Федерального Банка г. С.-Луиса www.stlouisfed.org

№1. Рассмотрим месячные данные по денежной массе M2 для US с 1990 г. по н.в. Ряд x_t – логарифм M2.

1. Постройте график x_t . Какие можно сделать выводы?
2. Тестируйте ряда x_t на единичный корень.

3. Какой порядок интегрирования у ряда x_t ?
4. Какую модель ARIMA выбрать для ряда x_t ?

№2. Рассмотрим квартальные данные по ВВП US с 1985 г. по н.в. Постройте ряд x_t – логарифм ВВП.

1. Постройте график x_t . Какие можно сделать выводы?
2. Тестируйте ряда x_t на единичный корень.
3. Какой порядок интегрирования у ряда x_t ?
4. Какую модель ARIMA выбрать для ряда x_t ?

№3. Рассмотрим месячные данные по краткосрочной ставке 3-х месячных Treasury Bills с 1990 г. пл н.в. (ряд x_t)

1. Постройте график x_t . Какие можно сделать выводы?
2. Тестируйте ряда x_t на единичный корень.
3. Какой порядок интегрирования у ряда x_t ?
4. Какую модель ARIMA выбрать для ряда x_t ?

№4. Рассмотрим месячные данные по долгосрочной ставке 10-ти летних бондов с 1990 г. по н.в. (ряд x_t)

1. Постройте график x_t . Какие можно сделать выводы?
2. Тестируйте ряда x_t на единичный корень.
3. Какой порядок интегрирования у ряда x_t ?
4. Какую модель ARIMA выбрать для ряда x_t ?

5.8 Многомерные модели временных рядов

5.8.1 VAR и коинтеграция

№1. Рассмотрим модель VAR(1)

$$\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + \mathbf{u}_t \quad \mathbf{x}_t = \begin{pmatrix} x_t \\ y_t \end{pmatrix} \quad \mathbf{u}_t = \begin{pmatrix} u_t \\ v_t \end{pmatrix},$$

где $u_t \sim \text{WN}(0, \sigma_u^2)$, $v_t \sim \text{WN}(0, \sigma_v^2)$, $\text{cov}(u_t, v_t) = \sigma_{uv}$. Проверьте условие стационарности для матриц

$$\mathbf{A} = \begin{pmatrix} 0.5 & 1 \\ 0 & 0.3 \end{pmatrix} \quad \begin{pmatrix} 0 & 0.5 \\ -0.5 & 0 \end{pmatrix} \quad \begin{pmatrix} 1 & 3 \\ 0 & 0.2 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$$

№2. Рассмотрим модели временных рядов

$$\begin{cases} x_t = x_{t-1} + u_t \\ y_t = y_{t-1} + v_t \end{cases} \quad \begin{cases} x_t = x_{t-1} + u_t \\ y_t = cx_t + v_t \end{cases}$$

$$\begin{cases} x_t = x_{t-1} + u_t \\ y_t = x_t + x_{t-1} + v_t \end{cases} \quad \begin{cases} x_t = 3x_{t-1} - 7y_{t-1} + u_t \\ y_t = x_{t-1} - 2.5y_{t-1} + v_t \end{cases}$$

1. Запишите эти модели в виде моделей VAR.
2. Проверьте условие стационарности.
3. Какие временные ряды коинтегрированы?
 - Если ряды коинтегрированы, то найдите коинтеграционное соотношение, корректирующую ошибку и запишите модель VECM.
 - Если ряды неинтегрированы, то постройте VAR-модель для первых разностей.

№3. Рассмотрим модель VAR(1)

$$\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + \mathbf{u}_t \quad \mathbf{x}_t = \begin{pmatrix} x_t \\ y_t \\ z_t \end{pmatrix} \quad \mathbf{u}_t = \begin{pmatrix} u_t \\ v_t \\ w_t \end{pmatrix} \sim \text{WN}$$

Проверьте условие стационарности для матриц

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 3 \\ -1 & 0 & -2 \\ 0 & 0 & 0.5 \end{pmatrix} \quad \begin{pmatrix} 0 & 0 & 0.5 \\ 0.5 & 0 & 0 \\ 0 & 0.5 & 0 \end{pmatrix}$$

№4. Рассмотрим модель VAR(2)

$$\mathbf{x}_t = \mathbf{A}_1 \mathbf{x}_{t-1} + \mathbf{A}_2 \mathbf{x}_{t-2} + \mathbf{u}_t \quad \mathbf{x}_t = \begin{pmatrix} x_t \\ y_t \end{pmatrix} \quad \mathbf{u}_t = \begin{pmatrix} u_t \\ v_t \end{pmatrix} \sim \text{WN}$$

Проверьте условие стационарности для матриц

$$\begin{aligned} 1) \mathbf{A}_1 &= \begin{pmatrix} 2 & 3 \\ 0 & 1 \end{pmatrix} & \mathbf{A}_2 &= \begin{pmatrix} -1 & 3 \\ 0 & -0.25 \end{pmatrix} \\ 2) \mathbf{A}_1 &= \begin{pmatrix} 0 & 0.5 \\ 0.5 & 0 \end{pmatrix} & \mathbf{A}_2 &= \begin{pmatrix} 0 & -0.25 \\ -0.25 & 0 \end{pmatrix} \end{aligned}$$

№5. Рассмотрим модели временных рядов

$$\begin{cases} x_t = 2x_{t-1} - x_{t-2} + u_t \\ y_t = 1.5y_{t-1} - 0.5y_{t-2} + v_t \end{cases} \quad \begin{cases} x_t = 1.5x_{t-1} + y_{t-1} - 0.5x_{t-2} - y_{t-2} + u_t \\ y_t = -x_{t-1} - 0.5y_{t-1} + x_{t-2} + 1.5y_{t-2} + v_t \end{cases}$$

$$\begin{cases} x_t = x_{t-1} + u_t \\ y_t = x_t + x_{t-1} + v_t \\ z_t = x_t + y_{t-1} + w_t \end{cases} \quad \begin{cases} x_t = y_{t-1} + u_t \\ y_t = z_{t-1} + v_t \\ z_t = x_{t-1} + w_t \end{cases}$$

1. Запишите эти модели в виде моделей VAR.
2. Проверьте условие стационарности.
3. Какие временные ряды коинтегрированы?
 - Если ряды коинтегрированы, то найдите коинтеграционное соотношение, корректирующую ошибку и запишите модель VECM.
 - Если ряды некойнтегрированы, то постройте VAR-модель для первых разностей.

5.9 Из пособия

№1. Является ли временной ряд, заданный авторегрессионным разностным уравнением, стационарным?

1. $y_t = 7 + 0.5y_{t-1} + \varepsilon_t$.
2. $y_t = 10 + 0.25y_{t-2} + \varepsilon_t$.

3. $y_t = 10 + y_{t-1} - 0.25y_{t-2} + \varepsilon_t$.
4. $y_t = \frac{3}{2}y_{t-1} - \frac{3}{4}y_{t-2} + \frac{1}{8}y_{t-3} + \varepsilon_t$.
5. $y_t = 3 + 0.4y_{t-1} - 0.04y_{t-2} + \varepsilon_t$.
6. $y_t = 5 - 3y_{t-1} - 3y_{t-2} - y_{t-3} + \varepsilon_t$.
7. $y_t = -2y_{t-1} + 1.25y_{t-2} - 0.25y_{t-3} + \varepsilon_t$.

№2. Написать формулы для прогноза

1. на $l = 3$ шага для процесса $y_t = 5 + 0.5y_{t-1} + \varepsilon_t$.
2. на $l = 4$ шага для процесса $y_t = 2 + 0.25y_{t-2} + \varepsilon_t$.
3. на $l = 3$ шага для процесса $y_t = \frac{1}{27}y_{t-3} + \varepsilon_t$.
4. на $l = 5$ шагов для процесса $y_t = 0.5y_{t-2} - 0.05y_{t-3} + 0.001y_{t-4} + \varepsilon_t$.
5. на $l = 6$ шагов для процесса $y_t = 0.5y_{t-3} + 0.001y_{t-4} + \varepsilon_t$.

№3. Написать в общем виде уравнения Юла – Уолкера для моделей AR(2), AR(3) и AR(4).

№4. По временному ряду длины $n = 60$ были оценены следующие авторегрессионные модели:

1. $\hat{y}_t = 2 + 0.7y_{t-1}, s^2 = 2.1$;
2. $\hat{y}_t = 2.3 + 0.6y_{t-1} - 0.3y_{t-2}, s^2 = 1.9$;
3. $\hat{y}_t = 1.8 + 0.55y_{t-1} - 0.25y_{t-2} + 0.01y_{t-3}, s^2 = 1, 85$.

Какую модель вы выберете?

№5. Для временного ряда были вычислены коэффициенты автокорреляции

$$\rho(1) = 0.7; \quad \rho(2) = 0.4; \quad \rho(3) = -0.2$$

и выборочное среднее значение $\bar{y} = 1.7$. Найти оценки коэффициентов в модели

1. AR(1);

2. AR(2);

3. AR(3).

№6. Для модели временного ряда длины $n = 50$ были оценены несколько моделей и в каждой из моделей вычислены коэффициенты автокорреляции остатков. Исследовать адекватность этих моделей

1. $r_e(1) = 0.001$; $r_e(2) = 0.0006$; $r_e(3) = 0.0002$; $r_e(4) = 0.001$ в модели AR(2).

2. $r_e(1) = 0.04$; $r_e(2) = 0.02$; $r_e(3) = 0.006$ в модели AR(1).

3. $r_e(1) = 0.02$; $r_e(2) = 0.0008$; $r_e(3) = 0.003$; $r_e(4) = 0.001$ в модели AR(3).

Глава 6

Модели регрессии с бинарной зависимой переменной

6.1 LPM-модель

№1. Рассмотрим LPM модель

$$\widehat{Male} = 0.37 + \underset{[0.011]}{0.02} Educ - \underset{[0.046]}{0.1} City \quad n = 700$$

зависимости пола от уровня образования $Educ$ и места жительства $City$ (в квадратных скобках указаны стандартные ошибки в форме Уайта). Почему при оценке модели были использованы стандартные ошибки в форме Уайта (робастные НС-стандартные ошибки)? Кратко поясните. Проверьте значимость влияния каждого из факторов (уровень значимости 1%). Вычислите предсказанное значение для человека с образованием 15 лет и живущего в городе. Как следует интерпретировать предсказанное значение?

№2. Рассмотрим LPM-модель (male – пол, age – возраст, educ – уровень образования в годах, city – бинарная переменная, отвечающая за место жительства)

Зависимая переменная: male; объем выборки 1320

	const	educ	age	city
coefficient	0.47	0.002	-0.001	-0.81
НС-s.e.	0.02	0.0008	0.0004	0.13

- a) Дайте интерпретацию полученных оценок коэффициентов
- b) Почему почему при оценке модели были использованы стандартные ошибки в форме Уайта? Ответ поясните.
- c) Вычислите предсказанное значение для 35-ти летнего городского жителя с уровнем образования 12 лет и дайте его интерпретацию.
- d) Значимо ли уровень образования влияет на зависимую переменную? Сформулируйте и тестируйте соответствующую статистическую гипотезу

№3. Пусть *approve* – бинарная переменная, равная 1 если кредит заёмщику одобрен. Рассмотрим LPM-модель (*appinc* – доход заёмщика в \$1000, *mortno* – бинарный фактор, равен 1 если нет ипотечной кредитной истории, *unem* – уровень безработицы в отрасли в %, *dep* – количество иждивенцев у заёмщика, *male* – гендерный фактор, *married* – семейный статус, *yjob* – стаж на текущей работе)¹

Зависимая переменная: *approve*; объем выборки: 1971

	const	appinc	mortno	unem	dep	male	married	yjob
коэфф.	0.875	-0.0001	0.074	-0.008	-0.018	0.002	0.047	-0.0003
OLS-s.e.	0.023	0.00008	0.016	0.003	0.007	0.02	0.017	0.007
НС-s.e.	0.024	0.00011	0.015	0.0039	0.0075	0.021	0.0187	0.006

$$R^2 = 0.020483 \quad R^2_{adj} = 0.016990 \quad F - \text{стат} = 5.711436$$

- a) Проверьте значимость влияния дохода заёмщика.
- b) Уровень безработицы в отрасли равен 3.7%. Для неженатого мужчины с ипотечной кредитной истории, с уровнем дохода \$107000, работающего на текущем месте 2 года и без иждивенцев вычислите предсказанное значение и дайте его интерпретацию.
- c) Дайте интерпретацию коэффициента при *married*, *dep* & *unem*.

¹НС – робастные стандартные ошибки, устойчивые к гетероскедастичности

6.2 LPM-модель. Практика

Для решение следующих задач необходимо установить пакеты

```
lmtest, sandwich, wooldridge, AER, Ecdat
```

Для загрузки необходимого датасета используйте команду

```
data(dataset, package)
```

№1. Для набора данных `loanapp`, Таблица A.6, оцените LPM-модель `approve` на остальные факторы.

1. Проверьте значимость коэффициентов
2. Дайте интерпретацию R^2 и проверьте значимость регрессии
3. Дайте интерпретацию **значимым** коэффициентам модели.
4. Тестируйте совместную значимость `arrinc`, `ujob`, `self`, `unem`.
5. Уровень безработицы в отрасли равен 3.7%. Для неженатого самозанятого мужчины с ипотечной кредитной истории, с уровнем дохода \$107000, работающего на текущем месте 2 года и без иждивенцев вычислите предсказанное значение и дайте его интерпретацию.

№2. Для набора данных `mroz_Green`, Таблица A.9, оцените LPM-модель `LFP` на возраст, квадрат возраста, семейный доход, уровень образования, число детей, место жительства `CIT` и уровень безработицы.

1. Проверьте значимость коэффициентов
2. Дайте интерпретацию R^2 и проверьте значимость регрессии
3. Дайте интерпретацию **значимым** коэффициентам модели.
4. тестируйте значимость влияния числа детей.
5. тестируйте значимость влияния возраста.
6. рассмотрим двух женщин со следующими характеристиками

Переменная	#1	#2
WA	35	40
FAMINC	12500	9800
WE	15	12
KL6	2	1
K618	0	2
CIT	1	0
UN	5	7.5

Вычислите прогноз по оценённой модели и дайте его интерпретацию

№3. Для набора данных SwissLabor, Таблица A.7, оцените LPM-модель participation² на остальные факторы + квадрат возраста

1. Проверьте значимость коэффициентов
2. Дайте интерпретацию R^2 и проверьте значимость регрессии
3. Дайте интерпретацию **значимым** коэффициентам модели.
4. тестируйте значимость влияния числа детей.
5. тестируйте значимость влияния возраста.
6. рассмотрим трёх индивидов со следующими характеристиками

Переменная	#1	#2	#3
income	0	1200	500
age	27	35	31
education	12	15	13
youngkids	0	1	1
oldkids	0	0	1
foreign	no	no	yes

Вычислите прогноз по оценённой модели и дайте его интерпретацию

²Т.к. participation – факторная переменная, то зависимую переменную в функции `lm` нужно указать как `as.numeric(participation)-1`

6.3 probit/logit-модель

№1. Рассмотрим probit-модель

$$P(y = 1) = \Phi\left(0.45 + \underset{(0.2)}{0.3x_1} + \underset{(1.2)}{2.1x_2} - \underset{(0.2)}{0.3x_3} + \underset{(0.9)}{1.2x_4}\right)$$

$$\ln \hat{\mathcal{L}} = -1202.3 \quad \ln \hat{\mathcal{L}}_0 = -1382.2 \quad n = 1050$$

Проверьте значимость коэффициентов регрессии. Проверьте значимость регрессии «в целом».

№2. Рассмотрим probit-модель (объем выборки 920)

$$P(\text{Male} = 1) = \Phi\left(0.02 + \underset{(0.009)}{0.013educ} - \underset{(0.046)}{0.1city} - \underset{(0.011)}{0.029age} - \underset{(0.07)}{0.22unem}\right)$$

зависимости пола от уровня образования *educ*, места жительства *city*, возраста *age* и статуса работающий/безработный *unem* (в скобках указаны стандартные ошибки).

- Напишите формулу для функции $\Phi(x)$.
- Проверьте значимость влияния каждого из факторов (уровень значимости 10%).
- Дайте интерпретацию коэффициентов модели.
- Как интерпретируются предсказанные значения в этой модели регрессии?
- Вычислите предельное значение $P(\text{Male} = 1)$ по *educ*.

№3. Рассмотрим logit-модель, оцененную в EViews

Dependent variable: crime86
Sample 1 2725

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-0.655477	0.119205	-5.498726	0.0000
pcnv	-0.924175	0.124452	-7.425948	0.0000
avgsen	0.018503	0.035298	0.524185	0.6001
totime	-0.010810	0.028172	-0.383710	0.7012
ptime86	-0.123048	0.031627	-3.890563	0.0001
qemp86	0.066483	0.045187	1.471275	0.1412
inc86	-0.009057	0.001290	-7.021432	0.0000
durat	0.019617	0.010214	1.920591	0.0548
black	0.776393	0.118454	6.554379	0.0000
hispan	0.498160	0.109949	4.530833	0.0000
born60	0.018356	0.094461	0.194319	0.8459

$$\ln \hat{\mathcal{L}} = -1481.458 \quad \ln \hat{\mathcal{L}}_0 = -1608.184$$

- Какие из факторов значимо, а какие незначимо влияют на зависимую переменную? Уровень значимости 10%.
- Вычислите $R_{McFadden}^2$
- Тестируйте значимость регрессии «в целом» при уровне значимости 1%.
- Была оценена *Logit*-модель без факторов black & hispan и получено значение логарифма функции правдоподобия $\ln \hat{\mathcal{L}}_r = -1506.727$. Тестируйте совместную значимость факторов black & hispan при уровне значимости 5%.

№4. Пусть *City* – бинарная переменная, равная 1 если человек живет в городе. Рассмотрим *logit*-модель (*educ* – уровень образования, *age* – возраст, *female* – пол человека)

Зависимая переменная: *City*; объем выборки 1023

	const	educ	age	female
coefficient	0.02	0.004	0.0018	0.04
Std. Error	0.13	0.0013	0.0002	0.023

$$\ln \hat{\mathcal{L}} = -1281 \quad \ln \hat{\mathcal{L}}_0 = -1508$$

Уровень значимости 5%

- a) Запишите уравнение регрессии и формулу для logit-функции.
- b) Дайте интерпретацию коэффициентов регрессии.
- c) Значимо ли влияние уровня образования?
- d) Вычислите $R_{McFadden}^2$
- e) Тестируйте значимость регрессии «в целом» и сформулируйте проверяемую статическую гипотезу.
- f) Как интерпретируются предсказанные значения в этой модели регрессии?
- g) Для мужчины возраста 30 лет и с образованием 15 лет вычислите предсказанное значение и дайте его интерпретацию.
- h) Укажите формулу для предельных значений по educ и по age.
- i) Для мужчин возраста 25 лет и 35 лет и с образованием 15 лет (каждый) вычислите предельные значения по age и дайте их интерпретацию.
- j) Пусть LR-статистика для проверки совместной значимости факторов age и female равна 13.5. Проверьте совместную значимость этих факторов и сформулируйте проверяемую статистическую гипотезу.

№5. Для набора данных `loanapp`, Таблица A.6, были оценены несколько логит-регрессий. В Таблице B.14 приведены результаты оценивания. В скобках указаны стандартные ошибки коэффициентов, $\ln \hat{\mathcal{L}}_0 = -737.979$

1. Дайте интерпретацию **значимым** коэффициентам первой регрессии.
2. Тестируйте значимость «в целом» регрессии №7.
3. Тестируйте совместную значимость факторов `mortno`, `dep`, `male & married`.
4. Тестируйте совместную значимость `appinc`, `male & ujob`.
5. Для каждой модели вычислите показатель «качества подгонки»
6. Какая из альтернативных моделей предпочтительней?
7. На Рис. 6.1 приведены графики ROC-кривых для моделей №1, №2 и №4. Какая модель предпочтительней?

№6. Для набора данных `loanapp`, Таблица A.6, были оценены несколько регрессий. В Таблице B.15 приведены результаты оценивания. В скобках указаны стандартные ошибки коэффициентов (для LPM-модели OLS и НС), $\ln \hat{\mathcal{L}}_0 = -737.979$.

1. Для линейной модели тестируйте значимость влияния семейного статуса.
2. Для неженатого мужчины с ипотечной кредитной историей, с уровнем дохода \$107000, работающего на текущем месте 2 года и имеющего одного иждивенца вычислите предсказанное значение по каждой оценённой модели и дайте его интерпретацию.
3. Запишите **теоретическую** модель LPM, logit и probit-регрессии.
4. Дайте интерпретацию значимым коэффициентам logit- и probit-модели
5. Вычислите коэффициент «качества подгонки» для logit и probit-регрессии.

6. На Рис. 6.2 приведены графики ROC-кривых для оценённых моделей. Какая модель предпочтительней?

№7. Для набора данных `default`, Таблица A.14, были оценены несколько logit-регрессий. В Таблице B.16 приведены результаты оценивания. В скобках указаны стандартные ошибки коэффициентов, $\ln \hat{\mathcal{L}}_0 = -2137.198$.

1. Дайте интерпретацию **значимым** коэффициентов второй регрессии.
2. Тестируйте значимость «в целом» регрессии №6.
3. Тестируйте совместную значимость факторов дохода и возраста.
4. Какая регрессия предпочтительней? Ответ обоснуйте
5. На Рис. 6.3 приведены ROC-кривые для регрессий №1, №2 и №4. Какие можно сделать выводы?

6.4 probit/logit-модель. Практика

№1. Для набора данных `loanapp`, Таблица A.6, оцените logit-модель `approve` на остальные факторы.

1. Чему равен показатель «качества подгонки»
2. Проверьте значимость коэффициентов
3. Проверьте значимость регрессии
4. Дайте интерпретацию **значимым** коэффициентам модели.
5. Тестируйте совместную значимость `arrinc`, `ujob`, `self`, `unem`.
6. Уровень безработицы в отрасли равен 3.7%. Для неженатого самозанятого мужчины с ипотечной кредитной историей, с уровнем дохода \$107000, работающего на текущем месте 2 года и без иждивенцев вычислите предсказанное значение и дайте его интерпретацию.
7. Сравните эти прогнозы с прогнозами LPM-модели

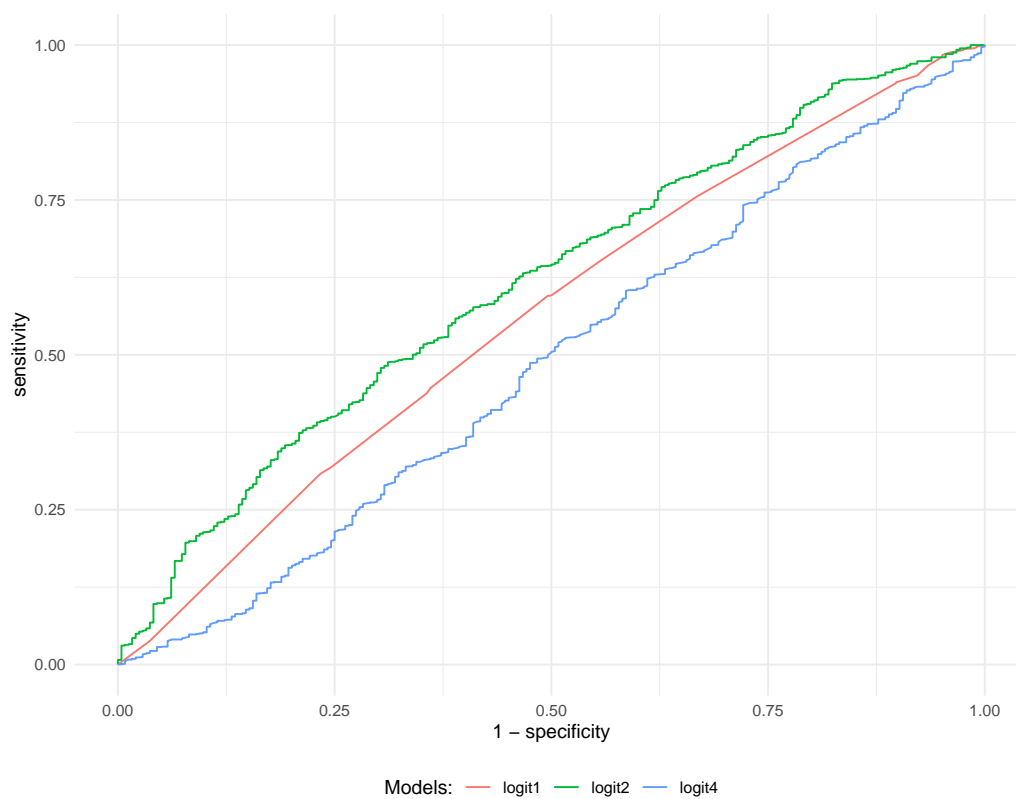


Рис. 6.1: ROC-кривые для моделей 1, 2 и 4

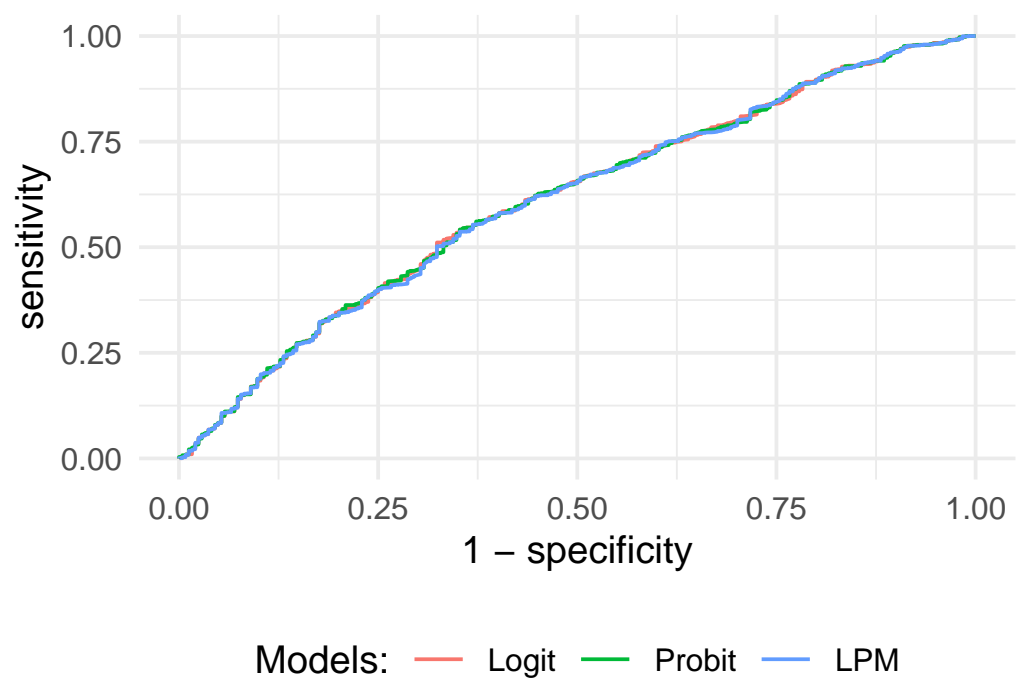


Рис. 6.2: ROC-кривые для logint-, probit- и LPM-моделей

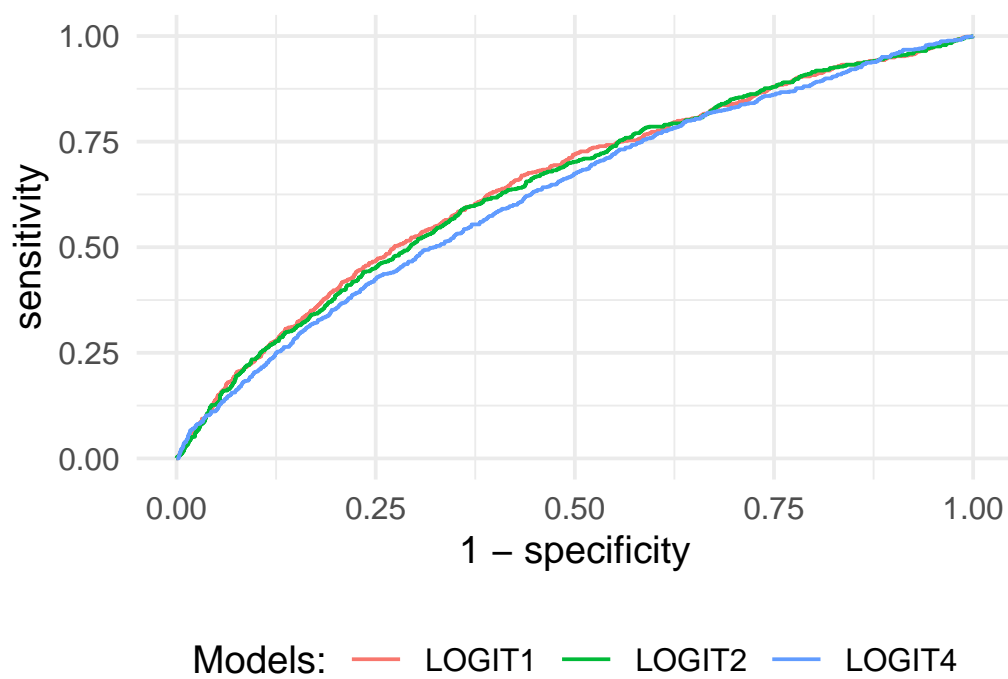


Рис. 6.3: ROC-кривые для моделей 1 и 3

8. Постройте ROC-кривую

№2. Для набора данных `mroz_Green`, Таблица **A.9**, оцените probit-модель LFP на возраст, квадрат возраста, семейный доход, уровень образования, число детей, место жительства CIT и уровень безработицы.

1. Чему равен показатель «качества подгонки»
2. Проверьте значимость коэффициентов
3. Проверьте значимость регрессии
4. Дайте интерпретацию **значимым** коэффициентам модели.
5. тестируйте значимость влияния возраста.
6. рассмотрим двух женщин со следующими характеристиками

Переменная	#1	#2
WA	35	40
FAMINC	12500	9800
WE	15	12
KL6	2	1
K618	0	2
CIT	1	0
UN	5	7.5

Вычислите прогноз по оценённой модели и дайте его интерпретацию

7. Постройте ROC-кривую

Глава 7

Панельные данные

7.1 Модели панельных данных

№1 (More Guns, Less Crime?). На основе панельных данных `Guns`, Таблица [A.17](#), оценим влияние на уровень преступности `violent` закона об ограничении оружия `law` (?). Рассмотрим несколько регрессий: модель пула, модель фиксированных эффектов и модель случайных эффектов. Результаты оценивания регрессий с индивидуальными эффектами приведены в Таблице [B.17](#), в скобках указаны робастные стандартные ошибки (ошибки Arellano-Bond).

Результаты диагностических тестов:

- i) Тест Вулдриджа на ненаблюдаемые эффекты:

```
Wooldridge's test for unobserved individual effects
```

```
data: formula
z = 3.8804, p-value = 0.0001043
alternative hypothesis: unobserved effect
```

- ii) F-теста для индивидуальных эффектов

```
F test for individual effects
```

```
data: guns.formula
F = 142.57, df1 = 50, df2 = 1114, p-value < 2.2e-16
alternative hypothesis: significant effects
```

iii) Тест Хаусмана

Hausman Test

```
data: guns.formula
chisq = 31.859, df = 8, p-value = 9.872e-05
alternative hypothesis: one model is inconsistent
```

iv) LM-тест на индивидуальные эффекты

Lagrange Multiplier Test - (Honda) for balanced panels

```
data: guns.formula
normal = 83.024, p-value < 2.2e-16
alternative hypothesis: significant effects
```

Ответьте на следующие вопросы:

1. Напишите спецификации моделей регрессии
2. Какая модель предпочтительней?
3. Что можно сказать о значимости влияния закона на уровень преступности?

№2 (More Guns, Less Crime?). На основе панельных данных `Guns`, Таблица [A.17](#), оценим влияние на уровень преступности `violent` закона об ограничении оружия `law` (?). Рассмотрим несколько регрессий: модель пула, модель фиксированных эффектов и модель случайных эффектов. Результаты оценивания регрессий с двунаправленными эффектами приведены в Таблице [B.18](#), в скобках указаны робастные стандартные ошибки (ошибки Arellano-Bond).

Результаты диагностических тестов:

i) Тест Вулдриджа на ненаблюдаемые эффекты:

Wooldridge's test for unobserved time effects

```
data: formula
z = 1.3693, p-value = 0.1709
alternative hypothesis: unobserved effect
```

ii) F-теста для индивидуальных эффектов

```
F test for twoways effects
```

```
data: guns.formula
F = 135.65, df1 = 72, df2 = 1092, p-value < 2.2e-16
alternative hypothesis: significant effects
```

iii) Тест Хаусмана

```
Hausman Test
```

```
data: guns.formula
chisq = 217.35, df = 8, p-value < 2.2e-16
alternative hypothesis: one model is inconsistent
```

iv) LM-тест на индивидуальные эффекты

```
Lagrange Multiplier Test - time effects (Honda) for balanced panels
```

```
data: guns.formula
normal = 4.1075, p-value = 2e-05
alternative hypothesis: significant effects
```

Ответьте на следующие вопросы:

1. Напишите спецификации моделей регрессии
2. Какая модель предпочтительней?
3. Что можно сказать о значимости влияния закона на уровень преступности?

№3 (Raising revenues and reducing cigarette consumption). На основе панельных данных *Cigar*, Таблица A.19, оценим влияние на объём продаж сигарет нескольких факторов. Рассмотрим несколько регрессий: модель пула, модель фиксированных эффектов и модель случайных эффектов. Результаты оценивания регрессий с индивидуальными эффектами приведены в Таблице B.19, в скобках указаны робастные стандартные ошибки (ошибки Arellano-Bond).

Результаты диагностических тестов:

i) Тест Вулдриджа на ненаблюдаемые эффекты:

Wooldridge's test for unobserved individual effects

```
data: formula
z = 2.2709, p-value = 0.02315
alternative hypothesis: unobserved effect
```

ii) F-теста для индивидуальных эффектов

F test for individual effects

```
data: cigar.formula
F = 98.631, df1 = 45, df2 = 1328, p-value < 2.2e-16
alternative hypothesis: significant effects
```

iii) Тест Хаусмана

Hausman Test

```
data: cigar.formula
chisq = 355.99, df = 6, p-value < 2.2e-16
alternative hypothesis: one model is inconsistent
```

iv) LM-тест на индивидуальные эффекты

Lagrange Multiplier Test - (Honda) for balanced panels

```
data: cigar.formula
normal = 91.163, p-value < 2.2e-16
alternative hypothesis: significant effects
```

Ответьте на следующие вопросы:

1. Напишите спецификации моделей регрессии
2. Какая модель предпочтительней?
3. Что можно сказать о значимости влияния дохода на потребление сигарет?

7.2 Модели панельных данных. Практика

Приложение А

Описание наборов данных

В этом разделе приведены описания основных наборов данных, используемых для решения задач.

Для работы с нужным датасетом необходимо загрузить его командой

```
data(dataset, package)
```

Таблица А.1: Набор данных `Labour` из пакета `Ecdat` (569 наблюдений) с данными о бельгийских фирмах за 1996 г.

<code>capital</code>	капитал (в млн евро)
<code>labour</code>	число сотрудников
<code>output</code>	выпуск (в млн евро)
<code>wage</code>	зарплата на одного сотрудника (в тыс евро)

Таблица А.2: Набор данных *Electricity* из пакета *Ecdat* (158 наблюдений) с данными о производителях электроэнергии в US за 1970 г.

cost	общие издержки за год
q	общий выпуск электроэнергии
pl	уровень зарплата (wage rate)
pk	цена привлечения капитала (capital price index)
pf	цена на топливо (fuel price)

Таблица А.3: Набор данных *sleep75* из пакета *wooldridge* (706 наблюдений). Основные переменные. Источник данных [5]

age	возраст (в годах)
educ	уровень образования (в года)
inlf	бинарная, 1 если участник рынка труда
leis1	нерабочее время, sleep-totwrk
smsa	бинарная, 1 если живёт в мегаполисе
male	гендерный фактор (бинарная, 1 если мужчина)
marr	семейный статус (бинарная, 1 если женат/замужем)
prot	бинарная, 1 если протестант
selfe	бинарная, 1 если самозанятый
sleep	продолжительность сна (мин/нед)
south	географический фактор (бинарная, 1 если живёт на юге)
spsepay	доход супруга/супруги
spwrk75	бинарная, 1 если супруг(а) работает
totwrk	занятость (мин/нед)
union	бинарная, 1 если член профсоюза
yngkid	бинарная, 1 если есть дети младше 3 лет
yrsmarr	сколько лет женат/замужем
hrwage	почасовая оплата

Таблица А.4: Набор данных `wage2` из пакета `wooldridge` (935 наблюдений). Основные переменные. Источник данных [6]

<code>wage</code>	месячная зарплата
<code>hours</code>	недельная занятость в часах
<code>IQ</code>	результаты теста IQ
<code>KWW</code>	результаты теста knowledge of world work
<code>educ</code>	уровень образования (в годах)
<code>exper</code>	опыт работы в годах
<code>tenure</code>	стаж работы на текущем месте
<code>age</code>	возраст (в годах)
<code>married</code>	семейный статус (бинарная, 1 если женат/замужем)
<code>south</code>	географический фактор (бинарная, 1 если живёт на юге)
<code>urban</code>	место жительства (1 если живет в городе)
<code>sibs</code>	число братьев/сестёр
<code>brthord</code>	какой по счёту ребёнок в семье
<code>meduc</code>	уровень образования матери (в годах)
<code>feduc</code>	уровень образования отца (в годах)

Таблица А.5: Набор данных `wage1` из пакета `wooldridge` (526 наблюдений). Основные переменные.

<code>wage</code>	средняя почасовая оплата
<code>educ</code>	уровень образования (в годах)
<code>female</code>	гендерный фактор
<code>exper</code>	опыт работы
<code>tenure</code>	стаж на текущем месте работы
<code>married</code>	семейный статус
<code>smsa</code>	живёт ли в мегаполисе (бинарная)
<code>south</code>	географический фактор (бинарная)
<code>west</code>	географический фактор (бинарная)
<code>northcen</code>	географический фактор (бинарная)

Таблица А.6: Набор данных `loanapp` из пакета `wooldridge` (1989 наблюдений). Основные переменные. Источник данных [9]

<code>approve</code>	бинарная, 1 если кредитная заявка одобрена
<code>arrinc</code>	доход заявителя (в \$1000)
<code>mortno</code>	бинарная, 1 если нет ипотечной кредитной истории
<code>unem</code>	уровень безработицы в отрасли в %
<code>dep</code>	количество иждивенцев
<code>male</code>	гендерный фактор
<code>married</code>	семейный статус
<code>yjob</code>	стаж на текущей работе
<code>self</code>	бинарная, 1 если самозанятый

Таблица А.7: Набор данных `SwissLabor` о рынке труда Швейцарии из пакета `AER` (872 наблюдений). Источник данных [7]

<code>participation</code>	Является ли участником рынка труда? (фактор, "yes"/"no")
<code>income</code>	логарифм дополнительного дохода (<code>nonlabor income</code>)
<code>age</code>	возраст (в десятилетиях)
<code>education</code>	уровень образования
<code>youngkids</code>	число маленьких детей (младше 7 лет)
<code>oldkids</code>	число старших детей (старше 7 лет)
<code>foreign</code>	является ли иностранцем? (фактор, "yes"/"no")

Таблица А.8: Набор данных `Consumption` из пакета `Ecdat` об индивидуальных доходах и расходах в Канаде (квартальные данные с 1947Q1 по 1996Q4).

<code>yd</code>	индивидуальный располагаемый доход в ценах 1986
<code>se</code>	индивидуальные расходы на потребление в ценах 1986

Таблица A.9: Набор данных из файла `mroz_Greene.csv` с сайта <http://meit.mgimo.ru/node/237> содержит данные о рынке труда замужних женщин. Основные переменные.

LFP	бинарная, 1 женщина работает
WHRS	уровень занятости (в часах)
KL6	число детей моложе 6 лет в семье
K618	число детей от 6 до 18 лет в семье
WA	возраст
WE	уровень образования (в годах)
WW	средняя почасовая оплата
HHRS	занятость мужа
HA	возраст мужа
HE	уровень образования мужа (в годах)
HW	зарплата мужа
FAMINC	доход домашнего хозяйства
WMED	уровень образования матери
WFED	уровень образования отца
UN	уровень безработицы в стране проживания
CIT	бинарная, 1 если живет в мегаполисе
AX	предыдущий опыт работы (в годах)

Таблица A.10: Набор данных `diamonds` из пакета `ggplot2` с данными о бриллиантах (53940 наблюдений). Основные переменные.

price	цена бриллианта
carat	вес бриллианта (в каратах)
cut	качество огранки (упорядоченный фактор с уровнями Fair<Good< Very Good<Premium<Ideal)
color	цвет (упорядоченный фактор с уровнями J<I<H<G<F<E<D)
clarity	прозрачность (упорядоченный фактор с уровнями I1<SI2<SI1<VS2<VS1<VVS2<VVS1<IF)
x	длина (в мм)
y	ширина (в мм)
z	глубина (в мм)

Таблица А.11: Набор данных `Diamond` из пакета `Ecdat` с данными о бриллиантах (308 наблюдений). Основные переменные. Источник данных [4]

<code>carat</code>	вес бриллианта (в каратах)
<code>colour</code>	цвет (фактор с уровнями D,E,F,G,H,I)
<code>clarity</code>	прозрачность (фактор с уровнями IF,VVS1,VVS2,VS1,VS2)
<code>certification</code>	орган по сертификации (фактор с уровнями GIA,IGI,HRD)
<code>price</code>	цена в Сингапуре

Таблица А.12: Набор данных `Icescream` из пакета `Ecdat` о потреблении мороженого в США (недельные данные с 1951–03–18 по 1953–07–11, всего 30 наблюдений). Источник данных [8]

<code>cons</code>	потребление мороженого (в пинтах)
<code>income</code>	средний недельный доход семьи (в \$)
<code>price</code>	цена мороженого (за пинту)
<code>temp</code>	средняя температура (по Фаренгейту)

Таблица А.13: Набор данных `nlsw88` из пакета `Counterfactual` о занятости женщин в США (2246 наблюдений). Основные переменные. Источник данных: сайт Stata <http://www.stata-press.com/data/r10/g.html>

<code>hours</code>	недельная занятость (в часах)
<code>married</code>	семейный статус
<code>ttl_exp</code>	общий стаж работы
<code>smsa</code>	бинарная, 1 если живёт в мегаполисе
<code>south</code>	географический фактор (бинарная, 1 если живёт на юге)
<code>wage</code>	почасовая оплата (в \$)
<code>age</code>	возраст (в годах)
<code>grade</code>	уровень образования (в годах)

Таблица A.14: Набор данных из файла `default.csv` с сайта <http://meit.mgimo.ru/node/237> о банкротствах по студенческим займам (6778 наблюдений)

<code>default</code>	бинарная переменная, равная 1 если индивид признал себя банкротом по студенческому займу
<code>age</code>	возраст
<code>adepcnt</code>	количество иждивенцев у индивида плюс 1
<code>acadmos</code>	количество месяцев, которые индивид прожил по текущему адресу
<code>majordrg</code>	количество зарегистрированных серьёзных правонарушений у этого индивида
<code>minordrg</code>	количество зарегистрированных мелких правонарушений у этого индивида
<code>ownrent</code>	1, если индивид живёт в собственном доме, и 0, если снимает
<code>income</code>	месячный доход в \$
<code>spending</code>	среднемесячный расход по кредитной карте
<code>inc_per</code>	<code>income</code> , делённая на количество иждивенцев
<code>exp_inc</code>	доля месячных расходов по кредитной карте в годовой заработной плате
<code>selfempl</code>	1, если индивид самозанятый, и 0 иначе

Таблица A.15: Набор данных `stockton3` из пакета `PoEdata` с данными о стоимости домов (2610 наблюдений). Основные переменные. Источник данных <https://github.com/ccolonescu/PoEdata>

<code>sprice</code>	цена продажи дома (в \$)
<code>livarea</code>	жилая площадь (кв.футы)
<code>pool</code>	наличие бассейна (бинарная)
<code>lgelot</code>	размер участка (бинарный фактор, 1 если участок больше 5 акров)
<code>age</code>	возраст (в годах)
<code>beds</code>	число спален

Таблица A.16: Набор данных из файла `applications.csv` с сайта <http://meit.mgimo.ru/node/237> о поступивших на магистерские и PhD-программы (400 наблюдений)

<code>admit</code>	бинарный фактор, 1 если заявка одобрена
<code>GPA</code>	средняя оценка за время обучения
<code>GRE</code>	балл за экзамен graduate record exam
<code>rank</code>	категориальная переменная, обозначающая престиж университета (1 – высший престиж, 4 – низший престиж)

Таблица A.17: Панель `Guns` из пакета `AER` с данными по 51 штату США с 1977 по 1999 гг. (всего 1173 наблюдения). Основные переменные. Источник данных [1]

<code>state</code>	factor indicating state
<code>year</code>	factor indicating year
<code>violent</code>	violent crime rate (incidents per 100,000 members of the population)
<code>murder</code>	murder rate (incidents per 100,000).
<code>robbery</code>	robbery rate (incidents per 100,000)
<code>prisoners</code>	incarceration rate in the state in the previous year (sentenced prisoners per 100,000 residents; value for the previous year)
<code>afam</code>	percent of state population that is African-American, ages 10 to 64
<code>cauc</code>	percent of state population that is Caucasian, ages 10 to 64
<code>male</code>	percent of state population that is male, ages 10 to 29
<code>population</code>	state population, in millions of people.
<code>income</code>	real per capita personal income in the state (US dollars)
<code>density</code>	population per square mile of land area, divided by 1,000
<code>law</code>	factor. Does the state have a shall carry law in effect in that year?

Таблица A.18: Панель `LaborSupply` из пакета `plm`, `Ecdat` с данными по 532 индивидуумам с 1979 по 1988 гг. (всего 5320 наблюдений). Основные переменные. Источник данных [10]

<code>lnhr</code>	логарифм годовой занятости в часах
<code>lnwg</code>	логарифм почасовой оплаты
<code>kids</code>	число детей
<code>age</code>	возраст
<code>disab</code>	бинарная, 1 если плохое здоровье

Таблица А.19: Панель **Cigar** из пакета **plm** с данными по 46 штатам США с 1963 по 1992 гг. (всего 1380 наблюдений). Основные переменные. Источник данных [2]

state	state abbreviation
year	the year
price	price per pack of cigarettes
pop	population
pop16	population above the age of 16
spi	consumer price index (1983=100)
ndi	per capita disposable income
sales	cigarette sales in packs per capita
pimin	minimum price in adjoining states per pack of cigarettes

Таблица А.20: Панель **Gasoline** из пакета **plm**, **Ecdat** с данными о потреблении бензина по 18 странам ОЕСД с 1960 по 1978 гг. (всего 342 наблюдений). Основные переменные. Источник данных [3]

lgasrcar	логарифм потребления бензина
lincomer	логарифм реального дохода на душу населения
lprmg	логарифм реальной цены на бензин
lcarrcar	логарифм объёма рынка машин

Приложение В

Результаты оценивания регрессий

Таблица В.1: Результаты оценивания регрессий

	Зависимая переменная				
	sleep				
	(1)	(2)	(3)	(4)	(5)
Constant	3597.451*** (42.265)	3650.351*** (219.497)	3600.613*** (42.728)	3549.857*** (56.801)	3213.749*** (40.666)
totwrk	-0.154*** (0.017)	-0.150*** (0.017)	-0.154*** (0.017)	-0.153*** (0.017)	
south	92.635** (41.909)	91.673** (41.872)	93.617** (41.974)	97.880** (42.130)	
smsa	-54.549 (33.147)	-55.934* (33.072)	-55.274* (33.194)	-50.840 (33.374)	
age		-6.271 (11.224)			
I(age^2)		0.112 (0.134)			
yngkid			-24.367 (47.183)	-32.194 (47.655)	-27.218 (50.437)
marr				55.057 (41.786)	65.617 (44.104)
union				12.675 (38.314)	10.126 (40.549)
Observations	706	706	706	706	706
R ²	0.116	0.123	0.116	0.118	0.003
Adjusted R ²	0.112	0.116	0.111	0.111	-0.001

Note:

*p<0.1; **p<0.05; ***p<0.01

Таблица В.2: Результаты оценивания регрессий

	Зависимая переменная				
	sleep				
	(1)	(2)	(3)	(4)	(5)
Constant	3586.377*** (38.912)	3674.950*** (219.484)	3650.351*** (219.497)	3624.347*** (220.422)	3240.833*** (43.940)
totwrk	-0.151*** (0.017)	-0.147*** (0.017)	-0.150*** (0.017)	-0.149*** (0.017)	
age		-8.102 (11.268)	-6.271 (11.224)	-7.605 (11.269)	
I(age ²)		0.133 (0.135)	0.112 (0.134)	0.128 (0.135)	
south			91.673** (41.872)	95.578** (42.001)	
smsa			-55.934* (33.072)	-51.175 (33.266)	-56.915* (34.241)
marr				52.717 (41.413)	56.522 (43.766)
union				18.382 (38.241)	8.352 (40.478)
Observations	706	706	706	706	706
R ²	0.103	0.110	0.123	0.125	0.007
Adjusted R ²	0.102	0.106	0.116	0.116	0.003

Note:

*p<0.1; **p<0.05; ***p<0.01

Таблица В.3: Результаты оценивания регрессий (Тест Чоу)

	Зависимая переменная		
	sleep/60		
	общая	только муж	только жен
	(1)	(2)	(3)
Constant	61.249*** (3.658)	58.839*** (4.744)	66.394*** (5.831)
totwrk	-0.002*** (0.0003)	-0.003*** (0.0004)	-0.002*** (0.0005)
age	-0.135 (0.188)	0.091 (0.237)	-0.437 (0.305)
I(age^2)	0.002 (0.002)	-0.0004 (0.003)	0.006 (0.004)
RSS	34420.266	17896.293	16066.166
Observations	706	400	306
R ²	0.110	0.147	0.089
Adjusted R ²	0.106	0.141	0.080
Residual Std. Error	7.002	6.723	7.294
F Statistic	28.944***	22.831***	9.837***

Note:

*p<0.1; **p<0.05; ***p<0.01

Таблица В.4: Результаты оценивания регрессий (Тест Чоу)

	Зависимая переменная		
	log(wage)		
	общая	только жен	только не жен
	(1)	(2)	(3)
Constant	5.087*** (0.166)	5.141*** (0.175)	5.332*** (0.512)
age	0.023*** (0.004)	0.023*** (0.004)	0.007 (0.014)
south	-0.096*** (0.027)	-0.100*** (0.028)	-0.087 (0.083)
urban	0.170*** (0.028)	0.172*** (0.029)	0.205** (0.092)
IQ	0.008*** (0.001)	0.008*** (0.001)	0.009*** (0.002)
Observations	935	835	100
R ²	0.178	0.182	0.185
Adjusted R ²	0.175	0.178	0.151
Residual Std. Error	0.383	0.378	0.383
F Statistic	50.402***	46.256***	5.396***

Note:

*p<0.1; **p<0.05; ***p<0.01

Таблица В.5: Результаты оценивания регрессий

	Зависимая переменная			
	sleep			
	(1)	(2)	(3)	(4)
Constant	3597.061*** (219.572)	3288.045*** (31.154)	3594.521*** (231.461)	3527.065*** (53.605)
totwrk	-0.090* (0.048)			-0.097** (0.048)
I(totwrk^2)	-0.00002* (0.00001)			-0.00002* (0.00001)
age	-7.016 (11.170)		-19.739* (11.734)	
I(age^2)	0.119 (0.133)		0.282** (0.140)	
male	90.384*** (34.162)	-29.845 (33.678)	-29.346 (33.532)	94.476*** (34.184)
south	90.241** (41.906)	74.770* (44.334)	71.915 (44.122)	91.834** (41.920)
smsa	-60.545* (32.957)	-46.437 (35.064)	-48.418 (34.865)	-59.192* (33.021)
Observations	706	706	706	706
R ²	0.135	0.010	0.024	0.129
Adjusted R ²	0.126	0.006	0.017	0.123
F Statistic	15.578***	2.305*	3.477***	20.720***

Note:

*p<0.1; **p<0.05; ***p<0.01

Таблица В.6: Показатели VIF для регрессий из Таблицы В.5

	VIF			
	(1)	(2)	(3)	(4)
totwrk	8.555			8.474
I(totwrk ²)	8.296			8.280
age	65.593		64.347	
I(age ²)	65.726		64.313	
male	1.173	1.001	1.004	1.169
south	1.080	1.061	1.064	1.076
smsa	1.066	1.060	1.061	1.066
Observations	706	706	706	706

Note: *p<0.1; **p<0.05; ***p<0.01

Таблица В.7: Корреляционная матрица для регрессоров из Таблицы В.5

	sleep	totwrk	I(totwrk ²)	age	I(age ²)	male	south	smsa
sleep	1	-0.32	-0.32	0.09	0.10	-0.04	0.08	-0.07
totwrk	-0.32	1	0.94	-0.05	-0.07	0.38	0.05	-0.04
I(totwrk ²)	-0.32	0.94	1	-0.05	-0.06	0.35	0.02	-0.05
age	0.09	-0.05	-0.05	1	0.99	0.03	-0.02	0.02
I(age ²)	0.10	-0.07	-0.06	0.99	1	0.03	-0.01	0.02
male	-0.04	0.38	0.35	0.03	0.03	1	-0.03	0.01
south	0.08	0.05	0.02	-0.02	-0.01	-0.03	1	-0.24
smsa	-0.07	-0.04	-0.05	0.02	0.02	0.01	-0.24	1

Таблица В.8: Результаты оценивания регрессий

	Зависимая переменная			
	sleep			
	(1)	(2)	(3)	(4)
Constant	3469.951*** (91.497)	3469.201*** (68.118)	3435.860*** (72.245)	3500.328*** (87.363)
totwrk	-0.143*** (0.026)	-0.149*** (0.017)	-0.144*** (0.026)	-0.164*** (0.018)
age	1.562 (2.088)	2.924** (1.397)	2.498* (1.409)	1.715 (2.083)
male	101.150 (141.467)		170.967** (82.493)	11.768 (116.589)
I(male *totwrk)	-0.041 (0.036)		-0.042 (0.036)	
I(male *age)	1.720 (2.830)			1.862 (2.827)
Observations	706	706	706	706
R ²	0.119	0.109	0.118	0.117
Adjusted R ²	0.112	0.106	0.113	0.112
F Statistic	18.853***	42.931***	23.494***	23.247***

Note:

*p<0.1; **p<0.05; ***p<0.01

Таблица В.9: Показатели VIF для регрессий из Таблицы В.8

	VIF			
	(1)	(2)	(3)	(4)
totwrk	2.492	1.002	2.481	1.195
age	2.255	1.002	1.028	2.245
male	19.792		6.736	13.438
I(male *totwrk)	9.949		9.929	
I(male *age)	14.413			14.384
Observations	706	706	706	706

Note: *p<0.1; **p<0.05; ***p<0.01

Таблица В.10: Корреляционная матрица для регрессоров из Таблицы В.8

	sleep	totwrk	age	male	I(male *totwrk)	I(male *age)
sleep	1	-0.32	0.09	-0.04	-0.16	0.01
totwrk	-0.32	1	-0.05	0.38	0.65	0.29
age	0.09	-0.05	1	0.03	-0.04	0.33
male	-0.04	0.38	0.03	1	0.88	0.92
I(male *totwrk)	-0.16	0.65	-0.04	0.88	1	0.77
I(male *age)	0.01	0.29	0.33	0.92	0.77	1

Таблица В.11: Результаты оценивания регрессий

	Зависимая переменная			
	log(wage)			
	(1)	(2)	(3)	(4)
Constant	5.997*** (0.143)	5.884*** (0.174)	6.000*** (0.258)	5.857*** (0.286)
age	0.021*** (0.004)	0.025*** (0.005)	0.021*** (0.008)	0.026*** (0.009)
south	-0.151*** (0.028)	0.192 (0.300)	-0.151*** (0.028)	0.196 (0.302)
urban	0.174*** (0.029)	0.177*** (0.029)	0.170 (0.307)	0.212 (0.309)
I(south *age)		-0.010 (0.009)		-0.010 (0.009)
I(urban *age)			0.0001 (0.009)	-0.001 (0.009)
Observations	935	935	935	935
R ²	0.097	0.098	0.097	0.098
Adjusted R ²	0.094	0.094	0.093	0.093
F Statistic	33.179***	25.221***	24.858***	20.158***

Note:

*p<0.1; **p<0.05; ***p<0.01

Таблица В.12: Результаты оценивания регрессий

	Зависимая переменная			
	sleep			
	(1)	(2)	(3)	(4)
Constant	3437.003*** (75.969)	3401.466*** (81.167)	3461.216*** (122.745)	3472.730*** (94.228)
totwrk	-0.165*** (0.018)	-0.141*** (0.026)	-0.175*** (0.046)	-0.163*** (0.018)
age	2.728* (1.394)	2.469* (1.409)	2.699* (1.400)	1.735 (2.084)
male	79.840** (35.015)	172.506** (82.517)	79.448** (35.073)	8.444 (116.698)
marr	33.442 (41.937)	39.229 (42.180)	6.953 (113.477)	32.859 (41.965)
I(totwrk *male)		-0.045 (0.037)		
I(totwrk *marr)			0.012 (0.049)	
I(age *male)				1.814 (2.829)
Observations	706	706	706	706
R ²	0.117	0.119	0.117	0.118
Adjusted R ²	0.112	0.113	0.111	0.112
F Statistic	23.304***	18.965***	18.631***	18.710***

Note:

*p<0.1; **p<0.05; ***p<0.01

Таблица В.13: Результаты оценивания регрессий

	Зависимая переменная		
	log(price)		
	(1)	(2)	(3)
Constant	5.866*** (0.049)	5.975*** (0.042)	5.686*** (0.063)
carat	5.667*** (0.135)	5.404*** (0.121)	5.650*** (0.198)
I(carat^2)	-2.153*** (0.099)	-1.983*** (0.094)	-2.187*** (0.145)
colourE	-0.076** (0.031)	-0.072** (0.032)	
colourF	-0.152*** (0.029)	-0.143*** (0.030)	
colourG	-0.227*** (0.029)	-0.214*** (0.030)	
colourH	-0.313*** (0.030)	-0.305*** (0.031)	
colourI	-0.408*** (0.031)	-0.399*** (0.032)	
certificationHRD	0.042*** (0.015)		0.031 (0.022)
certificationIGI	0.085*** (0.019)		0.068** (0.027)
Observations	308	308	308
R ²	0.984	0.983	0.965
Adjusted R ²	0.983	0.982	0.964

Note: *p<0.1; **p<0.05; ***p<0.01

Таблица В.14: Результаты оценивания регрессий

	Зависимая переменная						
	approve						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Constant	1.766*** (0.158)	1.942*** (0.199)	1.951*** (0.169)	2.283*** (0.154)	1.867*** (0.160)	2.095*** (0.157)	1.899*** (0.165)
appinc		-0.001* (0.001)	-0.001* (0.001)	-0.001 (0.001)		-0.001* (0.001)	-0.001 (0.001)
mortno		0.773*** (0.172)	0.773*** (0.172)		0.752*** (0.171)	0.793*** (0.169)	
unem		-0.063** (0.029)	-0.062** (0.029)	-0.065** (0.028)	-0.061** (0.029)	-0.063** (0.028)	
dep	-0.146** (0.064)	-0.162** (0.065)	-0.162** (0.064)		-0.170*** (0.064)		
male	-0.018 (0.183)	0.022 (0.186)					0.148 (0.171)
married	0.516*** (0.161)	0.427*** (0.162)	0.433*** (0.154)		0.422*** (0.154)		
yjob		-0.004 (0.066)		0.007 (0.063)		-0.001 (0.065)	0.012 (0.063)
AUC	0.563	0.616	0.615	0.486	0.616	0.583	0.522
Observations	1971	1971	1971	1971	1971	1971	1971
Log Likelihood	-731.828	-717.417	-717.426	-734.985	-718.676	-722.585	-737.098
Akaike Inf. Crit.	1471.656	1450.834	1446.852	1477.970	1447.352	1455.169	1482.195
Bayesian Inf. Crit.	1494.001	1495.525	1480.370	1500.315	1475.284	1483.101	1504.541

Note:

*p<0.1; **p<0.05; ***p<0.01

Таблица В.15: Результаты оценивания регрессий

	Зависимая переменная			
	approve			
	<i>OLS</i>		<i>logistic</i>	<i>probit</i>
	OLS	HC		
(1)	(2)	(3)	(4)	
Constant	0.848*** (0.019)	0.848*** (0.021)	1.719*** (0.170)	1.021*** (0.092)
appinc	-0.0001 (0.0001)	-0.0001 (0.0001)	-0.001 (0.001)	-0.001 (0.0004)
mortno	0.075*** (0.016)	0.075*** (0.015)	0.783*** (0.172)	0.400*** (0.086)
dep	-0.018** (0.007)	-0.018** (0.008)	-0.166*** (0.064)	-0.085** (0.035)
male	-0.003 (0.020)	-0.003 (0.021)	-0.027 (0.184)	-0.010 (0.099)
married	0.048*** (0.018)	0.048** (0.019)	0.434*** (0.162)	0.228*** (0.087)
yjob	-0.0002 (0.007)	-0.0002 (0.006)	-0.002 (0.067)	0.003 (0.035)
AUC	0.609	0.609	0.611	0.611
Observations	1971	1971	1971	1971
R ²	0.018	0.018		
Log Likelihood			-719.660	-720.029
Akaike Inf. Crit.	1195.597	1195.597	1453.320	1454.058
Bayesian Inf. Crit.	1240.288	1240.288	1492.424	1493.162

Note:

*p<0.1; **p<0.05; ***p<0.01

Таблица В.16: Результаты оценивания logit-регрессий

	Зависимая переменная					
	default					
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	-0.334 (0.317)	-1.394*** (0.141)	-0.589* (0.318)	-0.563* (0.300)	-2.083*** (0.071)	-1.395*** (0.141)
income	-0.0003*** (0.0001)	-0.0003*** (0.0001)	-0.0003*** (0.0001)	-0.0003*** (0.00005)		-0.0003*** (0.0001)
age	-0.058*** (0.017)		-0.042** (0.017)	-0.041** (0.017)		
I(age^2)	0.001*** (0.0002)		0.0004* (0.0002)	0.0004* (0.0002)		
adepcnt	0.200*** (0.035)	0.173*** (0.034)			0.139*** (0.034)	
ownrent	-0.185* (0.097)	-0.265*** (0.093)			-0.366*** (0.092)	
spending	-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)		-0.003*** (0.0005)	-0.001 (0.001)
selfempl	-0.101 (0.209)	-0.126 (0.208)	-0.078 (0.208)		-0.225 (0.207)	-0.116 (0.207)
exp_inc	0.835 (1.177)	0.959 (1.174)	0.810 (1.184)		5.255*** (0.914)	0.954 (1.176)
AUC	0.651	0.646	0.631	0.625	0.632	0.63
Observations	6777	6777	6777	6777	6777	6777
Log Likelihood	-2068.422	-2075.803	-2083.942	-2086.815	-2093.395	-2089.149
Akaike Inf. Crit.	4154.845	4165.606	4181.883	4181.630	4198.789	4188.298
Bayesian Inf. Crit.	4216.236	4213.355	4229.632	4208.915	4239.717	4222.404

Note:

*p<0.1; **p<0.05; ***p<0.01

Таблица В.17: Результаты оценивания панельных регрессий (индивидуальные эффекты)

	Зависимая переменная		
	pool	log(violent)	
		FE	RE
	(1)	(2)	(3)
lawyes	-0.368*** (0.112)	-0.046 (0.041)	-0.070* (0.038)
prisoners	0.002*** (0.001)	-0.0001 (0.0002)	0.0002 (0.0002)
density	0.027 (0.041)	-0.172 (0.136)	0.066 (0.043)
income	0.00000 (0.00002)	-0.00001 (0.00001)	-0.00001 (0.00001)
population	0.043*** (0.012)	0.012 (0.014)	0.023** (0.011)
afam	0.081 (0.070)	0.104*** (0.032)	0.107*** (0.027)
cauc	0.031 (0.034)	0.041*** (0.013)	0.040*** (0.013)
male	0.009 (0.034)	-0.050** (0.020)	-0.038** (0.018)
Constant	2.982 (2.138)		3.525*** (0.768)
Observations	1173	1173	1173
R ²	0.564	0.218	0.225
Adjusted R ²	0.561	0.177	0.219

Note: *p<0.1; **p<0.05; ***p<0.01

Таблица В.18: Результаты оценивания панельных регрессий (двунаправленные эффекты)

	Зависимая переменная		
	pool	log(violent)	
		FE	RE
	(1)	(2)	(3)
lawyes	−0.368*** (0.112)	−0.028 (0.040)	−0.036 (0.037)
prisoners	0.002*** (0.001)	0.0001 (0.0002)	0.0003** (0.0001)
density	0.027 (0.041)	−0.092 (0.121)	0.062 (0.048)
income	0.00000 (0.00002)	0.00000 (0.00002)	0.00001 (0.00001)
population	0.043*** (0.012)	−0.005 (0.015)	0.019* (0.011)
afam	0.081 (0.070)	0.029 (0.048)	0.108*** (0.034)
cauc	0.031 (0.034)	0.009 (0.023)	0.040*** (0.015)
male	0.009 (0.034)	0.073 (0.051)	−0.009 (0.024)
Constant	2.982 (2.138)		2.794*** (0.919)
Observations	1173	1173	1173
R ²	0.564	0.056	0.116
Adjusted R ²	0.561	−0.013	0.110

Note: *p<0.1; **p<0.05; ***p<0.01

Таблица В.19: Результаты оценивания панельных регрессий (индивидуальные эффекты)

	Зависимая переменная		
	pool	sales FE	RE
	(1)	(2)	(3)
price	-1.528*** (0.473)	-0.647*** (0.239)	-0.666*** (0.228)
pop	-0.003 (0.006)	-0.006 (0.005)	-0.006 (0.006)
pop16	0.002 (0.008)	0.009 (0.008)	0.008 (0.009)
cpi	0.144 (0.247)	0.940*** (0.162)	0.911*** (0.153)
ndi	0.006*** (0.002)	-0.005* (0.003)	-0.005* (0.003)
pimin	0.595 (0.445)	0.218 (0.370)	0.223 (0.359)
Constant	138.621*** (10.115)		126.221*** (5.805)
Observations	1380	1380	1380
R ²	0.246	0.443	0.426
Adjusted R ²	0.243	0.421	0.424

Note: *p<0.1; **p<0.05; ***p<0.01

Таблица В.20: Результаты оценивания панельных регрессий (двунаправленные эффекты)

	Зависимая переменная		
		sales	
	pool	FE	RE
	(1)	(2)	(3)
price	-1.528*** (0.473)	-0.821*** (0.244)	-0.693*** (0.226)
pop	-0.003 (0.006)	-0.003 (0.005)	-0.005 (0.006)
pop16	0.002 (0.008)	0.004 (0.008)	0.006 (0.008)
cpi	0.144 (0.247)		0.988*** (0.167)
ndi	0.006*** (0.002)	-0.006** (0.003)	-0.005* (0.003)
pimin	0.595 (0.445)	-0.035 (0.404)	0.218 (0.370)
Constant	138.621*** (10.115)		124.428*** (6.653)
Observations	1380	1380	1380
R ²	0.246	0.230	0.336
Adjusted R ²	0.243	0.184	0.333

Note: *p<0.1; **p<0.05; ***p<0.01

Литература

- [1] Ayres, I., and Donohue, J.J. (2003). Shooting Down the ‘More Guns Less Crime’ Hypothesis. *Stanford Law Review*, 55, 1193–1312
- [2] Baltagi B, Levin D (1992). Cigarette taxation: Raising revenues and reducing consumption. *Structural Change and Economic Dynamics*, 3(2), 321-335
- [3] Baltagi, B.H. and Y.J. Griggin (1983) “Gasoline demand in the OECD: an application of pooling and testing procedures”, *European Economic Review*, 22.
- [4] Chu, Singfat (2001) “Pricing the C’s of Diamond Stones”, *Journal of Statistics Education*, 9(2).
- [5] J.E. Biddle and D.S. Hamermesh (1990), “Sleep and the Allocation of Time,” *Journal of Political Economy* 98, 922-943.
- [6] M. Blackburn and D. Neumark (1992), “Unobserved Ability, Efficiency Wages, and Interindustry Wage Differentials”, *Quarterly Journal of Economics* 107, 1421-1436.
- [7] Gerfin, M. (1996). Parametric and Semi-Parametric Estimation of the Binary Response Model of Labour Market Participation. *Journal of Applied Econometrics*, 11, 321–339.
- [8] Hildreth, C. and J. Lu (1960) Demand relations with autocorrelated disturbances, Technical Bulletin No 2765, Michigan State University.
- [9] W.C. Hunter and M.B. Walker (1996), “The Cultural Affinity Hypothesis and Mortgage Lending Decisions”, *Journal of Real Estate Finance and Economics* 13, 57-70

- [10] Ziliak, Jim (1997) "Efficient Estimation With Panel Data when Instruments are Predetermined: An Empirical Comparison of Moment-Condition Estimators", *Journal of Business and Economic Statistics*, 419-431