Ю. В. Зелепухин

ЭКОНОМЕТРИКА

Учебно-методическое пособие

DirectMEDIA

Ю. В. Зелепухин

ЭКОНОМЕТРИКА

Учебно-методическое пособие

DirectMEDIA

Москва
Берлин
2020

УДК 330.43(075)
ББК 65в631я7
З-48

Рецензенты:
Мазная Е. А. — кандидат экономических наук, доцент (СГСПУ);
Щербаков И. В. — кандидат экономических наук, доцент (СамГУПС)

Зелепухин, Ю. В.
З-48 Эконометрика : учебно-методическое пособие / Ю. В. Зелепухин. — Москва ; Берлин : Директ-Медиа, 2020. — 122 с.

ISBN 978-5-4499-0573-4

Учебно-методическое пособие предназначено для бакалавров и студентов ссузов экономических и других направлений (специальностей) очной и заочной форм обучения.

В учебно-методическом пособии представлен краткий курс лекций по основным темам учебной дисциплины «Эконометрика». Студентам предлагаются также контрольные вопросы, задачи и тесты которые позволяют им самостоятельно выяснить, как они усвоили основные задачи, принципы и методы эконометрики, особенности и границы их применения.

Текст приводится в авторской редакции.

УДК 330.43(075)
ББК 65в631я7

ISBN 978-5-4499-0573-4

ВВЕДЕНИЕ

Учебно-методическое пособие составлено в соответствии с требованиями федеральных государственных образовательных стандартов (ФГОС 3++).

Оно содержит обязательный минимум знаний по эконометрическим исследованиям, который должен усвоить будущий специалист.

СОДЕРЖАНИЕ РАЗДЕЛОВ ДИСЦИПЛИНЫ «ЭКОНОМЕТРИКА»

Тема 1. Основные аспекты эконометрического моделирования

Введение в эконометрическое моделирование. Основные этапы и проблемы эконометрического моделирования. Задачи эконометрики в области социально-экономических исследований. Информационные технологии на базе ПЭВМ в эконометрических исследованиях. Классификация переменных в эконометрических моделях.

Тема 2. Элементы теории вероятностей и математической статистики

Случайные величины и их числовые характеристики. Функции распределения случайной величины. Многомерные случайные величины. Закон больших чисел. Точечные и интервальные оценки параметров. Проверка статистических гипотез.

Тема 3. Парный регрессионный анализ. Показатели качества регрессии

Функциональная, статистическая и корреляционная зависимости. Линейная парная регрессия. Коэффициент корреляции. Основные положения регрессионного анализа. Интервальная оценка функции регрессии и ее параметров. Оценка значимости уравнения регрессии.

Тема 4. Линейная модель множественной регрессии

Классическая нормальная линейная модель множественной регрессии. Матричная форма модели множественной регрессии.

Предпосылки для множественного регрессионного анализа. Оценка значимости множественной регрессии.

Тема 5. Метод наименьших квадратов (МНК)

Метод наименьших квадратов. Допущения классической линейной модели регрессии. Теорема Гаусса — Маркова.

Тема 6. Свойства оценок МНК

Свойства оценок: состоятельность, несмещенность и эффективность.

Тема 7. Линейные регрессионные модели с гетероскедастичными и автокорреляционными остатками

Последствия нарушения допущений классической модели линейной регрессии. Гомоскедастичность и гетероскедастичность. Тесты на гетероскедастичность: Голфельда — Квандта, Уайта, Глейзера. Устранение гетероскедастичности.

Автокорреляция регрессионных остатков. Проверка уравнения регрессии на автокорреляцию: тесты Дарбина — Уотсона, Бреуша — Годфри, Льюинга — Бокса. Устранение автокорреляции.

Тема 8. Обобщенный метод наименьших квадратов (ОМНК)

Обобщенная линейная модель множественной регрессии (ОЛММР) и обобщенный метод наименьших квадратов (ОМНК). Теорема Айткена. ОЛММР с гетероскедастичными остатками. Сравнение ОМНК и МНК: оценки в моделях регрессии с гетероскедастичными остатками. ОЛММР с автокоррелированными остатками. Искажения характеристик точности МНК: оценки, обусловленные автокоррелированностью остатков.

Тема 9. Вопросы практического использования регрессионных моделей. Регрессионные модели с переменной структурой. Фиктивные переменные

Мультиколлинеарность: полная и частичная. Методы устранения мультиколлинеарности. Использование регрессионных моделей с переменной структурой. Фиктивные переменные, их влияние на оценку регрессионной модели. Критерий Г. Чоу. Частная корреляция.

Тема 10. Нелинейные модели регрессии и их линеаризация

Некоторые виды нелинейных зависимостей, поддающиеся непосредственной линеаризацией: экспоненциальные, логарифмические, гиперболические, степенные.

Тема 11. Характеристики временных рядов

Временной ряд и этапы его анализа. Составляющие временного ряда: тренд, сезонная, циклическая, случайная компоненты. Аналитическое выравнивание временного ряда. Прогнозирование на основе моделей временных рядов.

Тема12. Модели стационарных и нестационарных временных рядов, их классификация

Модели стационарных временных рядов и их идентификация: модели авторегрессии порядка «p» ($AR(p)$), скользящего среднего порядка «q» ($MA(q)$) и авторегрессионные модели скользящими средними в остатках ($ARMA$ (p, q, k)).

Тема 13. Система линейных одновременных уравнений. Косвенный, двухшаговый и трехшаговый метод наименьших квадратов

Система одновременных уравнений (СОУ). Косвенный метод наименьших квадратов.

Структурная и приведенная формы модели систем одновременных уравнений. Рекурсивные системы одновременных уравнений. Модель спроса — предложения как пример системы одновременных уравнений. Основные структурные характеристики моделей. Условия идентифицируемости уравнений системы. Идентификация рекурсивных систем.

Статистическое оценивание неизвестных значений параметров. Двухшаговый метод наименьших квадратов (2 МНК) оценивания структурных параметров отдельного уравнения системы.

Трехшаговый метод наименьших квадратов (3 МНК) одновременного оценивания всех параметров системы уравнений.

ТЕМА № 1
ОСНОВНЫЕ АСПЕКТЫ ЭКОНОМЕТРИЧЕСКОГО МОДЕЛИРОВАНИЯ

План

1. Введение в эконометрическое моделирование.
2. Основные этапы и проблемы эконометрического моделирования.

1.1. Введение в эконометрическое моделирование

Последние десятилетия эконометрика как научная дисциплина стремительно развивается. Свидетельством признания эконометрики является присуждение за наиболее выдающиеся разработки в этой области Нобелевских премий по экономике Р. Фришу и Я. Тинбергену, Л. Клейну, Дж. Тобину, Р. Лукасу, Дж. Хекману и Д. Мак-Фаддену, Р. Энглу и К. Грейнджеру.

«Современное экономическое образование, — утверждает директор ЦЭМИ РАН академик В. Л. Макаров, — *держится на трех китах: макроэкономике, микроэкономике и эконометрике»*.

Дадим определение эконометрики:

> Эконометрика — это самостоятельная научная дисциплина, объединяющая совокупность теоретических результатов, приемов, методов и моделей, предназначенных для того, чтобы на базе экономической теории, экономической статистики, экономических измерений и математико-статистического инструментария придавать конкретное количественное выражение общим (качественным) закономерностям, обусловленным экономической теорией.

Таким образом, эконометрика занимается разработкой и применением статистических методов для измерений взаимосвязей между экономическими переменными.

В любой эконометрической модели зависимая переменная разбивается на две части: объясненную и случайную. Задача эконометрического моделирования может быть сформулирована следующим образом: *на основании экспериментальных данных* *определить объясненную часть и, рассматривая случайную составляющую как случайную величину, получить оценки параметров ее распределения*.

Эконометрическая модель имеет следующий вид:

$$Y = f(X) + e, \tag{1.1}$$

где: Y — наблюдаемое значение зависимой переменной;

$f(X)$ — объясненная часть, зависящая от значений объясняющих переменных;

e — случайная составляющая.

1.2. Основные этапы и проблемы эконометрического моделирования

Выделяют шесть основных этапов эконометрического моделирования:

1. *Постановочный*.

На этом этапе формируется *цель* исследования и *набор* участвующих в модели экономических переменных.

В качестве практических целей эконометрической модели (1.1) можно выделить:

* анализ исследуемого экономического объекта (процесса);

* прогноз экономических показателей, характеризующих состояние и развитие анализируемой системы;

* имитация возможных сценариев социально-экономического развития анализируемой системы, когда статистически выявленные взаимосвязи между ее различными характеристиками используются для прослеживания того, как возможные изменения тех или иных параметров повлияют на значения интересующих нас характеристик;

* выработка управленческих решений.

При работе с экономическими показателями необходимо учитывать то, что многие из них неотрицательны (поэтому их надо описывать неотрицательными случайными величинами), неопределенны (поэтому их надо описывать в терминах теории нечетких множеств, математики и статистики интервальных данных), имеют нечисловую природу (поэтому к ним больше применимы методы статистики объектов нечисловой природы).

При выборе экономических переменных необходимо теоретическое обоснование каждой переменной, они не должны быть связаны функциональной или корреляционной зависимостью.

2. Априорный.

На данном этапе проводится *анализ сущности изучаемого объекта*, формирование и формализация априорной (известной до начала моделирования) информации.

3. Параметризация.

На этом этапе осуществляется *моделирование*, то есть выбор общего вида функции f(X) в эконометрической модели (1.1), выявление входящих в нее связей.

4. Информационный.

На этом этапе осуществляется *сбор* необходимой статистической информации — наблюдаемых значений экономических переменных:

$$\left(x_{i_1}, x_{i_2} \ldots x_{i_p}\ \ y_{i_1}, y_{i_2}, \ldots . y_{i_q}\right), \text{где } i = 1, 2 \ldots n .$$

Статистическая информация может быть получена в условиях активного или пассивного эксперимента.

5. Идентификация модели.

На этом этапе осуществляется *статистический анализ модели и оценка ее параметров*.

6. Верификация модели.

На данном этапе проводится *проверка адекватности модели*. Выясняется, насколько удачно решены проблемы идентификации модели, какова точность расчетов по данной модели, насколько построенная модель (1.1) соответствует моделируемому реальному экономическому объекту (процессу).

Контрольные вопросы к теме № 1

1. Что такое эконометрика.
2. Каков математический инструментарий эконометрики.
3. Каковы специфические особенности экономических данных.
4. Назовите прикладные цели эконометрического исследования.
5. Какие ученые внесли наибольший вклад в эконометрику.

6. Сформулируйте задачу эконометрического моделирования.

7. Назовите основные этапы эконометрического моделирования и дайте их характеристику.

Задача № 1.1. Торговое предприятие имеет сеть, состоящую из 12 магазинов, информация о деятельности которых представлена в таблице № 1.

Таблица № 1

№ магазина	Годовой товарооборот, млн руб.	Торговая площадь, тыс. $м^2$	Среднее число посетителей в день, тыс. чел.
1	19,76	0,24	8,25
2	38,09	0,31	10,24
3	40,95	0,55	9,31
4	41,08	0,48	11,01
5	56,29	0,78	8,54
6	68,51	0,98	7,51
7	75,01	0,94	12,36
8	89,05	1,21	10,81
9	91,13	1,29	9,89
10	91,26	1,12	13,72
11	99,84	1,29	12,27
12	108,55	1,49	13,92

Нужно:

1. Построить диаграммы рассеяния годового товарооборота (Y) в зависимости от торговой площади (X_1) и среднего числа посетителей в день (X_2).

2. Определить форму связи между результирующим показателем (Y) и каждым из факторов (X_1 и X_2).

Задача № 1.2. На основании информации, приведенной в таблице № 1, построено двухфакторное уравнение годового товарооборота (Y) в зависимости от торговой площади (X_1) и среднего числа посетителей в день (X_2), которое имеет вид:

$$Y = -10,8153 + 61,6583 X_1 + 2,2748 X_2 .$$

Нужно:

Дать экономическую интерпретацию коэффициентов уравнения.

ТЕМА № 2
ЭЛЕМЕНТЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

План

1. Случайные величины и их числовые характеристики.
2. Функция распределения случайной величины.
3. Многомерные случайные величины.
4. Закон больших чисел.
5. Точечные и интервальные оценки параметров.
6. Проверка статистических гипотез.

2.1. Случайные величины и их числовые характеристики

Приведем основные понятия теории вероятностей и математической статистики, которые используются в эконометрике.

Случайный эксперимент — процесс регистрации наблюдения на единице обследуемой совокупности.

Каждый из возможных исходов случайного эксперимента называется *элементарным событием*, а совокупность таких исходов — *пространством элементарных событий*:

$$\Omega = \left(\omega_1, \omega_2 \ldots \varpi_n\right) \tag{2.1}$$

Случайным событием «A» называют любое подмножество ($\omega_1, \omega_2, ..., \omega_k$) пространства элементарных событий, то есть

$$A = \left(\omega_1, \omega_2, \ldots, \omega_k\right) \tag{2.2}$$

Это означает, что осуществление любого из элементарных событий, входящих в «*A*», влечет за собой осуществление тия «*A*».

Каждому элементу ω_i пространства элементарных событий соответствует некоторая неотрицательная числовая характеристика p_i шансов его появления, называемая вероятностью события ω_i, причем:

$$p_1 + p_2 + ... + p_n = \sum p = 1 \qquad (2.3)$$

Вероятностью P(A) события «A» называется численная мера степени объективной возможности появления этого события. Вероятность события *A* равна:

$$P\left(A\right) = m/n\,, \qquad (2.4)$$

где m — число случаев, благоприятствующих событию «*A*»;
n — общее число случаев.

При определенных условиях в качестве оценки вероятности события $P(A)$ используется статистическая вероятность $P^*(A)$, то есть частность $W(A)$ появления события «*A*» в «*n*» произведенных испытаниях.

Случайной величиной называется переменная, которая в результате испытания в зависимости от случая принимает одно из возможного множества своих значений. Можно сказать, что случайная величина (X) это функция, определенная на множестве элементарных исходов, ее возможные значения и их общее число определяются структурой соответствующего пространства Ω элементарных событий, то есть

$$X = f\left(\omega\right) \qquad (2.5)$$

где ω — элементарное событие, принадлежащее пространству Ω, то есть $\omega \subset \Omega$.

Случайные величины бывают дискретные и непрерывные. Для *дискретной* случайной величины множество возможных значений конечно или счетно, для *непрерывной* — бесконечно и несчетно.

Случайные величины описываются законами распределения. *Законом распределения* случайной величины называется всякое соотношение, устанавливающее связь между возможными значениями случайной величины и соответствующими им вероятностями. Например,

X_1	X_2	………………	X_i	…………….	X_n
P_1	P_2	………………	P_i	………………	P_n

Такая таблица называется *рядом распределения* дискретной случайной величины.

Закон распределения дискретной случайной величины позволяет вычислить вероятности любых событий, связанных со случайной величиной, однако он не всегда удобен для анализа.

Поэтому для описания случайных величин часто используют их числовые характеристики. *Числовые характеристики* — числа, в сжатой форме выражающие наиболее существенные черты распределения случайной величины.

А) *Числовые характеристики центра группирования значений случайной величины.*

∗ *Математическое ожидание* (*среднее значение*) дискретной случайной величины (X) представляет собой сумму произведений ее значений (x_i) на соответствующие им вероятности (p_i):

$$M(x) = \sum x_i \times p_i$$

Для непрерывного случая:

$$M(x) = \int xf\left(x\right)dx \qquad (2.7)$$

Свойства математического ожидания:

1) $M\left(c\right) = c$, где c — постоянная величина.

2) $M\left(cx\right) = cM\left(x\right)$.

3) $M\left(x_1 + x_2 + ... + x_n\right) = M\left(x_1\right) + M\left(x_2\right) + ... + M\left(x_n\right)$.

4) $M\left(xy\right) = M\left(x\right)M\left(y\right)$, где $\left(x; y\right)$ — независимые случайные величины.

5) $M\left(x + c\right) = M\left(x\right) + c$.

• *Медиана* ($X_{мед}$) случайной величины (X). Представляет значение случайной величины, обладающее свойством: вероятность того, что случайная величина окажется больше ($X_{мед}$) равна вероятности того, что она окажется меньше ($X_{мед}$):

$$P(X \geq X_{мед}) = P(X \leq X_{мед}) = 0{,}5 \qquad (2.8)$$

• *Мода* ($X_{мод}$) случайной величины (X) — наиболее часто наблюдаемое ее значение.

В) *Числовые характеристики степени рассеяния значений случайной величины.*

Они дают представление о том, как сильно могут отклоняться от своего центра группирования значения случайной величины.

• *Дисперсия* $D(X)$ *случайной величины* (X) есть математическое ожидание квадрата ее отклонения от математического ожидания:

$$D(x) = M\left(X - M(X)\right)^2;$$

$$D(x) = \sum\left(x_i - M(x)\right)^2 \cdot p_i \text{ (для дискретного случая);} \quad (2.9)$$

$$D(x) = \int\left(x - M(x)\right)^2 f(x) d(x) \text{ (для непрерывного случая)} \quad (2.10)$$

Свойства дисперсии:

1) $D(c) = 0$, где c — постоянная величина.

2) $D(cx) = c^2 D(x)$.

3) $D(x + y) = D(x) + D(y)$ где $(x; y)$ — независимые случайные величины.

4 $D(x) = M(x^2) - M(x)^2$.

5) $D(c + bx) = b^2 D(x)$.

* *Среднее квадратическое отклонение* δ_x случайной величины (X) представляет собой арифметическое значение корня квадратного из ее дисперсии:

$$\delta_x = \sqrt{D(x)}. \quad (2.11)$$

* *Коэффициент вариации* V_x:

$$V_x = \sqrt{\frac{D(x)}{M(x)}} \cdot 100\%. \quad (2.12)$$

С) Показатели связи между двумя переменными.

Наиболее распространенными показателями линейной связи между двумя переменными являются:

* *Ковариация* cov(*x, y*) случайных величин (*x*) и (*y*). Представляет собой математическое ожидание произведения отклонений этих величин от своих математических ожиданий:

$$\operatorname{cov}(x,y)=M\left(\left[x-M(x)\right]\left[y-M(y)\right]\right). \tag{2.13}$$

Отметим основные свойства ковариации двух случайных величин:

1) $\operatorname{cov}(x,y)=0$, если (*x*), (*y*) независимы.

2) $\operatorname{cov}(a+вx,\ c+дy)=вд\operatorname{cov}(x,y)$, где *a*, *в*, *c*, *д* — константы.

• *Коэффициент корреляции* P_{xy} двух случайных величин (*x*) и (*y*). Представляет собой отношение их ковариации к произведению средних квадратических отклонений этих величин:

$$P_{xy}=\frac{\operatorname{cov}(x,y)}{\delta_x\delta_y}. \tag{2.14}$$

Отметим свойства коэффициента корреляции:

1) $-1\le P_{xy}\le 1$.

2) $P_{xy}=0$, если случайные величины (*x*) и (*y*) независимы.

3) Если $P_{xy}=1$, то между случайными величинами (*x*) и (*y*) существует линейная функциональная зависимость.

2.2. Функция распределения случайной величины

Формами закона распределения случайной величины (*x*) являются *функция и плотность распре деления.* Плотность распределения $\varphi(x)$ существует только для непрерывных случайных величин.

Функция распределения случайной величины (*x*) представляет собой функцию F(*x*), выражающую для каждого значения (*x*) вероятность того, что случайная величина (*x*) примет значение, меньше (*x*):

$$F(x)=P(X\le x). \tag{2.15}$$

Назовем основные свойства функции распределения случайной величины:

1) Функция распределения случайной величины есть неубывающая функция на всей числовой оси, то есть при $x_2 \geq x_1$

$$F(x_2) \geq F(x_1).$$

2) Функция распределения случайной величины есть неотрицательная функция, заключенная между нулем и единицей:

$$0 \leq F(x) \leq 1.$$

3) Вероятность попадания случайной величины (X) в интервал (x_1, x_2) равна приращению ее функции распределения на этом интервале, то есть:

$$P(x_1 \leq X \leq x_2) = F(x_2) - F(x_1).$$

Случайная величина (X) называется непрерывной, если ее функция распределения непрерывна в любой точке и дифференцируема всюду, кроме, быть может, отдельных точек.

Плотностью вероятности $\varphi(x)$ непрерывной случайной величины (X) называется производная ее функции распределения:

$$\varphi(x) = F'(x). \qquad (2.16)$$

2.3. Многомерные случайные величины

Упорядоченный набор $x = (x_1, x_2, \ldots x_n)$ случайных величин называется *многомерной случайной величиной*. Пример, показатель «качество жизни россиян» в определенный период времени может быть охарактеризован многомерной случайной величиной $x = (x_1, x_2, \ldots, x_n)$ где x_1 — продолжительность жизни, x_2 — уровень потребления социально значимых благ и услуг, x_3 — уровень развития науки и т. д.

Функцией распределения многомерной случайной величины $(X_1, \ldots, X_n)$ называется функция $F(x_1, x_2, \ldots x_n)$, выражающая вероятность совместного выполнения (n) неравенств $X_1 \le x_1,\ X_2 \le x_2, \ldots, X_n \le x_n$, то есть

$$F\left(x_1,\, x_2, \ldots x_n\right) = P(X_1 \le x_1,\, X_2 \le x_2, \ldots,\, X_n \le x_n)\,. \tag{2.17}$$

Плотностью распределения непрерывной двумерной случайной величины (X, Y) называется вторая смешанная частная производная ее функции распределения, то есть

$$\varphi\left(x, y\right) = F''_{x,y}\left(x, y\right). \tag{2.18}$$

Условным законом распределения одной из одномерных составляющих двумерной случайной величины (X, Y) называется ее закон распределения, вычисленный при условии, что другая составляющая приняла определенное значение. Условные плотности распределения двумерной случайной величины (X, Y) определяются по формулам:

$$\varphi_y\left(x\right) = \frac{\varphi\left(x, y\right)}{\int\limits_{-\infty}^{+\infty} \varphi\left(x, y\right) d\left(x\right)};$$
$$\varphi_x\left(y\right) = \frac{\varphi\left(x, y\right)}{\int\limits_{-\infty}^{\infty} \varphi\left(x, y\right) d\left(y\right)}. \tag{2.19}$$

Условное математическое ожидание $M_x(Y)$ случайной величины (Y) при $X = x$ есть функция от (x), называемая *функцией регрессии* (Y) по (X).

2.4. Закон больших чисел

Рассмотрим некоторые результаты теории вероятностей, используемые в эконометрике.

4.1. Закон больших чисел

Под *законом больших чисел* понимается принцип, согласно которому совокупное действие большого числа случайных факторов приводит (при общих условиях) к результату, почти не зависящему от случая.

4.2. Теорема Бернулли

Частость события в (n) повторных независимых испытаниях, в каждом из которых оно может произойти с одной и той же вероятностью (p), при неограниченном увеличении числа (n) сходится по вероятности к вероятности (p) этого события в отдельном испытании. То есть

$$\lim P = \left((\frac{m}{n} - p) \leq \varepsilon \right) = 1\,, \qquad (2.20)$$

где ε — сколь угодно малая положительная величина.

4.3. Теорема Ляпунова

Если независимые случайные величины $x_1, x_2, .., x_n$ имеют конечные математические ожидания и дисперсии, по своему значению ни одна из этих случайных величин резко не выделяется среди остальных, то при $n \to \infty$ закон распределения их суммы $\sum_{i=1}^{n} X_i$ неограниченно приближается к нормальному.

В частности, если $x_1, x_2, .., x_n$ одинаково распределены, то закон распределения их суммы $\sum_{i=1}^{n} x_i$ при $n \to \infty$ неограниченно приближается к нормальному.

2.5. Точечные и интервальные оценки параметров

Теперь рассмотрим точечные и интервальные оценки случайных величин.

Оценкой ($\tilde{Q}_n$) *параметра* (Q) называют всякую функцию результатов наблюдений над случайной величиной (X), с помощью которой судят о значениях параметра (Q). Оценка ($\tilde{Q}_n$) есть величина случайная, и она должна обладать наименьшим разбросом относительно оцениваемого параметра (Q).

Оценка ($\tilde{Q}_n$) параметра (Q) называется *несмещенной*, если ее математическое ожидание равно оцениваемому параметру, то есть М($\tilde{Q}_n$) = (Q). В противном случае оценка называется *смещенной.*

Оценка ($\tilde{Q}_n$) параметра (Q) называется *состоятельной*, если она удовлетворяет закону больших чисел, то есть сходится по вероятности к оцениваемому параметру:

$$\lim P\left(\left|\tilde{Q}_n - Q\right| \le \varepsilon\right) = 1 . \tag{2.21}$$

Несмещенная оценка ($\tilde{Q}_n$) параметра (Q) называется *эффективной*, если она имеет наименьшую дисперсию среди всех возможных несмещенных оценок параметра (Q), вычисленных по выборкам одного и того же объема (*n*).

Полный набор данных, относящихся к исследуемому явлению, называется *генеральной совокупностью*. В реальности редко возможны ситуации, когда исследователю доступны все данные для исследования. Обычно приходится иметь дело с ограниченным их количеством — *выборкой*. Статистические показатели, полученные на основе выборок, называются *выборочными показателями*. Для решения задач статистического оценивания рассчитывают показатели на основе некоторой совокупности выборок, в результате получается *выборочное распределение выборочного показателя*.

Для нахождения оценок параметров генеральной совокупности по данным выборки используется *метод максимального правдоподобия*. Основу этого метода составляет *функция правдоподобия*, которая выражает плотность вероятности совместного появления результатов выборки $\left(x_1, x_2 \dots x_n\right)$:

$$L\left(x_1, x_2 \dots x_n, Q\right) = \prod_{i=1}^{n} \varphi\left(x_i, Q\right) . \tag{2.22}$$

Согласно этого метода в качестве оценки неизвестного параметра (Q) принимается такое значение ($\tilde{Q}_n$), которое максимизирует функцию (*L*).

Достоинством метода максимального правдоподобия является то, что получаемые им оценки состоятельны, эффективны и имеют нормальное распределение. Несмещенной, состоятельной и эффективной оценкой математического ожидания (α) является *выборочная средняя:*

$$\overline{x} = \frac{\sum x_i n_i}{n}, \qquad (2.23)$$

где x_i, n_i — частоты значений x_i.

Несмещенной, состоятельной оценкой дисперсии (δ^2) является *выборочная дисперсия*:

$$s^2 = \frac{\sum_{i=1}^{n} \left(x_i - \overline{x}\right)^2 n_i}{n}. \qquad (2.24)$$

Наряду с точечными оценками рассматривают интервальные оценки параметров.

Интервальная оценка параметра (Q) представляет собой числовой интервал (Q_1, Q_2), который с заданной вероятностью (γ) накрывает неизвестное значение параметра (Q). Такой интервал называется *доверительным*, а вероятность (γ) — *надежностью* оценки.

Величина доверительного интервала зависит от объема выборки (n) и от значения надежности.

Доверительный интервал для *генеральной* средней (Хо) (математическое ожидание средней всех выборочных средних) на уровне значимости (α) определяется по формуле:

$$\left(\overline{x} - t_{1-\alpha,\, n-1} \frac{s}{\sqrt{n-1}}, \overline{x} + t_{1-\alpha,\, n-1} \frac{s}{\sqrt{n-1}}\right). \qquad (2.25)$$

Доверительный интервал для генеральной дисперсии на уровне значимости (α) определяется по формуле:

$$\left(\frac{ns^2}{\chi_{\alpha/2,\, n-1}{}^2}, \frac{ns^2}{\chi_{1-\alpha/2, n-1}{}^2}\right). \qquad (2.26)$$

2.6. Проверка статистических гипотез

Исследователь часто располагает априорными догадками относительно величины параметров генеральной совокупности.

Введем следующие понятия:

• *Статистическая гипотеза* — это предположение о величине параметра распределения генеральной совокупности.

Процесс проверки гипотезы базируется на формировании двух типов гипотез: нулевой и альтернативной.

∗ *Нулевая гипотеза* (H_0) — допущение, которое считается верным до тех пор, пока не будет доказано обратное, исходя из результатов статистической проверки.

∗ *Альтернативная гипотеза* (H_1) — гипотеза, которая принимается, если в результате проверки отвергается нулевая гипотеза.

Статистическая проверка гипотезы состоит в использовании *стандартизированного статистического критерия, вычисляемого по данным выборки.*

В процессе проверки гипотезы существует вероятность того, что нулевая гипотеза (H_0) будет отвергнута, когда в действительности она должна быть принята. Это называется *ошибкой 1-го рода*. *Ошибка 2-го рода* имеет место при принятии нулевой гипотезы (H_0) в то время, когда она должна быть отвергнута.

Следует помнить, что принятие гипотезы (H_0) расценивается не как абсолютно верный содержащийся в ней факт, а лишь как достаточное правдоподобное, не противоречащее опыту утверждение.

Контрольные вопросы к теме № 2

1. Дайте определение случайной величины.
2. Назовите числовые характеристики случайной величины.
3. В чем состоит смысл закон больших чисел и предельных теорем Бернулли и Ляпунова.
4. В чем суть проверки статистической гипотезы.

Задача 2.1. Дан ряд распределения случайной величины (*X*):

X	0	1	2	3
P	0,05	0,30	0,45	0,22

Необходимо: а) найти математическое ожидание M(*X*), дисперсию D(*X*), среднее квадратическое отклонение (δ) случайной величины (*X*).

Задача 2.2. Случайная величина (*X*) сосредоточена на интервале (–1,3), задана функцией распределения $F(x)=\frac{1}{4}x+\frac{1}{4}$. Найти вероятность попадания случайной величины (*X*) в интервал (0, 2).

Задача 2.3. Случайная величина (*X*) задана функцией распределения:

$$F(x)=\begin{matrix} 0 & \text{при } x\le 0 \\ x^2 & \text{при } 0\le x\le 1 \\ 1 & \text{при } x\ge 1 \end{matrix}$$

Найти плотность распределения $\varphi(x)$.

Задача 2.4. Дан ряд распределения случайной величины:

X	1	4	5
P	0,4	0,1	0,5

Найти и изобразить графически функцию ее распределения.

ТЕМА № 3 ПАРНЫЙ РЕГРЕССИОННЫЙ АНАЛИЗ. ПОКАЗАТЕЛИ КАЧЕСТВА РЕГРЕССИИ

План

1. Функциональная, статистическая и корреляционная зависимости.
2. Линейная парная регрессия.
3. Коэффициент корреляции.
4. Основные положения регрессионного анализа.
5. Интервальная оценка функции регрессии и ее параметров.
6. Оценка значимости уравнения регрессии.

3.1. Функциональная, статистическая и корреляционная зависимости

Регрессионный анализ занимает центральное место в математическом аппарате эконометрики. *Задачами регрессионного анализа* являются установление формы зависимости между переменными, оценка функции регрессии, прогноз значений зависимой переменной. Различают функциональную, статистическую и корреляционную зависимости. При *функциональной* зависимости каждому значению одной переменной соответствует вполне определенное значение другой. При *статистической* зависимости каждому значению одной переменной соответствует определенное распределение другой переменной (например, зависимость производительности труда на заводе от его энерговооруженности). При *корреляционной* зависимости каждому значению одной переменной соответствует определенное математическое ожидание другой.

Регрессионный анализ применим тогда, когда изучаемые зависимости:

1) имеют *стохастическую* (*вероятностную*) природу;

2) выявляются на основании *статистического* наблюдения за анализируемыми событиями.

Корреляционная зависимость представляется в виде

$$M_x(Y) = \varphi(x), \tag{3.1}$$

где (Y) — эндогенная переменная (результирующая, объясняемая):

X — экзогенная переменная (объясняющая, регрессор).

Эндогенная переменная характеризует некий результат функционирования анализируемой экономической системы. В регрессионном анализе эндогенная переменная выступает в роли *функции*, значения которой всегда стохастичны по своей природе. Экзогенная переменная описывает функционирование изучаемой экономической системы и задается как бы «извне». В регрессионном анализе она играет роль *аргумента* той функции, в качестве которой рассматривается эндогенная переменная. По своей природе она может быть как случайной, так и неслучайной.

Уравнение (3.1) называют *уравнением регрессии*, а $\varphi(x)$ — *функцией регрессии*.

С помощью уравнения регрессии (3.1) можно решить следующие практические задачи:

1) установить факт наличия (отсутствия) связи между переменными, проверить статистическую значимость этой связи;

2) осуществить прогноз неизвестных значений эндогенной переменной по заданным значениям экзогенных переменных;

3)выявить причинно-следственные связи между эндогенными и экзогенными переменными;

4) осуществить регулирование значений экзогенных показателей с целью управления эндогенными показателями.

Выбор вида функции регрессии — наиболее важная и наименее теоретически обоснованная часть регрессионного анализа. В практике регрессионного анализа сложились определенные традиции, которых обычно придерживаются при выборе вида искомой функции:

1) прежде всего, необходимо опираться на имеющуюся априорную информацию об изучаемом явлении;

2) исследователь проводит визуальный анализ данных, использует разные приемы математической статистики при обработке полученных данных;

3) главное правило при построении регрессионной модели — это движение от простого к сложному. Поэтому анализ принято начинать с простейшего случая — линейной модели.

Рассмотрим ситуацию, когда в регрессионной модели только одна экзогенная переменная. Такая зависимость называется *парной* или *простой* регрессией. Она имеет вид:

$$y_i = b_0 + b_1 x_i + u_i, \qquad (3.2)$$

где $i = 1, 2.., n$ — номер наблюдения;
y — эндогенная переменная;
x — экзогенная переменная;
u — регрессионный остаток;
b_0, b_1 — коэффициенты регрессии.

Дело в том, что на (y), помимо (x), влияет множество других факторов. Все их учесть невозможно, поэтому считают, что член (u) отражает влияние всех прочих факторов, влияющих на показатель (y).

Оценкой модели парной регрессии по выборке является выборочное уравнение регрессии:

$$\hat{y}_i = \tilde{b}_0 + \tilde{b}_1 x_i,$$

где $\tilde{b}_0, \tilde{b}_1$ — оценки коэффициентов регрессии.

Алгоритм регрессионного анализа представлен в виде следующей последовательности действий:

1. По фактическим данным оценивают b_0, b_1 — коэффициенты регрессии. Обозначим $\tilde{b}_0, \tilde{b}_1$ — оценки коэффициентов регрессии.

2. Рассчитывают ожидаемое значение ($\hat{y}_i$) из выражения:

$$\hat{y}_i = \tilde{b}_0 + \tilde{b}_1 x_i. \qquad (3.3)$$

То есть для каждого значения (x_i) существует фактическое значение (y_i), но появляется также *оценочное* значение $(\hat{y}_i)$.

$$u_i = \hat{y}_i - y_i. \qquad (3.4)$$

3. Для статистической проверки взаимосвязи между зависимыми и независимыми переменными находят значения b_0, b_1 — коэффициентов регрессии, и (u_i).

Фактические и оценочные данные регрессионной модели можно представить графически (*рис. 3.1*), где точками изображены фактические данные, а прямая линия представляет собой их аппроксимацию, то есть описывается уравнением (3.3).

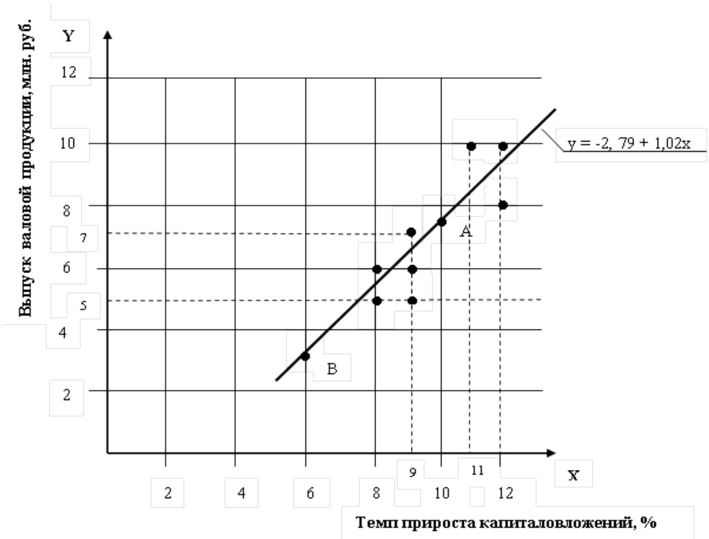


Рис. 3.1. Линейная стахостическая зависимость

Следует обратить внимание, что аппроксимирующее уравнение (3.3) является строгой функцией, поэтому в ней отсутствует регрессионный остаток (u).

Оценочные значения коэффициентов B_0 и B_1 получают с помощью метода наименьших квадратов, о котором пойдет речь в теме № 5.

3.3. Коэффициент корреляции

Для оценки тесноты корреляционной зависимости (y) и (x) служит *коэффициент корреляции* (r)*:*

$$r = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{\sqrt{n\sum_{i=1}^{n} x^2{}_i - \left(\sum_{i=1}^{n} x_i\right)^2} \times \sqrt{n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2}}. \qquad (3.5)$$

Коэффициент корреляции (r) обладает следующими свойствами:

1. Принимает значения на отрезке (–1, 1), чем ближе к единице, тем теснее связь. Если ($r > 0$), то корреляционная связь меду переменными называется *прямой*, если ($r < 0$) то *обратной*.

2. При $r = 1$ корреляционная зависимость является функциональной

3. При $r = 0$ линейная корреляционная связь отсутствует.

3.4. Основные положения регрессионного анализа

В основе линейной регрессионной модели (3.2) лежат следующие допущения:

а) истинная форма взаимосвязи между эндогенными и экзогенными переменными является линейной;

б) x — есть величина неслучайная;

в) математическое ожидание регрессионных остатков равно нулю:

$$M\left(u_i\right) = 0\,).$$

То есть математическое ожидание зависимой переменной (y_i) равно линейной функции регрессии

$$M\left(y_i\right) = b_0 + b_1 x_1\,;$$

г) дисперсия регрессионных остатков постоянна и конечна для всех значений (x):

$$D\left(u_i\right) = const\,;$$

д) регрессионные остатки статистически независимы друг от друга:

$$\operatorname{cov}\left(u_i, u_j\right) = 0\,;$$

е) регрессионные остатки и экзогенные переменные независимы друг от друга:

$$\operatorname{cov}\left(u_i, x_i\right) = 0\,.$$

При соблюдении этих допущений а) — з) модель (3.2) называется *классической нормальной линейной регрессионной моделью*.

Эти допущения необходимы для того, чтобы полученные с помощью метода наименьших квадратов оценки истинных коэффициентов (B_0) и (B_1) обладали некоторыми желательными свойствами, которые будут рассмотрены в теме № 6.

Воздействие неучтенных случайных факторов в модели (3.2) определяются с помощью *остаточной дисперсии, оценкой которой является выборочная остаточная дисперсия*:

$$S^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-2}; \qquad S^2 = \frac{\sum_{i-1}^{n} (\hat{y}_i - y_i)^2}{n-2},$$

где $\hat{y}_i$ — групповая средняя, найденная по уравнению регрессии;

$u_i = e_i = \hat{y}_i - y_i$ — регрессионный остаток (выборочная оценка возмущений).

Поскольку вероятностные модели представляют только лишь оценки коэффициентов регрессии (B_0) и (B_1), то важно проверить насколько представительными являются эти оценки относительно истинных значений коэффициентов. Для этого используют методы проверки статистической значимости коэффициентов регрессии и самой функции.

3.5. Интервальная оценка функции регрессии и ее параметров

∗ *Доверительный интервал для функции регрессии* (*условного математического ожидания*), который с заданной надежностью (доверительной вероятностью) $(1-\alpha)$ накрывает неизвестное значение $M_x(Y)$ определяется по формуле:

$$\hat{y} - t_{1-\alpha,k} \cdot s_{\hat{y}} \le M_x(Y) \le \hat{y} + t_{1-\alpha,k} \cdot s_{\hat{y}} \qquad (3.10)$$

$$s^2_{\hat{y}} = s^2 \left(\frac{1}{n} + \frac{\left(x - \overline{x}\right)^2}{\sum_{i=1}^{n} \left(x_i - \overline{x}\right)^2} \right)$$

$$s = \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n-2}.$$

Доверительная область для $M_x(Y)$ определяет местоположение линии регрессии, но не отдельных возможных значений зависимой переменной, которые отклоняются от средней. Поэтому при определении *доверительного интервала для индивидуальных значений* $У_0$ *зависимой переменной* необходимо учитывать рассеяние вокруг линии регрессии, то есть в оценку суммарной дисперсии (s_y) следует включить величину (s). Оценка дисперсии индивидуальных значений ($У_0$) при $X = X_0$ равна:

$$S_{\hat{y}}{}^2 = S^2\left(1 + \frac{1}{n} + \frac{\left(x_0 - \overline{x}\right)^2}{\sum\left(x_i - \overline{x}\right)^2}\right). \qquad (3.12)$$

* *Доверительный интервал для прогнозов индивидуальных значений* y_0^* определяется по формуле:

$$\hat{y}_0 - t_{1-\alpha,n-2}S_{\hat{y}_0} \le y_0^* \le \hat{y}_0 + t_{1-\alpha,n-2}S_{\hat{y}_0}. \qquad (3.13)$$

Доверительный интервал для коэффициента регрессии (b)

$$\hat{b} - t_{kp}\left(\alpha, k\right)\cdot s_{\hat{b}} \le b \le \hat{b} + t_{kp}\left(\alpha, k\right)\cdot s_{\hat{b}};$$

$$s_{\hat{b}} = \sqrt{\frac{s^2{}_e}{\sigma_x^2 \times n}},$$

где s_e^2 — *несмещенная оценка остаточной дисперсии* σ_e^2.

$$s_e{}^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2}$$

$$\sigma_x{}^2 = \frac{1}{n}\sum x_i{}^2 - \left(\frac{1}{n}\sum x_i\right)^2.$$

Доверительный интервал для свободного члена уравнения по формуле:

$$\hat{\alpha} - t_{kp}\left(\alpha, k\right) \cdot s_{\hat{\alpha}} \le \alpha \le \hat{\alpha} + t_{kp}\left(\alpha, k\right) \cdot s_{\hat{\alpha}}\,;$$

$$s_{\hat{\alpha}} = s_{\hat{b}} \sqrt{\frac{\sum x_i^{\,2}}{n}}\,.$$

3.6. Оценка значимости уравнения регрессии.

Регрессионная модель (3.2) показывает, что вариация (*Y*) может быть объяснена вариацией независимой переменной (*X*) и значением ошибки (*u*). Проверить значимость регрессионной модели — это значит установить, соответствует ли математическая модель, выражающая зависимость между переменными экспериментальным данным.

Мерой качества уравнения регрессии является *коэффициент детерминации*, определяемый по формуле:

$$R^2 = \frac{Q_R}{Q} = 1 - \frac{Q_e}{Q} \qquad (3.16)$$

Значения (Q, Q_R, Q_e) представлены в *табл. 3.1*.

Таблица 3.1

Компоненты дисперсии	Сумма квадратов	Число степеней свободы	Средние квадраты
Регрессия	$Q_R = \sum_{i=1}^{n}\left(\hat{y}_i - \overline{y}\right)^2$	$m-1$	$S^2{}_R = \frac{Q_R}{m-1}$
Остаточная	$Q_e = \sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2$	$n-m$	$S^2 = \frac{Q_e}{n-m}$
Общая	$Q = \sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2$	$n-1$	

Величина R^2 $\left(0 \le R^2 \le 1\right)$ показывает, какая часть вариации зависимой переменной обусловлена вариацией объясняющей (экзогенной) переменной. Чем ближе R к единице, тем лучше регрессия аппроксимирует эмпирические данные. Если $R^2 = 1$, то эмпирические точки $\left(x_i, y_i\right)$ лежат на линии регрессии и между переменными (Y) и (X) существует линейная функциональная зависимость. Если $R^2 = 0$, то вариация зависимой переменной полностью обусловлена воздействием неучтенных факторов, и линия регрессии параллельна оси абсцисс.

В случае парной линейной регрессионной модели коэффициент детерминации равен квадрату корреляции, то есть $R^2 = r^2$

Если известен коэффициент детерминации, то критерий значимости уравнения регрессии может быть записан в виде:

$$F = \frac{R^2\left(n-m\right)}{\left(1-R^2\right)\left(m-1\right)} > F_{\alpha,k,k_{21}}, \tag{3.17}$$

где F_{α,k_1k_2} — табличное значение F-критерия Фишера, определенное на уровне значимости $\left(\alpha\right)$ при $k_1 = m-1$ и $k_2 = n-m$ степенях свободы.

Несмотря на преимущества коэффициента детерминации (легко вычисляется, имеет четкую интерпретацию), при его использовании могут возникнуть следующие проблемы:

— нельзя сравнивать величины R^2 для моделей с различными зависимыми величинами;

— R^2 малопригоден для оценки качества моделей временных рядов;

— R^2 никогда не уменьшается в случае добавления экзогенных переменных в модель. Дело в том, что R^2 всегда будет увеличиваться по мере включения новых переменных в модель, что создает у исследователя стимул необоснованно включать дополнительные переменные в модель, поэтому становится проблематичным определить, улучшает ли дополнительная переменная качество модели.

Данная проблема решается путем использования *скорректированного коэффициента детерминации:*

$$R^2 = 1 - \frac{n-1}{n-k}\left(1 - R^2\right), \qquad (3.18)$$

где n — число наблюдений;
k — число независимых переменных.

Контрольные вопросы к теме № 3

1. Назовите основные задачи регрессионного анализа.
2. Что такое классическая модель линейной регрессии.
3. Какие допущения лежат в основе классической модели линейной регрессии.
4. Что такое коэффициент детерминации, какие проблемы могут возникнуть при использовании коэффициента детерминации.
5. Что такое скорректированный коэффициент детерминации.
6. Как проводится оценка качества уравнения регрессии.
7. Что такое коэффициент корреляции, какими свойствами он обладает.
8. Каким образом осуществляется проверка регрессионной модели.

Задача 3.1. Исследуется зависимость затрат на рекламу (Y) от годового оборота (X) в некоторой области. Для этого собрана информация n = 20 предприятиям о годовом обороте (X_i) и соответствующих расходах на рекламу (Y_i). Из выборки получены следующие данные: $S^2 = 0,7021$; $\overline{x} = 17,3$; $\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 = 3264,2$; $\sum\left(x^2{}_i\right) = 9250$; $\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2 = 12,6372$. Уравнение зависимости затрат на рекламу от годового оборота имеет вид:

$$Y = -1,6042 + 0,1621X.$$

Ошибка уравнения распределена нормально с нулевым математическим ожиданием и постоянной дисперсией.

Требуется:

1) Определить 95 % прогнозный интервал математического ожидания целевой переменной (Y_0) при X_0 = 30.

2) Оценить 95 % прогнозный интервал для отдельного значения целевой переменной при X_0 =30 и сравнить его с прогнозным интервалом в п. 1

Задача 3.2. По данным задачи 3.1:

1) Оценить дисперсию ошибки уравнения регрессии.

2) Оценить дисперсии оценок параметров регрессии.

Задача 3.3

Имеются следующие данные об уровне механизации работ (*X*, %) и производительности труда (*Y*, т/час.) для 14 предприятий:

Х	32	30	36	**40**	**41**	**47**	**56**	**54**	**60**	**55**	**61**	**67**	**69**	**76**
У	20	24	28	30	31	33	34	37	38	40	41	43	45	48

Необходимо оценить тесноту и направление связи между переменными с помощью коэффициента корреляции.

ТЕМА № 4 ЛИНЕЙНАЯ МОДЕЛЬ МНОЖЕСТВЕННОЙ РЕГРЕССИИ

План

1. Классическая нормальная модель множественной регрессии.
2. Матричная форма модели множественной регрессии.
3. Предпосылки для множественного регрессионного анализа.
4. Оценка значимости множественной регрессии.

4.1. Классическая нормальная модель множественной регрессии.

Экономические явления определяются большим числом одновременно и совокупно действующих факторов. Перед исследователем стоит задача установления зависимости одной переменной (Y) от нескольких экзогенных переменных. Эта задача решается с помощью *множественного регрессионного анализа*.

Множественная регрессия представляет собой уравнение вида (4.1):

$$y_i = b_0 + b_1 X_{i_1} + b_2 X_{i_2} + \ldots + b_p X_{i_p} + \varepsilon_i, \qquad (4.1)$$

где $i = 1, 2, ..., n$.

Здесь $\boldsymbol{B} = (b_0, b_1 \ldots b_p)$ — вектор размерности $(p + 1)$ неизвестных параметров b_j, $j = 0, 1, 2 ... p$, называется j-ым теоретическим коэффициентом регрессии (частичным коэффициентом регрессии). Он характеризует чувствительность величины Y к изменению X_j. Другими словами, он отражает влияние на условное математическое ожидание $M\left(Y / _{X_1 = x_1,\, X_2 = x_2\, \ldots\, X_p = x_p}\right)$ зависимой переменной Y объясняющей переменной X_j при условии, что все другие объясняющие переменные модели остаются постоянными. b_0 — свободный член, определяющий значение Y в случае, когда все объясняющие переменные X_j равны нулю.

Модель (4.1), регрессионные остатки которой удовлетворяют приведенным выше предпосылкам (Тема № 3) — (*а* — *з*), называется *классической нормальной линейной моделью множественной регрессии*.

4.2. Матричная форма модели множественной регрессии

При анализе уравнение (4.1) записывается в матричной форме:

$$Y = BX + U$$

$$Y = \begin{pmatrix} y_1 \\ y_n \end{pmatrix} \quad B = \begin{pmatrix} b_0 \\ b_k \end{pmatrix} \quad U = \begin{pmatrix} u_1 \\ u_n \end{pmatrix}, \qquad (4.2)$$

$$X = \begin{pmatrix} 1 & x_{11} & \dots & x_{k1} \\ 1 & x_{1n} & \dots & x_{kn} \end{pmatrix}$$

где Y — матрица-столбец (вектор) значений зависимой переменной размера $n \times 1$

X — матрица значений экзогенных переменных размера $n \times k$

B — матрица — столбец (вектор) параметров размера $k \times 1$

U — матрица — столбец (вектор) возмущений (ошибок, регрессионных остатков) размера $n \times 1$

Оценкой этой модели по выборке является уравнение:

$$Y = \tilde{b}X + e$$

$$e = \begin{pmatrix} e_1 & e_n \end{pmatrix}' \quad \tilde{b} = \begin{pmatrix} \tilde{b}_0 & \tilde{b}_1 & \tilde{b}_k \end{pmatrix}',$$ транспонированные матрицы.

Матричное описание регрессии облегчает как теоретические концепции анализа, так и необходимые расчетные процедуры.

4.3. Предпосылки для множественного регрессионного анализа.

Предпосылки для множественного регрессионного анализа:

1. U — случайный вектор, а X — неслучайная матрица.
2. $M(U) = 0_n$, 0_n — нулевой вектор размера (n).

3. $\sum U = \delta^2 E_n$, E_n единичная матрица (n)-го порядка.

4. U — нормально распределенный случайный вектор, то есть $U \approx N\left(0, \delta^2 E_n\right)$.

5. $r(X) = p + 1 < n$, где $r(X)$ — ранг матрицы.

Для оценки коэффициентов регрессии (B) применяется метод наименьших квадратов, о котором пойдет речь в Теме № 5. Согласно методу наименьших квадратов:

$$\tilde{b} = \left(X'X\right)^{-1}\left(X'Y\right), \qquad (4.3)$$

где $\left(X'X\right)^{-1}$ — *матрица, обратная матрице* $\left(X'X\right)$.

$$\left(X'X\right) = \begin{pmatrix} n & \sum x_{1_i} & \dots & \sum x_{k_i} \\ \sum x_{1_i} & \sum x_{1_i}{}^2 & \dots & \sum x_{1_i} x_{k_i} \\ \dots & \dots & \dots & \dots \\ \sum x_{k_i} & \sum x_{1_i} x_{k_i} & .. & \sum x_{k_i}^2 \end{pmatrix},$$

где $\left(X'Y\right)$ — *матрица столбец или вектор ее свободных членов* .

$$\left(X'Y\right) = \begin{pmatrix} \sum y_i \\ \sum y_i x_{1_i} \\ \dots \\ \sum y_i x_{k\,i} \end{pmatrix}.$$

На практике часто необходимо сравнение влияния на эндогенную переменную различных экзогенных переменных, когда они выражаются разными единицами измерения. В этом случае используют *стандартизированные коэффициенты регрессии* $\left(\tilde{b}'_j\right)$ и *коэффициенты эластичности* (Ej) (j =1, 2 … p):

$$\tilde{b}'_j = \tilde{b}_j \frac{s_{x_j}}{s_y}. \qquad (4.4)$$

$$E_j = \tilde{b}_j \frac{\overline{x}_j}{\overline{y}} \qquad (4.5)$$

Стандартизированный коэффициент регрессии $\left(\tilde{b}'_j\right)$ показывает, на сколько величин (s_y) изменится в среднем зависимая переменная (Y) при увеличении только j-й экзогенной переменной на (s_{x_j}), а коэффициент эластичности (E_j) — на сколько процентов (от средней) изменится в среднем (Y) при увеличении только (X_j) на 1 %.

4.4. Оценка значимости множественной регрессии

Вариация оценок параметров определяет точность уравнения множественной регрессии. Для их измерения в многомерном регрессионном анализе рассматривают *ковариационную матрицу вектора оценок параметров*, *являющуюся матричным аналогом дисперсии одной переменной*:

$$\sum\nolimits_{\tilde{b}} = \begin{matrix} \delta_{00} & \delta_{01} & \dots & \delta_{0p} \\ \delta_{10} & \delta_{11} & \dots & \delta_{1p} \\ \dots & \dots & \dots & \dots \\ \delta_{p0} & \delta_{p1} & \dots & \delta_{pp} \end{matrix},$$

где элементы (δ_{ij}) — ковариации оценок параметров $\left(b_i\right)$ и $\left(b_j\right)$.

Ковариация характеризует как степень рассеяния значений двух переменных относительно их математических ожиданий, так и взаимосвязь этих переменных.

Оценка дисперсии коэффициентов множественной регрессии рассчитывается следующим образом:

$$\sum\nolimits_{\tilde{b}} = \delta^2 \left(X'X\right)^{-1} \qquad (4.6)$$

Оценочное значение дисперсии ошибок (возмущений) в случае множественной регрессии находится по формуле:

$$S^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-p-1} \qquad (4.7)$$

Перейдем к оценке значимости коэффициентов и параметров модели множественной регрессии:

— *Значимость коэффициентов множественной регрессии* (b_j) можно проверить по t-критерию Стьюдента. Поэтому $(\tilde{b}_j)$ значимо отличается от нуля на уровне значимости (α) ,если $|t| = \frac{|\tilde{b}_j|}{S_{b_j}} > t_{1-\alpha, n-p-1}, t_{1-\alpha, n-p-1}$ — табличное значение t-критерия Стьюдента, определенное на уровне значимости (α) при числе степеней свободы $K = n - p - 1$. В общей постановке гипотеза (H_0) о равенстве параметра (b_j) заданному числу (b_{j_0}) отвергается, если:

$$|t| = \frac{|\tilde{b}_j - b_{j0}|}{S_{b_j}} > t_{1-\alpha, n-p-1} \qquad (4.8)$$

Поэтому *доверительный интервал для параметра* (b_j) есть:

$$\tilde{b}_j - t_{(1-\alpha,\, n-p-1)} \cdot s_{bj} \le b_j \le \tilde{b}_j + t_{(1-\alpha, n-p-1)} \cdot s_{bj}, \qquad (4.9)$$

где
$$s_{bj} = s \cdot \sqrt{\left[\left(X\ X\right)^{-1}\right] jj};$$

$s^2 = \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n-p-1}$ — выборочная остаточная дисперсия;

p — число экзогенных переменных.

— *Доверительный интервал для функции множественной регрессии* (*условного математического ожидания*) имеет вид:

$$\hat{y} - t_{1-\alpha,k} \times s_{\hat{y}} \le M(Y) \le \hat{y} + t_{1-\alpha,k} \times s_{\hat{y}},$$

где $\hat{y}$ – групповая средняя, определяемая по уравнению регрессии;

$s_{\hat{y}}$ – стандартная ошибка групповой средней:

$$s_{\hat{y}} = s \cdot \sqrt{\left(X_0'(X'X)^{-1}X_0\right)}$$

$$X_0' = (1, x_{01}, x_{02} \ldots x_{0k})$$

$$s^2 = \frac{\sum\left(\hat{y}_i - y_i\right)^2}{n - p - 1},$$

где s^2 – выборочная остаточная дисперсия.

— *Доверительный интервал для индивидуальных значений зависимой переменной (y_0^*) в множественной регрессии имеет вид:*

$$\hat{y}_0 - t_{1-\alpha,n-p-1} \cdot s_{\hat{y}_0} \le y_0^{\bullet} \le \hat{y}_0 + t_{1-\alpha,n-p-1} \cdot s_{\hat{y}_0}$$

$$s_{\hat{y}_0} = s \cdot \sqrt{1 + X_0'(X'X)^{-1}X_0}$$

$$s^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n - p - 1}$$

Доверительный интервал для параметра δ^2 в множественной регрессии имеет вид:

$$\frac{ns^2}{\chi_{\frac{\alpha}{2},n-p-1}} \le \delta^2 \le \frac{ns^2}{\chi_{1-\frac{\alpha}{2},n-p-1}}.$$

Уравнение множественной регрессии значимо, если:

$$F = \frac{Q_R(n-p-1)}{Q_e p} = \frac{R^2(n-p-1)}{(1-R^2)p} \succ F_{\alpha,k_1,k_2}$$

$$k_1 = p$$

$$k_2 = n - p - 1$$

$$Q_R = Q - Q_e = b'X'Y - n\overline{y}^2$$

$$Q_e = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = Y'Y - b'X'Y$$

$$Q = \sum_{i=1}^{n} (y_i - \overline{y})^2 = Y'Y - n\overline{y}^2,$$

где $F_{\alpha,k,k_{21}}$ — табличное значение F-критерия Фишера — Снедекора.

В теме № 3 был введен коэффициент детерминации $\left(R^2\right)$ как оценка адекватности регрессионной модели. *Коэффициент множественной детерминации* определяется по формуле:

$$R^2 = \frac{Q_R}{Q} = 1 - \frac{Q_e}{Q} = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2}.$$

Недостатком коэффициента множественной детерминации является то, что он увеличивается при добавлении новых экзогенных переменных, хотя это и не обязательно означает улучшение качества регрессионной модели. Поэтому предпочтительней использовать скорректированный коэффициент детерминации

$$\hat{R}^2 = 1 - \frac{n-1}{n-p-1}(1 - R^2) \tag{4.18}$$

Контрольные вопросы к теме № 4

1. Что представляет уравнение множественной регрессии.
2. Почему при рассмотрении множественной регрессии необходимо пользоваться матричной записью.
3. Каким образом производится оценка параметров в уравнении множественной регрессии.
4. Назовите предпосылки для множественного регрессионного анализа.

5. Для чего служат стандартизированные коэффициенты регрессии и коэффициенты эластичности.

Задача № 4.1. Имеется линейная модель множественной регрессии:

$$Y = 0{,}21 + 0{,}0030\,X_{1i} + 0{,}0092\,X_{2i} + U_i$$
$$(0,045) \qquad (0{,}0016) \qquad (0{,}0050)$$

В скобках указаны стандартные отклонения оценок коэффициентов.

$$I = 1, 2 \ldots 15.$$

Требуется:

1. Проверить статистическую значимость коэффициентов уравнения при $\alpha = 0{,}05$.
2. Определить, является ли константа значимо меньше 0,31.
3. Проверить совместную статистическую значимость переменных X_1 и X_2, если сумма квадратов ошибок составляет 0,0084, а дисперсия наблюдаемой переменной $Y = 0{,}0011$.

Задача № 4.2. Имеется уравнение регрессии в виде:

$$Y = -3{,}54 + 0{,}854\,X_1 + 0{,}367\,X_2.$$

Требуется:

1. Оценить математическое ожидание $M_x(Y)$ при $X_0 = (1\ 8\ 6)$.

Задача 4.3. Имеется уравнение регрессии в виде:

$$Y = -3{,}54 + 0{,}854\,X_1 + 0367\,X_2$$

А также известно $\sum_{i=1}^{10} e_i = 6{,}329$ (остаточная дисперсия), $n = 10$,

$$\left(X'X\right)^{-1} = \frac{1}{3738}\begin{pmatrix} 15027 & -1209 & -522 \\ -1209 & 201 & -108 \\ -522 & -108 & 244 \end{pmatrix}$$ — матрица, обратная матрице сумм первых степеней, квадратов и попарных произведений (n) наблюдений экзогенных переменных.

Требуется:

1. Найти 95%-ые доверительные интервалы для индивидуального и среднего значений эндогенной переменной.

2. Найти интервальную оценку для дисперсии эндогенной переменной.

3. Проверить значимость коэффициентов регрессии (B_1) и (B_2)

Задача 4.4. Изменение спроса на некоторое благо (Y) у домашних хозяйств можно объяснить с помощью цены этого блага (X_1) и дохода домохозяйств (X_2). Информация дана в таблице.

Y	31,4	30,4	32,1	31	30,5	29,8	31,1	31,7	30,7	29,7
X_1	4,1	4,2	4,0	4,6	4,0	5,0	3,9	4,4	4,5	4,8
X_2	1050	1010	1070	1060	1000	1040	1030	1080	1050	1020

Требуется:

1) Оценить параметры линейного двухфакторного уравнения $Y = B_0 + B_1 X_{1_i} + B_2 X_{2_i} + U_i$ и интерпретировать оценки.

2) Оценить дисперсию ошибки $\left(\delta_u^{\,2}\right)$.

3) Рассчитать оценку математического ожидания (Y) при $X_1 = 5{,}5$ и $X_2 = 980$.

ТЕМА № 5
МЕТОД НАИМЕНЬШИХ КВАДРАТОВ

План

1. Метод наименьших квадратов.
2. Допущения классической линейной модели регрессии Теорема Гаусса — Маркова.

5.1. Метод наименьших квадратов

Классическая модель линейной регрессии (для случая множественной регрессии) имеет вид (4.1):

$$y_i = b_0 + b_1 X_{i_1} + b_2 X_{i_2} + \ldots + b_p X_{i_p} + \varepsilon_i .$$

Для простоты ограничимся рассмотрением парной линейной регрессии, то есть модели вида (3.2):

$$y_i = b_0 + b_1 x_i + u_i .$$

Напомним, что оценкой этой модели является уравнение регрессии

$$\hat{y}_i = \tilde{b}_0 + \tilde{b}_1 x_i ,$$

где $\tilde{b}_0, \tilde{b}_1$ – оценки коэффициентов регрессии .

Для расчета неизвестных параметров $\left(\tilde{b}_0\right)$ и $\left(\tilde{b}_1\right)$ применяют *метод наименьших квадратов* (*МНК*). В основе данного метода лежит поиск таких значений коэффициентов $\left(\tilde{b}_0\right)$ и $\left(\tilde{b}_1\right)$, при которых сумма квадратов ошибок была бы наименьшей:

$$s^2 = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 = \sum_{i=1}^{n} (\tilde{b}_0 - \tilde{b}_1 - y_i)^2 \to \min . \qquad (5.1)$$

Согласно методу наименьших квадратов:

$$\hat{y} - \overline{y} = \tilde{b}_1 (x - \overline{x}) ; \qquad (5.2)$$

$$\tilde{b}_1 = \frac{\overline{xy} - \overline{x} \times \overline{y}}{\overline{x^2} - \left(\overline{x}\right)^2}; \qquad (5.3)$$

$$\tilde{b}_0 = \overline{y} - \tilde{b}_1 \overline{x}. \qquad (5.4)$$

5.2. Допущения классической линейной модели регрессии Теорема Гаусса — Маркова

Возникает вопрос, являются ли оценки $\tilde{b}_i$ параметров b_i наилучшими. Ответ на этот вопрос дает теорема Гаусса — Маркова: Если регрессионная модель (3.2) удовлетворяет предпосылкам (а — з) (тема № 3), то оценки $\tilde{b}_0$ (5.4) и $\tilde{b}_1$ (5.3) имеют наименьшую дисперсию в классе всех линейных несмещенных оценок.

Свойства оценок параметров классической линейной регрессионной модели изложены в теме № 6.

Контрольные вопросы к теме № 5

1. В чем заключается метод наименьших квадратов.
2. Какие допущения лежат в основе классической модели линейной регрессии.

Для чего нужны эти допущения.

3. Сформулируйте теорему Гаусса — Маркова.

Задача 5.1. По данным задачи 3.3:

А) найти уравнение регрессии (*Y*) по (*X*).

В) найти коэффициент детерминации и пояснить его смысл.

Задача 5.2. Имеется зависимость между сменной добычей угля на одного рабочего (*Y*) (т) и мощностью пласта (*X*)(м), полученная по *n* =10 шахтам:

X_i	8	11	12	9	8	8	9	9	8	12
Y_i	5	10	10	7	5	6	6	5	6	8

Требуется найти:

1) уравнение регрессии (*Y*) по (*X*) и построить ее график;

2) коэффициент корреляции между (*Y*) и (*X*);

3) сменную среднюю добычу угля на одного рабочего для шахт с мощностью пласта 8 м;

4) 95%-ые доверительные интервалы для индивидуального и среднего значений сменной добычи угля на одного рабочего для шахт с мощностью пласта 8м;

5) с надежностью 0,95 % интервальные оценки коэффициента регрессии (B_1) и дисперсии $\left(\delta^2\right)$.

Задача 5.3. Имеется информация о n = 10 парах наблюдений экзогенной переменной (X) и эндогенной переменной (Y):

X	15,8	8,4	14,5	8,6	11,8	19,5	21,4	4,7	9,8	13,5
Y	18,3	10,1	16,9	11,4	14,9	19,9	22,8	7,8	10,3	16,6

Требуется с помощью МНК оценить параметры линейного однофакторного уравнения регрессии

$$Y = A_0 + A_1 X_i + U_i,\ i = 1, 2 \ldots n.$$

Задача 5.4. Имеется информация о взаимосвязи экзогенной (X) и эндогенной (Y) переменной, данные представлены в таблице:

X_i	40	50	60	80	100	110	120	130	150	160	180	200	310
Y_i	14	14	17	19	17	20	24	22	25	24	18	20	26

Требуется с помощью МНК оценить параметры регрессионного уравнения

$$Y_i = B_0 + A_1 X_i + U_i,\ i = 1, 2 \ldots 13.$$

ТЕМА № 6
СВОЙСТВА ОЦЕНОК МНК

План

Свойства оценок параметров классической линейной регрессионной модели:

- состоятельность;
- несмещенность:
- эффективность.

Если выполняются допущения (*а* — *з*, Тема № 3), лежащие в основе модели линейной регрессии, то оценки коэффициентов $\left(\tilde{b}_i\right)$, полученные с помощью МНК, являются BLUE, что означает:

- **B** (best) — наилучшая.
- **L** (linear) — линейная.
- **U** (unbiased) — несмещенная.
- **E** (estimator) — оценка.

Рассмотрим эти свойства.

Линейная — свойство линейной функциональной зависимости оценки параметров от выборочных наблюдений.

Несмещенная — математическое ожидание оценки параметров равно параметру генеральной совокупности, то есть оценка неизвестного параметра называется несмещенной, если при любом объеме выборки результат ее осреднения по всем возможным выборкам данного объема равен истинному значению параметра:

$$\begin{aligned} M(\tilde{b}_0) &= b_0 \\ M(\tilde{b}_1) &= b_1 \end{aligned}.$$

Наилучшая (эффективная) — оценка, имеющая наименьшую дисперсию из всех возможных несмещенных оценок. Означает, что если оценка эффективна, то распределение значений сосредоточено вокруг истинной величины оцениваемого параметра.

Состоятельность. Оценка неизвестного параметра называется состоятельной, если по мере роста числа наблюдений она стремится по вероятности к оцениваемому параметру:

$$\lim_{n\to\infty} P\left[(\tilde{b}-b) \triangleright \delta\right] = 0$$

где δ – сколь угодно малая величина.

Контрольные вопросы к теме № 6

Какие свойства имеют оценки параметров классической линейной регрессионной модели.

ТЕМА № 7 ЛИНЕЙНЫЕ РЕГРЕССИОННЫЕ МОДЕЛИ С ГЕТЕРОСКЕДАСТИЧНЫМИ И АВТОКОРРЕЛЯЦИОННЫМИ ОСТАТКАМИ

План

1. Последствия нарушения допущений классической модели линейной регрессии.
2. Гетероскедастичность: обнаружение и устранение.
3. Автокорреляция регрессионных остатков: тестирование и устранение.

7.1. Последствия нарушения допущений классической модели линейной регрессии

В классической модели линейной регрессии делается ряд допущений, для того, чтобы оценки коэффициентов, полученные методом наименьших квадратов, имели ряд желательных свойств — состоятельность, несмещенность, эффективность.

Однако при моделировании реальных экономических процессов возможны ситуации, в которых условия классической линейной модели регрессии оказываются нарушенными. В итоге все результаты моделирования потеряют свою легитимность.

Рассмотрим нарушения каждого из допущений в отдельности:

1. Если математическое ожидание регрессионных остатков не равно нулю, то возможны следующие проблемы: оценки коэффициентов регрессии b_i могут быть смещенными, а коэффициенты детерминации R^2 и $\hat{R}^2$ могут иметь негативные (бессмысленные) значения. Чтобы избежать подобных проблем, следует включать в линейную модель регрессии свободный член b_0

2. Допущение о постоянстве дисперсии регрессионных остатков известно как допущение о *гомоскедастичности* (*равноизменчивости*). Если дисперсия регрессионных остатков не является постоянной, то говорят, что оценки *гетероскедастичны*. Гетероскедастичность приводит к тому, что результаты, связанные

с анализом точности модели, оценкой значимости и построением интервальных оценок ее коэффициентов, оказываются непригодными.

Обнаружить гетероскедастичность возможно с помощью теста:

1) *Тест Голдфелда — Квандта*

Предполагает, что гипотеза об отсутствии гетероскедастичности отвергается, если

$$F = \frac{\sum_{i=1}^{m} u_i^2}{\sum_{i=g-m+1}^{n} u_i^2} \succ F_{\alpha,m-p,m-p},$$

где p — число экзогенных переменных;

n — число наблюдений;

$F_{\alpha,m-p,m-p}$ — табличное значение критерия Фишера-Снедекора.

Эффективность теста максимальна, если $m = \frac{n}{3}$.

2) *Тест Глейзера*

Для изучения гетероскедастичности оценивается абсолютная величина регрессионных остатков, то есть модель вида

$$|e\ | = \varphi(x_i) + u_i, \quad \text{где } i = 1, 2, 3, \ldots n.$$

В качестве функции φ выбирается функция вида $\varphi = \alpha + \gamma x^{\delta}$. Регрессия $|e_i|$ осуществляется при разных значениях (δ), затем выбирается то значение, при котором коэффициент (γ) оказывается наиболее значимым, то есть имеет наибольшее значение t — статистики.

3) Тест Уайта

1. Имеется исходная модель вида

$$y_t = \beta_t + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t.$$

В результате оценки данной модели получаем регрессионные остатки $\hat{u}_t$.

7.2. Гетероскедастичность: обнаружение и устранение

Оцениваем вспомогательную регрессию вида

$$\hat{u}^2 = \alpha_1 + \alpha_2 x_{2t} + \alpha_3 x_{3t} + \alpha_4 x_{2t}{}^2 + \alpha_5 x_{3t}{}^2 + \alpha_6 x_{2t} x_{3t} + \nu_t \, ,$$

где ν_t — нормально распределенная ошибка, независимая от u_t.

3. Исследуем статистику

$$nR^2 \infty \chi_m{}^2 \, ,$$

где m — количество независимых переменных

n — количество наблюдений.

4. Проверяем нулевую гипотезу:

$$\alpha_2 = 0 \quad \alpha_3 = 0 \quad \alpha_4 = 0 \quad \alpha_5 = 0 \quad \alpha_6 = 0 \, .$$

Если фактические значения статистики превышают критические величины распределения χ^2, то нулевая гипотеза о гомоскедастичности остатков отвергается, то есть делается вывод о присутствии гетероскедастичности.

Если известна причина и форма гетероскедастичности, то для ее устранения применяют *обобщенный метод наименьших квадратов* (*ОМНК*)*,* речь о котором пойдет в Теме № 8.

7.3. Автокорреляция регрессионных остатков: тестирование и устранение

Регрессионная модель МНК позволяет получить несмещенную оценку с минимальной дисперсией при условии, что регрессионные остатки независимы друг от друга, то есть

$$cov\left(u_i, u_j\right) = 0 \, .$$

Если это условие нарушено, то имеет место *автокорреляция*, при которой коэффициенты регрессии несмещены, но стандартные ошибки недооценены, а проверка статистической значимости коэффициентов ненадежна. Автокорреляция бывает:

— *положительной*, когда завышенные значения в предыдущих наблюдениях приводят к их завышению в наблюдениях последующих.

— *отрицательной,* когда завышенные значения в предыдущих наблюдениях приводят к занижению их в наблюдениях последующих.

Для проверки автокорреляции существуют ряд тестов:

• *Тест Дарбина — Уотсона*

Определяет наличие автокорреляции между соседними членами регрессии. С этой целью рассчитывают критерий:

$$d = \frac{\sum_{t=2}^{n}(u_t - u_{t-1})^2}{\sum_{t=1}^{n} u^2{}_t} \approx 2(1 - r),$$

где u_t — остатки регрессии;

r — выборочный коэффициент корреляции.

Если

$d_в \prec d \prec 4 - d_в$, то гипотеза об отсутствии автокорреляции не отвергается;

$d_н \prec d \prec d_в$, то вопрос о принятии гипотезы остается открытым;

$0 \prec d \prec d_н$, то принимается альтернативная гипотеза о положительной автокорреляции;

$4 - d_н \prec d \prec 4$, то принимается альтернативная гипотеза об отрицательной автокорреляции,

где $d_н, d_в$ — пороговые значения критерия Дарбина — Уотсона.

Таким образом, близость наблюдаемого значения (d) к нулю должна означать наличие положительной автокорреляции, к четырем — отрицательной.

Данный тест применим в случае, если в регрессионной модели присутствует свободный член, экзогенные переменные являются нестохастическими. Кроме того, надо учитывать, что тест проверяет только наличие *автокорреляции между регрессионными остатками в последовательных наблюдениях*.

• *Тест Бреуша — Годфри*

Тест проводится согласно алгоритму.

1. Проводится оценка линейной регрессии с помощью МНК и рассчитываются регрессионные остатки.

2. Проводится оценка вспомогательной регрессии:

$$\hat{u}_t = \gamma_1 + \gamma_2 x_{2t} + \gamma_3 x_{3t} + \gamma_4 x_{4t} + \rho_1 \hat{u}_{t-1} + \rho_2 \hat{u}_{t-2} + \rho_r u_{t-r} + \nu_t$$

$$\nu_t \infty N(0, \delta_\nu{}^2).$$

На основе вспомогательной регрессии рассчитывается коэффициент детерминации $\left(R^2\right)$.

3. Рассчитывается статистика $\left(n-r\right)R^2 \infty \chi_r{}^2$, где n — объем выборки.

Если значение статистики превосходит критическую величину $\left(\chi^2\right)$, то отвергается нулевая гипотеза об отсутствии автокорреляции.

• *Тест Льинга — Бокса*

Статистика Льинга — Бокса имеет вид:

$$Q_p = n(n+2)\sum_{\tau=1}^{p} \frac{r^2(\tau)}{n-\tau}.$$

Если верна гипотеза H_0 о равенстве нулю всех коэффициентов корреляции $\rho(e_t, e_{t-\tau})$, где $\tau = 1, 2, \dots p$, то статистика Q_p имеет распределение χ^2 с (ρ) степенями свободы.

Заметим, что провести «вручную» вышеуказанные тесты достаточно трудно, поэтому в большинстве компьютерных пакетов тесты выполняются специальными командами.

Для устранения автокорреляции можно воспользоваться обобщенным методом наименьших квадратов (ОМНК) (Тема № 8).

Контрольные вопросы к теме № 7

1. Какие последствия имеют нарушения допущений классической модели линейной регрессии.

2. Каковы последствия нарушения допущения $M(u_i) = 0$.

3. Каковы последствия нарушения допущения $D(u_i)$ — const.
4. Каким образом можно обнаружить гетероскедастичность.
5. Каким образом можно устранить гетероскедастичность.
6. Каковы последствия нарушения допущения $\operatorname{cov}(u_i, u_j) = 0$.
7. Как проводится проверка уравнения на автокорреляцию.
8. Каким образом можно устранить автокорреляцию.

Задача 7.1. По данным n =150 наблюдений о доходе человека (Y), уровне его образования (X1) и возрасте (X2) выяснить, можно ли считать на уровне значимости a =0.05 линейную модель

$$Y = -3{,}06 + 3{,}25\,X_1 + 0{,}48\,X_2$$

$$(-1{,}4) \quad (5{,}96) \qquad (8{,}35)$$

(в скобках указаны значения t-статистик коэффициентов регрессии) гетероскедастичной. Исходные данные:

$$\sum_{i=1}^{50} e^2{}_i = 894{,}1 \qquad \sum_{i=101}^{150} e_i^2 = 3918{,}2.$$

Задача 7.2. В таблице приведены данные по 18 наблюдениям модели пространственной выборки:

I	X_i	e_i	I	X_i	e_i
1	21,3	2,3	10	71,5	23,8
2	22,6	5,6	11	75,7	45,7
3	32,7	12,8	12	76	34,7
4	41,9	10,1	13	78,9	56,9
5	43,8	14,6	14	79,8	56,8
6	49,7	13,9	15	80,7	49,8
7	56,9	24	16	80,8	58,9
8	59,7	21,9	17	96,9	87,8
9	67,8	19,7	18	97	87,5

Предполагая, что ошибки регрессии представляют собой нормально распределенные случайные величины, проверить гипотезу о гомоскедастичности, используя тест Голдфелда — Квандта.

Задача 7.3. Для линейного однофакторного уравнения регрессии $Y_t = A_0 + A_1 X_t + u_t$ (t =1,2, … ,T) имеется T =20 пар наблюдений целевой переменной (Y) и экзогенной переменной (X), которые представлены в таблице.

X_t	5,5	8,5	20,1	24,5	17	22	19	16	5	13,4
Y_t	4,5	10	18,5	20	18,5	25	8,5	13	7,4	15,6
X_t	3,0	6,1	22,2	20,1	8	12	14	19,5	18	15,1
Y_t	5,5	5,2	18,5	18,0	8	9,8	12	14,8	15,2	12

Автокорреляция отсутствует, ошибки регрессии распределены нормально. Имеется подозрение на гетероскедастичность. Для уровня $\alpha = 0{,}05$ проверить гипотезу, что дисперсии ошибки для первых и последних 10 пар наблюдений различны.

ТЕМА № 8
ОБОБЩЕННЫЙ МЕТОД НАИМЕНЬШИХ КВАДРАТОВ

План

1. Обобщенная линейная модель множественной регрессии.
2. Теорема Айткена Обобщенный метод наименьших квадратов.

8.1. Обобщенная линейная модель множественной регрессии

Имеет вид:

$$Y = XB + U$$

$$Y = \begin{pmatrix} y_1 \\ y_n \end{pmatrix} X = \begin{pmatrix} 1 & x_{11} & x_{k1} \\ 1 & x_{1n} & x_{kn} \end{pmatrix} B = \begin{pmatrix} b_0 \\ b_k \end{pmatrix} U = \begin{pmatrix} u_1 \\ u_n \end{pmatrix}.$$

Описывается системой условий:

U — случайный вектор, X — неслучайная матрица;

$M(u) = 0_n$, 0_n — нулевая матрица-столбец размера (n);

$\sum_u = \Omega$, Ω — положительно определенная матрица;

$r(X) = p + 1 \prec n$, где p — число экзогенных переменных; n — число наблюдений.

Обобщенная линейная модель множественной регрессии отличается от классической модели тем, что в ней ковариации и дисперсии могут быть произвольными.

8.2. Теорема Айткена. Обобщенный метод наименьших квадратов

Обычный метод наименьших квадратов в обобщенной линейной регрессионной модели дает *смещенную* оценку ковариационной матрицы $\sum_b (4.6)$ вектора оценок (B). Оценка (B), определенная по (4.3) будет *состоятельной, но не оптимальной* в смысле теоремы Гаусса — Маркова. Для получения эффективной оценки нужно использовать оценку, полученную *обобщенным методом наименьших квадратов* (*ОМНК*).

Вопрос об эффективности несмещенной оценки вектора (B) для обобщенной модели регрессии решается с помощью *теоремы Айткена*: ***В классе линейных несмещенных оценок вектора** (B) **для обобщенной регрессионной модели оценка** $b^* = \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}Y$ **имеет наименьшую ковариационную матрицу.***

Оценка дисперсии вектора (B) рассчитывается по формуле

$$\mathrm{S}_*^{\ 2} = \frac{(Y - Xb^*)'\Omega_0^{\ -1}(Y - Xb^\bullet)}{n - p - 1};$$

$$\Omega_0 = \begin{pmatrix} 1 & p & p^{n-1} \\ p & 1 & p^{n-2} \\ p^{n-1} & p^{n-2} & 1 \end{pmatrix},$$

где (p) — неизвестный параметр, который нужно оценить (коэффициент авторегрессии).

В модели с гетероскедастичными остатками ковариационная матрица $\sum_u = \Omega$ вектора возмущений имеет следующую структуру:

$$\Omega = \begin{pmatrix} \delta_1^{\ 2} & 0 & 0 \\ 0 & \delta_2^{\ 2} & 0 \\ 0 & 0 & \delta_n^{\ 2} \end{pmatrix},$$

где $\delta_i^{\ 2}$ — дисперсии ошибок.

В модели с автокорреляционными остатками ковариационная матрица вектора ошибок имеет вид

$$\Omega = \frac{\delta_0^{\ 2}}{1 - p^2} \begin{pmatrix} 1 & p & p^{n-1} & \\ p & 1 & p^{n-2} & \\ p^{n-1} & & p^{n-2} & 1 \end{pmatrix},$$

где p — коэффициент авторегрессии; $\delta_0^{\ 2}$ — нулевая дисперсия.

Таким образом, для применения обобщенного метода наименьших квадратов необходимо знание ковариационной матрицы вектора возмущений $\left(\sum_U = \Omega\right)$, что встречается редко в практике эконометрического моделирования. Поэтому часто применяется *реализуемый* (*доступный*) *обобщенный метод наименьших квадратов*.

Оценкой доступного обобщенного метода наименьших квадратов вектора (B) есть:

$$\hat{b}^* = \left(X'\hat{\Omega}_0^{-1}X\right)^{-1} X'\Omega^{-1}{}_0 Y .$$

Контрольные вопросы к теме № 8

1. Является ли коэффициент детерминации удовлетворительной мерой качества обобщенной регрессионной модели.

2. Почему для практической реализации обобщенного метода наименьших квадратов необходимо вводить дополнительные условия на структуру ковариационной матрицы вектора возмущений (ошибок).

3. Какова структура ковариационной матрицы вектора ошибок в модели с автокорреляционными и гетероскедастичными остатками.

Задача 8.1. Для линейного однофакторного уравнения регрессии $Y_t = A_0 + A_1X_t + E_t$; t = 1, 2, ..., T.

Имеется T = 12 пар наблюдений целевой переменной (Y) и экзогенной переменной (X), которые представлены в таблице

X_t	5	2,5	1,8	6,8	9	3,8	6,5	9	1	3,5	7,1	10
Y_t	5	4,8	3,1	8,2	8,6	5,5	6,5	11,1	2,1	4,5	8,9	11,8

Для ошибки уравнения (E_t) выполняются предпосылки авторегрессии первого порядка с известными значениями p = 1 и $\delta_t^2 = 1$.

Требуется оценить параметры уравнения (A_0) и (A_1) с помощью ОМНК.

Задача 8.2. Имеется гетероскедастичная однофакторная регрессионная модель $Ct = A_0 + A_1Y_t + E_t$; t = 1, 2, ..., T., где C — потребление домохозяйства определенной структуры, Y — доход этого домохозяйства. Ошибки попарно не коррелированы, дисперсии ошибки при доходе от 50 до 100 единиц в 2 раза больше, чем при доходе до 50 единиц. Имеется выборка объемом 9 наблюдений:

Y_t	30	35	35	45	50	60	70	90	160
C_t	30	30	35	35	40	50	70	80	120

Требуется: 1) Определить ковариационную матрицу ошибок для этой модели.

2) Оценить параметры уравнения при помощи ОМНК.

Задача 8.3. Для линейного однофакторного уравнения регрессии $Y_t = A_0 + A_1X_t + E_t$; t = 1, 2, ..., T.

Имеется T = 18 пар наблюдений переменной (Y) и экзогенной переменной (X), которые представлены в таблице:

X_t	0,10	0,15	0,2	0,25	0,3	0,35	0,4	0,45	0,5
Y_t	0,019	0,019	0,027	0,051	0,093	0,136	0,171	0,198	0,267
X_t	0,55	0,6	0,65	0,7	0,75	0,8	0,85	0,9	0,95
Y_t	0,314	0,365	0,396	0,482	0,569	0,627	0,71	0,835	0,913

Для ошибки уравнения (u) выполняется предпосылка авторегрессии первого порядка.

Требуется определить оценку параметра авторегрессии ошибки.

ТЕМА № 9
ВОПРОСЫ ПРАКТИЧЕСКОГО ИСПОЛЬЗОВАНИЯ РЕГРЕССИОННЫХ МОДЕЛЕЙ. РЕГРЕССИОННЫЕ МОДЕЛИ С ПЕРЕМЕННОЙ СТРУКТУРОЙ. ФИКТИВНЫЕ ПЕРЕМЕННЫЕ

План

1. Мультиколлинеарность.
2. Частная корреляция.
3. Линейные регрессионные модели с переменной структурой. Фиктивные переменные.

9.1. Мультиколлинеарность

Мультиколлинеарность — высокая взаимная коррелированность экзогенных переменных. При мультиколлинеарности коэффициенты регрессии ненадежны и уравнение регрессии, как правило, не имеет реального смысла.

Точных количественных критериев для обнаружения мультиколлинеарности нет. Тем не менее, есть некоторые рекомендации по ее выявлению

1) Анализируют матрицу парных коэффициентов корреляции (ту ее часть, которая относится к экзогенным переменным). Считается, что если в ней содержатся коэффициенты корреляции, превышающие по абсолютной величине значения 0,75–0,8, то это свидетельствует о присутствии мультиколлинеарности.

2) Исследуют матрицу ($X'X$). Если определитель матрицы det ($X'X$) или ее минимальное собственное значение близки к нулю, то это говорит о наличии мультиколлинеарности.

3) О присутствии мультиколлинеарности свидетельствуют некоторые внешние признаки модели:

- некоторые из коэффициентов имеют неверные с позиции экономической теории знаки или неоправданно большие значения;
- небольшое изменение исходных данных приводит к существенному изменению оценок модели;

- при проверке с помощью t-критерия большинство коэффициентов регрессии оказываются статистически незначимыми, в то время как в действительности они имеют отличные от нуля значения, а модель в целом оказывается статистически значимой при проверке с помощью F-теста.

Для уменьшения мультиколлинеарности используют следующие меры:

1) Увеличивают объем выборки, что позволяет получить МНК — оценки с меньшей дисперсией.

2) Исключают переменные, которые высококоррелированы с остальными. При этом, какую переменную оставить, а какую убрать из анализа, решают на основании экономических соображений.

3) Используют пошаговые процедуры отбора наиболее информативных переменных. Процедура введения новых переменных продолжается до тех пор, пока будет увеличиваться соответствующий коэффициент детерминации.

9.2. Частная корреляция

Если переменные коррелируют друг с другом, то на значении коэффициента корреляции сказывается влияние других переменных. Поэтому возникает необходимость исследования *частной корреляции* между переменными при исключении влияния одной или нескольких переменных. *Частным коэффициентом корреляции* между переменными (X_i) и (X_j) при фиксированных значениях остальных ($p - 2$) переменных называется выражение

$$r_{ij1,2,3\dots p} = \frac{g_{ij}}{\sqrt{g_{ii} g_{jj}}},$$

где g_{ii} и g_{jj} — алгебраические дополнения элементов r_{ii} и r_{jj} матрицы выборочных коэффициентов корреляции

$$g_p = \begin{pmatrix} 1 & r_{12} & r_{1p} \\ r_{21} & 1 & r_{2p} \\ r_{p1} & r_{2p} & 1 \end{pmatrix}.$$

Если число переменных $n = 3$, то

$$r_{ij,k} = \frac{r_{ij} - r_{ik} r_{jk}}{\sqrt{\left(1 - r^2{}_{ik}\right)\left(1 - r^2{}_{jk}\right)}}.$$

9.3. Линейные регрессионные модели с переменной структурой. Фиктивные переменные

На практике часто возникает необходимость исследования влияния качественных признаков, имеющих два или несколько уровней (образование, пол, сезонность и т. д.).

Качественные признаки влияют на структуру линейных связей между переменными и приводят к скачкообразному изменению параметров регрессионной модели. В этом случае строится *регрессионная модель с переменной структурой*.

Построение модели с переменной структурой предполагает введение в нее *фиктивных переменных или манекенов*. Фиктивные переменные вводятся в регрессионную модель для отражения влияния на эндогенную переменную $y(Y)$ сопутствующих качественных переменных.

В качестве фиктивных переменных используют бинарные переменные, которые принимают только одно из двух значений — «0» или «1».

В практике исследователя возникают случаи, когда имеются две выборки пар значений зависимой и объясняющей переменных и необходимо выяснить, возможно ли объединение двух выборок в одну и исследование единой модели регрессии (Y) по (X). Для ответа на этот вопрос служит тест (критерий) Г. Чоу. Согласно данному тесту по каждой выборке строятся две линейные регрессионные модели

$$y_i = \beta'_o + \sum_{j=1}^{p} \beta'_j x_{ij} + e' \quad i = 1, 2, \ldots n_1$$

$$y_i = \beta''_0 + \sum_{j=1}^{p} \beta''_j x_{ij} + e''_i \quad i = n_1 + 1, \ldots n_1 + n_2$$

Проверяемая нулевая гипотеза имеет вид

$$H_0 : \beta' = \beta'' \quad D = D\left(e''\right) = \sigma^2 ,$$

где β' и β'' — векторы параметров двух моделей;

e', e'' — их случайные возмущения.

Если нулевая гипотеза верна, то две регрессионные модели можно объединить в одну объема $n_1 + n_2 = n$

$$y_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + e_i ; \quad i = 1 \ldots n$$

Нулевая гипотеза (H_0) отвергается на уровне значимости (α), если статистика

$$F = \frac{\left(\sum_{i=1}^{n} e^2 i - \sum_{i=1}^{n1} e^2 i - \sum_{i=n1}^{n} e^2 i\right)(n - 2p - 2)}{\left(\sum_{i=1}^{n1} e^2 i + \sum_{i=n1}^{n} e^2 i\right)(p+1))} > F(\alpha, p+1, n-2p-2) ,$$

где $\sum_{i=1}^{n} e^2 i \quad \sum_{i=1}^{n1} e^2 i \quad \sum_{i=n1}^{n} e^2 i$ — остаточные суммы квадратов соответственно для объединенной, первой, второй выборок.

Контрольные вопросы к теме № 9

1. Что такое мультиколлинеарность.
2. По каким признакам можно обнаружить мультиколлинеарность.
3. Каким образом можно уменьшить мультиколлинеарность.
4. В каких случаях используются модели с переменной структурой.
5. Что такое фиктивные переменные.
6. Какое влияние оказывают фиктивные переменные на оценку модели.
7. Для чего служит критерий (тест) Г. Чоу.
8. Что позволяет исследовать частный коэффициент корреляции.

Задача 9.1. С целью исследования влияния факторов X_1 — среднемесячного количества профилактических наладок автоматической линии и X_2 — среднемесячного числа обрывов нити на показатель Y — среднемесячную характеристику качества ткани (в баллах) по данным 37 предприятий легкой промышленности были вычислены парные коэффициенты корреляции ry_1 = 0,105, ry_2 = 0.024, r_{12} = 0,996. Оценить частные коэффициенты корреляции ry_{12} и ry_{21} и оценить их значимость на 5%-ном уровне.

Задача 9.2. Для исследования зависимости между производительностью труда X1, возрастом X2 и производственным стажем X3 была произведена выборка из 100 рабочих одной и той же специальности. Вычисленные парные коэффициенты корреляции оказались значимыми и составили: r12 = 0.2, r13 = 0.41. r23 = 0.82. Вычислить частные коэффициенты корреляции и оценить их значимость на уровне α =0,05.

Задача 9.3. Для линейного трехфакторного уравнения регрессии $Y_t = A_0 + A_1X_{1t} + A_2X_{2t} + A_3X_{3t} + E_t$ (t =1, 2, ..., T) имеются данные

X_{1t}	10,3	14,6	11,4	17,1	10,6
X_{2t}	20,8	28	23	30,5	21,7
X_{3t}	4,1	20,3	9,8	8,1	17,7
Y_t	40	80	55	58	70

Требуется:

1. Определить корреляционную матрицу (R) и содержащийся в этих данных размер коллинеарности как det(R).
2. Рассчитать размер коллинеарности, в случае если из уравнения выводится переменная (X_2).

Задача 9.4. Для переменной « заработная плата» представлена модель: $Y_t = A_0 + A_1X_{1t} + A_2X_{2t} + E_t$, где Y_t — логарифм совокупной заработной платы; $X1_t$ — количество лет обучения; $X2_t$ — опыт работы.

Выборка составлена таким образом, что номера от 1 до 100 соответствуют женщинам, а со 101 по 300 — мужчинам.

Требуется:

1. Предложить два способа представления нулевой гипотезы, что заработная плата мужчины для данного уровня образования и опыта работы выше, чем у женщины с такими же характеристиками

2. Проверить гипотезу, что коэффициенты уравнений типа (9.3), построенных отдельно для подвыборок мужчин и женщин, совпадают. Известно, что в модели для женщин сумма квадратов остатков равна 0,13, а для мужчин 0,33. Оценка МНК по всей выборке дает сумму квадратов остатков 0,6.

ТЕМА № 10
НЕЛИНЕЙНЫЕ МОДЕЛИ РЕГРЕССИИ И ИХ ЛИНЕАРИЗАЦИЯ

План

1. Два подхода для оценки параметров нелинейных моделей.
2. Нелинейные модели регрессии.

10.1. Два подхода для оценки параметров нелинейных моделей

Соотношение между социально-экономическими явлениями и процессами не всегда можно выразить линейными функциями. Примерами нелинейных зависимостей являются: зависимость между объемом продукции и факторами производства, между спросом на товар и ценой и т.д.

Для оценки параметров нелинейных моделей используют два подхода:

1) Попытка подобрать такое преобразование к анализируемым переменным, которое позволило бы представить существующую зависимость в виде линейной функции (*линеаризация модели*).

2) Если линеаризация невозможна, то тогда к исследуемой зависимости применяют методы нелинейной регрессии (*нелинейная оптимизация*).

10.2. Нелинейные модели регрессии

Линеаризации поддаются следующие зависимости:

1) *Экспоненциальная*

$$y = \beta_0 e^{\beta_1 x + u} .$$

Ее можно представить в линейном виде путем преобразований

$$y^* = \beta_0{}^* + \beta_1 x + u$$
$$\beta^*{}_0 = \ln \beta_0 \quad y^* = \ln y .$$

2) *Логарифмическая*

$$y = \beta_0 + \beta_1 \ln X + u .$$

Переход к линейной зависимости осуществляется с помощью преобразования экзогенной переменной $X^* = \ln X$.

3) *Гиперболическая*

$$y = \beta_0 + \beta_1 \frac{1}{x} + u .$$

Приводится к линейному виду с помощью преобразования экзогенной переменной $x^* = \frac{1}{x}$.

В результате получим линейную модель вида $y = \beta_0 + \beta_1 x^* + u$.

4) *Степенная*

$$y = \beta_0 x^{\beta_1} .$$

Для перехода к линейной модели нужно произвести логарифмирование обоих частей уравнения

$$\log y = \log \beta_0 + \beta_1 \log x .$$

Обозначим $y^* = \log y \quad \beta_0^* = \log \beta_0 \quad x^* = \log x$,
получим линейную модель $y^* = \beta_0^* + \beta_1 x^*$.

5) *Показательная*

$$y = \beta_0 + \beta_1^{\ x} .$$

Для перехода к линейной модели нужно произвести логарифмирование обоих частей уравнения

$$\log y = \log \beta_0 + x \log \beta_1 .$$

Обозначим $y^* = \log y \quad \beta_0^* = \log \beta_0 \quad \beta_1^* = \log \beta_1$,
получим линейную модель $y^* = \beta_0^* + x\beta_1^*$.

Контрольные вопросы к теме № 10

1. Что делать, если исследуемая функция регрессии нелинейная.

2. Какие виды нелинейных зависимостей поддаются линеаризации.

ТЕМА № 11 ХАРАКТЕРИСТИКИ ВРЕМЕННЫХ РЯДОВ

План

1. Временной ряд и этапы его анализа.
2. Составляющие временного ряда: тренд, сезонная, циклическая, случайная компоненты.
3. Аналитическое выравнивание временного ряда. Прогнозирование на основе моделей временного ряда.

11.1. Временной ряд и этапы его анализа

Временной ряд представляет собой последовательность наблюдений некоторого признака (Y) в последовательные моменты времени. Анализ временных рядов базируется на следующей идее: *результирующая переменная* (Y) *складывается под влиянием большого числа факторов, многие из которых не поддаются идентификации и непосредственному наблюдению и измерению*. Поэтому лучшим источником информации о совокупности влияния всех этих факторов являются значения самой исследуемой переменной (Y) в прошлые моменты времени, а также текущие и прошлые значения случайных ошибок.

Основное прикладное значение моделей временных рядов состоит в прогнозировании экономических показателей. Использование временных рядов позволяет строить кратко- и среднесрочные экономические прогнозы, которые используются при решении следующих задач:

• планирование в производстве и торговле;

• управление и оптимизация социально-экономических процессов в обществе;

• частичное управление важными параметрами демографических процессов;

• принятие решений в бизнесе.

Выделяют *основные этапы* анализа временного ряда:

• графическое представление и описание поведения временного ряда;

• выделение и удаление закономерных составляющих временного ряда (U_t, C_t, V_t);

• сглаживание и фильтрация (удаление низко- или высокочастотных составляющих временного ряда);

• исследование (E_t), построение и проверка адекватности модели для ее описания;

• прогнозирование развития изучаемого процесса на основе имеющегося временного ряда;

• исследование взаимосвязи между различными временными рядами.

11.2. Составляющие временного ряда: тренд, сезонная, циклическая, случайная компоненты

$$Y_t = U_t + V_t + C_t + E_t, \ (t = 1, 2, \ldots, n),$$

где U_t — *тренд, плавно меняющаяся компонента*, описывающая длительную тенденцию изменения анализируемого показателя;

V_t — *сезонная компонента*, отражающая повторяемость экономических процессов в течении не очень длительного периода (год, месяц, неделя);

C_t — *циклическая компонента*, отражающая повторяемость экономических процессов в течении длительных периодов4

E_t — *случайная компонента*, отражающая влияние не поддающихся учету и регистрации случайных факторов.

11.3. Аналитическое выравнивание временного ряда. Прогнозирование на основе моделей временного ряда

Распространенными методами анализа временного ряда являются *корреляционный*, *спектральный анализ*, *модель авторегрессии и скользящей средней*, речь о них пойдет в следующих темах.

Важнейшей задачей при исследовании временного ряда является выявление *основной тенденции* изучаемого процесса, выраженной неслучайной составляющей $Y(t)$ (U_t, V_t, C_t). Для решения этой задачи прежде всего необходимо выбрать вид функции $f(t)$. Часто используются следующие функции:

• линейная

$$y_t = bt \ ;$$

• полиномиальная

$$y_t = b_0 + b_1 t^1 + \ldots + b_m t^m ;$$

• экспоненциальная

$$y_t = b_0 b^t ;$$

• степенная

$$y_t = b_0 t^b ;$$

• гиперболическая

$$y_t = b_0 + \frac{b}{t} .$$

При выборе функции f(t) используют *содержательный анализ* (устанавливающий характер динамики процесса), *визуальные наблюдения* (на основе графического изображения временного ряда). Предпочтение следует отдавать более простым функциям. Для выявления основной тенденции временного ряда применяют МНК. При этом значения временного ряда рассматриваются эндогенная переменная, а время (t) как экзогенная:

$$y_t = \varphi(t) + e_t ,$$

где e_t — возмущения, удовлетворяющие основным предпосылкам регрессионного анализа.

Если функция $\varphi(t)$ линейная, то согласно МНК, уравнение регрессии можно представить

$$\hat{y}_t - \overline{y} = b\left(t - \overline{t}\right)$$

$$\overline{t} = \frac{\sum_{i=1}^{n} t}{n} = \frac{1+n}{2}; \quad \overline{y} = \frac{\sum_{i=1}^{n} y_t}{n},$$

а коэффициент регрессии $\left(b_1\right)$ найти по формуле:

$$b_1 = \frac{\overline{y_t t} - \overline{y}_t \times \overline{t}}{\overline{t^2} - \left(\overline{t}\right)^2},$$

$$\overline{y_t t} = \frac{\sum_{t=1}^{n} y_t t}{n}; \quad \overline{y}_t = \frac{\sum_{t=1}^{n} y_t}{n};$$

$$\overline{t} = \frac{(1+n)}{2}; \quad \overline{t^2} = \frac{(n+1)(2n+1)}{6}.$$

В случае если функция $\varphi(t)$ нелинейная, то можно применить методы линеаризации или специальные процедуры оценивания.

Другим методом выявления основной тенденции временного ряда является *метод скользящих средних*. Он основан на переходе от начальных значений членов ряда к их средним значениям на интервале времени, длина которого определена заранее. Для усреднения могут быть использованы средняя арифметическая, медиана и др.

Одна из важных задач анализа временного ряда состоит в *прогнозировании* на его основе развития изучаемого процесса. Задача формулируется так: имеется временной ряд Yt ($t = 1, 2, \dots n$) и требуется дать прогноз уровня этого ряда на момент $n + \tau$.

Если рассматривать временной ряд как регрессионную модель изучаемого признака по переменной «время», то к нему можно применить рассмотренные в темах № 3, 4 методы анализа. Следует только помнить, что основной предпосылкой регрессионного анализа является независимость возмущения e_t $(1, 2, \dots n)$ и $M(e_t) = 0$, то есть соблюдаются условия нормальной классической регрессионной модели.

Контрольные вопросы к теме № 11

1. Что такое анализ временных рядов.
2. Каковы области практического применения анализа временных рядов.
3. Каковы основные компоненты, составляющие временной ряд.
4. Назовите основные этапы анализа временных рядов.
5. Назовите методы выявления основной тенденции развития изучаемого процесса.

6. Какие функции используют при выявлении основной тенденции изучаемого процесса.

Задача 11.1. В таблице приведены данные, отражающие спрос на товар за восьмилетний период:

Год, t	1	2	3	4	5	6	7	8
Спрос, Y_t усл. ед.	213	171	291	309	317	362	351	361

Найти уравнение тренда для временного ряда Y_t, полагая тренд линейным, проверить значимость уравнения тренда по F — критерию на 5%-ном уровне значимости.

Задача 11.2. Проверить сглаживание временного ряда Y_t по данным задачи 11.1 методом скользящих средних, используя простую среднюю арифметическую с интервалом сглаживания $m = 3$ года.

Задача 11.3. По данным задачи 11.1 дать точечную и с надежностью 0,95 интервальную оценки прогноза среднего и индивидуального значений спроса на товар на момент $t = 9$ (девятый год). Тренд линейный, возмущения удовлетворяют требованиям классической модели.

Задача 11.4. В таблице представлены данные, отражающие динамику роста доходов на душу населения (Y_t) (ден. ед.) за восьмилетний период:

T	1	2	3	4	5	6	7	8
Yt	1133	1222	1354	1389	1342	1377	1491	1684

Полагая, что тренд линейный и условия классической модели выполнены:

а) найти уравнение тренда и оценить его значимость на уровне 0,05;

в) дать точечный и с надежностью 0,95 интервальный прогнозы среднего и индивидуального значений доходов на девятый год.

Задача 11.5. Имеются следующие данные об урожайности озимой пшеницы (Y_t) (ц/га) за 10 лет:

t	1	2	3	4	5	6	7	8	9	110
Y	16,3	20,2	17,1	7,7	15,3	16,3	19,9	14,4	18,7	20,7

Требуется найти:

а) найти уравнение тренда, полагая, что он линейный;

в) провести сглаживание тренда временного ряда (*Y*) методом скользящих средних, используя простую среднюю арифметическую с интервалом сглаживания: 1) m = 3; 2) m =5.

ТЕМА № 12 МОДЕЛИ СТАЦИОНАРНЫХ И НЕСТАЦИОНАРНЫХ ВРЕМЕННЫХ РЯДОВ, ИХ ИДЕНТИФИКАЦИЯ

План

1. Стационарные временные ряды и их характеристики.
2. Авторегрессионные модели и модели скользящей средней.
3. Нестационарные временные ряды.

12.1. Стационарные временные ряды и их характеристики

Стационарный временной ряд *это такой ряд, который имеет постоянную среднюю, а значения ряда колеблются вокруг этой средней с некоторой постоянной дисперсией.*

Различают строго стационарный и слабо стационарный временной ряд. Ряд $\left(x_t\right)$ называется *строго стационарным*, если совместное распределение вероятностей *m* наблюдений $\left(x_{t1}, x_{t2}, \ldots x_{tm}\right)$ такое же как и для *m* наблюдений $\left(x_{t_1+\tau}, x_{t_2+\tau}, \ldots, x_{t_m+\tau}\right)$, для любых m, $\left(t_1, t_2, \ldots, t_m\right)$ и $\left(\tau\right)$, то есть закон распределения и числовые характеристики такого ряда не зависят от (*t*).

Ряд $\left(x_t\right)$ называется *слабо стационарным*, если он имеет постоянные математическое ожидание, дисперсию и автоковариацию (ковариацию, измеренную для различных значений одного и того же временного ряда), то есть для него выполняются условия:

$$1.\ M(y_t) = \eta$$

$$2.\ M\left[\left(y_t - \eta\right)\left(y_t - \eta\right)\right] = \sigma^2 \prec \infty$$

$$3.\ M\left[(y_{t1} - \eta)(y_{t2} - \eta)\right] = \gamma_{t2-t1}$$

Автоковариация определяется соотношением:

$$M\left[(y_t-\eta_t)(y_{t-s}-\eta_{t-s})\right],$$

где (y_t) — слабо стационарный процесс, $M(y_t)=\eta$, $\eta\prec\infty$ и $M\left[(y_t-\eta)^2\right]=\sigma_y{}^2\prec\infty$ для всех (t). Пусть $\overline{y}_t=\frac{1}{n}\sum_{t=1}^{n}y_t$ — средняя по времени. Если $(\overline{y}_t)$ сходится по вероятности к (η) при $n\to\infty$, то говорят, что ряд (y_t) обладает свойством *эргодичности* по отношению к средней.

Математическое ожидание и дисперсия стационарного временного ряда оцениваются по формулам:

$$\overline{y}_t=\frac{\sum_{t-1}^{n}y_t}{n};$$

$$s_t{}^2=\frac{\sum_{t=1}^{n}\left(y_t-\overline{y}_t\right)^2}{n}.$$

Простейшим примером стационарного временного ряда является «*белый шум»*. «Белый шум» — стационарный временной ряд, обладающий следующими свойствами:

1. $M(y_t)=\eta;$
2. $D(y_t)=\sigma^2;$
3. $\gamma_{t-\tau}=\begin{pmatrix}\sigma^2, t=r\\ 0, t\neq r\end{pmatrix}.$

Это значит, что «белый шум» имеет постоянное математическое ожидание и дисперсию, а случайные величины, соответствующие наблюдениям процесса «белого шума» в разные моменты времени, некоррелированы.

Степень тесноты связи между наблюдениями временного ряда $(Y_1,Y_2,\ldots,Y_n)$ и $(Y_{1+\tau},Y_{2+\tau},\ldots,Y_{n+\tau})$, то есть сдвинутых относительно друг друга лагом (τ), может быть определена с помощью *коэффициента корреляции* (*автокорреляции*):

$$p_r = \frac{M\left[(y_t - a)(y_{t+\tau} - a)\right]}{\sigma_x(t) \times \sigma_x(t+\tau)} = \frac{M\left[(y_t - a)(y_{t+\tau} - a)\right]}{\sigma^2},$$

$$M(y_t) = M(y_{t+\tau}) = a \quad ; \quad \sigma_y(t) = \sigma_y(t+\tau) = \sigma.$$

Зависимость называют *автокорреляционной функцией*.

Статистической оценкой (p_r) является *выборочный коэффициент автокорреляции* $(r(\tau))$, определяемый по формуле:

$$r(\tau) = \frac{(n-\tau)\sum_{t=1}^{n-\tau} y_t y_{t+\tau} - \sum_{t=1}^{n-\tau} y_t \sum_{t=1}^{n-\tau} y_{t+\tau}}{\sqrt{(n-\tau)\sum_{t=1}^{n-\tau} y^2{}_t - (\sum_{t=1}^{n-\tau} y_t)^2} \times \sqrt{(n-\tau)\sum_{t=1+\tau}^{n} y^2{}_{t+\tau} - (\sum_{t=1}^{n-\tau} y_{t+\tau})^2}}.$$

Функцию $(r(\tau))$ называют выборочной автокорреляционной функцией

При исследовании временных рядов рассматривают также *частную автокорреляционную функцию* $p_{част}(\tau)$, где $p_{част}(\tau)$ есть частный коэффициент корреляции между членами временного ряда (y_t), $y_{t+\tau}$, и ее статистическую оценку — *выборочную частную автокорреляционную функцию* $r_{част}(\tau)$, где $r_{част}(\tau)$ — *выборочный частный коэффициент корреляции*, определяемый по формуле

$$r = \frac{r_{ij} - r_{ik} \cdot r_{jk}}{\sqrt{\left(1 - r_{ik}{}^2\right)\left(1 - r_{jk}{}^2\right)}}.$$

При определении наличия автокорреляции между соседними членами временного ряда используется следующая связь коэффициента Дарбина — Уотсона с выборочным коэффициентом корреляции:

$$d = 2\ (1 - r).$$

12.2. Авторегрессионные модели и модели скользящей средней

Для анализа временных рядов используют следующие линейные модели:

1. Модель скользящей средней порядка (q) (модель $MA(q)$) — модель временного ряда, в которой моделируемая величина задается функцией от прошлых ошибок (e_t):

$$y_t = e_t + \gamma_1 e_{t-1} + ... + \gamma_q e_{t-q} .$$

2. Модель авторегрессии порядка (p) (модель $AR(p)$) — модель временного ряда, в которой текущее значение моделируемой переменной задается функцией от прошлых значений самой переменной:

$$y_t = \beta_0 + \beta_1 y_{t-1} + ... + \beta_p y_{t-p} + e_t .$$

3. Авторегрессионная модель скользящей средней порядков (p) и (q) (модель $ARMA(p, q)$) — модель, сочетающая авторегрессионные процессы с процессами скользящей средней:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \ldots + \beta_p y_{t-p} + e_t + \gamma_1 e_{t-1} + \ldots + \gamma_q e_{t-q} .$$

Систематический подход к построению ARMA-моделей предложен Боксом и Дженкинсом и включает три этапа:

1. *Идентификация* — определение порядка (p, q) модели, то есть строится ARMA-модель, в которой остатки (E_t) представляют собой «белый шум», а экзогенные переменные значимы.

С этой целью часто применяется *метод элементарного подбора*, то есть пробуются разные модели, начиная с самых простых, которые усложняются до момента идентификации.

Полезную информацию можно получить и при помощи *выборочных автокорреляционной и частной автокорреляционной функциях*: если все значения выборочной частной автокорреляционной функции порядка выше (p) незначимо отличаются от нуля, временной ряд следует идентифицировать с помощью модели, порядок авторегрессии которой не выше (p). Или: если все значения выборочной автокорреляционной функции порядка выше (q)

незначимо отличаются от нуля, временной ряд следует идентифицировать с помощью модели скользящей средней, порядок которой не выше (q).

Для идентификации ARMA-модели применяют также *информационные критерии*. Эти критерии имеет две компоненты: функцию от остаточной суммы квадратов и «наказание» за потерю степеней свободы в результате включения в модель дополнительных переменных. Целью исследования является выбор такого количества параметров, при котором величина информационных критериев будет минимальна.

Обычно используют следующие критерии:

а) Информационный критерий Акаике. Согласно нему, среди альтернативных значений (k) выбирается то, которое минимизирует величину:

$$AIC = \ln\left(\hat{\sigma}^2\right) + \frac{2k}{n},$$

где $\hat{\sigma}^2$ — остаточная дисперсия;

$k = p + q + 1$ — общее количество оцениваемых параметров;

n — размер выборки.

в) Байесовский информационный критерий Шварца:

$$SBIC = \ln\left(\hat{\sigma}^2\right) + \frac{k}{n}\ln\left(n\right).$$

c) Информационный критерий Хеннана — Куинна:

$$HQIC = \ln\left(\hat{\sigma}^2\right) + \frac{2k}{n}\ln(\ln(n)).$$

2. *Оценка параметров* — заключается в нахождении конкретных числовых параметров модели, сформированной на этапе идентификации.

Для оценки параметров АМА-модели применяют обобщенный метод наименьших квадратов (ОМНК), рассмотренный выше или метод максимального правдоподобия (ММП).

3. *Диагностика модели* — установление, является ли полученная модель адекватной для описания изучаемого явления.

Диагностика модели предполагает использование разных вариантов Q-статистики — инструмента, предназначенного для проверки гипотезы о том, что наблюдаемые данные являются реализацией «белого шума»:

а) Q-статистика Бокса — Пирса

$$Q = n\sum_{k=1}^{m} \hat{p}_k^{\;2} ,$$

где n — объем выборки;

m — максимальное количество лагов;

$\hat{p}_k$ — коэффициент автокорреляции.

При больших значениях выборки Q-статистика приближенно имеет распределение $\left(\chi_m^{\;2}\right)$, то есть:

$$Q \infty \chi_m^{\;2};$$

в случае верности нулевой гипотезы о том, что все (m) коэффициентов автокорреляции равны нулю. При уровне значимости, равным 0,05, гипотеза отвергается при выполнении неравенства

$$Q \succ \chi^2_{0,95(m)} .$$

б) Q -статистика Льюинга — Бокса

При малом объеме выборки Q-статистики Бокса — Пирса плохо приближается распределение $\chi_{\mathrm{m}}^{\;2}$.Поэтому для малых выборок лучше использовать другой вариант этого показателя, известный как статистика Льюинга — Бокса:

$$Q^* = n(n+2)\sum_{k=1}^{m} \frac{\hat{p}_k^{\;2}}{n-k} \infty \chi_m^{\;2} .$$

Значения Q-статистики Льюинга — Бокса интерпретируются аналогично статистики Бокса — Пирса.

12.3. Нестационарные временные ряды

Рассмотрим временной ряд:

$$y_t = \mu + \alpha\, y_{t-1} + e_t .$$

В зависимости от (α) имеем:

1. Если $|\alpha| \prec 1$, то временной ряд стационарный.

2. Если $|\alpha| \succ 1$, то временной ряд нестационарный.

При $|\alpha| \succ 1$, имеет место *взрывной* временной ряд, однако в реальных экономических задачах он никогда не возникает.

В экономической практике рассматривают два типа нестационарных временных рядов:

• Временной ряд со стохастическим трендом (случайное блуждание):

$$y_t = \mu + y_{t-1} + e_t .$$

Практика показывает, что чаще всего в экономических исследованиях имеет место временной ряд со стохастическим трендом.

• Временной ряд с детерминистическим трендом

$$y_t = \alpha + \beta_t + e_t$$

Таким образом *вопрос о нестационарности временного ряда* (Y_t) *сводится к следующему: верно ли, что в регрессии* $y_t = \mu + \alpha y_{t-1} + e_t$ *истинное значение параметра* (*a*) *равно единице.* Эта задача называется *проблемой единичного корня.*

Проблему единичного корня решают с помощью теста Дики — Фуллера. В рамках данного теста исследуется уравнение вида:

$$y_t = \mu + \alpha y_{t-1} + e_t .$$

Проводится проверка нулевой и альтернативной гипотез:

H_0: $\alpha = 1$ — уравнение содержит единичный корень;

H_1: $\alpha = 1$ — ряд стационарен.

На практике для упрощения вычислений оценивается модель вида:

$$\Delta y_t = \psi\, y_{t-1} + e_t .$$

Гипотеза $\alpha = 1$ эквивалентна гипотезе $\psi = 0$.

Тест Дики — Фуллера имеет три разновидности:

1. В качестве нулевой гипотезы рассматривается «случайное блуждание», а в качестве альтернативной — авторегрессионный процесс первого порядка *AR*(1) с нулевым постоянным членом

$$H_1 : y_t = \alpha y_{t-1} + e_t, \alpha \prec 1;$$
$$H_0 : y_t = y_{t-1} + e_t.$$

2. В качестве нулевой гипотезы рассматривается «случайное блуждание», а в качестве альтернативной — авторегрессионный процесс первого порядка $AR(1)$ с ненулевым постоянным членом:

$$H_0 :\ y_t = y_{t-1} + e_t;$$
$$H_1 = y_t = \alpha y_{t-1} + \mu + e_t, \quad \alpha \prec 1.$$

3. В качестве нулевой гипотезы рассматривается «случайное блуждание», а в качестве альтернативной — авторегрессионный процесс первого порядка $AR(1)$ с ненулевым постоянным членом и детерминистическим трендом

$$H_0 : y_t = y_{t-1} + e_t;$$
$$H_1 : y_t = \alpha y_{t-1} + \mu + \lambda t + e_t, \quad \alpha \prec 1.$$

Нулевую гипотезу формулируют в виде:

$$\Delta y_t = e_t,$$

где $\Delta \mathrm{y}_{\mathrm{t}} = y_t - y_{t-1}$.

Альтернативную гипотезу формулируют в виде:

$$\Delta y_t = \psi\, y_{t-1} + \mu + \lambda t + e_t,$$

где $\mu = \lambda = 0$ в первом случае;

$\lambda = 0$ во втором случае;

$\psi = \alpha - 1$ во всех трех случаях

По результатам оценки уравнения $\Delta y_t = \psi\, y_{t-1} + \mu + \lambda t + e_t$ исследуется статистика:

$$test\ statistic = \frac{\hat{\psi}}{SE(\hat{\psi})},$$

где $SE(\hat{\psi})$ стандартная ошибка коэффициента (ψ).

Критические значения статистики теста представлены в *табл. 12.1*.

Таблица 12.1

Критические значения для теста Дики — Фуллера

Уровень значимости	10 %	5 %	1 %
Критические значения для случая с константой, но без тренда	–2,57	–2,86	–3,43
Критические значения для случая с трендом и константой	–3,12	–3,41	–3,96

Нулевая гипотеза о наличии единичного корня отвергается, если фактическая величина статистики превышает по модулю ее критическое значение.

Способ приведения нестационарного временного ряда к стационарному зависит от типа нестационарного временного ряда.

А) Если временной ряд имеет детерминистический тренд, то для достижения стационарности ряда нужно удалить тренд из данных. После этого с ним можно работать методами, используемыми для стационарных рядов.

В) Если ряд имеет стохастический тренд, то к нему применяется метод взятия разностей.

Рассмотрим ряд вида

$$y_t = \mu + y_{t-1} + e_t .$$

В результате вычитания из обеих частей уравнения члена $\left(y_{t-1}\right)$ получаем:

$$y_t - y_{t-1} = \mu + e_t;$$

$$\left(1-t\right) y_t = \mu + 1;$$

$$\Delta y_t = \mu + 1.$$

Мы получили новый ряд $\left(\Delta y_t\right)$, который, в отличие от исходного $\left(y_t\right)$, является стационарным. Иногда изъятие разностей приходится применять несколько раз, прежде чем ряд становится стационарным. Если для получения стационарного процесса *ARMA* (p, q) к исходному ряду (Y_t) необходимо (d) раз применить взятие разностей, то такой процесс называется авторегрессионной интегрированной моделью скользящей средней — *ARIMA* (p, d, q), обладающей тремя параметрами: p — порядок регрессии, d — требуемый порядок предварительно определяемых разностей и q — порядок скользящей средней в модели.

Подробные вопросы, связанные с нестационарными временными рядами, изложены в [6].

Традиционная эконометрика оперирует в основном линейными моделями временных рядов, что связано с удобством их анализа. Однако очевидной становится ограниченная применимость линейных моделей для описания сложных экономических явлений. В рамках линейной модели каждое действие вызывает приблизительно пропорциональную реакцию. Однако более реалистична ситуация, когда на некоторое воздействие возникает экспоненциальная суперреакция. В этом и заключается сущность нелинейности. Особенно заметна нелинейность в финансовых временных рядах: периоды затишья, когда финансовые показатели лишь незначительно колеблются вокруг среднего, чередуются с периодами всплеска, характеризующимися широким размахом значений тех же показателей.

Среди *нелинейных моделей временного ряда* выделяют:

• *модель авторегрессионной условной гетероскедастичности* (*ARCH*)

Напомним, в классической модели линейной регрессии ошибки считаются гомоскедастичными, то есть имеют постоянные дисперсии. Однако это допущение далеко от реальности при анализе динамики доходности финансовых активов — дисперсия ошибок оказывается непостоянной и коррелирует с прошлыми значениями. Поэтому была разработана модель *ARCH*, которая учитывает эти тенденции. Однофакторная модель *ARCH* имеет вид:

$$\sigma_t^2 = \alpha_0 + \alpha_1 e^2_{t-1}$$
$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_p x_{pt} + e_t,$$

где σ_t^2 — условная дисперсия случайной переменной, обусловленная информацией о других случайных переменных.

Многофакторная модель $ARCH\,(q)$ соответствует случаю, когда условная дисперсия зависит от (q) лаговых значений квадратов ошибок:

$$\sigma_t^{\,2} = \alpha_0 + \alpha_1 e^2_{\;t-1} + \alpha_2 e^2_{\;t-2} + \ldots + \alpha_q e^2_{\;t-q}\,.$$

Тест на обнаружение возможных ARCH-эффектов проводится следующим образом:

1. Проводится оценка основного уравнения модели:

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \ldots + \beta_p x_{pt} + e_t\,.$$

2. На основе оценочных значений ошибок, полученных из данного проводится оценка модели вида:

$$\hat{e}_t^2 = \gamma_0 + \gamma_1 e^2_{\;t-1} + \gamma_2 e^2_{\;t-2} + \ldots + \gamma_q e^2_{\;t-q} + \nu_t\,,$$

где ν_t — остаточный член.

На основе данного уравнения рассчитывается коэффициент детерминации R^2.

3. Формулируется нулевая и альтернативная гипотеза:

$$H_0: \gamma_1 = 0 \quad \gamma_2 = 0 \quad \gamma_q = 0;$$
$$H_1: \gamma \neq 0 \ \text{ или } \ \gamma_2 \neq 0 \ \text{ или } \ \gamma_q \neq 0.$$

4. Рассчитывается статистика:

$$nR^2 \propto \chi_q^{\,2}\,,$$

где n — объем выборки.

Если значение статистики превосходит критическую величину распределения (χ^2), то нулевая гипотеза отвергается, то есть признается наличие ARCH-эффектов.

• *Модель обобщенной авторегрессионной условной гетероскедастичности* (*GARCH*)

Модель соответствует случаю, когда текущее значение условной дисперсии зависит от (q) лаговых значений квадратов ошибок и (p) собственных лаговых значений:

$$\sigma^2{}_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i e^2{}_{t-i} + \sum_{j=1}^{p} \beta_j \sigma^2{}_{t-j}$$

Контрольные вопросы к теме № 12

1. Что такое стационарные временные ряды.
2. Что понимают под строго стационарным и слабостационарным временным рядом.
3. Что такое «белый шум».
4. Что такое модели скользящей средней.
5. Что такое модели авторегрессии.
6. Что такое авторегрессионные модели скользящей средней.
7. Каким образом строятся модели ARMA.
8. Что такое информационные критерии, назовите наиболее популярные.
9. Каковы основные типы нестационарных временных рядов.
10. Каким образом можно привести нестационарные временные ряды к стационарным.
11. Каким образом определяется наличие единичного корня.
12. Чем объясняется необходимость в нелинейных моделях.
13. Что такое модели ARCH и GARCH.

Задача 12.1. По данным таблицы (см. задачу 11.1) для временного ряда (*Y*) найти среднее значение, среднее квадратическое отклонение, коэффициенты автокорреляции (для лагов = 1, 2) и частный коэффициент автокорреляции 1-го порядка.

Задача 12.2. В таблице представлена динамика курса акций корпорации «Омега».

Требуется построить и проверить на адекватность модель *ARCH*(1).

Таблица

Период	Курс	Период	Курс	Период	Курс	Период	Курс	Период	Курс	Период	Курс
1	532	9	548	17	513	25	547	33	630	41	560
2	539	10	537	18	507	26	568	34	634	42	630
3	548	11	548	19	510	27	578	35	667	43	650
4	546	12	544	20	526	28	578	36	680	44	620
5	564	13	534	21	543	29	581	37	696	45	603
6	571	14	542	22	542	30	633	38	675	46	613
7	570	15	521	23	521	31	600	39	650	47	640
8	566	16	509	24	509	32	601	40	604	48	680

ТЕМА № 13 СИСТЕМА ЛИНЕЙНЫХ ОДНОВРЕМЕННЫХ УРАВНЕНИЙ. КОСВЕННЫЙ, ДВУХШАГОВЫЙ, ТРЕХШАГОВЫЙ МЕТОД НАИМЕНЬШИХ КВАДРАТОВ

План

1. Система одновременных уравнений (СОУ).
2. Проблема идентифицируемости системы одновременных уравнений. Оценка системы одновременных уравнений.

13.1. Система одновременных уравнений (СОУ)

Системой одновременных уравнений (*СОУ*) называется набор взаимосвязанных регрессионных моделей, в которых одни и те же переменные в различных уравнениях системы могут *одновременно* играть роль *эндогенных* (результирующих) и *экзогенных* (объясняющих) переменных.

Приведем общий вид системы одновременных уравнений. Пусть $Y_1....Y_m$ — эндогенные переменные, $X_1.....X_l$ — экзогенные переменные. Введем блочные матрицы (В), и (Г):

$$B = \begin{pmatrix} \beta_{11} & \beta_{1m} \\ \beta_{m1} & \beta_{mm} \end{pmatrix} \quad \Gamma = \begin{pmatrix} \gamma_{11} & \gamma_{1l} \\ \gamma_{m1} & \gamma_{ml} \end{pmatrix}.$$

Тогда общий вид системы одновременных уравнений представляется в матричной форме как:

$$BY + \Gamma X = e$$

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_m \end{pmatrix} \quad X = \begin{pmatrix} X_1 \\ X_2 \\ X_l \end{pmatrix} \quad e = \begin{pmatrix} e_1 \\ e_2 \\ e_m \end{pmatrix}. \tag{13.1}$$

Мы ограничимся рассмотрением случая двух уравнений с двумя эндогенными переменными. Это не приведет ни к какой потери — все необходимые аспекты теории можно проследить на этом простом случае. В то же время такое ограничение позволит избежать излишней громоздкости в вычислениях. Уравнение (13.1) можно записать в виде:

$$Y_1 = \alpha_1 + \beta_1 X_1 + \gamma_1 Y_2 + e_1 . \tag{13.1}$$

$$Y_2 = \alpha_2 + \beta_2 X_2 + \gamma_2 Y_2 + e_2 . \tag{13.2}$$

Наборы переменных X_1 и X_2 могут быть произвольными. Если применить к уравнениям (13.2–13.3) обычный МНК, то получатся несостоятельные оценки параметров (α, β, γ). Таким образом, оценивание систем одновременных уравнений требует специальных методов.

13.2. Проблема идентифицируемости системы одновременных уравнений. Оценка системы одновременных уравнений

Наиболее распространенными инструментами анализа систем одновременных уравнений являются:

- косвенный метод наименьших квадратов;
- двухшаговый метод наименьших квадратов (2МНК);
- трехшаговый метод наименьших квадратов (3МНК).

Рассмотрим их.

- *Косвенный метод наименьших квадратов*

Решая уравнения (13.2–13.3) относительно Y_1, Y_2, запишем уравнение в виде:

$$\begin{aligned} Y_1 &= a_1 + b_1 X_1 + c_1 X_2 + \nu_1 \\ Y_2 &= a_2 + b_2 X_1 + c_2 X_2 + \nu_2 \end{aligned}, \tag{13.4}$$

где $a_1 = \dfrac{\alpha_1 + \gamma_1 \alpha_2}{1 - \gamma_1 \gamma_2}$; $a_2 = \dfrac{\alpha_2 + \gamma_2 \alpha_1}{1 - \gamma_1 \gamma_2}$; $b_1 = \dfrac{\beta_1}{1 - \gamma_1 \gamma_2}$; $b_2 = \dfrac{\gamma_2 \beta_1}{1 - \gamma_1 \gamma_2}$;

$c_1 = \dfrac{\gamma_1 \beta_2}{1 - \gamma_1 \gamma_2}$; $c_2 = \dfrac{\beta_2}{1 - \gamma_1 \gamma_2}$; $\nu_1 = \dfrac{\gamma_1 \varepsilon_2 + \varepsilon_1}{1 - \gamma_1 \gamma_2}$; $\nu_2 = \dfrac{\gamma_2 \varepsilon_1 + \varepsilon_2}{1 - \gamma_1 \gamma_2}$.

Применив к (13.4) МНК, получим оценки параметров (b, c):

$$\hat{b}_1 = \frac{(X_2X_2)(X_1Y_1)-(X_1X_2)(X_2Y_1)}{(X_1X_1)(X_2X_2)-(X_1X_2)^2};$$

$$\hat{c}_1 = \frac{(X_1X_1)(X_2Y_1)-(X_1X_2)(X_1Y_1)}{(X_1X_1)(X_2X_2)-(X_1X_2)^2};$$

$$\hat{b}_2 = \frac{\left(X_1X_1\right)\left(X_2Y_1\right)-\left(X_1X_2\right)\left(X_2Y_2\right)}{\left(X_1X_1\right)(X_2X_2)-(X_1X_2)^2}; \qquad (13.7)$$

$$c_2 = \frac{(X_1X_1)(X_2Y_2)-(X_1X_2)(X_1Y_2)}{(X_1X_1)(X_2X_2)-(X_1X_2)^2},$$

где $\left(X_iX_j\right)=\sum_{t=1}^{n}x_{ti}x_{tj}$; $\left(Y_iY_j\right)=\sum_{t=1}^{n}y_{ti}y_{tj}$; $\left(X_iY_j\right)=\sum_{t=1}^{n}x_{ti}y_{tj}$;
$x_{ti}, x_{tj}, y_y, y_{tj}$ — значения переменных X_i, X_j, Y_iY_j.

Можно выразить исходные параметры α, β, γ через a, b, c

$$\hat{\beta}_1 = \frac{\hat{b}_1\hat{c}_2-\hat{b}_2c_1}{\hat{c}_2}; \quad \hat{\beta}_2 = \frac{\hat{b}_1\hat{c}_2-\hat{b}_2\hat{c}_1}{\hat{b}_1};$$

$$\hat{\gamma}_1 = \frac{\hat{c}_2}{\hat{c}_1}; \qquad \hat{\gamma}_2 = \frac{\hat{b}_2}{\hat{b}_1}. \qquad (13.8)$$

Окончательно получим (13.9):

$$\hat{\beta}_1 = \frac{(X_1Y_1)(X_2Y_2)-(X_2Y_1)(X_1Y_2)}{(X_1X_1)(X_2Y_2)-(X_1X_2)(X_1Y_2)}; \quad \hat{\beta}_2 = \frac{\left(X_1Y_1\right)\left(X_2Y_2\right)-\left(X_2Y_1\right)\left(X_1Y_2\right)}{\left(X_2X_2\right)\left(X_1Y_1\right)-\left(X_1X_2\right)\left(X_2Y_1\right)};$$

$$\hat{\gamma}_1 = \frac{\left(X_1X_1\right)\left(X_2Y_1\right)-\left(X_1X_2\right)\left(X_1Y_1\right)}{\left(X_1X_1\right)\left(X_2Y_2\right)-\left(X_1X_2\right)\left(X_1Y_2\right)}; \quad \hat{\gamma}_2 = \frac{\left(X_2X_2\right)\left(X_1Y_2\right)-\left(X_1X_2\right)\left(X_2Y_2\right)}{\left(X_2X_2\right)\left(X_2Y_1\right)-\left(X_1X_2\right)\left(X_2Y_1\right)}.$$

Часто экономическая модель, описывающая какую — либо экономическую теорию, может быть выражена в виде системы уравнений, где каждое уравнение представляет собой некоторое соотношение между экзогенными, эндогенными переменными и параметрами. Такая система уравнений называется *структурной моделью*. Форма (13.1) будет являться структурной формой системы одновременных уравнений. Параметры структурной формы называются *структурными параметрами*. Именно структурная форма раскрывает экономический механизм формирования значений эндогенных переменных. Оценивать эту модель с помощью МНК нельзя, так как оценки коэффициентов получаются смещенными (в связи с тем, что экзогенные переменные коррелированны с ошибками). Форма (13.4) называется *приведенной формой* системы одновременных уравнений. Параметры ее оцениваются с помощью рассмотренного выше косвенного МНК.

Проблема *идентифицируемости* является центральной при работе с системами одновременных уравнений. Структурный параметр называется *идентифицируемым*, если он может быть однозначно оценен с помощью косвенного МНК. Уравнение идентифицируемо, если идентифицируемы все входящие в него структурные параметры. Структурный параметр называется *неидентифицируемым*, если его значение невозможно получить, даже зная точные значения параметров приведенной формы. Структурный параметр называется *сверхидентифицируемым*, если косвенный МНК дает несколько различных его оценок. Нужно отметить, что проблема сверхидентифицируемости это *проблема количества наблюдений*: с увеличением объема выборки все разные состоятельные оценки параметра стремятся к одному и тому же истинному значению. А проблема неидентифицируемости — это проблема структуры *модели.* Неидентифицируемость не исчезает с ростом количества наблюдений и означает, что существует бесконечное число структурных моделей, имеющих одну и туже приведенную форму.

• *Двухшаговый МНК*

По сути метод наименьших квадратов применяется дважды: сначала для получения набора экзогенных переменных (*X*), затем для получения оценок параметра (β). Процедура 2МНК следующая.

Необходимо подобрать новые переменные *Zj* (*j* = 1, …, n), которые бы тесно коррелировали с *Xj* и не коррелировали с (*e*) в уравнении:

$$Y = X\beta + e, \tag{13.10}$$

где X — неслучайная матрица;
e — случайный вектор.

Набор переменных Zj может включать те экзогенные переменные, которые не коррелируют с (e), а также другие величины. Такие переменные Zj называются *инструментальными*. Они позволяют получить состоятельную оценку параметра (β) модели (13.10). Такая оценка имеет вид:

$$\tilde{\beta}_{iv} = \left(Z'X\right)^{-1} Z'X = \left(\frac{1}{n} Z'X\right)^{-1} \left(\frac{1}{n} Z'Y\right). \qquad (13.11)$$

Процедура 2МНК реализована в большинстве компьютерных регрессионных пакетов.

• *Трехшаговый МНК*

Заключается в том, что на первом шаге к исходной модели (13.1) применяется обобщенный МНК с целью устранения корреляции случайных членов. Затем к полученным уравнениям применяется 2МНК.

На экономических курсах подробно изучаются следующие классические примеры систем одновременных уравнений:

1. Кейнсианская модель формирования доходов:

$$C_t = \alpha + \beta Y_t + e_t; \qquad (13.12)$$

$$Y_t = C_t + I_t, \qquad (13.13)$$

где Y, C, I — соответственно совокупный выпуск, объем потребления и инвестиций. I — экзогенная переменная, а Y — эндогенная переменная.

Модель идентифицируема, ее приведенная форма имеет вид:

$$Y = \frac{\alpha}{1-\beta} + \frac{1}{1-\beta} I + \frac{1}{1-\beta} \varepsilon \qquad (13.14)$$

2. Модель формирования спроса и предложения

$$\begin{aligned} Q_d &= \beta_1 + \beta_2 P + \beta_3 I + \varepsilon_1 \\ Qs &= \beta_4 + \beta_5 P + \varepsilon_2, \end{aligned} \qquad (13.15)$$

где I — доход потребителя. P — цена товара. Q_d — спрос. Q_s — предложение. P и Q — эндогенные переменные, а I — экзогенная переменная.

Контрольные вопросы к теме № 13

1. Что такое системы одновременных уравнений.
2. Что такое структурная и приведенная форма моделей.
3. Что такое идентифицируемость.
4. Каким образом производится оценка систем одновременных уравнений.

Задача 13.1. Имеется макроэкономическая модель:

$$C_t = \gamma_1 e_t + \beta_1 + e_{1t}\,; \qquad (13.1.1)$$

$$I_t = \gamma_2 y_t + \beta_2 y_{t-1} + \beta_3 + e_{2t}\,; \qquad (13.1.2)$$

$$Y_t = C_t + I_t G_t\,, \qquad (13.1.3)$$

где C_t — потребление, I_t — инвестиции, G_t — государственные расходы; Y_t — валовый национальный продукт в период t.

Требуется: 1. Определить типы уравнений и типы переменных, входящих в модель (13.1.1–3.1.3).

2. Представить структурные уравнения в матричной форме.

ЛИТЕРАТУРА

А) ОСНОВНАЯ ЛИТЕРАТУРА

1. *Путко, Б. А.* Эконометрика: учебник / Б. А. Путко, Н. Ш. Кремер; ред. Н. Ш. Кремер. — 3-е изд., перераб. и доп. — Москва: Юнити-Дана, 2012. — 329 с. — (Золотой фонд российских учебников). — ISBN 978-5-238-01720-4; То же [Электронный ресурс]. — URL: http://biblioclub.ru/index.php?page=book&id=118251.

2. Эконометрика: учебник / В. Н. Афанасьев, Т. В. Леушина, Т. Лебедева, А. П. Цыпин; под ред. В. Н. Афанасьева; Министерство образования и науки Российской Федерации, Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Оренбургский государственный университет». — Оренбург: Оренбургский государственный университет, 2012. — 402 с.: табл., схем. — Библиогр.: с. 376–380.; То же [Электронный ресурс]. — URL: http://biblioclub.ru/index.php?page=book&id=260747.

Б) ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

3. *Грибанова, Е. Б.* Эконометрика: учебное пособие / Е. Б. Грибанова; Министерство образования и науки Российской Федерации, Томский Государственный Университет Систем Управления и Радиоэлектроники (ТУСУР). — Томск: ТУСУР, 2014. — 156 с.: схем. — Библиогр.: с. 132; То же [Электронный ресурс]. — URL: http://biblioclub.ru/index.php?page=book&id=480470.

4. *Шиловская, Н. А.* Эконометрика: учебное пособие / Н. А. Шиловская; Министерство образования и науки Российской Федерации, Северный (Арктический) федеральный университет имени М. В. Ломоносова. — Архангельск: САФУ, 2013. — 239 с.: схем., табл. — Библиогр. в кн. — ISBN 978-5-261-00743-2; То же [Электронный ресурс]. — URL: http://biblioclub.ru/index.php?page=book&id=436404.

5. Эконометрика для бакалавров: учебник / В. Н. Афанасьев, Т. В. Леушина, Т. В. Лебедева, А. П. Цыпин; под ред. В. Н. Афанасьева; Министерство образования и науки Российской Федерации. — Издание третье, переработанное и дополненное. — Оренбург: ООО ИПК «Университет», 2014. — 434 с.: схем., табл. — Библиогр.:

с. 406–412; То же [Электронный ресурс]. — URL: http://biblioclub.ru/index.php?page=book&id=330491.

6. Прикладная эконометрика: научно-практический журнал / гл. ред. С. А. Айвазян — Москва: 4. Синергия ПРЕСС, 2014. — № 1(33). — 145 с.: схем., табл., ил. — Библиогр. в кн. — ISSN 1993-7601; То же [Электронный ресурс]. — URL: http://biblioclub.ru/index.php?page=book&id=429916.

В) ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ

7. Пакет прикладных программ Microsoft Office 2016 Professional Plus.

Г) БАЗЫ ДАННЫХ, ИНФОРМАЦИОННО-СПРАВОЧНЫЕ И ПОИСКОВЫЕ СИСТЕМЫ

8. Компьютерная справочная правовая система «Консультант Плюс» http://www.consultant.ru/.

9. Справочная правовая система «ГАРАНТ-Аналитик».

10. Электронная библиотечная система «Университетская библиотека онлайн» (базовая часть) // http://www.biblioclub.r.

11. Электронная библиотека «e-LIBRARY.RU» // http://elibrary.ru.

ТЕСТЫ ДЛЯ ИТОГОВОГО КОНТРОЛЯ

ВАРИАНТ 1

1. Эконометрическая модель имеет вид
   а) $\hat{y} = f(x)$
   б) $\hat{y} = a + b_1x + b_2x^2$
   в) $y = f(x) + \varepsilon$
   г) $y = f(x)$

2. Уравнение линейной множественной регрессии
   а) $\hat{y} = a + bx$
   б) $\hat{y} = a + b_1x_1 + b_2x_2 + \cdots + b_px_p$
   в) $\hat{y} = ax_1^{b_1}x_2^{b_2}\ldots x_p^{b_p}$
   г) $y_t = T_t + S_t + E_t$

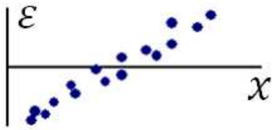
3. Верные утверждения о включении в уравнение линейной множественной регрессии факторов (несколько вариантов ответа)
   а) включение фактора в модель приводит к заметному возрастанию коэффициента множественной детерминации;
   б) коэффициент парной корреляции для фактора и результативной переменной меньше 0,3;
   в) значение *t*-критерия Стьюдента для коэффициента регрессии при факторе меньше табличного значения;
   г) фактор должен объяснять поведение изучаемого показателя согласно принятым положениям экономической теории.

4. Уравнение множественной регрессии в естественной форме имеет вид $y = 20 + 0{,}7x_1 + 0{,}5x_2 + \varepsilon$. На результативный признак оказывает большое влияние:
   а) x_1
   б) x_1 и x_2
   в) x_2
   г) нельзя сделать вывод

5. Верные утверждения относительно коэффициента множественной корреляции (несколько вариантов ответа)
   а) Чем ближе значение к единице $R_{yx_1...x_p}$, тем теснее связь результативного признака со всеми факторами
   б) Чем ближе значение к нулю $R_{yx_1...x_p}$, тем теснее связь результативного признака со всеми факторами
   в) $R_{yx_1...x_p}$ принимает значения из промежутка [0, 1]
   г) $R_{yx_1...x_p}$ принимает значения из промежутка [– 1, 1]

6. Оценка статистической значимости уравнения линейной множественной регрессии в целом осуществляется с помощью
   а) критерия Стьюдента
   б) критерия Фишера
   в) критерия Дарбина — Уотсона
   г) критерия Фостера — Стюарта

7. Укажите выводы, которые соответствуют графику зависимости остатков (несколько вариантов ответа)



   а) нарушена предпосылка МНК о независимости остатков друг от друга
   б) имеет место автокорреляция остатков
   в) отсутствует закономерность в поведении остатков
   г) отсутствует автокорреляция остатков

8. Регрессии, нелинейные по объясняющим переменным, но линейные по оцениваемым параметрам (несколько вариантов ответа)
   а) $y = a + b_1x + b_2x^2 + \varepsilon$
   б) $y = a \cdot x^b \cdot \varepsilon$
   в) $y = a + \frac{b}{x} + \varepsilon$
   г) $y = a + bx + \varepsilon$
   д) $y = a \cdot b^x \cdot \varepsilon$
   е) $y = e^{a+bx} \cdot \varepsilon$

9. Модель $y = a \cdot x^b \cdot \varepsilon$ относится к классу … эконометрических моделей нелинейной регрессии
   а) степенных
   б) обратных
   в) показательных
   г) линейных

10. Под изменением, определяющим общее направление развития, основную тенденцию временного ряда, понимается …
   а) тренд
   б) сезонная компонента
   в) циклическая компонента
   г) случайная компонента

11. Построена аддитивная модель временного ряда, где Y_t — временной ряд, T_t — трендовая компонента, S_t — сезонная компонента, E_t — случайная компонента. Если $Y_t = 15$, то правильно найдены значения компонент ряда …
   а) $T_t = 8,\ S_t = 5,\ E_t = 0$
   б) $T_t = 8,\ S_t = 5,\ E_t = 2$
   в) $T_t = 15,\ S_t = 5,\ E_t = 0$
   г) $T_t = 15,\ S_t = -5,\ E_t = 2$

12. В результате сглаживания временного ряда 6, 2, 7, 5, 12 простой четырехчленной скользящей средней первое сглаженное значение равно …

13. Известны значения коэффициентов автокорреляции $r_1 = 0{,}8$, $r_2 = 0{,}2$, $r_3 = 0{,}3$, $r_4 = 0{,}9$. Укажите верные утверждения…(несколько вариантов ответа)
   а) временной ряд содержит линейный тренд
   б) временной ряд содержит тренд в виде полинома 4 порядка
   в) временной ряд содержит циклические колебания с периодом 2
   г) временной ряд содержит циклические колебания с периодом 4

14. Для экспоненциального сглаживания используется формула
   а) $S_t = \alpha y_t + (1-\alpha) y_{t-1}$
   б) $S_t = \alpha y_t + (1-\alpha) S_{t-1}$
   в) $y_t = k + a \cdot b^t, a < 0, b < 1$
   г) $Y_t = T_t + S_t + E_t$

15. Сельскохозяйственное предприятие занимается выращиванием пшеницы, кукурузы, ячменя, гречихи. Построена эконометрическая модель, описывающая урожайность каждой культуры в зависимости от вносимых доз удобрений и количества влаги. Эта модель принадлежит к классу систем … уравнений
   а) одновременных
   б) независимых
   в) рекурсивных
   г) нормальных

16. В системе одновременных уравнений экзогенными переменными являются (несколько вариантов ответа)

$$\begin{cases} y_1 = b_{12} y_2 + a_{11} x_1 + \varepsilon_1, \\ y_2 = b_{21} y_1 + a_{22} x_2 + \varepsilon_2 \end{cases}$$

   а) x_1
   б) x_2
   в) y_1
   г) y_2
   д) ε_1
   е) ε_2

17. Приведенная форма модели, соответствующая структурной форме системы одновременных уравнений (несколько вариантов ответа)

$$\begin{cases} y_1 = b_{12} y_2 + a_{11} x_1 + \varepsilon_1, \\ y_2 = b_{21} y_1 + a_{22} x_2 + \varepsilon_2 \end{cases}$$

включает в себя уравнения
   а) $y_1 = a_{11} x_1 + \varepsilon_1$
   б) $y_2 = a_{22} x_2 + \varepsilon_2$
   в) $y_1 = \delta_{11} x_1 + u_1$

г) $y_2 = \delta_{22}x_2 + u_2$
д) $y_1 = \delta_{11}x_1 + \delta_{12}x_2 + u_1$
е) $y_2 = \delta_{21}x_1 + \delta_{22}x_2 + u_2$

18. Используя необходимое условие идентификации для модели динамики цены и заработной платы, укажите верные утверждения …(несколько вариантов ответа)

$$\begin{cases} y_1 = b_{12}y_2 + a_{11}x_1 + \varepsilon_1, \\ y_2 = b_{21}y_1 + a_{22}x_2 + a_{23}x_3 + \varepsilon_2, \end{cases}$$

где y_1 — темп изменения месячной зарплаты,
y_2 — темп изменения цен,
x_1 — процент безработных,
x_2 — темп изменения постоянного капитала,
x_3 — темп изменения цен на импорт сырья

а) оба уравнения являются точно идентифицируемыми
б) оба уравнения являются не идентифицируемыми
в) оба уравнения являются сверх идентифицируемыми
г) первое уравнение является сверх идентифицируемым
д) второе уравнение являются точно идентифицируемым

19. Оценки, полученные на экзамене: 5 — отлично; 4 — хорошо; 3 — удовлетворительно; 2 — неудовлетворительно. Данный показатель измерен в …
а) шкале наименований
б) порядковой шкале
в) интервальной шкале
г) шкале отношений

ВАРИАНТ 2

1. Установите соответствие

а) регрессионная модель

б) система одновременных уравнений

в) модель временного ряда

1) $x - 1 = \begin{cases} 0, x = 0 \\ x - 1, x > 0 \end{cases}$

2) $\begin{cases} R = a_1 + b_{11}M + b_{12}Y + \varepsilon_1, \\ Y = a_2 + b_{21}R + \varepsilon_2, \end{cases}$

3) $y = a + b_1x_1 + b_2x_2 + \varepsilon$

4) $y_t = T_t + S_t + E_t$

2. Для линейного уравнения множественной регрессии установите соответствие

$y = a + b_1x_1 + b_2x_2 + \varepsilon$

а) Факторные переменные

б) Результативная переменная

в) Параметры

г) Случайная компонента

1) y

2) a

3) a, ε

4) x_1, x_2

5) ε

6) a, b_1, b_2

3. Верные утверждения о включении в уравнение линейной множественной регрессии факторов

а) включение фактора в модель приводит к заметному возрастанию коэффициента множественной детерминации;

б) коэффициент парной корреляции для фактора и результативной переменной меньше 0,3;

в) значение t-критерия Стьюдента для коэффициента регрессии при факторе меньше табличного значения;

г) фактор должен объяснять поведение изучаемого показателя согласно принятым положениям экономической теории.

4. Уравнение множественной регрессии в стандартизованном масштабе имеет вид $t_y = 20 + 0{,}9t_{x_1} + 0{,}5t_{x_2} + \varepsilon$. На результативный признак оказывает большое влияние:

а) x_1

б) x_1 и x_2

в) x_2
г) нельзя сделать вывод

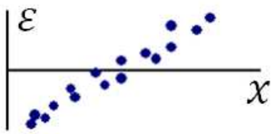
5. Установите соответствие

а) общая сумма квадратов отклонений TSS	1) $\sum(y-\bar{y})^2$
б) регрессионная сумма квадратов отклонений RSS	2) $\sum(y-\bar{x})^2$
в) остаточная сумма квадратов отклонений ESS	3) $\sum(y-\hat{y})^2$
	4) $\sum(\hat{y}-\bar{y})^2$

6. Для общей (TSS), регрессионной (RSS) и остаточной (ESS) суммы квадратов отклонений и коэффициента детерминации R^2 выполняется равенство …
   а) $R^2 = \frac{RSS}{TSS}$
   б) $R^2 = 1 - \frac{ESS}{TSS}$
   в) $R^2 = \frac{ESS}{TSS}$
   г) $R^2 = 1 - \frac{RSS}{TSS}$
   д) $R^2 = \frac{RSS}{TSS} + \frac{ESS}{TSS}$

7. Оценка статистической значимости коэффициентов линейной множественной регрессии осуществляется с помощью
   а) критерия Стьюдента
   б) критерия Фишера
   в) критерия Дарбина — Уотсона
   г) критерия Фостера — Стюарта

8. Укажите выводы, которые соответствуют графику зависимости остатков



а) нарушена предпосылка МНК о независимости остатков друг от друга
б) имеет место автокорреляция остатков
в) отсутствует закономерность в поведении остатков
г) отсутствует автокорреляция остатков

9. Для отражения влияния качественной сопутствующей переменной, имеющей m состояний, обычно включают в модель … фиктивную переменную
а) $m+1$
б) $(m+1)^2$
в) $m-1$
г) $(m-1)^2$

10. Укажите верные утверждения по поводу модели
$y = a + \frac{b}{x} + \varepsilon$
а) линеаризуется линейную модель множественной регрессии
б) линеаризуется линейную модель парной регрессии
в) относится к классу нелинейных моделей по объясняющим переменным, но линейных по оцениваемым параметрам
г) относится к классу линейных моделей

11. Было замечено, что при увеличении количества вносимых удобрений урожайность также возрастает, однако, по достижении определенного значения фактора моделируемый показатель начинает убывать. Для исследования данной зависимости можно использовать спецификацию уравнения регрессии…
а) $y = a + bx + cx^2 + \varepsilon$
б) $y = a + b_1x_1 + b_2x_2 + \varepsilon$
в) $y = a + \frac{b}{x} + \varepsilon$
г) $y = a + x^b + \varepsilon$

12. Регулярными компонентами временного ряда являются
а) тренд
б) сезонная компонента

в) циклическая компонента
г) случайная компонента

13. Построена аддитивная модель временного ряда, где Y_t — временной ряд, T_t — трендовая компонента, S_t — сезонная компонента, E_t — случайная компонента. Если $Y_t = 15$, то правильно найдены значения компонент ряда …

а) $T_t = 8,\ S_t = 5,\ E_t = 0$

б) $T_t = 8,\ S_t = 5,\ E_t = 2$

в) $T_t = 15,\ S_t = 5,\ E_t = 0$

г) $T_t = 15,\ S_t = -5,\ E_t = 2$

14. В результате сглаживания временного ряда 6, 2, 7, 5, 12 простой трехчленной скользящей средней первое сглаженное значение равно …
Ответ: 5

15. Автокорреляционная функция …
а) зависимость коэффициента автокорреляции от первых разностей уровней временного ряда
б) зависимость уровня временного ряда от коэффициента корреляции с его номером
в) последовательность коэффициентов автокорреляции, расположенных по возрастанию их порядка
г) последовательность коэффициентов автокорреляции, расположенных по возрастанию их значений

16. Модель временного ряда считается адекватной, если значения остатков …
а) имеют нулевое математическое ожидание
б) значение фактическое значение F-критерия меньше табличного
в) подчиняются нормальному закону распределения
г) подчиняются равномерному закону распределения
д) положительны
е) являются случайными и независимыми

17. Постоянная сглаживания α в модели экспоненциального сглаживания $S_t = \alpha y_t + (1 - \alpha)S_{t-1}$ принимает значения
   а) 0,2 или 0,3
   б) от 0,7 до 0,9
   в) [0;1]
   г) произвольные

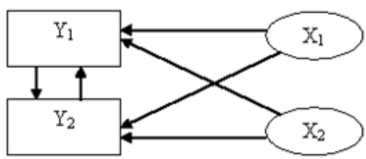
18. Сельскохозяйственное предприятие занимается выращиванием пшеницы, кукурузы, ячменя, гречихи. Построена эконометрическая модель, описывающая урожайность каждой культуры в зависимости от вносимых доз удобрений и количества влаги. Эта модель принадлежит к классу систем … уравнений
   а) одновременных
   б) независимых
   в) рекурсивных
   г) нормальных

19. В системе одновременных уравнений эндогенными переменными являются

$$\begin{cases} y_1 = b_{12}y_2 + a_{11}x_1 + \varepsilon_1, \\ y_2 = b_{21}y_1 + a_{22}x_2 + \varepsilon_2 \end{cases}$$

   а) x_1
   б) x_2
   в) y_1
   г) y_2
   д) ε_1
   е) ε_2

20. Уравнения, которые необходимо включить в систему для указанной схемы взаимосвязей между переменными



   а) $Y_1 = b_{12}Y_2 + a_{11}X_1 + a_{12}X_2 + \varepsilon_1$
   б) $Y_2 = b_{21}Y_1 + a_{21}X_1 + a_{22}X_2 + \varepsilon_2$

в) $Y_1 = a_{11}X_1 + a_{12}X_2 + \varepsilon_1$

г) $Y_2 = a_{21}X_1 + a_{22}X_2 + \varepsilon_2$

д) $Y_1 = b_{12}Y_2 + a_{11}X_1 + \varepsilon_1$

е) $Y_2 = b_{21}Y_1 + a_{21}X_1 + \varepsilon_2$

21. Приведенная форма для модели динамики цены и заработной платы

$$\begin{cases} y_1 = b_{12}y_2 + a_{11}x_1 + \varepsilon_1, \\ y_2 = b_{21}y_1 + a_{22}x_2 + a_{23}x_3 + \varepsilon_2, \end{cases}$$

где y_1 — темп изменения месячной зарплаты,
y_2 — темп изменения цен,
x_1 — процент безработных,
x_2 — темп изменения постоянного капитала,
x_3 — темп изменения цен на импорт сырья,
имеет вид …

а) $\begin{cases} y_1 = \delta_{11}x_1 + \varepsilon_1, \\ y_2 = \delta_{22}x_2 + \delta_{23}x_3 + \varepsilon_2 \end{cases}$

б) $\begin{cases} y_1 = \delta_{12}y_2 + \delta_{11}x_1 + \varepsilon_1, \\ y_2 = \delta_{21}y_1 + \delta_{22}x_2 + \delta_{23}x_3 + \varepsilon_2 \end{cases}$

в) $\begin{cases} y_1 = \delta_{12}y_2 + \varepsilon_1, \\ y_2 = \delta_{21}y_1 + \varepsilon_2 \end{cases}$

г) $\begin{cases} y_1 = \delta_{11}x_1 + \delta_{12}x_2 + \delta_{13}x_3 + \varepsilon_1, \\ y_2 = \delta_{21}x_1 + \delta_{22}x_2 + \delta_{23}x_3 + \varepsilon_2 \end{cases}$

22. Пусть *D* — число экзогенных переменных, которые содержатся в системе, но не содержатся в данном уравнении. Для первого уравнения модели динамики цены и заработной платы значение *D* равно …

$$\begin{cases} y_1 = b_{12}y_2 + a_{11}x_1 + \varepsilon_1, \\ y_2 = b_{21}y_1 + a_{22}x_2 + a_{23}x_3 + \varepsilon_2, \end{cases}$$

23. Установите соответствие для счетного правила необходимого условия идентификации, если *H* — число эндогенных переменных в системе, *D* — число экзогенных переменных, которые содержатся в системе, но не содержатся в данном уравнении

а) уравнение не идентифицируемо	1) D+1<H
б) уравнение сверх идентифицируемо	2) D+1=H
	3) D+1>H

24. Семейное положение выражается следующими категориями: 1 — холост; 2- женат; 3 — вдовец; 4 — разведен. Данный показатель измерен в …
    а) шкале наименований
    б) порядковой шкале
    в) интервальной шкале
    г) шкале отношений

ВАРИАНТ 3

1. Регрессия — это
   а) зависимость значений результативной переменной от значений объясняющих переменных (факторов)
   б) правило, согласно которому каждому значению одной переменной ставится в соответствие единственное значение другой переменной
   в) правило, согласно которому каждому значению независимой переменной ставится в соответствие значение зависимой переменной
   г) зависимость среднего значения результативной переменной от значений объясняющих переменных (факторов)

2. Проблема спецификации регрессионной модели включает в себя … (несколько вариантов ответа)
   а) отбор факторов, включаемых в уравнение регрессии
   б) оценка параметров уравнения регрессии
   в) оценка надежности результатов регрессионного анализа
   г) выбор вида уравнения регрессии

3. При построении модели множественной регрессии методом пошагового включения переменных на первом этапе рассматривается модель с …
   а) одной объясняющей переменной, которая имеет с зависимой переменной наименьший коэффициент корреляции
   б) одной объясняющей переменной, которая имеет с зависимой переменной наибольший коэффициент корреляции
   в) несколькими объясняющими переменными, которые имеют с зависимой переменной коэффициенты корреляции по модулю больше 0,5
   г) полным перечнем объясняющих переменных

4. Уравнение множественной регрессии в естественной форме имеет вид $y = 20 + 0{,}7x_1 + 0{,}5x_2 + \varepsilon$. На результативный признак оказывает большое влияние:

а) x_1
б) x_1 и x_2
в) x_2
г) нельзя сделать вывод

5. Коэффициент множественной корреляции для линейной зависимости можно рассчитать по формуле…(несколько вариантов ответа)

а) $R_{yx_1\ldots x_p} = \sqrt{\sum \beta_i r_{yx_i}}$

б) $R_{yx_1\ldots x_p} = \sum \beta_i r_{yx_i}$

в) $r = \dfrac{\overline{x \cdot y} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y}$

г) $R_{yx_1\ldots x_p} = \sqrt{1 - \frac{\sum(y-\hat{y})^2}{\sum(y-\bar{y})^2}}$

6. Отношение остаточной дисперсии к общей дисперсии равно 0,05. Это означает … (несколько вариантов ответа)
   а) коэффициент детерминации $R^2 = 0{,}95$
   б) коэффициент детерминации $R^2 = 0{,}05$
   в) разность $(1 - R^2) = 0{,}95$, где R^2 — коэффициент детерминации
   г) разность $(1 - R^2) = 0{,}05$, где R^2 — коэффициент детерминации

7. Если коэффициент регрессии является существенным, то для него выполняются условия…(несколько вариантов ответа)
   а) фактическое значение *t*-критерия Стьюдента меньше критического
   б) фактическое значение *t*-критерия Стьюдента больше критического
   в) доверительный интервал проходит через ноль
   г) стандартная ошибка не превышает половины значения параметра

8. При выполнении предпосылок метода наименьших квадратов (МНК) остатки уравнения регрессии, как правило, характеризуются … (несколько вариантов ответа)
   а) нулевой средней величиной
   б) гетероскедстичностью
   в) случайным характером
   г) высокой степенью автокорреляции

9. Регрессии, нелинейные по объясняющим переменным, но линейные по оцениваемым параметрам…(несколько вариантов ответа)
   а) $y = a + b_1 x + b_2 x^2 + \varepsilon$
   б) $y = a \cdot x^b \cdot \varepsilon$
   в) $y = a + \frac{b}{x} + \varepsilon$
   г) $y = a + bx + \varepsilon$
   д) $y = a \cdot b^x \cdot \varepsilon$
   е) $y = e^{a+bx} \cdot \varepsilon$

10. Модель $y = a \cdot b^x \cdot \varepsilon$ относится к классу … эконометрических моделей нелинейной регрессии
   а) степенных
   б) обратных
   в) показательных
   г) линейных

11. Для получения оценок параметров степенной регрессионной модели $\hat{y} = a \cdot x^b$ …
   а) метод наименьших квадратов неприменим
   б) требуется подобрать соответствующую подстановку
   в) необходимо выполнить логарифмическое преобразование
   г) необходимо выполнить тригонометрическое преобразование

12. Если период циклических колебаний уровней временного ряда не превышает одного года, то их называют …
   а) годичными
   б) конъюнктурными
   в) сезонными
   г) многолетними

13. Определить наличие тренда во временном ряду можно …
    а) по графику временного ряда
    б) по объему временного ряда
    в) по отсутствию случайной компоненты
    г) с помощью статистической проверки гипотезы о существовании тренда

14. В результате сглаживания временного ряда 6, 2, 7, 5, 12 простой четырехчленной скользящей средней первое сглаженное значение равно …

15. Если наиболее высоким оказался коэффициент автокорреляции 4 порядка, то временной ряд имеет
    а) линейный тренд
    б) случайную компоненту
    в) тренд в виде полинома 4 порядка
    г) циклические колебания с периодом 4

16. Независимость остатков модели временного ряда может быть проверена с помощью
    а) критерия Дарбина — Уотсона
    б) критерия Пирсона
    в) критерия Фишера
    г) анализа автокорреляционной функции остатков

17. Выбор оптимального значения постоянной сглаживания α в модели экспоненциального сглаживания $S_t = \alpha y_t + (1-\alpha)S_{t-1}$ осуществляется
    а) всегда используется значение $\alpha = 0{,}3$
    б) всегда используется значение $\alpha = 0{,}7$
    в) оптимальным считается такое значение α, при котором получена наименьшая дисперсия ошибки
    г) оптимальным считается такое значение α, при котором получена наибольшая дисперсия ошибки

18. Состояние закрытой экономики описывается следующими характеристиками: *Y* — валовой внутренний продукт (ВВП), *C* — уровень потребления, *I* — величина инвестиций, *G* — государственные расходы, *T* — величина налогов, *R* — реальная ставка процента. Спецификация модели

основана на следующих положениях экономической теории: 1) потребление объясняется величиной располагаемого дохода ($Y — T$); 2) уровень инвестиций определяется величиной ВВП и ставкой процента; 3) потребление, инвестиции и государственные расходы в сумме равны ВВП. Соответствующая система взаимосвязанных уравнений будет иметь вид:

а) $\begin{cases} C = a_0 + a_1 \cdot Y + \varepsilon_1, \\ I = b_0 + b_1 \cdot Y + b_2 \cdot R + \varepsilon_2, \\ Y = C + I + G \end{cases}$

б) $\begin{cases} C = a_0 + a_1 \cdot (Y - T) + \varepsilon_1, \\ I = b_0 + b_1 \cdot Y + \varepsilon_2, \\ Y = C + I + G \end{cases}$

в) $\begin{cases} C = a_0 + a_1 \cdot (Y - T) + \varepsilon_1, \\ I = b_0 + b_1 \cdot Y + b_2 \cdot R + \varepsilon_2, \\ Y = c_0 + c_1 \cdot C + c_2 \cdot I + c_3 \cdot G + \varepsilon_3 \end{cases}$

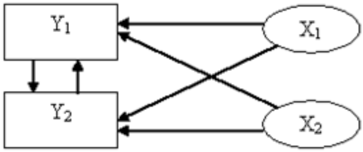
г) $\begin{cases} C = a_0 + a_1 \cdot (Y - T) + \varepsilon_1, \\ I = b_0 + b_1 \cdot Y + b_2 \cdot R + \varepsilon_2, \\ Y = C + I + G \end{cases}$

19. В системе одновременных уравнений экзогенными переменными являются…(несколько вариантов ответа)

$$\begin{cases} y_1 = b_{12} y_2 + a_{11} x_1 + \varepsilon_1, \\ y_2 = b_{21} y_1 + a_{22} x_2 + \varepsilon_2 \end{cases}$$

а) x_1
б) x_2
в) y_1
г) y_2
д) ε_1
е) ε_2

20. Уравнения, которые необходимо включить в систему для указанной схемы взаимосвязей между переменными (несколько вариантов ответа)



а) $Y_1 = b_{12}Y_2 + a_{11}X_1 + a_{12}X_2 + \varepsilon_1$

б) $Y_2 = b_{21}Y_1 + a_{21}X_1 + a_{22}X_2 + \varepsilon_2$

в) $Y_1 = a_{11}X_1 + a_{12}X_2 + \varepsilon_1$

г) $Y_2 = a_{21}X_1 + a_{22}X_2 + \varepsilon_2$

д) $Y_1 = b_{12}Y_2 + a_{11}X_1 + \varepsilon_1$

е) $Y_2 = b_{21}Y_1 + a_{21}X_1 + \varepsilon_2$

21. Единственность соответствия между приведенной и структурной формами модели системы одновременных уравнений составляет проблему …
    а) мультиколлинеарности факторов
    б) идентификации
    в) гетероскедастичности остатков
    г) неоднородности данных

22. Пусть D — число экзогенных переменных, которые содержатся в системе, но не содержатся в данном уравнении. Для второго уравнения модели динамики цены и заработной платы значение D равно …

$$\begin{cases} y_1 = b_{12}y_2 + a_{11}x_1 + \varepsilon_1, \\ y_2 = b_{21}y_1 + a_{22}x_2 + a_{23}x_3 + \varepsilon_2, \end{cases}$$

23. Обычный МНК успешно применяется для оценки структурных коэффициентов … (несколько вариантов ответа)
    а) систем неидентифицируемых уравнений
    б) систем рекурсивных уравнений (треугольных моделей)
    в) систем взаимосвязанных или одновременных уравнений
    г) систем уравнений-тождеств
    д) систем независимых уравнений

24. Оценки, полученные на экзамене: 5 — отлично; 4 — хорошо; 3 — удовлетворительно; 2 — неудовлетворительно. Данный показатель измерен в …
    а) шкале наименований
    б) порядковой шкале
    в) интервальной шкале
    г) шкале отношений

ВАРИАНТ 4

1. Метод наименьших квадратов …
а) позволяет получить оценки параметров линейной регрессии, исходя из условия $\sum_{i=1}^{n}\left(y_i - \widehat{y_i}\right)^2 \to min$
б) позволяет получить оценки параметров регрессии, исходя из условия $\ln\left(\prod_{i=1}^{n} f(y_i,\right) \to max$
в) позволяет проверить статистическую значимость параметров регрессии

2. Позволяет получить оценки параметров нелинейной регрессии, исходя из условия $\sum_{i=1}^{n}\left(\overline{y} - \widehat{y_i}\right)^2 \to min$. Требования к факторам, включаемым в модель линейной множественной регрессии…(несколько вариантов ответа)
   а) число факторов должно быть в 6 раз меньше объема совокупности
   б) факторы должны представлять временные ряды
   в) факторы должны иметь одинаковую размерность
   г) между факторами не должно быть высокой корреляции

3. Параметры при факторах в линейной множественной регрессии $\hat{y} = a + b_1x_1 + b_2x_2 + \ldots + b_px_p$ характеризуют
   а) долю дисперсии результативной переменной, объясненную регрессией в его общей дисперсии
   б) тесноту связи между результативной переменной и соответствующим фактором, при устранении влияния других факторов, включенных в модель
   в) среднее изменение результативной переменной с изменением соответствующего фактора на единицу, при неизменном значении других факторов, закрепленных на среднем уровне
   г) на сколько процентов в среднем изменяется результативная переменная с изменением соответствующего фактора на 1 %

4. К свойствам уравнения регрессии в стандартизированном виде относятся … (несколько вариантов ответа)
   а) коэффициенты регрессии при объясняющих переменных равны между собой
   б) постоянный параметр (свободный член уравнения) регрессии отсутствует
   в) стандартизированные коэффициенты регрессии несравнимы между собой
   г) входящие в состав уравнения переменные являются безразмерными

5. Верные утверждения относительно коэффициента множественной корреляции… (несколько вариантов ответа)
   а) чем ближе значение к единице $R_{yx_1 \dots x_p}$, тем теснее связь результативного признака со всеми факторами
   б) чем ближе значение к нулю $R_{yx_1 \dots x_p}$, тем теснее связь результативного признака со всеми факторами
   в) $R_{yx_1 \dots x_p}$ принимает значения из промежутка [0, 1]
   г) $R_{yx_1 \dots x_p}$ принимает значения из промежутка [– 1, 1]

6. Для устранения систематической ошибки остаточной дисперсии для оценки качества модели линейной множественной регрессии используется
   а) коэффициент множественной детерминации
   б) коэффициент множественной корреляции
   в) скорректированный коэффициент множественной детерминации
   г) скорректированный коэффициент частной корреляции

7. Если уравнение регрессии является существенным, то фактическое значение F-критерия …
   а) больше критического
   б) меньше критического
   в) близко к единице
   г) близко к нулю

8. К методам обнаружения гетероскедастичности остатков относятся…(несколько вариантов ответа)

а) критерий Дарбина — Уотсона
б) тест Голдфелда — Квандта
в) графический анализ остатков
г) метод наименьших квадратов

9. Регрессии, нелинейные по оцениваемым параметрам…(несколько вариантов ответа)
а) $y = a + b_1x + b_2x^2 + \varepsilon$
б) $y = a \cdot x^b \cdot \varepsilon$
в) $y = a + \frac{b}{x} + \varepsilon$
г) $y = a + bx + \varepsilon$
д) $y = a \cdot b^x \cdot \varepsilon$
е) $y = e^{a+bx} \cdot \varepsilon$

10. Модель $y = a \cdot x^b \cdot \varepsilon$ относится к классу … эконометрических моделей нелинейной регрессии
а) степенных
б) обратных
в) показательных
г) линейных

11. С помощью метода наименьших квадратов нельзя оценить значения параметров уравнения регрессии …
а) $y = a + \frac{b}{x} + \varepsilon$
б) $y = a + bx^c + \varepsilon$
в) $y = a + bx + cx^2 + \varepsilon$
г) $y = a + b_1x_1 + b_2x_2 + \varepsilon$

12. Пусть Y_t — временной ряд, T_t — трендовая компонента, S_t — сезонная компонента, E_t — случайная компонента. Аддитивная модель временного ряда имеет вид …
а) $Y_t = T_t + S_t + E_t$
б) $Y_t = T_t \cdot S_t + E_t$
в) $Y_t = T_t + S_t \cdot E_t$
г) $Y_t = T_t \cdot S_t \cdot E_t$

13. Определить наличие циклических (сезонных) колебаний во временном ряду можно …(несколько вариантов ответа)

а) в результате анализа автокорреляционной функции
б) по графику временного ряда
в) по объему временного ряда
г) с помощью критерия Фостера — Стюарта

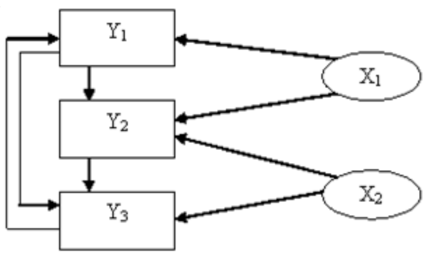
14. Для описания тенденции временного ряда используется кривая роста с насыщением …
   а) $y = a + b_1 t + b_2 t^2$
   б) $y = a + b_1 t + b_2 t^2 + b_3 t^3$
   в) $y = a \cdot b^t, b > 1$
   г) $y = k + a \cdot b^t, a < 0, b < 1$

15. Известны значения коэффициентов автокорреляции $r_1 = 0{,}8$, $r_2 = 0{,}2$, $r_3 = 0{,}3$, $r_4 = 0{,}9$. Укажите верные утверждения …(несколько вариантов ответа)
   а) временной ряд содержит линейный тренд
   б) временной ряд содержит тренд в виде полинома 4 порядка
   в) временной ряд содержит циклические колебания с периодом 2
   г) временной ряд содержит циклические колебания с периодом 4

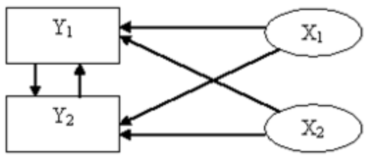
16. Случайность остатков модели временного ряда может быть проверена с помощью (несколько вариантов ответа)
   а) анализа автокорреляционной функции остатков
   б) критерия Пирсона
   в) проверки гипотезы о наличии тренда
   г) расчета асимметрии и эксцесса

17. Параметр адаптации $\alpha = 0{,}3$, $y_5 = 8$, $y_6 = 7$, $S_4 = 6$. Значение S_6, полученное в результате экспоненциального сглаживания временного ряда по формуле $S_t = \alpha y_t + (1 - \alpha) S_{t-1}$, равно…

18. В структурной форме модели, построенной по указанной схеме взаимосвязей между переменными, количество экзогенных переменных равно …



19. Количество уравнений системы для указанной схемы взаимосвязей между переменными равно …



20. Приведенная форма модели, соответствующая структурной форме системы одновременных уравнений (несколько вариантов ответа)

$$\begin{cases} y_1 = b_{12}y_2 + a_{11}x_1 + \varepsilon_1, \\ y_2 = b_{21}y_1 + a_{22}x_2 + \varepsilon_2 \end{cases}$$

включает в себя уравнения

а) $y_1 = a_{11}x_1 + \varepsilon_1$

б) $y_2 = a_{22}x_2 + \varepsilon_2$

в) $y_1 = \delta_{11}x_1 + u_1$

г) $y_2 = \delta_{22}x_2 + u_2$

д) $y_1 = \delta_{11}x_1 + \delta_{12}x_2 + u_1$

е) $y_2 = \delta_{21}x_1 + \delta_{22}x_2 + u_2$

21. Установите соответствие между типом структурной модели и соответствием структурных и приведенных коэффициентов …

а) идентифицируема	1) число приведенных коэффициентов меньше числа структурных коэффициентов

б) частично идентифицируема	2) число приведенных коэффициентов больше числа структурных коэффициентов
в) сверх идентифицируема	3) все структурные коэффициенты определяются однозначно по приведенным коэффициентам
г) не идентифицируема	

22. Пусть H — число эндогенных переменных в системе, D — число экзогенных переменных, которые содержатся в системе, но не содержатся в данном уравнении. Для первого уравнения модели динамики цены и заработной платы значение (H — D) равно …

$$\begin{cases} y_1 = b_{12}y_2 + a_{11}x_1 + \varepsilon_1, \\ y_2 = b_{21}y_1 + a_{22}x_2 + a_{23}x_3 + \varepsilon_2, \end{cases}$$

23. Для идентифицируемой структурной формы системы одновременных уравнений при оценке параметров применяется …
    а) обычный метод наименьших квадратов
    б) косвенный метод наименьших квадратов
    в) двухшаговый метод наименьших квадратов
    г) трехшаговый метод наименьших квадратов

24. Количество значений, которые может принимать результативный показатель в моделях бинарного выбора, равен …

ОГЛАВЛЕНИЕ

Научное издание

Зелепухин Юрий Валентинович

Эконометрика

Учебно-методическое пособие

Ответственный редактор *С. Краснова*
Верстальщик *Е. Семенова*

Издательство «Директ-Медиа»
117342, Москва, ул. Обручева, 34/63, стр. 1
Тел/факс + 7 (495) 334-72-11
E-mail: manager@directmedia.ru
www.biblioclub.ru

Отпечатано в ООО «ПАК ХАУС»
142172, г. Москва, г. Щербинка,
ул. Космонавтов, д. 16