

Федеральное агентство по образованию
Государственное образовательное учреждение высшего
профессионального образования
Ульяновский государственный технический университет

Н. И. Шанченко

ЛЕКЦИИ ПО ЭКОНОМЕТРИКЕ

Учебное пособие
для студентов высших учебных заведений, обучающихся
по специальности «Прикладная информатика (в экономике)»

Ульяновск
2008

УДК 330.43 (075.8)

ББК 65в6я73

Ш 20

Рецензенты:

Доктор физико-математических наук, профессор кафедры информационной безопасности и управления УлГУ, А. С. Андреев;

Кафедра общепрофессиональных дисциплин УВАУГА

*Утверждено редакционно-издательским советом университета
в качестве учебного пособия*

Шанченко, Н. И.

Ш 20 Лекции по эконометрике : учебное пособие для студентов высших учебных заведений, обучающихся по специальности «Прикладная информатика (в экономике)» / Н. И. Шанченко. – Ульяновск : УлГТУ, 2008. – 139 с.

ISBN 978-5-9795- 0504-6

Содержит краткий курс лекций по дисциплине «Эконометрика», включая описание основных задач эконометрики и методов, применяемых для их решения. Предназначено для студентов экономических и информационных специальностей.

УДК 330.43 (075.8)

ББК 65в6я73

© Н. И. Шанченко, 2008

© Оформление. УлГТУ, 2008

ISBN 978-5-9795-0504-6

СОДЕРЖАНИЕ

СОДЕРЖАНИЕ	3
Введение.....	7
1. Предмет и методы эконометрики.....	10
1.1. Предмет и методы эконометрики.....	10
1.2. Характеристика взаимосвязей.....	12
1.3. Основные этапы построения эконометрической модели.....	13
1.4. Выбор вида эконометрической модели.....	16
1.5. Методы отбора факторов.....	18
1.6. Оценка параметров моделей.....	20
1.7. Примеры эконометрических моделей.....	21
Контрольные вопросы.....	22
2. Парный регрессионный анализ.....	23
2.1. Понятие парной регрессии.....	23
2.2. Построение уравнения регрессии.....	24
2.2.1. Постановка задачи.....	24
2.2.2. Спецификация модели.....	25
2.3. Оценка параметров линейной парной регрессии.....	26
2.4. Оценка параметров нелинейных моделей.....	28
2.5. Качество оценок МНК линейной регрессии. Теорема Гаусса-Маркова.....	29
2.6. Проверка качества уравнения регрессии. F -критерий Фишера.....	30
2.7. Коэффициенты корреляции. Оценка тесноты связи.....	32
2.8. Точность коэффициентов регрессии. Проверка значимости.....	33
2.9. Точечный и интервальный прогноз по уравнению линейной регрессии.....	35
2.10. Коэффициент эластичности.....	36
Контрольные вопросы.....	37
3. Множественный регрессионный анализ.....	38
3.1. Понятие множественной регрессии.....	38
3.2. Отбор факторов при построении множественной регрессии.....	39
3.2.1. Требования к факторам.....	39
3.2.2. Мультиколлинеарность.....	40
3.3. Выбор формы уравнения регрессии.....	42
3.4. Оценка параметров уравнения линейной множественной регрессии.....	43
3.5. Качество оценок МНК линейной множественной регрессии. Теорема Гаусса-Маркова.....	46
3.6. Проверка качества уравнения регрессии. F -критерий Фишера.....	47
3.7. Точность коэффициентов регрессии. Доверительные интервалы....	49
3.8. Частные уравнения регрессии. Частная корреляция.....	50
3.9. Обобщенный метод наименьших квадратов. Гетероскедастичность.....	52
3.9.1. Обобщенный метод наименьших квадратов.....	52

3.9.2. Обобщенный метод наименьших квадратов в случае гетероскедастичности остатков	53
3.10. Проверка остатков регрессии на гетероскедастичность	55
3.11. Построение регрессионных моделей при наличии автокорреляции остатков.....	56
3.12. Регрессионные модели с переменной структурой. Фиктивные переменные.....	58
3.12.1. Фиктивные переменные	58
3.12.2. Тест Чоу	59
3.11. Проблемы построения регрессионных моделей.....	59
Контрольные вопросы	60
4. Системы эконометрических уравнений.....	61
4.1. Структурная и приведенная формы модели.....	61
4.2. Оценка параметров структурной формы модели	65
4.3. Косвенный метод наименьших квадратов.....	66
4.4. Двухшаговый метод наименьших квадратов	68
4.5. Трехшаговый метод наименьших квадратов	69
Контрольные вопросы	70
5. Моделирование одномерных временных рядов и прогнозирование.....	71
5.1. Составляющие временного ряда	71
5.2. Автокорреляция уровней временного ряда.....	72
5.3. Моделирование тенденции временного ряда.....	73
5.3.1. Методы определения наличия тенденции	73
5.3.2. Сглаживание временного ряда по методу скользящей средней.....	74
5.3.3. Метод аналитического выравнивания	76
5.3.4. Выбор вида тенденции	77
5.3.5. Оценка адекватности и точности модели тенденции.....	79
5.4. Моделирование периодических колебаний	82
5.4.1. Выделение периодической компоненты по методу скользящей средней.....	82
5.4.2. Моделирование сезонных колебаний с помощью фиктивных переменных	83
5.4.3 Моделирование сезонных колебаний с помощью гармонического анализа.....	83
5.5. Прогнозирование уровней временного ряда на основе кривых роста.	84
5.5.1. Метод аналитического выравнивания	84
5.6. Адаптивные модели прогнозирования	86
5.6.1. Понятие адаптивных методов прогнозирования	86
5.6.2. Экспоненциальное сглаживание	87
5.6.3. Использование экспоненциальной средней для краткосрочного прогнозирования.....	88
5.6.4. Адаптивные полиномиальные модели.....	88
5.7. Исследование взаимосвязи двух временных рядов.....	89

5.8. Коинтеграция временных рядов	91
Контрольные вопросы	92
6. Линейные модели стохастических процессов	93
6.1. Стационарные стохастические процессы	93
6.1.1. Основные понятия	93
6.1.2. Параметрические тесты стационарности	94
6.1.3. Непараметрические тесты стационарности	96
6.2. Линейные модели стационарных временных рядов.	
Процессы ARMA	97
6.2.1. Модели авторегрессии (AR)	97
6.2.2. Модели скользящего среднего (MA)	98
6.2.3. Модели авторегрессии-скользящего среднего (ARMA)	99
6.3. Автокорреляционные функции	99
6.3.1. Автокорреляционная функция	99
6.3.2. Частная автокорреляционная функция	100
6.4. Прогнозирование ARMA-процессов	101
6.4.1. AR-процессы	101
6.4.2. MA-процессы	102
6.4.3. ARMA-процессы	103
6.5. Нестационарные интегрируемые процессы	103
6.5.1. Нестационарные стохастические процессы.	
Нестационарные временные ряды	103
6.5.2. Тесты Дики-Фуллера	104
6.5.3. Модификации теста Дики-Фуллера для случая автокорреляции	104
6.5.4. Метод разностей и интегрируемость	105
6.6. Модели ARIMA	105
6.6.1. Определение и идентификация модели	105
6.6.2. Прогнозирование ARIMA-процессов	106
Контрольные вопросы	107
7. Динамические эконометрические модели	108
7.1. Общая характеристика динамических моделей	108
7.2. Модели с распределенным лагом	109
7.2.1. Оценка параметров модели с распределенным лагом методом Койка	109
7.2.2. Оценка параметров модели с распределенным лагом методом Алмон	110
7.2.3. Интерпретация параметров	111
7.3. Модели авторегрессии	112
7.3.1. Интерпретация параметров	112
7.3.2. Оценка параметров моделей авторегрессии	113
7.4. Модель частичной корректировки	114
7.5. Модель адаптивных ожиданий	115
Контрольные вопросы	116
8. Информационные технологии эконометрических исследований	117

8.1. Электронные таблицы Excel	118
8.2. Статистический пакет общего назначения STATISTICA.....	119
8.3. Эконометрические программные пакеты. Matrixer 5.1	120
8.4. Анализ временных рядов в системе ЭВРИСТА	122
Контрольные вопросы	124
Глоссарий	125
Приложения	131
1. Нормированная функция Лапласа.....	131
2. Значения критических уровней $t_{\alpha,k}$ для распределения Стьюдента...	132
3. Значения F -критерия Фишера на уровне значимости $\alpha = 0,05$	133
4. Значения F -критерия Фишера на уровне значимости $\alpha = 0,01$	134
5. Значения $\chi^2_{\alpha;k}$ критерия Пирсона.....	135
6. Значения статистик Дарбина-Уотсона $d_L d_U$	136
7. Критические значения f -критерия для DF-, ADF- и PP-тестов, рассчитанные по Маккиннону	137
8. Критические значения коинтеграционного ADF-критерия.....	137
Библиографический список	138
Интернет-ресурсы	138

Введение

Развитие экономики, усложнение экономических процессов и повышение требований к принимаемым управленческим решениям в области макро и микроэкономики потребовало более тщательного и объективного анализа реально протекающих процессов на основе привлечения современных математических и статистических методов.

С другой стороны, проблема нарушения предпосылок классических статистических методов при решении реальных экономических задач привели к необходимости развития и совершенствования классических методов математической статистики и уточнения постановок соответствующих задач.

В результате этих процессов осуществилось выделение и формирование новой отрасли знания под названием Эконометрика, связанной с разработкой и применением методов количественной оценки экономических явлений и процессов и их взаимосвязей.

Основным методом исследования в эконометрике является экономико-математическое моделирование. Правильно построенная модель должна давать ответ на вопрос о количественной оценке величины изменения изучаемого явления или процесса в зависимости от изменений внешней среды. Например, как скажется увеличение или уменьшение уровня инвестиций на совокупном валовом продукте, какие дополнительные ресурсы понадобятся для запланированного увеличения выпуска продукции и т. п.

Практическая значимость эконометрики определяется тем, что применение ее методов позволяет выявить реально существующие связи между явлениями, дать обоснованный прогноз развития явления в заданных условиях, проверить и численно оценить экономические последствия принимаемых управленческих решений.

Построение эконометрических моделей приходится осуществлять в условиях, когда нарушаются предпосылки классических статистических методов, и учитывать наличие таких явлений, как:

- мультиколлинеарность объясняющих переменных;
- закрытость механизма связи между переменными в изолированной регрессии;
- эффект гетероскедастичности, т. е. отсутствия нормального распределения остатков для регрессионной функции;
- автокорреляция остатков;
- ложная корреляция.

Разработка методов, преодолевающих эти трудности, составляет теоретическую основу эконометрики.

Наряду с логически правильным формальным применением имеющегося математического и статистического инструментария важными составляющими успеха эконометрического исследования являются экономически адекватная постановка задачи и последующая экономическая интерпретация полученных результатов.

Огромный толчок развитию эконометрических методов и их широкому внедрению в практику дало развитие средств вычислительной техники и особенно появление персональных и портативных компьютеров. Разработка программных пакетов, реализующих методы построения и исследования эконометрических моделей привело к тому, что выполнение эконометрических процедур становится доступным самому широкому кругу аналитиков, экономистов и менеджеров. В настоящее время основные усилия прикладного исследователя сводятся к подготовке качественных исходных данных, к правильной постановке проблемы и экономически обоснованной интерпретации результатов исследования. Вместе с тем, от исследователя требуется четкое понимание областей применимости используемых методов и сложности и неочевидности процесса перенесения полученных теоретических результатов на реальную действительность.

Настоящее пособие отражает содержание односеместрового курса лекций, читаемых на факультете информационных систем и технологий УлГТУ студентам специальности «Прикладная информатика (в экономике)» и соответствует Государственному образовательному стандарту по дисциплине «Эконометрика». Пособие состоит из восьми глав и приложения.

В первой главе дается характеристика предмету эконометрики и применяемым методам, освещаются основные аспекты эконометрического моделирования, применяемые методики и виды используемых переменных.

Во второй главе рассмотрены вопросы построения парных регрессионных моделей: постановка задачи, спецификация и оценка параметров моделей, оценка качества полученных моделей, получение точечного и интервального прогнозных значений, экономическая интерпретация модели.

Третья глава посвящена построению множественных регрессионных моделей. Подробно рассмотрены вопросы спецификации и оценки параметров модели, оценки качества полученной модели и ее статистической значимости. Приведены условия, обеспечивающие эффективность метода наименьших квадратов (теорема Гаусса-Маркова). Описан обобщенный метод наименьших квадратов, позволяющий получать эффективные оценки параметров в условиях мультиколлинеарности факторов и автокорреляции остатков. Рассмотрены регрессионные модели с переменной структурой.

Четвертая глава посвящена построению моделей в виде системы эконометрических уравнений. Изложены особенности моделей, возникающие трудности применения классических методов и описаны наиболее широко применяемые методы оценки параметров, такие как косвенный, двухшаговый и трехшаговый методы наименьших квадратов.

В пятой главе рассмотрены вопросы моделирования одномерных временных рядов и прогнозирования: структура временного ряда, явление автокорреляции, моделирование тенденции и периодической составляющей ряда, прогнозирование уровней ряда. Отдельное внимание уделено адаптивным методам прогнозирования и моделированию коинтегрируемых временных рядов.

В шестой главе освещены вопросы построения линейных моделей стохастических процессов: AR, MA и ARMA-моделей стационарных процессов,

ARIMA-моделей нестационарных процессов. Описаны методы проверки временных рядов на стационарность.

В седьмой главе излагаются модели и методы, применяемые для исследования эконометрических моделей, описывающих динамику развития экономических процессов. Рассмотрены модели авторегрессии и модели с распределенным лагом. Описаны применяемые для оценки параметров моделей, такие как методы инструментальных переменных, методы Койка и Алмон.

Восьмая глава посвящена информационным технологиям эконометрических исследований. Изложены общие требования к программному обеспечению и возможности программных пакетов Excel, STATISTICA, ЭВРИСТА, Matrixer 5.1.

В приложении даны часто используемые статистические таблицы.

Пособие предназначено студентам экономических и информационных специальностей. Изложение материала ориентировано на читателя, обладающего знаниями в пределах курсов высшей математики и математической статистики, читаемых студентам экономических и информационных специальностей. Пособие будет также полезно всем желающим познакомиться с основными задачами, моделями и методами эконометрики.

1. Предмет и методы эконометрики

1.1. Предмет и методы эконометрики

Эконометрика как наука возникла в первой половине 20-го века в результате активного использования для решения задач экономической теории математических и статистических методов.

Термин эконометрика введен в научную литературу в 1930 году норвежским статистиком Рагнарсом Фришем. Он первым определил эконометрику, как научную дисциплину, базирующуюся на синтезе экономической теории, статистики и математики.

В дословном переводе слово эконометрика означает «экономические измерения». Это очень широкое толкование данного понятия. Как правило, термин эконометрика применяется в более узком смысле. А именно, под эконометрикой понимается раздел науки, изучающий конкретные количественные и качественные взаимосвязи экономических объектов и процессов с помощью математических и статистических методов и моделей (БСЭ).

Можно сказать, что главной задачей эконометрики является количественная оценка имеющихся взаимосвязей между экономическими явлениями и процессами.

Экономические явления взаимосвязаны и взаимообусловлены. Следствием этого является то, что значения соответствующих экономических показателей изменяются во времени с учетом этих взаимосвязей. Так, например, известно, что совокупный спрос зависит от уровня цен, потребление – от располагаемого дохода, инвестиции – от процентной ставки и так далее. Перед исследователем стоит задача выявления таких связей, количественная их оценка и изучение возможности использования выявленных связей в экономическом анализе и прогнозировании. Разработкой соответствующего инструментария и его применением для решения конкретных практических экономических задач как раз и занимается эконометрика.

В основе любого эконометрического исследования лежит построение экономико-математической модели, адекватной изучаемым реальным экономическим явлениям и процессам.

Процесс построения эконометрических моделей начинается с качественного исследования проблемы методами экономической теории, формулируются цели исследования, выделяются факторы, влияющие на изучаемый показатель, и формулируются предположения о характере предполагаемой зависимости.

На этой основе изучаемые зависимости выражаются в виде математических формул и соотношений.

Следует отметить, что ввиду невозможности одновременно учесть большое количество факторов, влияющих на изучаемый показатель, предполагаемые зависимости между переменными будут выполняться не точно, а с определенной погрешностью. Кроме того, экономическим явлениям присуща внутренняя неопределенность, связанная с целенаправленной деятельностью субъектов экономики.

Вышесказанное обуславливает применение статистических методов, с помощью которых осуществляется отбор значимых факторов, определяется наличие и степень тесноты связи между изучаемыми показателями, дается количественная оценка параметров предполагаемых зависимостей и исследуется степень их соответствия реальной действительности.

Основным инструментом математической статистики, используемым для построения эконометрических моделей, являются методы корреляционного и регрессионного анализа.

Корреляционный анализ ставит своей целью проверку наличия и значимости линейной зависимости между переменными без разделения переменных на зависимые и объясняющие. Ответ на эти вопросы дается с помощью вычисления показателей (коэффициентов) корреляции.

Регрессионный анализ направлен на выражение изучаемой зависимости в виде аналитической формулы с предварительным выделением зависимых и объясняющих переменных.

Регрессионный анализ призван ответить на такие вопросы, как:

- какие переменные определяют поведение других величин и, следовательно, могут использоваться как объясняющие переменные?
- какова формула зависимости и каков экономический смысл ее коэффициентов?

Результатом проведения регрессионного анализа является построение, так называемого, уравнения регрессии.

После построения уравнения регрессии осуществляется проверка его статистического качества, включающая:

- проверку статистической значимости коэффициентов уравнения регрессии;
- проверку общего качества уравнения регрессии;
- проверку наличия свойств данных, предполагавшихся при оценивании уравнения регрессии.

Рассматривая эконометрическое исследование в целом, в нем можно выделить следующие этапы:

1. Постановка проблемы, т. е. определение цели и задач исследования, выделение зависимых (y_j) и независимых (x_k) экономических переменных на основе качественного анализа изучаемых взаимосвязей методами экономической теории.
2. Сбор необходимых исходных данных.
3. Построение эконометрической модели и оценка ее адекватности и степени соответствия исходным данным.
4. Использование модели для целей анализа и прогнозирования параметров исследуемого явления.
5. Качественная и количественная интерпретация полученных на основе модели результатов.
6. Практическое использование результатов.

В процессе экономической интерпретации результатов необходимо ответить на следующие вопросы:

– являются ли статистически значимыми объясняющие факторы, важные с теоретической точки зрения?

– соответствуют ли оценки параметров модели качественным представлениям?

Примером эконометрической модели может служить аналитическое выражение взаимосвязи показателей инфляции и безработицы, записанное без учета инфляционных ожиданий (1.1) и с учетом последних (1.2) [6]:

$$\pi = -\beta(u - u^*), \quad (1.1)$$

$$\pi = \pi^e - \beta(u - u^*), \quad (1.2)$$

где π – фактический и π^e – ожидаемый темпы инфляции (в процентах), u – фактический и u^* – естественный уровни безработицы (в процентах), β – постоянный параметр. При проведении исследования определяется, какая из этих зависимостей лучше соответствует реальной взаимосвязи между уровнями инфляции и безработицы, а также оценивается значение величины естественного уровня безработицы.

1.2. Характеристика взаимосвязей

Основная задача эконометрики заключается в исследовании и количественной оценке объективно существующих взаимосвязей и зависимостей между экономическими явлениями. Наибольший интерес для исследователя представляют причинно-следственные отношения между явлениями, что позволяет выявлять факторы, оказывающие основное влияние на вариацию изучаемых явлений и процессов.

Причинно-следственное отношение – это такая связь между явлениями, при которой изменение одного из них, называемого причиной, ведет к изменению другого, называемого следствием. Следовательно, причина всегда предшествует следствию.

Причинно-следственные связи в социально-экономических явлениях обладают следующими особенностями. Во-первых, причина X и следствие Y взаимодействуют не непосредственно, а через промежуточные факторы, которые, как правило, при анализе опускаются. Формально это может быть выражено с помощью схемы $X \rightarrow X' \rightarrow X'' \rightarrow Y$, где X' и X'' изображают такие промежуточные факторы.

Во-вторых, социально-экономические явления развиваются и формируются в результате одновременного воздействия большого числа факторов. Поэтому одной из главных проблем при изучении этих явлений становится задача выявления главных, существенных причин и абстрагирование от второстепенных.

Признаки по их роли в изучаемой взаимосвязи делятся на два класса: *факторные* и *результативные*.

Факторными признаками (факторами) называются признаки, обуславливающие изменения других, связанных с ними признаков. Факторные признаки называются также независимыми, объясняющими или входными переменными.

Результативными называются признаки, изменяющиеся под действием факторных признаков. Результативные признаки называются также зависимыми, объясняемыми или выходными переменными.

По направлению изменения связи подразделяются на *прямые* (когда изменение результативного и факторного признаков происходит в одном направлении) и *обратные* (когда изменение результативного и факторного признаков происходит в противоположных направлениях).

По характеру проявления различают *функциональную* связь и *стохастическую* зависимость.

Функциональной называют такую связь, при которой определенному значению факторного признака соответствует одно и только одно значение результативного признака. Функциональная связь проявляется во всех случаях наблюдения и для каждой конкретной единицы исследуемой совокупности. Такие связи изучаются в основном в естественных науках.

В эконометрике в основном изучаются причинные зависимости, которые проявляется не в каждом отдельном случае, а в общем, среднем при большом числе наблюдений. То есть одним и тем же значениям факторных признаков, как правило, соответствуют различные значения результативного признака. Но, тем не менее, рассматривая всю совокупность наблюдений можно отметить наличие определенной зависимости между значениями признаков. Такие причинные зависимости называются *стохастическими*.

Частным случаем стохастической связи является *корреляционная* связь, при которой изменение среднего значения результативного признака обусловлено изменением факторных признаков.

По аналитическому выражению выделяют связи *линейные* и *нелинейные*.

Линейной называется связь, в которой изменение результативного признака прямо пропорционально изменению факторных признаков. В противном случае связь называется нелинейной. Аналитически линейная стохастическая связь между явлениями может быть представлена уравнением прямой линии на плоскости, либо уравнением гиперплоскости в n -мерном пространстве (при наличии n факторных переменных).

1.3. Основные этапы построения эконометрической модели

Построение эконометрической модели является основой эконометрического исследования. Оно основывается на предположении о реально существующей зависимости между признаками. От того, насколько хорошо полученная модель описывает изучаемые закономерности между экономическими процессами, зависит степень достоверности результатов анализа и их применимости.

Построение эконометрической модели начинается со спецификации модели, заключающейся в получении ответа на два вопроса: 1) какие экономические показатели (признаки) должны быть включены в модель; 2) какой вид имеет аналитическая зависимость между отобранными признаками.

В обобщенной форме эконометрическая модель, описывающая взаимосвязи между явлениями или закономерности их развития, представляется с помощью соотношения:

$$y = f(\mathbf{a}, \mathbf{x}) + \varepsilon, \quad (1.3)$$

где $f(\mathbf{a}, \mathbf{x})$ – функционал, выражающий вид и структуру взаимосвязей. Здесь величина y выражает уровень исследуемого явления и называется зависимой (объясняемой) переменной или результативным признаком; величина $\mathbf{x} = (x_1, x_2, \dots, x_n)$ представляет собой вектор значений независимых (объяс-

няющих) переменных x_i или факторных признаков (факторов); через $\alpha = (\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_n)$ обозначен вектор некоторых произвольных констант, называемых параметрами модели; ε – ошибка модели.

Ошибка модели ε характеризует отличие наблюдаемого (реализованного) значения переменной y от вычисленных согласно соотношения (1.3) в конкретных условиях (при конкретных значениях переменных факторов x_i) и рассматривается как случайная величина.

Для расчета численных значений параметров $\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_n$ используется предварительно накопленный массив наблюдений за совместным проявлением изучаемого процесса и рассматриваемых факторов. Одно наблюдение представляет собой множество значений $(y_t, x_{1t}, x_{2t}, \dots, x_{nt})$. Индекс t соответствует отдельному наблюдению.

Отдельные наблюдения могут характеризовать уровни изучаемого явления в различные моменты времени (табл. 1.1) либо его проявление для различных однородных объектов в один и тот же момент или период времени (табл. 1.2). В первом случае индекс t соответствует отдельному моменту времени, а во втором – отдельному объекту.

Таблица 1.1

Макроэкономические данные по России за период 1995-2006 гг.

Текущий период, t	ВВП, Y (млрд руб.)	Денежная масса, M (млрд руб.)	Внутренние инвестиции, I (млрд руб.)	Национальный доход, Y (млрд руб.)	Расходы на личное потребление, C (млрд руб.)	Валовая прибыль экономики, Q (млрд руб.)
1995 г.	1428,5	98,7	267,0	1412,7	1016,6	610,8
1996 г.	2007,8	220,8	376,0	1978,9	1435,9	699,4
1997 г.	2342,5	288,3	408,8	2292,0	1776,1	783,3
...
2005 г.	21620,1	4363,3	3611,1	21079,5	14363,5	7908,1
2006 г.	26781,1	6044,7	4580,5	26009,7	17742,6	9606,9

Таблица 1.2

Сравнительные данные по предприятиям

Предприятие	Объем реализации Q , млн руб. / мес.	Среднемесячная численность чел.	Основные производственные фонды млн руб.	Затраты на производство млн руб.
1	100	300	250	80
2	120	310	190	90
3	150	420	310	110
4	90	200	150	50
5	40	80	100	30
6	200	400	420	150

Зависимую переменную y часто называют эндогенной (внутренней) переменной модели, отражая тот факт, что значения зависимой переменной y определяются только значениями независимых переменных x_i .

Независимые переменные (факторы) x_1, x_2, \dots, x_n называют экзогенными (внешними) переменными. Термин «внешний» говорит о том, что значения переменных x_i определяются вне рассматриваемой модели, для которой они являются заданными.

В эконометрике переменная y согласно (1.3) всегда рассматривается как случайная величина.

Независимые переменные x_i могут считаться как случайными или детерминированными. В классической эконометрической модели они рассматриваются как детерминированные величины. В этом случае при ошибке модели, обладающей свойствами «белого шума», функционал $f(\alpha, x)$ можно рассматривать как математическое ожидание условного распределения переменной y при заданных значениях $x_{1t}, x_{2t}, \dots, x_{nt}, t = 1, 2, \dots, T$.

Представление значений независимых переменных эконометрических моделей как проявлений случайных величин, как правило, не вносит существенных изменений в методы оценки параметров моделей.

В классических регрессионных моделях обычно предполагается, что факторы независимы между собой и с ошибкой модели, обладающей свойствами «белого шума». Вместе с тем, ряд ошибки может характеризоваться свойствами непостоянства дисперсии для различных наблюдений; наличием автокорреляционных связей между соседними значениями ε_t и ε_{t-1} (для упорядоченных значений факторной переменной) и т. д. Могут иметь место корреляционные связи с экзогенными переменными x_i и др.

В моделях, описывающих динамику процессов или явлений, т. е. в моделях, когда состояние явления в последующие периоды времени зависит от состояний, достигнутых в предыдущие моменты времени, в качестве экзогенных переменных используются значения переменных (эндогенных или экзогенных) в предыдущие моменты времени ($y_{t-1}, y_{t-2}, \dots; x_{it-1}, x_{it-2}, \dots$), называемые лаговыми переменными.

В исследованиях, посвященных разработке методов прогнозирования таких финансовых показателей, как курсы валют, ценных бумаг, индексов широко применяются модели, основанные на предположении, что динамика этих процессов полностью определяется внутренними условиями. В этом случае модели соответствующих временных рядов включают в качестве факторов только лаговые значения результативного показателя y_{t-1}, y_{t-2}, \dots и (или) ошибки $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots$.

После выделения совокупности рассматриваемых переменных следующим этапом является определение конкретного вида модели, наилучшим образом соответствующего изучаемому явлению.

По характеру связей факторов с переменной y модели подразделяются на линейные и нелинейные.

По свойствам своих параметров модели подразделяются на модели с постоянной и переменной структурой.

Особый вид моделей составляют системы взаимосвязанных эконометрических уравнений, включающие несколько уравнений вида (1.3). Каждому уравнению соответствует своя зависимая переменная y_i , которая в других уравнениях системы может выступать в качестве независимого фактора.

Если на основе предварительного качественного анализа рассматриваемого явления не удастся однозначно выбрать наиболее подходящий тип модели, то рассматриваются несколько альтернативных моделей, среди которых в процессе исследования выбирается та, которая в наибольшей степени соответствует изучаемому явлению.

В общем случае процедуру построения эконометрической модели можно представить в виде следующих этапов:

1. Спецификация модели, т. е. выбор класса моделей, наиболее подходящих для описания изучаемых явлений и процессов. Этот этап предполагает решение двух задач:

- а) отбор существенных факторов для их последующего включения в модель;
- б) выбор типа модели, т. е. выбор вида аналитической зависимости, связывающей включенные в модель переменные.

2. Оценка параметров модели, т. е. получение численных значений констант модели. При этом используется предварительно полученный массив исходных данных.

3. Проверка качества построенной модели и обоснование возможности ее дальнейшего использования.

Наиболее сложным и трудоемким в эконометрическом исследовании является этап оценки параметров модели, где применяются методы теории вероятностей и математической статистики.

1.4. Выбор вида эконометрической модели

При решении проблемы выбора вида аналитической зависимости могут использоваться различные соображения:

- выводы аналитических исследований о качественном характере зависимости (направление изменения переменных и его особенности),
- описание свойств различных аналитических зависимостей,
- цели построения модели.

Выбор вида эконометрической модели основывается, прежде всего, на результатах предварительного качественного или содержательного анализа, проводимого методами экономической теории. По возможности характер предполагаемой зависимости обосновывается исходя из теоретически предположений о характере закономерности развития изучаемого явления или процесса.

Примером может служить зависимость между общими затратами на производство продукции (Z) и объемом производства (V)

$$Z = Z_{\text{пост}} + Z_{\text{уд.пер}} \cdot V,$$

где $Z_{\text{пост}}$ – постоянные затраты (не зависят от объема производства), $Z_{\text{уд.пер}}$ – удельные переменные затраты (переменные затраты на выпуск единицы продукции).

Другой подход основан на анализе массива исходных данных, который позволяет выявить некоторые характеристики предполагаемых зависимостей и на этой основе сформулировать, как правило, несколько предположений о виде аналитической связи. Построенная модель используется для формулирования предположений о характере закономерности в развитии изучаемого явления, которые проверяются в течение дальнейших исследований.

Приведем некоторые виды аналитических зависимостей, наиболее часто используемых при построении моделей:

1) линейная

$$y = a + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_p \cdot x_p + \varepsilon, \quad (1.4)$$

2) степенная

$$y = a \cdot x_1^{b_1} \cdot x_2^{b_2} \cdot \dots \cdot x_p^{b_p} \cdot \varepsilon, \quad (1.5)$$

3) полулогарифмическая

$$y = a + b_1 \ln x_1 + b_2 \ln x_2 + \dots + b_p \cdot \ln x_p, \quad (1.6)$$

4) гиперболическая

$$y = a + b_1 \frac{1}{x_1} + b_2 \frac{1}{x_2} + \dots + b_p \frac{1}{x_p} + \varepsilon, \quad (1.7)$$

5) экспоненциальная

$$y = e^{a+b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_p \cdot x_p + \varepsilon}. \quad (1.8)$$

Могут применяться также комбинации рассмотренных зависимостей. Например,

$$y = a + b_1 x_1 + b_2 \frac{1}{x_2} + b_3 \cdot \ln x_3.$$

При выборе вида аналитической зависимости важную роль играют требования простоты модели и наличия наглядной экономической интерпретации ее параметров. Исходя из этих соображений, наиболее часто используются линейная (1.4) и степенная (1.5) функции.

В линейной модели (1.4) параметры b_i при факторах x_i характеризуют величину среднего изменения зависимой переменной y с изменением соответствующего фактора x_i на единицу, в то время как значения остальных факторов остаются неизменными.

В степенной модели (1.5) параметры b_j при факторах x_i являются коэффициентами эластичности. Они показывают, на сколько процентов в среднем изменяется зависимая переменная y при изменении соответствующего фактора x_i на 1 % в условиях неизменности действия других факторов. Этот вид уравнения регрессии получил наибольшее распространение в производственных функциях, в исследованиях спроса и потребления.

При определении вида модели могут использоваться следующие соображения. Если изменение результативного признака y прямо пропорционально изменению значения фактора, то адекватной является линейная модель (1.4).

Если изменение результирующего признака y пропорционально значению фактора, то адекватной может быть либо степенная $y = a \cdot x^b \cdot \varepsilon$, либо экспоненциальная $y = e^{a+b \cdot x + \varepsilon}$ модели.

Если при увеличении значения факторов значение результирующего признака y монотонно стремится к конечному пределу, то можно использовать гиперболическую модель (1.7).

С целью отразить свойство оптимальности экономических переменных, т. е. наличия таких значений факторов x_i , на которых достигается минимаксное воздействие на зависимую переменную, в модель включают факторы x_i не только первой, но и второй степени

$$y = a + b_1x + b_2x^2. \quad (1.9)$$

Например, при увеличении возраста рабочих до определенного значения уровень производительности труда возрастает, а затем начинает снижаться.

Наибольшее применение в эконометрике нашли линейные модели. Это обусловлено несколькими причинами. Во-первых, существуют эффективные методы построения таких моделей.

Во-вторых, в небольшом диапазоне значений факторных признаков линейные модели с достаточной точностью могут аппроксимировать реальные нелинейные зависимости.

В-третьих, параметры модели имеют наглядную экономическую интерпретацию.

В-четвертых, прогнозы по линейным моделям, характеризуются, как правило, меньшим риском значительной погрешности прогноза.

1.5. Методы отбора факторов

Важной составляющей процесса построения эконометрической модели является отбор факторов, существенно влияющих на изучаемый показатель и подлежащих включению в разрабатываемую модель. Оптимальный набор факторов определяется на основе качественного и количественного анализа. Прежде всего, на этапе постановки задачи и содержательного экономического анализа экономической модели отбираются факторы, влияние которых должно быть учтено при построении модели. В ряде случаев набор факторов определяется однозначно или с большой степенью уверенности. Например, спрос на товар определяется в основном ценой и доходом.

В более сложных случаях на следующем этапе с помощью формальных статистических методов проверяется целесообразность включения в модель каждого фактора.

Прежде всего, факторы проверяются на наличие тесной линейной корреляционной зависимости между ними. Признаком наличия линейной корреляционной зависимости между факторами x_i и x_j является условие

$$|r_{x_i x_j}| \geq r_{кр}^1 \quad (1.10)$$

где $r_{x_i x_j}$ – выборочный линейный коэффициент корреляции, определяемый соотношением

$$r_{x_i x_j} = \frac{\frac{1}{n-1} \sum_{i=1}^n (x_{it} - \bar{x}_i)(x_{jt} - \bar{x}_j)}{\sigma_{x_i} \sigma_{x_j}}, \quad (1.11)$$

n – количество наблюдений, $r_{кр}^1$ – критическое значение $r_{кр}^1 \approx 0,8-0,9$ (определяется эмпирически).

Существование тесной корреляционной зависимости между факторами приводит к получению ненадежных оценок параметров модели.

Для преодоления сильной межфакторной корреляции применяется ряд подходов:

- исключение из модели одного или нескольких факторов. Из двух коррелирующих факторов исключаются тот, который более коррелирует с остальными факторами;

- преобразование факторов, при котором уменьшается корреляция между ними. Например, переходят от исходных переменных к их линейным комбинациям, не коррелированным друг с другом (метод главных компонент). При построении модели на основе рядов динамики переходят от первоначальных данных к первым разностям уровней ряда $\Delta y_t = y_t - y_{t-1}$, чтобы исключить влияние тенденции.

Одним из критериев включения факторов в модель является степень их изолированного влияния на результативный признак, определяемая с помощью коэффициента парной корреляции r_{yx_i} . Отбираются факторы x_i , удовлетворяющие условию

$$|r_{yx_i}| \geq r_{кр}^2, \quad (1.12)$$

где $r_{кр}^2 \approx 0,5-0,6$ (определяется эмпирически).

При определении «оптимального» набора факторов могут использоваться два метода:

- метод включения;
- метод исключения.

Согласно методу включения, сначала строится уравнение регрессии с одним наиболее влияющим фактором (фактор, для которого значение парного коэффициента корреляции с результативным признаком r_{yx_i} больше по модулю).

Затем в него последовательно вводятся следующие факторы и определяется пара наиболее влияющих факторов. На следующем к первым двум добавляется еще по одному фактору и определяется наилучшая тройка факторов и т. д. На каждом шаге строится модель регрессии и проверяется значимость факторов. В модель включают только значимые факторы. Для проверки значимости фактора могут использоваться либо критерий Стьюдента, либо частный критерий Фишера. Процесс заканчивается, когда не остается факторов, которые следует включить в модель.

Согласно методу исключения сначала строится уравнение регрессии с полным набором факторов, из числа которых затем последовательно исключаются незначимые (наименее значимые) факторы. На каждом шаге исключается только один фактор, так как после исключения какого-либо фактора другой фактор, бывший до этого незначимым, может стать значимым. Процесс заканчивается, когда не остается факторов, которые следует исключить из модели.

Методы включения и исключения не гарантируют определение оптимального набора факторов, но в большинстве случаев дают результаты либо оптимальные, либо близкие к ним.

Не рекомендуется включать в модель очень большое число факторов, так как это может затруднить выявление качественных закономерностей и возрастает опасность включения в модель несущественных случайных факторов.

Кроме того, для получения достаточно надежных оценок параметров желательно, чтобы количество наблюдений превышало количество определяемых параметров не менее чем в 6–7 раз.

1.6. Оценка параметров моделей

После отбора факторов и выбора вида аналитической зависимости осуществляется определение численных значений параметров α_i модели (1.3). Данная процедура носит название оценка параметров модели. Следует сказать, что сами полученные численные значения параметров α_i также называются оценка параметров. Путаницы не происходит, потому что то, в каком смысле используется этот термин, ясно из контекста.

При оценке параметров модели в качестве исходных данных используется заранее подготовленный массив наблюдений $\{(y_t, x_{1t}, x_{2t}, \dots, x_{nt}), t = 1, 2, \dots, n\}$. Так как исходные данные содержат проявления случайных величин, то и полученные оценки являются случайными величинами, зависящими от исходных данных и метода оценивания. Отсюда возникает задача отбора методов оценивания параметров, дающих оценки более высокого качества.

Согласно теории статистического оценивания качество оценок определяется наличием у них таких свойств как несмещенность, состоятельность и эффективность.

Оценка параметра называется *несмещенной*, если ее математическое ожидание равно оцениваемому параметру.

Оценка параметра называется *состоятельной*, если она сходится по вероятности к оцениваемому параметру при возрастании количества наблюдений.

Оценка параметра называется *эффективной*, если она имеет наименьшую дисперсию среди возможных несмещенных оценок параметра, вычисленных по выборкам одного и того же объема n .

Наиболее часто для оценки параметров применяются методы максимального правдоподобия и метод наименьших параметров. При выполнении определенных условий (относительно погрешностей модели ε_i) оценки параметров, полученные с помощью этих методов, обладают свойствами несмещенности, состоятельности и эффективности. Поэтому после получения оценок параметров необходимо проверить выполнение упомянутых условий, чтобы убедиться

в качестве полученных оценок. Если эти условия не выполняются, то следует скорректировать модель соответствующим образом.

Причины нарушения условий, налагаемых на погрешности модели ε_t , могут быть следующими:

- в модели не учтены существенные факторы;
- неправильно выбран вид модели.

1.7. Примеры эконометрических моделей

Модель ценообразования на основной капитал. Задается уравнением регрессии [2]

$$r - r_f = \alpha + \beta \cdot (r_m - r_f) + \varepsilon, \quad (1.13)$$

где r и r_f – прибыли рассматриваемой и безрисковой ценной бумаги;

r_m – прибыль общерыночного портфеля ценных бумаг;

α , β – константы ($\beta = \sigma/\sigma_m$, σ и σ_m – стандартные отклонения рассматриваемой ценной бумаги и рынка в целом);

ε – погрешность модели.

Производственная функция. Производственной функцией называется соотношение между входными факторами производства и выпуском продукции. Производственная функция часто применяется для оценки эластичности выпуска продукции по отдельным факторам производства. Например, производственная функция Кобба-Дугласа имеет вид

$$P = a \cdot L^\alpha \cdot K^{1-\alpha} \cdot \varepsilon, \quad (1.14)$$

где P – выпуск продукции;

L – затраты труда;

K – объем капитала;

$$0 < \alpha < 1.$$

Модель формирования спроса и предложения.

$$\begin{aligned} Q^d &= \beta_1 + \beta_2 \cdot P + \beta_3 I + \varepsilon_1, \\ Q^s &= \beta_4 + \beta_5 \cdot P + \varepsilon_2 \end{aligned}, \quad (1.15)$$

где Q^d – спрос на товар;

Q^s – предложение товара;

P – цена;

I – доход.

Макроэкономическая модель.

$$\begin{aligned} C_t &= a_1 + b_{11} \cdot Y_t + b_{12} \cdot Y_{t-1} + \varepsilon_{1t}, && \text{(функция потребления)} \\ I_t &= a_2 + b_{21} \cdot Y_t + \varepsilon_{2t}, && \text{(функция инвестиций)} \\ Y_t &= C_t + I_t + G_t. && \text{(тождество дохода)} \end{aligned} \quad (1.16)$$

где C_t – потребление;

Y_t – ВВП;

I_t – валовые инвестиции;

G_t – государственные расходы;

t – текущий период;

$t-1$ – предыдущий период.

Контрольные вопросы

1. Охарактеризуйте предмет эконометрики.
2. Укажите основные этапы эконометрического исследования.
3. Какие задачи решают корреляционный и регрессионный анализы?
4. Каковы особенности причинно-следственных отношений в социально-экономических явлениях?
5. Какие зависимости называются стохастическими?
6. Какие типы данных используются в эконометрическом исследовании?
7. Опишите основные этапы построения эконометрической модели.
8. Какие виды аналитических зависимостей, наиболее часто используются при построении моделей?
9. Какие методы используются для отбора факторов?
10. Какие методы используются для оценки параметров модели?
11. Какими свойствами характеризуется качество оценок параметров?

2. Парный регрессионный анализ

2.1. Понятие парной регрессии

Регрессией в теории вероятностей и математической статистике принято называть зависимость среднего значения какой-либо величины (y) от некоторой другой величины или от нескольких величин (x_i).

Парной регрессией называется модель, выражающая зависимость среднего значения зависимой переменной y от одной независимой переменной x

$$\hat{y} = f(x), \quad (2.1)$$

где y – зависимая переменная (результативный признак); x – независимая, объясняющая переменная (признак–фактор).

Парная регрессия применяется, если имеется доминирующий фактор, обуславливающий большую долю изменения изучаемой объясняемой переменной, который и используется в качестве объясняющей переменной.

Множественной регрессией называют модель, выражающую зависимость среднего значения зависимой переменной y от нескольких независимых переменных x_1, x_2, \dots, x_p

$$\hat{y} = f(x_1, x_2, \dots, x_p). \quad (2.2)$$

Множественная регрессия применяется в ситуациях, когда из множества факторов, влияющих на результативный признак, нельзя выделить один доминирующий фактор и необходимо учитывать одновременное влияние нескольких факторов.

Используя уравнение регрессии (2.1), соотношение между значениями переменными y и x (модель связи) можно записать как

$$y = f(x) + \varepsilon, \quad (2.3)$$

где первое слагаемое $f(x)$ можно интерпретировать как ту часть значения y , которая объяснена уравнением регрессии (2.1), а второе слагаемое ε как необъясненную часть значения y (или возмущение). Соотношение между этими частями характеризует качество уравнения регрессии, его способность представлять зависимость между переменными x и y . При построении уравнения регрессии ε рассматривается как ошибка модели, представляющая собой случайную величину, удовлетворяющую определенным предположениям.

Наличие составляющей ε обусловлено такими причинами, как наличие дополнительных факторов, оказывающих влияние на переменную y , неверный вид функциональной зависимости $f(x)$, ошибки измерения, выборочный характер исходных данных.

По виду аналитической зависимости различают линейные и нелинейные регрессии.

Линейная парная регрессия описывается уравнением:

$$\hat{y} = a + b \cdot x. \quad (2.4)$$

Примеры наиболее часто используемых нелинейных регрессий:

– полиномы разных степеней $\hat{y}_x = a + b_1 \cdot x + b_2 \cdot x^2 + b_3 \cdot x^3$,

- равносторонняя гипербола $\hat{y} = a + \frac{b}{x}$,
- степенная $\hat{y} = a \cdot x^b$
- экспоненциальная $\hat{y} = e^{a+b \cdot x}$,
- показательная $\hat{y} = a \cdot b^x$,
- логистическая $\hat{y} = \frac{K}{1 + a \cdot e^{-bt}}$.

2.2. Построение уравнения регрессии

2.2.1. Постановка задачи

Постановка задачи: по имеющимся данным n наблюдений за совместным изменением двух переменных показателей x и y $\{(x_i, y_i), i=1, 2, \dots, n\}$ необходимо определить аналитическую зависимость $\hat{y} = f(x)$, наилучшим образом описывающую данные наблюдений.

Результаты наблюдений удобно представлять в виде таблицы

Таблица 2.1

Данные наблюдений

	x	y
1	x_1	y_1
2	x_2	y_2
...
n	x_n	y_n

Каждая строка таблицы представляет собой результат одного наблюдения (x_i, y_i) .

Поясним понятие зависимости $\hat{y} = f(x)$, наилучшим образом описывающей данные наблюдений. Значения x_i, y_i из каждой строки можно рассматривать как координаты точки (x_i, y_i) на координатной плоскости xy . Совокупность всех точек составляют, так называемое, поле корреляций (рис. 2.1).

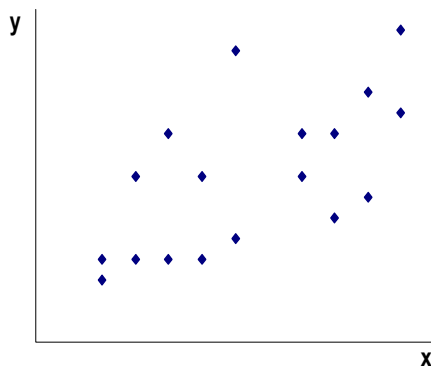


Рис. 2.1. Поле корреляций

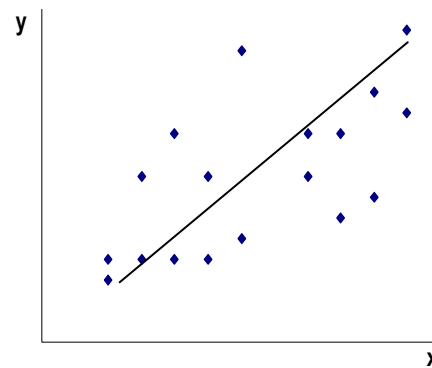


Рис. 2.2. Лучшая линейная регрессия

Зависимости $\hat{y} = f(x)$ соответствует некоторая кривая на плоскости. Чем ближе данная кривая подходит ко всем точкам поля корреляций, тем лучше зависимость $\hat{y} = f(x)$ описывает исходные данные.

Для формализации этого понятия рассмотрим разность между e_i расчетными (теоретическими, модельными) $\hat{y}_i = f(x_i)$ и наблюдаемыми y_i значениями

$e_i = \hat{y}_i - y_i$. Наилучшей будем считать такую зависимость, для которой сумма квадратов отклонений принимает минимальное значение, т. е.

$$S = \sum (\hat{y}_i - y_i)^2 = \min. \quad (2.5)$$

Построение уравнения регрессии предполагает решение двух задач (или, другими словами, осуществляется в два этапа):

- 1) спецификация модели (выбор вида аналитической зависимости $\hat{y} = f(x)$);
- 2) оценка параметров выбранной модели (определение численных значений параметров на основе массива наблюдений).

2.2.2. Спецификация модели

Парная регрессия применяется для моделирования зависимости, если имеется доминирующий фактор, который и используется в качестве объясняющей переменной.

Для выбора вида аналитической зависимости можно использовать следующие методы:

- графический (вид зависимости определяется на основе анализа поля корреляций);
- аналитический (на основе качественного анализа изучаемой взаимосвязи);
- экспериментальный (построение нескольких моделей различного вида с выбором наилучшей согласно применяемому критерию качества).

Визуальный анализ поля корреляций (рис. 2.1) позволяет определить форму кривой регрессии, ее особенности. Зная типичный вид графиков различных функций можно подобрать соответствующую аналитическую зависимость.

Примером применения аналитического метода может служить зависимость между затратами (y) и объемом производства (x). Считая, что затраты прямо пропорциональны объему производства, зависимость между ними можно представить в виде линейной функции

$$y = a + b \cdot x,$$

где a – часть затрат, не зависящая от объема производства, b – дополнительные затраты на производство единицы продукции.

Разделив обе части последнего уравнения на объем производства x , получим зависимость удельных затрат ($z = y/x$) на производство единицы продукции от объема производства

$$z = \frac{y}{x} = b + \frac{a}{x}.$$

При построении модели зависимости спроса товар от его цены при выборе вида зависимости следует учитывать, что при увеличении цены спрос падает. В этом случае могут использоваться следующие зависимости:

$$y = a - b \cdot x, \quad (b > 0);$$

$$y = \frac{1}{a + bx}, \quad (b > 0);$$

$$y = e^{a-bx}, \quad (b > 0).$$

Если из соображений экономической теории следует, что величина изменения зависимой переменной y пропорциональна значению независимой пере-

менной x , то можно выбрать полиномиальную, степенную или показательную зависимости (см. п. 2.1).

Если предполагается, что значение зависимой переменной y при увеличении значения независимой переменной x не может превысить некоторого предела, то можно выбрать гиперболическую $\hat{y} = a + \frac{b}{x}$ или логистическую

$$\hat{y} = \frac{K}{1 + a \cdot e^{-bt}}$$
 зависимости.

В случае, если в рассматриваемой области изменения фактора x результирующая переменная y принимает минимальное или максимальное значение, в уравнение регрессии включают переменные x не только первой, но и второй степени, например

$$y = a + b_1x + b_2x^2.$$

В качестве критерия качества модели может использоваться либо средняя квадратическая ошибка модели $\bar{\varepsilon}_{\hat{e}\hat{a}} = \sqrt{\frac{1}{n} \sum (\hat{y}_i - y_i)^2}$, либо остаточная дисперсия

$$D_{\hat{m}\hat{a}} = \frac{1}{n} \sum (\hat{y}_i - y_i)^2.$$

Этот подход легко реализуем при наличии соответствующих вычислительных средств. Но он не является определяющим, так как в эконометрике более важным является не способность модели соответствовать имеющемуся массиву данных наблюдений, а ее способность раскрывать существующие закономерности в экономических явлениях и процессах и интерпретация полученных с ее помощью результатов.

2.3. Оценка параметров линейной парной регрессии

Линейная парная регрессия описывается уравнением:

$$\hat{y} = a + b \cdot x, \quad (2.6)$$

согласно которому изменение Δy переменной y прямо пропорционально изменению Δx переменной x ($\Delta y = b \cdot \Delta x$).

Для оценки параметров a и b уравнения регрессии (2.6) воспользуемся методом наименьших квадратов (МНК). При определенных предположениях относительно ошибки ε МНК дает наилучшие оценки параметров линейной модели

$$y = a + b \cdot x + \varepsilon. \quad (2.7)$$

Согласно МНК, выбираются такие значения параметров a и b , при которых сумма квадратов отклонений фактических значений результирующего признака y_i от теоретических значений $\hat{y}_i = f(x_i)$ (при тех же значениях фактора x_i) минимальна, т. е.

$$S = \sum (\hat{y}_i - y_i)^2 = \min. \quad (2.8)$$

С учетом вида линейной парной регрессии (2.6) величина S является функцией неизвестных параметров a и b

$$S = \sum (y_i - a - b \cdot x_i)^2 = S(a, b). \quad (2.9)$$

Следовательно, оптимальные значения параметров a и b удовлетворяют условиям

$$\frac{\partial S}{\partial a} = 0; \quad \frac{\partial S}{\partial b} = 0. \quad (2.10)$$

Выполняя соответствующие вычисления, получим для определения параметров a и b следующую систему уравнений

$$\begin{aligned} \frac{\partial S}{\partial a} &= -2\sum (y_i - a - b \cdot x_i) = 0, \\ \frac{\partial S}{\partial b} &= -2b\sum (y_i - a - b \cdot x_i) = 0, \end{aligned}$$

откуда после некоторых преобразований получается система нормальных уравнений метода наименьших квадратов

$$\begin{cases} na + b \sum x_i = \sum y_i, \\ a \sum x_i + b \sum x_i^2 = \sum y_i x_i. \end{cases} \quad (2.11)$$

Используя соотношения $n\bar{x} = \sum x_i$, $n\bar{y} = \sum y_i$, $n \cdot \overline{x^2} = \sum x_i^2$, $n \cdot \overline{yx} = \sum y_i x_i$ из (2.8) получим

$$\begin{cases} a + b \cdot \bar{x} = \bar{y}, \\ a \cdot \bar{x} + b \cdot \overline{x^2} = \overline{yx}. \end{cases} \quad (2.12)$$

Откуда следуют следующие выражения для определения параметров a и b

$$a = \bar{y} - b \cdot \bar{x}, \quad b = \frac{\overline{y \cdot x} - \bar{y} \cdot \bar{x}}{\overline{x^2} - \bar{x}^2}. \quad (2.13)$$

Формулу для параметра b можно представить следующим образом

$$b = \frac{\text{cov}(x, y)}{\sigma_x^2} = \frac{\frac{1}{n} \sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sigma_x^2}. \quad (2.14)$$

Рассмотрим интерпретацию параметров уравнения линейной регрессии.

Коэффициент b при факторной переменной x показывает насколько изменится в среднем величина y при изменении фактора x на единицу. Например, допустим, что зависимость между затратами (тыс. руб.) и объемом выпуска продукции описывается соотношением

$$y = 35000 + 0,58 \cdot x.$$

В этом случае увеличение объема выпуска на 1 единицу потребует дополнительных затрат на 580 рублей.

Что касается свободного члена a в уравнении (2.6), то в случае, когда переменная x представляет собой время, он показывает уровень явления в начальный момент времени. В других случаях, параметр a может не иметь экономической интерпретации.

2.4. Оценка параметров нелинейных моделей

Нелинейные уравнения регрессии можно разделить на два класса:

– уравнения, которые с помощью замены переменных можно привести к линейному виду в новых переменных x' , y'

$$y' = a' + b' \cdot x'; \quad (2.15)$$

– уравнения, для которых это невозможно. Назовем их внутренне нелинейными.

В первом случае, уравнения регрессии преобразуются к линейному виду с помощью введения новых (линеаризующих) переменных x' , y' . При этом предварительно формируются массивы значений $\{(x'_i, y'_i), i = 1, \dots, n\}$. В последующем, после определения параметров линейного уравнения регрессии с помощью обратного преобразования можно получить параметры исходного уравнения регрессии, представляющие интерес для исследователя.

Линеаризующие преобразования для некоторых нелинейных моделей приведены в таблице 2.2.

Таблица 2.2

Линеаризующие преобразования

Зависимость	Формула	Преобразование	Зависимость между параметрами
Гиперболическая	$y = a + \frac{b}{x}$	$y' = y$ $x' = \frac{1}{x}$	$a = a'$ $b = b'$
Логарифмическая	$y = a + b \cdot \ln x$	$y' = y$ $x' = \ln x$	$a = a'$ $b = b'$
Степенная	$\hat{y} = a \cdot x^b$	$y' = \ln y$ $x' = \ln x$	$\ln a = a'$ $b = b'$
Экспоненциальная	$\hat{y} = e^{a+b \cdot x}$	$y' = \ln y$ $x' = x$	$a = a'$ $b = b'$
Показательная	$\hat{y} = a \cdot b^x$	$y' = \ln y$ $x' = x$	$\ln a = a'$ $\ln b = b'$

Для оценки параметров внутренне нелинейных зависимостей также можно применить метод наименьших квадратов и определять оптимальные значения параметров a и b исходя из условия (2.8) или (2.9). Но в данном случае условия (2.10) уже не являются линейными алгебраическими уравнениями относительно параметров a и b , поэтому величины параметров a и b удобнее определять непосредственно из условия (2.9) как значения, доставляющие минимум величине S .

Итерационную процедуру минимизации S в общем виде можно представить в виде следующих последовательных шагов.

1. Задаются некоторые «правдоподобные» начальные (исходные) значения a^0 и b^0 параметров a и b .

2. Вычисляются теоретические значения $\hat{y}_i = f(x_i)$ с использованием этих значений параметров.

3. Вычисляются остатки $e_i = \hat{y}_i - y_i$ и сумма квадратов остатков $S = \sum (\hat{y}_i - y_i)^2$.

4. Вносятся изменения в одну или более оценку параметров.

5. Вычисляются новые теоретические значения \hat{y}_i , остатки e_i и S .

6. Если произошло уменьшение S , то новые значения оценок используются в качестве новой отправной точки.

7. Шаги 4, 5 и 6 повторяются до тех пор, пока не будет достигнута ситуация, когда величину S невозможно будет улучшить (в пределах заданной точности).

8. Полученные на последнем шаге значения параметров a и b являются оценками параметров уравнения регрессии, полученными по нелинейным методом наименьших квадратов.

Конкретные методы минимизации S отличаются способом выбора новых измененных значений оценок параметров.

2.5. Качество оценок МНК линейной регрессии.

Теорема Гаусса-Маркова

При использовании полученных различными способами оценок параметров уравнения регрессии (2.6) важно быть уверенными, являются ли они «лучшими» среди всех остальных в некотором смысле. Ответ на этот вопрос дает теорема Гаусса-Маркова, согласно которой оценки параметров линейной регрессии, полученные методом наименьших квадратов, будут несмещенными и эффективными (т. е. будут иметь наименьшую дисперсию) в классе линейных несмещенных оценок при выполнении четырех условий, известных как условия Гаусса-Маркова.

Эти условия принимаются в качестве основных предпосылок регрессионного анализа.

1-е условие Гаусса-Маркова: математическое ожидание случайного члена ε_i равно нулю в любом наблюдении

$$M(\varepsilon_i) = 0. \quad (2.16)$$

2-е условие Гаусса-Маркова: дисперсия случайного члена ε_i постоянна для всех наблюдений

$$D(\varepsilon_i) = \sigma_\varepsilon^2. \quad (2.17)$$

3-е условие Гаусса-Маркова: значения случайного члена в любых наблюдениях ε_i и ε_j не коррелируют между собой

$$\text{Cov}(\varepsilon_i, \varepsilon_j) = 0 \quad (i \neq j). \quad (2.18)$$

Это условие с учетом того, что $M(\varepsilon_i) = M(\varepsilon_j) = 0$ принимает вид

$$M(\varepsilon_i, \varepsilon_j) = 0 \quad (i \neq j). \quad (2.19)$$

4-е условие Гаусса-Маркова: случайный член должен быть распределен независимо от объясняющих переменных x_i

$$\text{Cov}(x_i, \varepsilon_i) = M(x_i, \varepsilon_i) = 0, \quad (2.20)$$

где было учтено, что $M(\varepsilon_i) = 0$.

Следует сказать, что последнее условие заведомо выполняется, если объясняющие переменные x_i считаются детерминированными величинами.

Выполнение 4-го условия Гаусса-Маркова обеспечивает несмещенность оценки параметра b .

Выполнение 1-го и 4-го условий Гаусса-Маркова обеспечивает несмещенность оценки параметра a .

Нарушение одного из условий Гаусса-Маркова приводит к нарушению эффективности оценок, т. е. в классе несмещенных оценок можно найти такие, которые имеют меньшую дисперсию.

В регрессионном анализе обычно делается еще одна предпосылка о нормальности распределения случайного члена, что позволяет выполнить количественную оценку точности полученных оценок параметров (2.13).

После построения модели необходимо вычислить значения остатков e_i и проверить выполнение условий Гаусса-Маркова, так как их нарушение снижает качество модели. Если условия нарушаются, то следует модернизировать модель соответствующим образом. Эти вопросы будут рассмотрены в 3 разделе.

2.6. Проверка качества уравнения регрессии. F-критерий Фишера

Оценка качества полученного уравнения регрессии основывается на методах дисперсионного анализа.

Наблюдаемые значения результативного признака y_i можно представить в виде суммы двух составляющих \hat{y}_i и e_i

$$y_i = \hat{y}_i + e_i. \quad (2.21)$$

Величина $\hat{y}_i = a + b \cdot x_i$ представляет собой расчетное значение переменной y в наблюдении i . Остаток e_i есть расхождение между наблюдаемым и расчетными значениями переменной y , или необъясненная с помощью уравнения регрессии часть переменной y .

Из (2.21) следует следующее соотношение между дисперсиями наблюдаемых значений переменной $D(y)$, ее расчетных значений $D(\hat{y})$ и остатков $D(e)$ (остаточной дисперсией $D_{ocm} = D(e)$)

$$D(y) = D(\hat{y}) + D(e). \quad (2.22)$$

Учитывая соотношения $D(y) = \frac{1}{n} \sum (y_i - \bar{y})^2$, $D(\hat{y}) = \frac{1}{n} \sum (\hat{y}_i - \bar{y})^2$,

$D(e) = D_{ocm} = \frac{1}{n} \sum (\hat{y}_i - y_i)^2$ и $M(e) = 0$ равенство (2.21) можно записать в виде

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (\hat{y}_i - y_i)^2. \quad (2.23)$$

Отношение объясненной части $D(\hat{y})$ дисперсии переменной y ко всей дисперсии $D(y)$

$$R^2 = \frac{D(\hat{y})}{D(y)} \quad \text{или} \quad R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2.24)$$

называют *коэффициентом детерминации* и используют для характеристики качества уравнения регрессии или соответствующей модели связи.

Соотношение (2.23) можно представить в альтернативном виде

$$R^2 = 1 - \frac{D_{\text{ост}}}{D(y)} \quad \text{или} \quad R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}. \quad (2.25)$$

Коэффициент детерминации R^2 принимает значения в диапазоне от нуля до единицы

$$0 \leq R^2 \leq 1.$$

Коэффициент детерминации R^2 показывает, какая часть дисперсии результативного признака y объяснена уравнением регрессии. Например, значение $R^2 = 0,56$ говорит о том, что соответствующее уравнение регрессии объясняет 56 % дисперсии результативного признака.

Чем больше R^2 , тем большая часть дисперсии результативного признака y объясняется уравнением регрессии и тем лучше уравнение регрессии описывает исходные данные. При отсутствии зависимости между y и x коэффициент детерминации R^2 будет близок к нулю. Таким образом, коэффициент детерминации R^2 может применяться для оценки качества (точности) уравнения регрессии.

Возникает вопрос, при каких значениях R^2 уравнение регрессии следует считать статистически незначимым, что делает необоснованным его использование в анализе. Ответ на этот вопрос дает F -критерий Фишера.

Введем следующие обозначения:

$$TSS = \sum_{i=1}^n (y_i - \bar{y})^2 - \text{полная сумма квадратов отклонений};$$

$$ESS = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 - \text{объясненная сумма квадратов отклонений};$$

$$RSS = \sum_{i=1}^n (\hat{y}_i - y_i)^2 = \sum_{i=1}^n e_i^2 - \text{необъясненная сумма квадратов отклонений}.$$

Известно, что величина

$$F = \frac{\frac{ESS}{k}}{\frac{RSS}{n - k - 1}}, \quad (2.26)$$

где k – число независимых переменных в уравнении регрессии (для парной регрессии $k = 1$), в случае нормально распределенной ошибки ε_i является F -статистикой Фишера (случайная величина, распределенная по закону Фишера) с числом степеней свободы $k_1 = k$, $k_2 = n - k - 1$.

Согласно F -критерию Фишера, выдвигается «нулевая» гипотеза H_0 о статистической незначимости уравнения регрессии (т. е. о статистически незначимом отличии величины F от нуля). Эта гипотеза отвергается при выполнении условия $F > F_{\text{крит}}$, где $F_{\text{крит}}$ определяется по таблицам F -критерия Фишера (ПЗ,

П4) при числе степеней свободы $k_1 = k$, $k_2 = n - k - 1$ и заданному уровню значимости α .

Уровень значимости (обозначается α) в статистических гипотезах называется вероятностью отвергнуть верную гипотезу (это, так называемая, ошибка первого рода). Уровень значимости α обычно принимает значения 0,05 и 0,01, что соответствует вероятности совершения ошибки первого рода 5 % и 1 %.

Используя соотношение (2.24), величину F можно выразить через коэффициент детерминации R^2

$$F = \frac{R^2}{1 - R^2} \frac{n - k - 1}{k}. \quad (2.27)$$

Например, по данным 30 наблюдений было получено уравнение регрессии $y = 50,5 + 3,2x$ и $R^2 = 0,60$.

Необходимо проверить его значимость при уровне значимости $\alpha = 0,05$.

Определим величину F -статистики, учитывая, что $k = 1$

$$F = \frac{R^2}{1 - R^2} \cdot \frac{n - k - 1}{k} = \frac{0,6}{1 - 0,6} \cdot \frac{30 - 1 - 1}{1} = \frac{0,6}{0,4} \cdot 28 = 42.$$

По таблицам F -критерия Фишера при

$$k_1 = k = 1, k_2 = n - k - 1 = 30 - 1 - 1 = 28 \text{ и } \alpha = 0,05$$

находим $F_{\text{крит}} = 4,20$. Так как $F = 42 > F_{\text{крит}} = 4,20$, то делаем вывод о статистической значимости уравнения регрессии.

2.7. Коэффициенты корреляции. Оценка тесноты связи

Наряду с построением уравнения регрессии осуществляется оценка тесноты связи между явлениями (между переменными).

Тесноту связи в случае линейной зависимости характеризуют с помощью выборочного коэффициента корреляции r_{xy}

$$r_{xy} = \frac{\frac{1}{n} \sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sigma_x \sigma_y} = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y} \quad (2.28)$$

или

$$r_{xy} = \frac{\overline{yx} - \bar{y} \cdot \bar{x}}{\sigma_x \sigma_y}. \quad (2.29)$$

Здесь σ_x , σ_y – средние квадратические отклонения переменных x и y .

Выборочный коэффициент корреляции r_{xy} связан с коэффициентом линейной регрессии b соотношением

$$r_{xy} = b \frac{\sigma_x}{\sigma_y}. \quad (2.30)$$

Выборочный коэффициент корреляции r_{xy} принимает значения в диапазоне $-1 \leq r_{xy} \leq 1$.

Чем ближе величина $|r_{xy}|$ к единице, тем теснее линейная связь и тем лучше линейная зависимость согласуется с данными наблюдений. При $|r_{xy}| = 1$ связь становится функциональной, т. е. соотношение $\hat{y}_i = a + b \cdot x_i$ выполняется для всех наблюдений.

При $r_{xy} > 0$ связь является прямой, при $r_{xy} < 0$ – обратной.

Тесноту нелинейной связи (задаваемой уравнением нелинейной регрессии $\hat{y} = f(x)$) оценивают с помощью индекса корреляции R

$$R = \sqrt{R^2} = \sqrt{1 - \frac{D_{ост}}{D(y)}} = \sqrt{1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}}. \quad (2.31)$$

Индекс корреляции R принимает значения в диапазоне $0 \leq R \leq 1$.

Чем ближе величина R к единице, тем теснее данная связь, тем лучше зависимость $\hat{y} = f(x)$ согласуется с данными наблюдений. При $R = 1$ ($R^2 = 1$) связь становится функциональной, т. е. соотношение $y_i = f(x_i)$ выполняется для всех наблюдений.

В случае линейной зависимости выборочный коэффициент корреляции r_{xy} и коэффициент детерминации R^2 связаны соотношением

$$r_{xy}^2 = R^2. \quad (2.32)$$

Коэффициент детерминации R^2 может использоваться для оценки качества (точности) построенной модели регрессии. Чем выше этот показатель, тем лучше модель описывает исходные данные.

2.8. Точность коэффициентов регрессии. Проверка значимости

Полученные согласно формулам (2.13) оценки коэффициентов регрессии зависят от используемой выборки значений переменных x и y и являются случайными величинами. Представление о точности полученных оценок, о том насколько далеко они могут отклониться от истинных значений коэффициентов можно получить используя, так называемые «стандартные ошибки» коэффициентов регрессии.

Под стандартной ошибкой коэффициента регрессии понимается оценка стандартного отклонения функции плотности вероятности коэффициента.

Стандартные ошибки коэффициентов регрессии (s_a , s_b) определяются соотношениями

$$s_b = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2 / (n-2)}{\sum_{i=1}^n (x_i - \bar{x})^2}} = \sqrt{\frac{s_{ocm}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} = \frac{s_{ocm}}{\sigma_x \sqrt{n}}, \quad (2.33)$$

$$s_a = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{(n-2)} \cdot \frac{\sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2}} = \sqrt{s_{ocm}^2 \frac{\sum_{i=1}^n x_i^2}{n^2 \sigma_x^2}} = s_{ocm} \frac{\sqrt{\sum_{i=1}^n x_i^2}}{n \sigma_x}, \quad (2.34)$$

где $s_{\text{ост}}^2$ представляет собой несмещенную оценку остаточной дисперсии

$$s_{\text{ост}}^2 = \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{(n-2)}. \quad (2.35)$$

Сопоставляя оценки параметров и их стандартные ошибки можно сделать вывод о надежности (точности) полученных оценок.

Отношения

$$t_a = \frac{a - \tilde{a}}{s_b} \quad \text{и} \quad t_b = \frac{b - \tilde{b}}{s_b} \quad (2.36)$$

в случае нормально распределенной ошибки ε_i являются t -статистиками, т. е. случайными величинами, распределенными по закону Стьюдента с числом степеней свободы $n-2$. Через \tilde{a} и \tilde{b} обозначены точные значения коэффициентов регрессии.

Для оценки статистической значимости коэффициентов регрессии применяется t -критерий Стьюдента, согласно которому выдвигается «нулевая» гипотеза H_0 о статистической незначимости коэффициента уравнения регрессии (т. е. о статистически незначимом отличии величины a или b от нуля). Эта гипотеза отвергается при выполнении условия $t > t_{\text{крит}}$, где $t_{\text{крит}}$ определяется по таблицам t -критерия Стьюдента (П2) по числу степеней свободы $k_1 = n-k-1$ (k – число независимых переменных в уравнении регрессии) и заданному уровню значимости α .

t -критерий Стьюдента может использоваться и для оценки статистической значимости выборочного коэффициента корреляции r_{xy} , так как величина

$$t_{xy} = \frac{r_{xy}}{s_{r_{xy}}}, \quad (2.37)$$

где

$$s_{r_{xy}} = \sqrt{\frac{1 - r_{xy}^2}{n-2}} \quad (2.38)$$

распределена по закону Стьюдента с числом степеней свободы $n-2$. Через $s_{r_{xy}}$ обозначена стандартная ошибка коэффициента корреляции r_{xy} .

Проверка значимости оценок параметров ничего не говорит о том, насколько эти оценки могут отличаться от точных значений. Ответ на этот вопрос дает построение доверительных интервалов.

Под доверительным интервалом понимаются пределы, в которых лежит точное значение определяемого показателя с заданной вероятностью ($P = 1-\alpha$).

Доверительные интервалы для параметров a и b уравнения линейной регрессии определяются соотношениями:

$$a \pm t_{1-\alpha, n-2} \cdot s_a; \quad b \pm t_{1-\alpha, n-2} \cdot s_b. \quad (2.39)$$

Величина $t_{1-\alpha, n-2}$ представляет собой табличное значение t -критерия Стьюдента на уровне значимости α при числе степеней свободы $n-2$.

Если в границы доверительного интервала попадает ноль, т. е. нижняя граница отрицательна, а верхняя положительна, то оцениваемый параметр принимается равным нулю, так как он не может одновременно принимать и положительное, и отрицательное значения.

Для статистически значимого коэффициента корреляции r_{xy} интервальные оценки (доверительный интервал) получают с использованием Z -преобразования Фишера:

$$Z = Z(r_{xy}) = \frac{1}{2} \cdot \ln \frac{1+r_{xy}}{1-r_{xy}}. \quad (2.40)$$

Первоначально определяется интервальная оценка для z

$$z \in \left[z' \pm t_{1-\alpha/2} \cdot \sqrt{\frac{1}{n-3}} \right], \quad (2.41)$$

где $t_{1-\alpha/2}$ – квантиль стандартного нормального распределения порядка $1-\alpha/2$, $z' = Z(r_{xy})$ – значение Z -преобразования Фишера, соответствующее полученному значению коэффициента корреляции r_{xy} .

Граничные значения доверительного интервала (r^- , r^+) для r_{xy} получаются из граничных значений доверительного интервала (z^- , z^+) для z с помощью обратного Z -преобразования Фишера $r_{xy} = Z^{-1}(z)$

$$r^- = Z^{-1}(z^-); \quad r^+ = Z^{-1}(z^+). \quad (2.42)$$

2.9. Точечный и интервальный прогноз по уравнению линейной регрессии

Точечный прогноз заключается в получении прогнозного значения y_p , которое определяется путем подстановки в уравнение регрессии $\hat{y}_x = a + b \cdot x$ соответствующего (прогнозного) значения x_p

$$y_p = a + b \cdot x_p.$$

Интервальный прогноз заключается в построении доверительного интервала прогноза, т. е. нижней и верхней границ y_{pmin} , y_{pmax} интервала, содержащего точную величину для прогнозного значения y_p ($y_{pmin} < y_p < y_{pmax}$) с заданной вероятностью.

При построении доверительного интервала прогноза используется стандартная ошибка прогноза S_{y_p} , связанная с дисперсией ошибки прогноза $S_{y_p}^2$ соотношением

$$S_{y_p} = \sqrt{S_{y_p}^2}.$$

Дисперсия ошибки прогноза $s_{y_p}^2$ представляет собой сумму дисперсии $s_{\hat{y}_p}^2$ ошибки прогноза расчетного значения $\hat{y}_p = a + b \cdot x_p$ и остаточной дисперсии $s_{\text{ост}}^2$ (2.34)

$$s_{y_p}^2 = s_{\hat{y}_p}^2 + s_{\text{ост}}^2. \quad (2.43)$$

Величина дисперсии $s_{\hat{y}_p}^2$ находится из соотношения

$$\hat{y} = \bar{y} + b \cdot (x - \bar{x})$$

и составляет

$$s_{\hat{y}_p}^2 = s_{\text{ост}}^2 \cdot \left(\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right). \quad (2.44)$$

Соответственно стандартные ошибки прогноза расчетного значения по уравнению регрессии и индивидуального значения прогноза $S_{\hat{y}_p}$ и S_{y_p} определяются соотношениями

$$S_{\hat{y}_p} = s_{\text{ост}} \cdot \sqrt{\frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}, \quad (2.45)$$

$$S_{y_p} = s_{\text{ост}} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}. \quad (2.46)$$

Доверительные интервалы прогноза определяются соотношениями: для расчетного значения по уравнению регрессии \hat{y}_p

$$\hat{y}_p - t_{1-\alpha; n-2} \cdot S_{\hat{y}_p} \leq \hat{y}_p \leq \hat{y}_p + t_{1-\alpha; n-2} \cdot S_{\hat{y}_p}, \quad (2.47)$$

для индивидуального значения прогноза y_p

$$\hat{y}_p - t_{1-\alpha; n-2} \cdot S_{y_p} \leq \hat{y}_p \leq \hat{y}_p + t_{1-\alpha; n-2} \cdot S_{y_p}, \quad (2.48)$$

где величина $t_{1-\alpha; n-2}$ представляет собой табличное значение t -критерия Стьюдента на уровне значимости α при числе степеней свободы $n-2$.

2.10. Коэффициент эластичности

В экономических исследованиях широкое применение находит такой показатель, как коэффициент эластичности. Если зависимость между переменными x и y имеет вид $y = f(x)$, то коэффициент эластичности \mathcal{E} вычисляется по формуле

$$\mathcal{E} = f'(x) \frac{x}{y}. \quad (2.49)$$

Коэффициент эластичности \mathcal{E} показывает, на сколько процентов в среднем изменится результативный признак y при изменении фактора x на 1 % от своего номинального значения. Для линейной регрессии $y = a + b \cdot x$ коэффициент эластичности равен

$$\mathcal{E} = b \frac{x}{y}.$$

Коэффициент эластичности \mathcal{E} в общем случае зависит от величины x и является величиной переменной. Чтобы исключить эту зависимость применяется средний коэффициент эластичности $\bar{\mathcal{E}}$

$$\bar{\mathcal{E}} = f'(\bar{x}) \frac{\bar{x}}{\bar{y}} = b \frac{\bar{x}}{\bar{y}}, \quad (2.50)$$

который уже является величиной постоянной.

Средний коэффициент эластичности $\bar{\mathcal{E}}$ показывает, на сколько процентов в среднем по совокупности значений фактора x изменится результативный признак y при изменении фактора x на 1 %.

Для степенной регрессии $y = a \cdot x^b$ коэффициент эластичности равен b и является величиной постоянной. Отсюда следует интерпретация параметра b в уравнении степенной регрессии: параметр b показывает, на сколько процентов изменится результативный признак y при изменении фактора x на 1 %.

Контрольные вопросы

1. Что понимается под регрессией в теории вероятностей и математической статистике?
2. Какие задачи решаются при построении уравнения регрессии?
3. Какие методы применяются для выбора вида модели регрессии?
4. Какие функции чаще всего используются для построения уравнения парной регрессии?
5. Какой вид имеет система нормальных уравнений метода наименьших квадратов?
6. Как осуществляется оценка параметров нелинейных моделей?
7. Назовите условия Гаусса-Маркова. О чем говорит теорема Гаусса-Маркова?
8. Что при проверке статистических гипотез называют уровнем значимости?
9. Как проверяется значимость уравнения регрессии?
10. Как проверяется значимость коэффициентов уравнения регрессии?
11. Как вычисляется коэффициент детерминации R^2 ?
12. По какой формуле вычисляется выборочный коэффициент парной корреляции r_{xy} ?
13. Как проверяется значимость выборочного коэффициента парной корреляции?
14. Как строится доверительный интервал для линейного коэффициента парной корреляции?
15. Как вычисляется и что показывает индекс детерминации?
16. Как осуществляется построение доверительного интервала прогноза в случае линейной регрессии?
17. Как вычисляется и как интерпретируется коэффициент эластичности \mathcal{E} ?

3. Множественный регрессионный анализ

3.1. Понятие множественной регрессии

Множественной регрессией называют уравнение связи с несколькими независимыми переменными:

$$\hat{y} = f(x_1, x_2, \dots, x_p). \quad (3.1)$$

Переменная y называется зависимой, объясняемой или результативным признаком. x_1, x_2, \dots, x_p – независимые, объясняющие переменные или факторные признаки (факторы).

Соответствующая регрессионная модель имеет вид

$$y = f(x_1, x_2, \dots, x_p) + \varepsilon, \quad (3.2)$$

где ε – ошибка модели, являющаяся случайной величиной.

Множественная регрессия применяется в ситуациях, когда из множества факторов, влияющих на результативный признак, нельзя выделить один доминирующий фактор и необходимо учитывать влияние нескольких факторов. Например, объем выпуска продукции определяется величиной основных и оборотных средств, численностью персонала, уровнем менеджмента и т. д., уровень спроса зависит не только от цены, но и от имеющихся у населения денежных средств.

Основная цель множественной регрессии – построить модель с несколькими факторами и определить при этом влияние каждого фактора в отдельности, а также их совместное воздействие на изучаемый показатель.

Постановка задачи множественной регрессии: по имеющимся данным n наблюдений (табл. 3.1) за совместным изменением $p+1$ параметра y и x_j и $((y_i, x_{j,i}); j=1, 2, \dots, p; i=1, 2, \dots, n)$ необходимо определить аналитическую зависимость $\hat{y} = f(x_1, x_2, \dots, x_p)$, наилучшим образом описывающую данные наблюдений.

Таблица 3.1

Результаты наблюдений

	y	x_1	x_2	...	x_p
1	y_1	x_{11}	x_{21}	...	x_{p1}
2	y_2	x_{12}	x_{22}	...	x_{p2}
...
n	y_n	x_{1n}	x_{2n}	...	x_{pn}

Каждая строка таблицы содержит $p + 1$ число и представляет собой результат одного наблюдения. Наблюдения различаются условиями их проведения.

Вопрос о том, какую зависимость следует считать наилучшей, решается на основе какого-либо критерия. В качестве такого критерия обычно используется минимум суммы квадратов отклонений расчетных или модельных значений результативного показателя $\hat{y}_i = f(x_{1i}, x_{2i}, \dots, x_{pi})$ от наблюдаемых значений y_i

$$S = \sum (\hat{y}_i - y_i)^2 = \min.$$

Как и в случае парной регрессии, построение уравнения множественной регрессии предполагает решение двух задач (или, другими словами, осуществляется в два этапа):

- 1) спецификация модели;
 - 2) оценка параметров выбранной модели.
- В свою очередь, спецификация модели включает в себя решение двух задач:
- отбор p факторов x_j , подлежащих включению в модель;
 - выбор вида аналитической зависимости $\hat{y} = f(x_1, x_2, \dots, x_p)$.

3.2. Отбор факторов при построении множественной регрессии

3.2.1. Требования к факторам

Процесс отбора факторов в достаточно сложных ситуациях является итерационной процедурой, предполагающей, в частности, построение уравнений регрессии, и включает два этапа. Первоначально отбор факторов осуществляется на основе качественных соображений, исходя из представлений о природе взаимосвязи моделируемого показателя с другими экономическими показателями. На следующем этапе отобранные факторы подвергаются проверке на статистическую значимость. Окончательное решение о включении фактора в модель основывается на количественной оценке степени влияния фактора на изучаемый показатель.

К факторам, включаемым в модель, предъявляются следующие требования:

1. Факторы не должны быть взаимно коррелированы и, тем более, находиться в точной функциональной связи. Наличие высокой степени коррелированности между факторами может привести к неустойчивости и ненадежности оценок коэффициентов регрессии, а также к невозможности выделить изолированное влияние факторов на результативный показатель.

2. Включение фактора в модель должно приводить к существенному увеличению доли объясненной части в общей вариации зависимой переменной. Так как данная величина характеризуется таким показателем, как коэффициент детерминации R^2 , включение фактора в модель должно приводить к заметному изменению последнего. Формальная проверка существенности вклада фактора в модель выполняется с помощью оценки значимости соответствующего частного коэффициента корреляции либо значимости коэффициента в уравнении регрессии.

Если необходимо учесть влияние качественного фактора (не имеющего количественной оценки), то в модель включается соответствующая ему «фиктивная» переменная, имеющая конечное количество формально численных значений, соответствующих градациям качественного фактора. Например, если нужно учесть влияние уровня образования (на размер заработной платы), то в уравнение регрессии можно включить переменную z , принимающую значения $z = 0$ при начальном образовании, 1 – при среднем, 2 – при высшем.

Если для какого-либо показателя, который представляется важным для данного исследования, отсутствуют исходные данные, либо сам показатель четко не определен, то может быть полезно включить в модель некоторый ее «заменитель». Например, в качестве показателя качества образования можно использовать число преподавателей или расходы на одного студента. Такой подход основан на том факте, что неучет существенного показателя приводит к

смещенным оценкам параметров. Например, производственная функция Кобба-Дугласа, построенная по данным экономики США за период 1949–1978 гг., построенная с учетом времени в качестве замещающей переменной для показателя технического прогресса имеет вид [4]

$$\log \hat{Y} = -1,03 + 0,17 \log K + 0,93 \log L + 0,024t,$$

(2,33) (0,66) (0,17) 0,016)

а без учета имеет вид

$$\log \hat{Y} = -4,50 + 1,19 \log K + 0,77 \log L,$$

(0,57) (0,10) (0,15)

где Y – индекс объема выпуска частного сектора; K – индекс затрат капитала; L – индекс затрат труда; t – время, равное единице в 1948 г. и т. д. Без учета замещающей переменной коэффициент при $\log K$ неправдоподобно велик.

При отборе факторов в модель следует, по возможности, стремиться к минимизации количества факторов, так как неоправданное их увеличение приводит к затруднениям в интерпретации модели и снижению достоверности результатов.

3.2.2. Мультиколлинеарность

Под мультиколлинеарностью понимается высокая взаимная коррелированность объясняющих переменных. Следствием мультиколлинеарности является линейная зависимость между столбцами наблюдений x_{ij} в таблице 3.1 или между столбцами матрицы X (3.11). В результате, матрица $X'X$ становится плохо обусловленной, что приводит к неустойчивости оценок коэффициентов регрессии, когда незначительные изменения данных наблюдений приводят к значительным изменениям оценок.

Проверка наличия мультиколлинеарности основывается на анализе матрицы парных корреляций между факторами

$$R = \begin{bmatrix} r_{x_1x_1} & r_{x_1x_2} & \dots & r_{x_1x_p} \\ r_{x_2x_1} & r_{x_2x_2} & \dots & r_{x_2x_p} \\ \dots & \dots & \dots & \dots \\ r_{x_px_1} & r_{x_px_2} & \dots & r_{x_px_p} \end{bmatrix} = \begin{bmatrix} 1 & r_{x_1x_2} & \dots & r_{x_1x_p} \\ r_{x_2x_1} & 1 & \dots & r_{x_2x_p} \\ \dots & \dots & \dots & \dots \\ r_{x_px_1} & r_{x_px_2} & \dots & 1 \end{bmatrix} \quad (3.3)$$

Коэффициенты парной корреляции $r_{x_i x_j}$ между объясняющими переменными используются для выявления дублирующих факторов. Линейная зависимость между объясняющими переменными x_i и x_j считается установленной, если выполняется условие $r_{x_i x_j} \geq 0,8$, а сами факторы называются явно коллинеарными (эмпирическое правило). Один из факторов должен быть исключен из модели. Предпочтение при этом отдается тому фактору, который при достаточно тесной связи с результатом имеет наименьшую тесноту связи с другими факторами.

Наряду с парной коллинеарностью может иметь место линейная зависимость между более, чем двумя переменными. Для оценки мультиколлинеарности факторов в этом случае может использоваться величину определителя $\text{Det}|R|$

матрицы парных коэффициентов корреляции $r_{x_i x_j}$ между факторами либо ее минимального собственного значения.

Чем ближе к нулю определитель (минимальное собственное значение) матрицы межфакторной корреляции, тем сильнее мультиколлинеарность между факторами и тем ненадежнее результаты множественной регрессии.

Для оценки статистической значимости мультиколлинеарности факторов может быть использован тот факт, что величина $\left[n - 1 - \frac{1}{6}(2m + 5) \lg \text{Det} R \right]$ имеет приближенное распределение χ^2 с $df = \frac{1}{2} p(p - 1)$ степенями свободы.

Выдвигается гипотеза H_0 о независимости переменных, т. е. $\text{Det}|R| = 1$. Если фактическое значение χ^2 превосходит табличное (критическое) $\chi_{\text{факт}}^2 > \chi_{\text{табл}(df, \alpha)}^2$, то гипотеза H_0 отклоняется и мультиколлинеарность считается доказанной.

Для выявления мультиколлинеарности факторов можно использовать коэффициенты множественной детерминации $R_{x_1|x_2x_3\dots x_p}^2; R_{x_2|x_1x_3\dots x_p}^2 \dots$, полученные по уравнениям регрессии, в которых качестве зависимой переменной рассматривается один из факторов. Чем ближе значение коэффициента детерминации к единице, тем сильнее проявляется мультиколлинеарность факторов. Согласно эмпирическому правилу, при значении коэффициента множественной детерминации $R_{x_1|x_2x_3\dots x_p}^2 > 0,6$ мультиколлинеарность факторов считается установленной. Оставляя в уравнении регрессии факторы с минимальной величиной коэффициента множественной детерминации, можно исключить мультиколлинеарность факторов.

Для преодоления явления линейной зависимости между факторами используются такие способы, как:

- исключение одного из коррелирующих факторов;
- переход с помощью линейного преобразования к новым некоррелирующим независимым переменным. Например, переход к главным компонентам вектора исходных объясняющих переменных (что позволяет также уменьшить количество рассматриваемых факторов), переход к последовательным разностям во временных рядах $\Delta x_{it} = x_{it} - x_{it-1}$ и т. п.;

- переход к смещенным оценкам, имеющим меньшую дисперсию. В частности, при использовании «ридж-регрессии» применяются смещенные оценки вектора параметров $b_\tau = (X'X + \tau E_{p+1})^{-1} X'Y$ (п. 3.4), где τ – некоторое положительное число, E_{p+1} – единичная матрица порядка $p+1$. Такое преобразование увеличивает определитель матрицы системы нормальных уравнений и повышает устойчивость результатов (снижает дисперсию оценок, которые становятся смещенными).

Другие аспекты вопроса отбора факторов рассмотрены в п. 1.5.

Следует также учитывать ограничение, накладываемое на количество факторов, имеющимся числом наблюдений. Количество наблюдений должно превышать количество факторов более чем в 6-7 раз.

3.3. Выбор формы уравнения регрессии

Различают следующие виды уравнений множественной регрессии: линейные, нелинейные, сводящиеся к линейным, и нелинейные, не сводящиеся к линейным (внутренне нелинейные). В первых двух случаях для оценки параметров модели применяются классического линейного регрессионного анализа. В случае внутренне нелинейных уравнений для оценки параметров приходится применять методы нелинейной оптимизации.

Основное требование, предъявляемое к уравнениям регрессии, заключается в наличии наглядной экономической интерпретации модели и ее параметров.

Исходя из этих соображений, наиболее часто используются линейная и степенная зависимости.

Линейная множественная регрессия имеет вид

$$\hat{y} = a + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_p \cdot x_p. \quad (3.4)$$

Параметры b_i при факторах x_i называются коэффициентами «чистой» регрессии. Они показывают, на сколько единиц в среднем изменится результативный признак y за счет изменения соответствующего фактора на единицу при неизменном значении других факторов, закрепленных на среднем уровне.

Предположим, например, что зависимость спроса на товар (Q^d) от цены (P) и дохода (I) характеризуется следующим уравнением:

$$Q^d = 2,5 - 0,12P + 0,23 I.$$

Коэффициенты данного уравнения говорят о том, что при увеличении цены на единицу, спрос уменьшится в среднем на 0,12 единиц измерения спроса, а при увеличении дохода на единицу, спрос возрастет в среднем 0,23 единицы.

Параметр a в (3.14) не всегда может быть содержательно проинтерпретирован.

Степенная множественная регрессия имеет вид

$$\hat{y} = a \cdot x_1^{b_1} \cdot x_2^{b_2} \cdot \dots \cdot x_p^{b_p} \quad (3.5)$$

Параметры b_j (степени факторов x_i) являются коэффициентами эластичности. Они показывают, на сколько процентов в среднем изменится результативный признак y за счет изменения соответствующего фактора x_i на 1 % при неизменном значении остальных факторов.

Наиболее широкое применение этот вид уравнения регрессии получил в производственных функциях, а также при исследовании спроса и потребления.

Например, зависимость выпуска продукции Y от затрат капитала K и труда L

$$Y = 0,89 K^{0,23} L^{0,81}$$

говорит о том, что увеличение затрат капитала K на 1 % при неизменных затратах труда вызывает увеличение выпуска продукции Y на 0,23 %. Увеличение затрат труда L на 1 % при неизменных затратах капитала K вызывает увеличение выпуска продукции Y на 0,81 %.

Экономический смысл имеет также сумма коэффициентов b_i каждого фактора (сумма эластичностей) $b = \sum b_i$. Эта величина дает обобщенную характеристику эластичности производства.

Если значение $b > 1$, то говорят, что функция имеет возрастающий эффект от масштаба производства. Значение $b = 1$ говорит о постоянном масштабе производства. Если значение $b < 1$, то имеет место убывающий эффект от масштаба производства.

Примеры других зависимостей, используемых при построении регрессии, приведены в п. 1.4.

Если один и тот же фактор вводится в регрессию в разных степенях, то каждая степень рассматривается как самостоятельный фактор. Например, если в нелинейной модели с двумя факторами x_1, x_2

$$y = a + b_1 \cdot x_1 + b_2 \cdot x_2 + b_3 \cdot x_1^2 + b_4 \cdot x_1 \cdot x_2^2 + \varepsilon,$$

величины $x_1^2, x_1 \cdot x_2^2$ рассматривать как новые дополнительные факторы, то, используя замену переменных $z_1 = x_1, z_2 = x_2, z_3 = x_1^2, z_4 = x_1 x_2^2$, ее можно привести к линейному уравнению регрессии с четырьмя факторами:

$$y = a + b_1 \cdot z_1 + b_2 \cdot z_2 + b_3 \cdot z_3 + b_4 \cdot z_4 + \varepsilon.$$

3.4. Оценка параметров уравнения линейной множественной регрессии

Рассмотрим уравнение линейной множественной регрессии

$$y = a + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_p \cdot x_p + \varepsilon. \quad (3.6)$$

Для оценки параметров уравнения множественной регрессии обычно применяется метод наименьших квадратов (МНК), согласно которому следует выбирать такие значения параметров a и b_i , при которых сумма квадратов отклонений фактических значений результативного признака y_i от теоретических значений $\hat{y} = f(x_{1i}, x_{2i}, \dots, x_{pi})$ (при тех же значениях фактора x_{ij}) минимальна, т. е.

$$S = \sum (\hat{y}_i - y_i)^2 = \min.$$

С учетом (3.6) величина S является функцией неизвестных параметров a и b_i

$$S = \sum_{i=1}^n (y_i - a - b_1 \cdot x_1 - b_2 \cdot x_2 - \dots - b_p \cdot x_p)^2 = S(a, b_1, \dots, b_p). \quad (3.7)$$

Оптимальные значения параметров a и b_i удовлетворяют условиям

$$\frac{\partial S}{\partial a} = 0, \quad \frac{\partial S}{\partial b_1} = 0, \quad \frac{\partial S}{\partial b_2} = 0, \quad \dots \quad \frac{\partial S}{\partial b_p} = 0. \quad (3.8)$$

Выполняя соответствующие вычисления, получим для определения параметров a и b_i следующую систему уравнений

$$\begin{aligned}
\frac{\partial S}{\partial a} &= -2 \sum_{i=1}^n (y_i - a - b_1 \cdot x_1 - b_2 \cdot x_2 - \dots - b_p \cdot x_p), \\
\frac{\partial S}{\partial b_1} &= -2b_1 \sum_{i=1}^n (y_i - a - b_1 \cdot x_1 - b_2 \cdot x_2 - \dots - b_p \cdot x_p), \\
&\dots \\
\frac{\partial S}{\partial b_p} &= -2b_p \sum_{i=1}^n (y_i - a - b_1 \cdot x_1 - b_2 \cdot x_2 - \dots - b_p \cdot x_p),
\end{aligned} \tag{3.9}$$

откуда после некоторых преобразований получается система нормальных уравнений метода наименьших квадратов

$$\begin{aligned}
\sum y &= n \cdot a + b_1 \sum x_1 + b_2 \sum x_2 + \dots + b_p \sum x_p; \\
\sum yx_1 &= a \sum x_1 + b_1 \sum x_1^2 + b_2 \sum x_2x_1 + \dots + b_p \sum x_px_1; \\
&\dots \\
\sum yx_p &= a \sum x_p + b_1 \sum x_1x_p + b_2 \sum x_2x_p + \dots + b_p \sum x_p^2.
\end{aligned} \tag{3.10}$$

Решение системы (3.10) удобно записать с помощью матричных обозначений. Обозначим

$$B = \begin{bmatrix} a \\ b_1 \\ \dots \\ b_p \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1x_{11} \dots x_{p1} \\ 1x_{12} \dots x_{p2} \\ \dots \\ 1x_{1n} \dots x_{pn} \end{bmatrix}, \tag{3.11}$$

где B – матрица-столбец $(p+1 \times 1)$ из коэффициентов a и b_i ;

Y – матриц-столбец $(n \times 1)$ исходных значений зависимой переменной y ;

X – матрица $(p+1 \times n)$ исходных значений независимых переменных x_i , в которой первый столбец из единиц можно рассматривать как значения «фиктивной» переменной, соответствующей коэффициенту a .

В этих обозначениях система (3.10) примет вид

$$(X'X)B = X'Y, \tag{3.12}$$

где X' – транспонированная матрица X . Матрица $X'X$ является неособенной квадратной размерности $(p+1 \times p+1)$ при условии, что столбцы матрицы X линейно независимы.

Решение системы (3.12) определяется соотношением

$$B = (X'X)^{-1} X'Y. \tag{3.13}$$

Независимые переменные x_i имеют различный экономический смысл, разные единицы измерения и масштаб. Если нужно определить степень относительного влияния отдельных факторов x_i на изменение результативной переменной y , то переменные x_i следует привести к сопоставимому виду. Это можно осуществить, вводя, так называемые, «стандартизованные» переменные $t_y, t_{x_1}, \dots, t_{x_p}$ с помощью соотношений

$$t_y = \frac{y - \bar{y}}{\sigma_y}, \quad t_{xi} = \frac{x_i - \bar{x}_i}{\sigma_{x_i}}, \quad (i = 1, 2, \dots, p) \quad (3.14)$$

где \bar{y}, \bar{x}_i – средние значения, σ_y, σ_{x_i} – средние квадратические отклонения переменных y и x_i .

Стандартизованные переменные обладают следующими свойствами: 1) средние значения равны нулю $\bar{t}_y = \bar{t}_{x_i} = 0$; 2) средние квадратические отклонения равны единице $\sigma_{t_y} = \sigma_{t_{x_i}} = 1$.

Уравнения множественной регрессии в стандартизованных переменных принимает вид

$$t_y = \beta_1 \cdot t_{x_1} + \beta_2 \cdot t_{x_2} + \dots + \beta_p \cdot t_{x_p} + \varepsilon. \quad (3.15)$$

Величины β_i называются стандартизованными коэффициентами. Их связь коэффициентами множественной регрессии b_i задается соотношениями

$$b_i = \beta_i \frac{\sigma_y}{\sigma_{x_i}} \quad \text{или} \quad \beta_i = b_i \frac{\sigma_{x_i}}{\sigma_y} \quad (i = 1, 2, \dots, p). \quad (3.16)$$

Параметр a уравнения (3.6) можно определить из соотношения

$$a = \bar{y} - b_1 \bar{x}_1 - b_2 \bar{x}_2 - \dots - b_p \bar{x}_p. \quad (3.17)$$

Стандартизованные коэффициенты регрессии β_i показывают, на сколько сигм (средних квадратических отклонений) изменится в среднем результативный признак y за счет изменения соответствующего фактора на одну сигму при неизменном значении других факторов, закрепленных на среднем уровне.

Система нормальных уравнений МНК (3.10) в стандартизованных переменных принимает вид:

$$\begin{aligned} r_{yx_1} &= \beta_1 + \beta_2 r_{x_2 x_1} + \beta_3 r_{x_3 x_1} + \dots + \beta_p r_{x_p x_1}; \\ r_{yx_2} &= \beta_1 r_{x_1 x_2} + \beta_2 + \beta_3 r_{x_3 x_2} + \dots + \beta_p r_{x_p x_2}; \\ &\dots \\ r_{yx_p} &= \beta_1 r_{x_1 x_p} + \beta_2 r_{x_2 x_p} + \beta_3 r_{x_3 x_p} + \dots + \beta_p. \end{aligned} \quad (3.18)$$

Стандартизованные коэффициенты регрессии β_i сравнимы между собой, что позволяет ранжировать факторы по силе их воздействия на результат. Больше относительное влияние на изменение результативной переменной y оказывает тот фактор, которому соответствует большее по модулю значение коэффициента β_i .

Отметим, что в случае парной линейной регрессии стандартизованный коэффициент регрессии β совпадает с линейным коэффициентом корреляции r_{yx} .

Для оценки параметров нелинейных уравнений множественной регрессии предварительно осуществляется преобразование последних в линейную форму (с помощью замены переменных) и МНК применяется для нахождения параметров линейного уравнения множественной регрессии в преобразованных переменных.

В случае внутренне нелинейных зависимостей (которые невозможно привести к линейному виду) для оценки параметров по методу наименьших квадратов приходится применять методы нелинейной оптимизации (п. 2.4).

3.5. Качество оценок МНК линейной множественной регрессии. Теорема Гаусса-Маркова

В классическом множественном регрессионном анализе обычно делаются следующие предпосылки:

1. Математическое ожидание случайного члена ε_i равно нулю в любом наблюдении

$$M(\varepsilon_i) = 0. \quad (3.19)$$

2. Дисперсия случайного члена ε_i постоянна для всех наблюдений

$$D(\varepsilon_i) = \sigma^2. \quad (3.20)$$

3. Значения случайного члена в любых наблюдениях ε_i и ε_j не коррелируют между собой

$$\text{Cov}(\varepsilon_i, \varepsilon_j) = 0 \quad (i \neq j). \quad (3.21)$$

Это условие с учетом того, что $M(\varepsilon_i) = M(\varepsilon_j) = 0$ принимает вид

$$M(\varepsilon_i, \varepsilon_j) = 0 \quad (i \neq j). \quad (3.22)$$

4. Случайный член должен быть распределен независимо от объясняющих переменных x_i в одних и тех же наблюдениях

$$\text{Cov}(x_{it}, \varepsilon_i) = M(x_{it}, \varepsilon_i) = 0, \quad (3.23)$$

где было учтено, что $M(\varepsilon_i) = 0$.

Следует сказать, что последнее условие заведомо выполняется, если объясняющие переменные x_{it} считаются детерминированными величинами.

5. Матрица $X'X$ является неособенной, т. е. столбцы матрицы X линейно независимы.

6. Значения случайного члена ε_i распределены по нормальному закону.

Модель (3.6), удовлетворяющая приведенным предпосылкам 1–6, называется классической нормальной линейной моделью множественной регрессии.

Модель (3.6), удовлетворяющая приведенным предпосылкам 1–5, называется классической линейной моделью множественной регрессии.

Согласно теореме Гаусса-Маркова, при выполнении указанных предпосылок оценки параметров линейной множественной регрессии (3.13), полученные методом наименьших квадратов, будут несмещенными и эффективными (т. е. будут иметь наименьшую дисперсию) в классе линейных несмещенных оценок.

Нарушение одного из условий Гаусса-Маркова приводит к нарушению эффективности оценок, т. е. в классе несмещенных оценок можно найти такие, которые имеют меньшую дисперсию.

После построения модели необходимо вычислить значения остатков e_i и проверить выполнение предпосылок 1–6, так как их нарушение снижает качество модели. Если условия нарушаются, то следует модернизировать модель соответствующим образом. Эти вопросы будут рассмотрены далее.

3.6. Проверка качества уравнения регрессии. F -критерий Фишера

Как и в случае парной регрессии для оценки качества полученного множественной уравнения регрессии (3.6) можно использовать коэффициент детерминации, представляющий собой отношение объясненной части $D(\hat{y})$ дисперсии переменной y ко всей дисперсии $D(y)$

$$R^2 = \frac{D(\hat{y})}{D(y)} \quad \text{или} \quad R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (3.24)$$

где

$$D(y) = \frac{1}{n} \sum (y_i - \bar{y})^2, \quad D(\hat{y}) = \frac{1}{n} \sum (\hat{y}_i - \bar{y})^2, \quad D(e) = D_{\text{ост}} = \frac{1}{n} \sum (\hat{y}_i - y_i)^2.$$

Коэффициент детерминации R^2 принимает значения в диапазоне от нуля до единицы $0 \leq R^2 \leq 1$ и показывает, какая часть дисперсии результирующего признака y объяснена уравнением регрессии. Чем выше значение R^2 , тем лучше данная модель согласуется с данными наблюдений.

Оценка статистической значимости уравнения регрессии (а также коэффициента детерминации R^2) осуществляется с помощью F -критерия Фишера

$$F = \frac{\frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{p}}{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n - p - 1}} = \frac{R^2}{1 - R^2} \frac{n - p - 1}{p}, \quad (3.25)$$

где p – число независимых переменных в уравнении регрессии (3.6).

Согласно F -критерию Фишера, выдвигаемая «нулевая» гипотеза H_0 о статистической незначимости уравнения регрессии отвергается при выполнении условия $F > F_{\text{крит}}$, где $F_{\text{крит}}$ определяется по таблицам F -критерия Фишера (ПЗ, П4) по двум степеням свободы $k_1 = p$, $k_2 = n - p - 1$ и заданному уровню значимости α .

Для оценки тесноты связи факторов с исследуемым признаком, задаваемой построенным уравнением регрессии $\hat{y} = f(x_1, x_2, \dots, x_p)$, используется коэффициент множественной корреляции R

$$R = \sqrt{R^2} = \sqrt{1 - \frac{D_{\text{ост}}}{D(y)}} = \sqrt{1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}}. \quad (3.26)$$

Коэффициент множественной корреляции R принимает значения в диапазоне $0 \leq R \leq 1$.

Чем ближе величина R к единице, тем теснее данная связь, тем лучше зависимость $\hat{y} = f(x_1, x_2, \dots, x_p)$ согласуется с данными наблюдений. При $R = 1$

($R^2 = 1$) связь становится функциональной, т. е. соотношение $\hat{y} = f(x_1, x_2, \dots, x_p)$ точно выполняется для всех наблюдений.

Коэффициент множественной корреляции может использоваться как характеристика качества построенного уравнения регрессии $\hat{y} = f(x_1, x_2, \dots, x_p)$, точности построенной модели.

Величина коэффициента множественной корреляции не может быть меньше максимального парного индекса корреляции $R \geq \max r_{yx_i}$, ($i = 1, 2, \dots, p$).

В случае линейной зависимости (3.6) коэффициент корреляции R связан с парными коэффициентами корреляции r_{yx_i} соотношением

$$R = \sqrt{\sum_i \beta_i * r_{yx_i}}, \quad (3.27)$$

где β_i – стандартизованные коэффициенты регрессии (3.16).

Использование коэффициента множественной детерминации R^2 для оценки качества модели, обладает тем недостатком, что включение в модель нового фактора (даже несущественного) автоматически увеличивает величину R^2 .

Поэтому при большом количестве факторов предпочтительнее использовать, так называемый, скорректированный, улучшенный (adjusted) коэффициент множественной детерминации \bar{R}^2 , определяемый соотношением

$$\bar{R}^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2 : (n - p - 1)}{\sum_{i=1}^n (y_i - \bar{y})^2 : (n - 1)} = 1 - \frac{n - 1}{n - p - 1} (1 - R^2), \quad (3.28)$$

где p – число факторов в уравнении регрессии, n – число наблюдений.

Чем больше величина p , тем сильнее различия \bar{R}^2 и R^2 .

При использовании \bar{R}^2 для оценки целесообразности включения фактора в уравнение регрессии следует однако учитывать, что увеличение \bar{R}^2 при включении нового фактора не обязательно свидетельствует о его значимости, так как значение увеличивается \bar{R}^2 всегда, когда t -статистика больше единицы ($|t| > 1$).

При заданном объеме наблюдений и при прочих равных условиях с увеличением числа независимых переменных (параметров) скорректированный коэффициент множественной детерминации убывает. При небольшом числе наблюдений скорректированная величина коэффициента множественной детерминации R^2 имеет тенденцию переоценивать долю вариации результативного признака, связанную с влиянием факторов, включенных в регрессионную модель.

Отметим, что низкое значение коэффициента множественной корреляции и коэффициента множественной детерминации R^2 может быть обусловлено следующими причинами:

- в регрессионную модель не включены существенные факторы;
- неверно выбрана форма аналитической зависимости, не отражающая реальные соотношения между переменными, включенными в модель.

3.7. Точность коэффициентов регрессии. Доверительные интервалы

Оценки коэффициентов регрессии зависят от используемой выборки значений переменных x и y и являются случайными величинами. Для характеристики точности полученных оценок можно использовать стандартные ошибки коэффициентов регрессии.

Под стандартной ошибкой коэффициента регрессии понимается оценка стандартного отклонения функции плотности вероятности данного коэффициента.

Стандартные ошибки коэффициентов регрессии s_{b_i} определяются соотношениями

$$s_{b_i} = s_{ocm} [(X'X)^{-1}]_{ii}, \quad (3.29)$$

где s_{ocm}^2 представляет собой несмещенную оценку остаточной дисперсии

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

$[(X'X)^{-1}]_{ii}$ – диагональный элемент матрицы $(X'X)^{-1}$.

Величину $[(X'X)^{-1}]_{ii}$ можно вычислить как

$$[(X'X)^{-1}]_{ii} = \frac{A_{ii}}{\det(X'X)},$$

где A_{ii} – алгебраическое дополнение к элементу ii матрицы $(X'X)$.

Сопоставляя оценки параметров и их стандартные ошибки, можно сделать вывод о надежности (точности) полученных оценок.

Для оценки статистической значимости коэффициентов регрессии применяется t -критерий Стьюдента, основанный на том факте, что отношения

$$t_{b_i} = \frac{b_i - \tilde{b}_i}{s_{b_i}} \quad (i = 1, 2, \dots, p) \quad (3.30)$$

являются t -статистиками, т. е. случайными величинами, распределенными по закону Стьюдента с числом степеней свободы $n-p-1$. Через \tilde{b}_i обозначены точные значения коэффициентов регрессии.

Согласно t -критерию Стьюдента, выдвигается «нулевая» гипотеза H_0 о статистической незначимости коэффициента уравнения регрессии (т. е. о статистически незначимом отличии величины a или b_i от нуля). Эта гипотеза отвергается при выполнении условия $t > t_{крит}$, где $t_{крит}$ определяется по таблицам t -критерия Стьюдента (П2) по числу степеней свободы $k_1 = n-p-1$ (p – число независимых переменных в уравнении регрессии) и заданному уровню значимости α .

t -критерий Стьюдента применяется в процедуре принятия решения о целесообразности включения фактора в модель. Если коэффициент при факторе в уравнении регрессии оказывается незначимым, то включать данный фактор в модель не рекомендуется. Отметим, что это правило не является абсолютным и бывают ситуации, когда включение в модель статистически незначимого фактора определяется экономической целесообразностью.

Доверительные интервалы для параметров b_i уравнения линейной регрессии определяются соотношениями:

$$b_i - t_{1-\alpha, n-p-1} \cdot s_{bi} \leq \tilde{b}_i \leq b_i + t_{1-\alpha, n-p-1} \cdot s_{bi}. \quad (3.31)$$

Величина $t_{1-\alpha, n-2}$ представляет собой табличное значение t -критерия Стьюдента на уровне значимости α при степени свободы $n-2$.

Если в границы доверительного интервала попадает ноль, т. е. нижняя граница отрицательна, а верхняя положительна, то оцениваемый параметр принимается равным нулю, так как он не может одновременно принимать и положительное, и отрицательное значения.

Точность полученного уравнения регрессии можно оценить, анализируя доверительный интервал для функции регрессии, т. е. для среднего значения \tilde{y}_0 , зависимой переменной y при заданных значениях объясняющих переменных $x_1 = x_{10}, x_2 = x_{20}, \dots, x_p = x_{p0}$.

Доверительный интервал для функции регрессии определяется соотношениями

$$\hat{y}_0 - t_{1-\alpha, n-p-1} \cdot s_{\hat{y}} \leq \tilde{y}_0 \leq \hat{y}_0 + t_{1-\alpha, n-p-1} s_{\hat{y}}, \quad (3.32)$$

где \hat{y}_0 – групповая средняя, определяемая по уравнению регрессии (3.4) при заданных значениях объясняющих переменных $x_1 = x_{10}, x_2 = x_{20}, \dots, x_p = x_{p0}$;

$$s_{\hat{y}} = s_{\hat{m}\hat{o}} \sqrt{X'_0(X'X)X_0} - \text{ее стандартная ошибка}; \quad (3.33)$$

\tilde{y}_0 – точное значение групповой средней; X_0 – вектор, составленный из заданных значений независимых переменных $X_0 = (1, x_{10}, x_{20}, \dots, x_{p0})$.

Доверительный интервал для индивидуальных значений зависимой переменной y^*_0 определяется соотношениями

$$\hat{y}_0 - t_{1-\alpha, n-p-1} \cdot s_{\hat{y}0} \leq y^*_0 \leq \hat{y}_0 + t_{1-\alpha, n-p-1} s_{\hat{y}0}, \quad (3.34)$$

где

$$s_{\hat{y}0} = s_{\hat{m}\hat{o}} \sqrt{1 + X'_0(X'X)X_0} \quad (3.35)$$

есть стандартная ошибка индивидуальных значений зависимой переменной y^*_0 .

3.8. Частные уравнения регрессии. Частная корреляция

Уравнение линейной множественной регрессии $\hat{y} = a + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_p \cdot x_p$ характеризует совместное влияние факторов x_1, x_2, \dots, x_p на исследуемую переменную y . Уравнение парной регрессии $\hat{y}_{x_i} = a^i + b^i \cdot x_i$ показывает зависимость между y и x_i при игнорировании остальных факторов. Коэффициент b^i наряду с влиянием фактора x_i частично отражает влияние и остальных факторов.

Частные уравнения регрессии, характеризующие изолированное влияние одного из факторов x_i на результативную переменную y при исключении влияния остальных факторов, включенных в уравнение регрессии, получаются из общего уравнения линейной множественной регрессии (3.6) при закреплении всех факторов кроме x_i на их среднем уровне:

$$\hat{y}_{x_i \cdot p} = a + b_1 \cdot \bar{x}_1 + b_2 \cdot \bar{x}_2 + \dots + b_{i-1} \cdot \bar{x}_{i-1} + b_i \cdot x_i + b_{i+1} \cdot \bar{x}_{i+1} + \dots + b_p \cdot \bar{x}_p, \\ (i = 1, 2, \dots, p)$$

или

$$\hat{y}_{x_i \cdot p} = A_i + b_i \cdot x_i, \quad (i = 1, 2, \dots, p) \quad (3.36)$$

где $A_i = a + b_1 \cdot \bar{x}_1 + b_2 \cdot \bar{x}_2 + \dots + b_{i-1} \cdot \bar{x}_{i-1} + b_{i+1} \cdot \bar{x}_{i+1} + \dots + b_p \cdot \bar{x}_p$ и $A_i \neq a^i$.

На основе частных уравнений регрессии (3.36) определяют частные коэффициенты эластичности

$$\dot{Y}_{y x_i} = b_i \frac{x_i}{\hat{y}_{x_i \cdot p}}, \quad (i = 1, 2, \dots, p) \quad (3.37)$$

где b_i – коэффициенты регрессии для фактора x_i в уравнении множественной регрессии; $\hat{y}_{x_i \cdot n}$ – значение результативного фактора, полученное из частного уравнения регрессии при данном значении фактора x_i ,

Средние частные коэффициенты эластичности

$$\bar{Y}_{y x_i} = b_i \frac{\bar{x}_i}{\bar{\hat{y}}_{x_i \cdot p}}. \quad (i = 1, 2, \dots, p) \quad (3.38)$$

показывают, на сколько процентов в среднем по совокупности изменится результат y от своей величины при изменении фактора x на 1 % от своего значения при неизменных значениях других факторов, и могут использоваться для выделения факторов, наиболее влияющих на результат.

Если факторы x_i, x_j находятся в корреляционной связи, то это влияет на способность коэффициента парной корреляции $r_{y x_i}$ изолированно выявить степень тесноты связи между переменными y и x_i . В такой ситуации следует использовать частные коэффициенты корреляции $r_{y x_i \cdot p}$, характеризующие тесноту связи между переменными y и x_i при исключении влияния остальных $p - 1$ фактора (при фиксированных значениях остальных факторов), определяемые соотношениями

$$r_{y x_i \cdot p} = \frac{-q_{yi}}{\sqrt{q_{yy} \cdot q_{ii}}}, \quad (i = 1, 2, \dots, p) \quad (3.39)$$

где q_{yi} , q_{yy} и q_{ii} – алгебраические дополнения соответственно к элементам $r_{y x_i}$, r_{yy} и $r_{x_i x_i}$ матрицы

$$q = \begin{bmatrix} r_{yy} & r_{y x_1} & r_{y x_2} & \dots & r_{y x_p} \\ r_{x_1 y} & r_{x_1 x_1} & r_{x_1 x_2} & \dots & r_{x_1 x_p} \\ r_{x_2 y} & r_{x_2 x_1} & r_{x_2 x_2} & \dots & r_{x_2 x_p} \\ \dots & \dots & \dots & \dots & \dots \\ r_{x_p y} & r_{x_p x_1} & r_{x_p x_2} & \dots & r_{x_p x_p} \end{bmatrix}. \quad (3.40)$$

Значимость частных коэффициентов корреляции $r_{y x_i \cdot p}$ проверяется также, как и значимость парного коэффициента корреляции (2.37), (2.38) с заменой числа наблюдений n на $n' = n - p + 1$, т. е. статистика

$$t = \frac{r_{yx_i \cdot p} \sqrt{n-p-1}}{\sqrt{1-r_{yx_i \cdot p}^2}} \quad (3.41)$$

имеет t -распределение Стьюдента с $n-p-1$ степенями свободы. Если $t > t_{1-\alpha; n-p-1}$, то коэффициент считается значимым.

В случае только двух факторов x_1 и x_2 формула (3.39) принимает вид

$$r_{yx_1 \cdot x_2} = \frac{r_{yx_1} - r_{yx_2} r_{x_1 x_2}}{\sqrt{(1-r_{yx_2}^2)(1-r_{x_1 x_2}^2)}}. \quad (3.42)$$

Существенность влияния корреляционной связи проанализируем на примере. Рассмотрим переменную y и два фактора x_1 и x_2 , находящиеся в корреляционной связи, и предположим, что парные коэффициенты корреляции имеют следующие значения $r_{yx_1} = 0,54$, $r_{yx_2} = 0,1$, $r_{x_1 x_2} = 0,6$. Вычисления по формуле (3.42) дают

$$r_{yx_1 \cdot x_2} = \frac{0,54 - 0,1 \cdot 0,6}{\sqrt{(1-0,1^2)(1-0,6^2)}} = \frac{0,48}{\sqrt{0,99 \cdot 0,64}} = 0,60;$$

$$r_{yx_2 \cdot x_1} = \frac{0,1 - 0,54 \cdot 0,6}{\sqrt{(1-0,54^2)(1-0,6^2)}} = -\frac{0,224}{\sqrt{0,78 \cdot 0,64}} = -0,33.$$

Значения коэффициентов r_{yx_1} и $r_{yx_1 \cdot x_2}$ близки между собой, а значения коэффициентов r_{yx_2} и $r_{yx_2 \cdot x_1}$ отличаются по величине более, чем в три раза и имеют разные знаки.

Частные коэффициенты корреляции $r_{yx_i \cdot p}$ позволяют ранжировать факторы по степени влияния на результативный признак и находят применение в процедуре отбора факторов для включения их в уравнение регрессии (учитываются факторы, которым соответствуют значимые коэффициенты частной корреляции).

3.9. Обобщенный метод наименьших квадратов.

Гетероскедастичность

3.9.1. Обобщенный метод наименьших квадратов

Оценки (3.13) коэффициентов линейной множественной регрессии (3.6) являются эффективными (имеющими минимальную дисперсию в классе линейных несмещенных оценок) только при выполнении предпосылок п. 3.5. Нарушение второй и третьей предпосылок ведет к утере эффективности оценок (3.13), т. е. существуют оценки с меньшей дисперсией (с меньшим разбросом значений оценок).

Следствием предпосылок 2 и 3 является диагональная структура матрицы ковариаций Σ_ε случайного члена ε_i с одинаковыми диагональными элементами σ^2 (дисперсия случайного члена ε_i)

$$\Sigma_\varepsilon = \sigma^2 E_n, \quad (3.43)$$

Система уравнений (3.47) соответствует модели, определяемой соотношениями

$$\frac{y_i}{\sigma_i} = a \frac{1}{\sigma_i} + b_1 \cdot \frac{x_{1i}}{\sigma_i} + b_2 \cdot \frac{x_{2i}}{\sigma_i} + \dots + b_p \cdot \frac{x_{pi}}{\sigma_i} + u_i, \quad (i = 1, 2, \dots, n) \quad (3.48)$$

которые получаются, если исходное уравнение множественной регрессии (3.6), записанное для каждого наблюдения разделить на среднее квадратическое отклонение σ_i случайного члена ε_i в i -наблюдении. Случайный член $u_i = \frac{\varepsilon_i}{\sigma_i}$ в модели (3.48) имеет постоянную для всех наблюдений дисперсию $\sigma_{ui}^2 = 1$. Запись модели в виде (3.48) соответствует уравнению линейной множественной регрессии (без свободного члена)

$$y' = ax'_0 + b_1 \cdot x'_1 + b_2 \cdot x'_2 + \dots + b_p \cdot x'_p + u, \quad (3.49)$$

записанному в новых переменных $y', x'_0, x'_1, \dots, x'_p$, значения которых определяются по формулам

$$y'_i = \frac{y_i}{\sigma_i}, x'_{0i} = \frac{1}{\sigma_i}, x'_{1i} = \frac{x_{1i}}{\sigma_i}, x'_{2i} = \frac{x_{2i}}{\sigma_i}, \dots, x'_{pi} = \frac{x_{pi}}{\sigma_i}, u_i = \frac{\varepsilon_i}{\sigma_i}, \quad (i = 1, 2, \dots, n) \quad (3.50)$$

Следует сказать, что величины σ_i^2 практически никогда не известны и вместо них следует использовать состоятельные оценки $\hat{\sigma}_i^2$.

При практическом использовании ОМНК используется какое-либо предположение относительно зависимости дисперсии σ_i^2 случайного члена ε от наблюдения или величины факторов x_i .

Представим дисперсии σ_i^2 случайного члена в виде произведения некоторой функции K_i^2 от факторов на постоянную величину σ^2

$$\sigma_i^2 = K_i^2 \sigma^2. \quad (3.51)$$

Тогда соотношения (3.50) принимают вид

$$y'_i = \frac{y_i}{K_i}, x'_{0i} = \frac{1}{K_i}, x'_{1i} = \frac{x_{1i}}{K_i}, x'_{2i} = \frac{x_{2i}}{K_i}, \dots, x'_{pi} = \frac{x_{pi}}{K_i}, u_i = \frac{\varepsilon_i}{K_i}, \sigma_{ui} = \text{const} = \sigma^2. \quad (i = 1, 2, \dots, n) \quad (3.52)$$

Часто на практике можно с достаточным основанием предположить, что величины σ_i пропорциональны значениям какого-либо фактора x_α , т. е. $\sigma_i = x_{\alpha i} \cdot \sigma$ ($\sigma_i^2 = x_{\alpha i}^2 \cdot \sigma^2$). В этом случае модель (3.49) принимает вид

$$\frac{y}{x_\alpha} = a \frac{1}{x_\alpha} + b_1 \cdot \frac{x_1}{x_\alpha} + \dots + b_{\alpha-1} \cdot \frac{x_{\alpha-1}}{x_\alpha} + b_\alpha + b_{\alpha+1} \cdot \frac{x_{\alpha+1}}{x_\alpha} + \dots + b_p \cdot \frac{x_p}{x_\alpha} + u \quad (3.53)$$

и $\sigma_{ui}^2 = \sigma^2$.

Оценки параметров модели (3.53) являются оценками параметров исходного уравнения (3.6).

Если, вычислив значения новых переменных, мы запишем модель в стандартном виде

$$\frac{y}{x_\alpha} = a + b_1 \cdot \frac{x_1}{x_\alpha} + b_2 \cdot \frac{x_2}{x_\alpha} + \dots + b_p \cdot \frac{x_p}{x_\alpha} + u, \quad (3.54)$$

то это будет новая модель с переменными, имеющими иной смысл. Оценки ее параметров будут отличаться от оценок параметров исходной модели.

Рассмотрим случай парной регрессии

$$y = a + b \cdot x + \varepsilon \quad (3.55)$$

и предположим, что величины σ_i пропорциональны значениям фактора x , т. е. $\sigma_i = x_i \cdot \sigma$ ($\sigma_i^2 = x_i^2 \cdot \sigma^2$). Преобразуя согласно ОМНК уравнение регрессии (3.55) получим следующую модель

$$\frac{y}{x} = \frac{a}{x} + b + u, \quad (3.56)$$

оценки параметров которой будут эффективными оценками параметров исходной модели (3.55). Заметим, что в новой модели параметры a и b поменялись местами, т. е. свободный член стал коэффициентом и наоборот.

3.10. Проверка остатков регрессии на гетероскедастичность

Так как оценки параметров, полученные МНК, являются эффективными только при выполнении предпосылок МНК (п. 3.5), то после вычисления оценок и построения модели следует определить наблюдаемые отклонения $e_i = y_i - f(x_{1i}, x_{2i}, \dots, x_{pi})$ и проверить, удовлетворяются ли предпосылки МНК.

Рассмотрим методы, применяемые для проверки выполнения предпосылки о постоянстве дисперсий остатков (их гомоскедастичности).

Тест ранговой корреляции Спирмена проверяет наличие монотонной зависимости между дисперсией ошибки и величиной фактора. Наблюдения (значения фактора x_i и остатки e_i) упорядочиваются по величине фактора x и вычисляется коэффициент ранговой корреляции Спирмена

$$\rho_{x,e} = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n^2 (n - 1)}, \quad (3.57)$$

где d_i – разность между рангами значений x_i и e_i в i -наблюдении.

Коэффициент ранговой корреляции $\rho_{x,e}$ считается значимым на уровне значимости α при $n > 10$, если выполняется условие

$$|t| = \frac{|\rho_{x,e}| \sqrt{n-2}}{\sqrt{1-\rho_{x,e}^2}} > t_{1-\alpha, n-2}, \quad (3.58)$$

где $t_{1-\alpha, n-2}$ – табличное значение t -критерия Стьюдента на уровне значимости α и при числе степеней свободы $(n-2)$.

Тест Гольдфельда–Квандта. Применяется в предположении, что средние квадратические отклонения случайного члена σ_i пропорциональны значениям фактора x_i и случайный член распределен по нормальному закону. Процедура применения теста Гольдфельда–Квандта состоит из следующих шагов:

- 1) наблюдения упорядочиваются по мере возрастания фактора x_i ;

2) выделяются первые n' и последние n' наблюдений и исключаются из рассмотрения $n-2n'$ центральных наблюдений. При этом должно выполняться условие $n' > p$, где p – число оцениваемых параметров;

3) по каждой из групп оцениваются уравнения регрессии остатков ε_i по значимым факторам;

4) определяются остаточные суммы квадратов для первой ($S_1 = \sum e^2_i$) и второй ($S_2 = \sum e^2_i$) групп и находится их отношение: $R = S_2 : S_1$ ($S_2 > S_1$);

5) нулевая гипотеза о гомоскедастичности остатков отвергается, если выполнено условие

$$R > F_{\alpha, n'-p, n'-p} \quad (3.59)$$

где $F_{\alpha, n'-p, n'-p}$ – табличное значение F -критерия Фишера на уровне значимости α при числе степеней свободы $(n'-p)$ и $(n'-p)$.

Авторами метода рекомендовано для случая одного фактора при $n=30$ принимать $n'=11$, а при $n=60$ принимать $n'=22$.

Тест Глейзера. Позволяет не только выявить наличие гетероскедастичности остатков, но и сделать определенные выводы о характере зависимости дисперсии остатков σ_i от значений фактора x_i . В тесте проверяется существование функциональной зависимости следующего вида

$$\sigma_i = \alpha + \beta x_i^\gamma \quad (3.60)$$

По полученным остаткам уравнения регрессии осуществляются регрессии

$$|e_i| = \alpha + \beta x_i^\gamma \quad (3.61)$$

при различных значениях параметра γ (например, -1; 0,5; 1; 1,5; 2; ...) и выбирается зависимость с наиболее значимым коэффициентом β . Если все коэффициенты β не значимы, то нет оснований говорить о гетероскедастичности остатков.

Отобранная зависимость (с наиболее значимым коэффициентом β) используется в ОМНК для получения улучшенных оценок параметров исходной модели.

3.11. Построение регрессионных моделей при наличии автокорреляции остатков

Предположим, что нарушается только предпосылка 3 о независимости значений случайного члена ε_i и ε_j в различных наблюдениях $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$ ($i \neq j$). В этом случае говорят об автокорреляции остатков. Оценки параметров, полученные методом наименьших квадратов, остаются несмещенными, но теряют свою эффективность.

Автокорреляция обычно встречается в регрессионном анализе только при использовании исходных данных в виде временных рядов. Более подробно понятие автокорреляции изложено в 5 разделе, где также приведены методы, позволяющие определить наличие и характер авторреляции во временном ряде. Здесь же мы рассмотрим случай, когда имеет место зависимость только между соседними остатками.

Предположим, что остатки в уравнении линейной регрессии

$$y_t = a + b \cdot x_t + \varepsilon_t \quad (3.62)$$

образуют авторегрессионный процесс первого порядка

$$\varepsilon_t = \rho \varepsilon_{t-1} + u_t. \quad (3.63)$$

Для оценки величины ρ может использоваться статистика Дарбина-Уотсона d (см. п. 5.3.5)

$$\rho = 1 - d/2. \quad (3.64)$$

Преобразуем уравнение (3.62), чтобы исключить автокорреляцию в остатках. Для этого уравнение (3.62), записанное для момента времени $t-1$,

$$y_{t-1} = a + b \cdot x_{t-1} + \varepsilon_{t-1}$$

умножим на ρ и вычтем из исходного уравнения (3.62)

$$y_t - \rho \cdot y_{t-1} = a - a \cdot \rho + b \cdot (x_t - \rho \cdot x_{t-1}) + \varepsilon_t - \rho \cdot \varepsilon_{t-1}.$$

Вводя новые переменные y'_t и x'_t

$$y'_t = y_t - \rho \cdot y_{t-1}; \quad x'_t = x_t - \rho \cdot x_{t-1} \quad (3.65)$$

и используя обозначение

$$a' = a(1 - \rho), \quad (3.66)$$

приведем исходную модель регрессии (3.62) к линейному уравнению регрессии

$$y'_t = a' + b \cdot x'_t + u_t \quad (3.67)$$

со случайными независимыми остатками u_t .

Для оценки параметров преобразованного уравнения (3.67) можно применять обычный МНК. После определения параметров a' и b параметр a находится из соотношения (3.66).

Изложенная процедура предварительного преобразования переменных с последующим применением МНК к оценке параметров уравнения регрессии в преобразованных переменных является частным случаем обобщенного метода наименьших квадратов.

Если $\rho = 1$, то данный метод становится методом первых последовательных разностей, так как

$$y'_t = y_t - y_{t-1}; \quad x'_t = x_t - x_{t-1}.$$

Если $\rho = -1$, т. е. в остатках наблюдается полная отрицательная корреляция, то с учетом соотношений

$$y'_t = y_t - (-1) \cdot y_{t-1} = y_t + y_{t-1}; \quad x'_t = x_t - (-1) \cdot x_{t-1} = x_t + x_{t-1};$$

$$a' = a(1 - (-1)) = 2 \cdot a$$

изложенный выше метод (уравнение (3.67)) принимает следующий вид

$$y_t + y_{t-1} = 2 \cdot a + b \cdot (x_t + x_{t-1}) + u_t$$

или

$$(y_t + y_{t-1})/2 = a + b \cdot (x_t + x_{t-1})/2 + u_t/2.$$

Данная модель является моделью регрессии по скользящим средним.

3.12. Регрессионные модели с переменной структурой. Фиктивные переменные

3.12.1. Фиктивные переменные

При изучении экономических взаимосвязей возникает необходимость учесть в модели влияние качественного фактора (фактора, не имеющего количественного выражения), например пол потребителя, фактор сезонности, наличие государственных программ. Влияние качественных признаков может приводить к скачкообразному изменению параметров линейных регрессионных моделей, построенных для различных значений качественного признака. Такие модели называются регрессионными моделями с переменной структурой.

Чтобы учесть влияние качественного фактора в рамках одного регрессионного уравнения вводятся, так называемые, фиктивные переменные с двумя значениями 0 и 1. Например, изучается зависимость потребления товара y от величины дохода x с учетом пола потребителя. С использованием фиктивной переменной z

$$z = \begin{cases} 1, & \text{мужской пол} \\ 0, & \text{женский пол} \end{cases}$$

уравнение регрессии принимает вид

$$y = a + b \cdot x + c \cdot z + \varepsilon. \quad (3.68)$$

Вводя новый член регрессии $c \cdot z$, мы тем самым предполагаем, что пол потребителя влияет только на величину свободного члена уравнения (параметр a характеризует объем потребления). Чтобы учесть влияние пола потребителя на величину коэффициента регрессии b (характеризующего «склонность» к потреблению), следует в модель регрессии ввести дополнительное слагаемое $d \cdot z \cdot x$, что дает

$$y = a + b \cdot x + c \cdot z + d \cdot z \cdot x + \varepsilon. \quad (3.69)$$

Таким образом, модель (3.69) является объединением двух моделей для мужчин и женщин

$$y = a^1 + b^1 \cdot x + \varepsilon,$$

$$y = a^2 + b^2 \cdot x + \varepsilon,$$

где $a^1 = a + c$; $b^1 = b + d$; $a^2 = a$; $b^2 = b$.

Проверка значимости коэффициентов при фиктивных факторах z и $z \cdot x$ покажет значимость влияния качественного показателя на изучаемый признак и необходимость включения в уравнение регрессии соответствующего члена.

Если качественный признак имеет более двух градаций признака, то вводится несколько фиктивных переменных, число которых на единицу меньше числа градаций признака. Например, чтобы учесть сезонность, вводятся три фиктивные переменные

$$z_1 = \begin{cases} 1, & \text{весна,} \\ 0, & \text{не весна,} \end{cases} \quad z_2 = \begin{cases} 1, & \text{лето,} \\ 0, & \text{не лето,} \end{cases} \quad z_3 = \begin{cases} 1, & \text{осень,} \\ 0, & \text{не осень} \end{cases} \quad (3.70)$$

и уравнение регрессии примет вид

$$y = a + b \cdot x + c_1 \cdot z_1 + c_2 \cdot z_2 + c_3 \cdot z_3 + \varepsilon.$$

Если качественных признаков несколько, то фиктивные переменные вводятся для каждого признака по таким же правилам.

3.12.2. Тест Чоу

Предположим, что имеется две набора наблюдений за совместным изменением двух зависимой и объясняющей переменной (x_i, y_i) , полученные в различных условиях. Возникает вопрос можно ли считать две полученные выборки наблюдений частями одной объединенной выборки или принципиально различными, для которых уравнения регрессии должны строиться отдельно, как показано на рисунке 3.1 [4]. Ответ на этот вопрос дается с помощью теста Чоу.

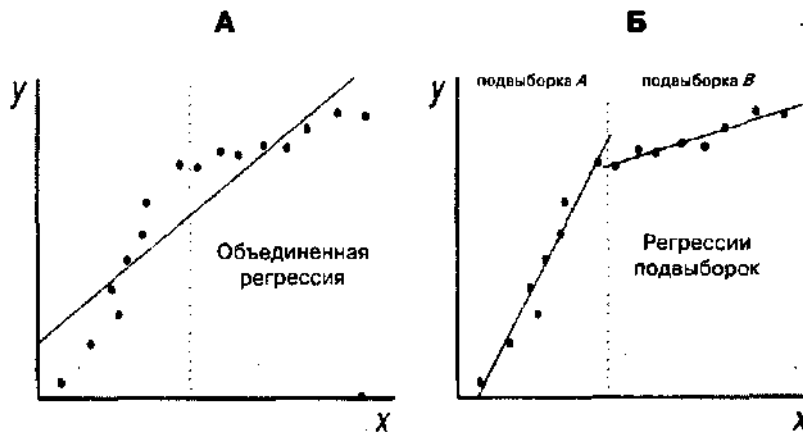


Рис. 3.1. Регрессии, оцениваемые для теста Чоу

Рассмотрим уравнения регрессии, построенные по первой, второй и объединенной выборкам

$$y_i = a^1 + b^1_1 \cdot x_{1i} + \dots + b^1_p \cdot x_{pi} + \varepsilon_i, \quad (i = 1, 2, \dots, n_1)$$

$$y_i = a^2 + b^2_1 \cdot x_{1i} + \dots + b^2_p \cdot x_{pi} + \varepsilon_i, \quad (i = 1, 2, \dots, n_2)$$

$$y_i = a + b_1 \cdot x_{1i} + \dots + b_p \cdot x_{pi} + \varepsilon_i. \quad (i = 1, 2, \dots, n = n_1 + n_2)$$

Обозначим суммы квадратов остатков регрессии, полученных по первой, второй и объединенной выборкам E^2_1, E^2_2, E^2 .

Согласно тесту Чоу, нулевая гипотеза H_0 о том, что две выборки являются частями одной объединенной выборки, отвергается при уровне значимости α , если выполняется условие

$$F = \frac{(E^2 - E^2_1 - E^2_2)(n - 2p - 2)}{(E^2_1 + E^2_2)(p + 1)} > F_{\alpha; p+1; n-2p-2}. \quad (3.71)$$

3.11. Проблемы построения регрессионных моделей

Последствия отсутствия в уравнении существенной независимой переменной. Если в уравнение регрессии не включена независимая переменная, оказывающая существенное влияние на результирующий признак, то в общем случае это приводит к смещению оценок коэффициентов регрессии. Смещение отсутствует только если ковариация отсутствующей переменной с переменными, включенными в модель, равна нулю. Стандартные ошибки коэффициентов

регрессии становятся некорректными, что приводит к неприменимости соответствующих t -тестов. Кроме того, возможно появление автокорреляции и гетероскедастичности остатков. Признаком отсутствия значимой переменной может служить несоответствие знаков коэффициентов теоретическим предположениям. Если нет возможности включить в уравнение регрессии такую переменную, то следует использовать замещающую переменную (п. 3.2.1).

Последствия включения в модель несущественной независимой переменной. Если в уравнение регрессии включена существенная независимая переменная, то в общем случае это не приводит к смещению оценок коэффициентов регрессии, но значения стандартных ошибок могут возрасти.

Последствия неправильной спецификации формы уравнения регрессии. Использование неверной формы уравнения регрессии приводит к смещенности и несостоятельности оценок параметров, низкому значению коэффициента детерминации R^2 . Возможно также появление автокорреляции и гетероскедастичности остатков.

Контрольные вопросы

1. Что понимается под множественной регрессией?
2. Какие задачи решаются при построении уравнения регрессии?
3. Какие задачи решаются при спецификации модели?
4. Какие требования предъявляются к факторам, включаемым в уравнение регрессии?
5. Что понимается под коллинеарностью и мультиколлинеарностью факторов?
6. Как проверяется наличие коллинеарности и мультиколлинеарности?
7. Какие подходы применяются для преодоления межфакторной корреляции?
8. Какие функции чаще используются для построения уравнения множественной регрессии?
9. Какой вид имеет система нормальных уравнений метода наименьших квадратов в случае линейной регрессии?
10. По какой формуле вычисляется коэффициент множественной корреляции?
11. Как вычисляются коэффициент множественной детерминации и скорректированный коэффициент множественной детерминации?
12. Что означает низкое значение коэффициента множественной корреляции?
13. Как проверяется значимость уравнения регрессии и его коэффициентов?
14. В каких случаях применяется Обобщенный МНК?
15. В чем отличие частных уравнений регрессии от уравнений парной регрессии?
16. Как вычисляются средние частные коэффициенты эластичности?
17. Что такое стандартизированные переменные?
18. Какой вид имеет уравнение линейной регрессии в стандартизированном масштабе?
19. Как оценивается значимость факторов?
20. Как вычисляются частные коэффициенты корреляции?
21. Что понимается под гомоскедастичностью остатков?
22. Как проверяется гипотеза о гомоскедастичности ряда остатков?
23. Каковы последствия неправильной спецификации модели?
24. К чему приводит отсутствие в уравнении существенной независимой переменной?

4. Системы эконометрических уравнений

4.1. Структурная и приведенная формы модели

Экономические процессы и явления, как правило, представляют собой сложные системы, характеризующиеся большим количеством параметров и сложными взаимосвязями. Использование отдельных изолированных уравнений регрессии для исследования экономических процессов является сильным упрощением. Оно предполагает, что факторы можно изменять независимо друг от друга и что изменение зависимой переменной (результативного признака) никак ни влияет на поведение изучаемой системы. В случае сложных экономических систем такое предположение, как правило, не может быть выполнено, так как изменение какого-либо признака повлечет за собой изменения во всей системе взаимосвязанных признаков. В таких ситуациях эконометрические модели строятся в виде систем эконометрических уравнений. Наиболее широко этот подход применяется в макроэкономических исследованиях, а также в исследованиях спроса и предложения.

Например, в рыночной экономике равновесные цены рассматриваются как результат взаимодействия спроса и предложения. При этом предложение товара в существенной степени зависит от сложившейся цены, а цена, в свою очередь, определяется величиной среднего дохода потребителя и имеющимся на рынке предложением товара. Соответствующая модель определяется системой из двух уравнений

$$\begin{aligned} Q_t &= a_{10} + b_{11} \cdot P_t + \varepsilon_{1t}, \\ P_t &= a_{20} + b_{21} \cdot Q_t + a_{11} \cdot I_t + \varepsilon_{2t}, \end{aligned} \quad (4.1)$$

где P_t – средняя цена за единицу товара, Q_t – объем предложения товара, I_t – средний уровень дохода, t – означает текущий период времени, a_{10} , a_{20} , b_{11} , b_{21} – постоянные параметры, ε_{1t} , ε_{2t} – ошибки уравнений.

В качестве другого примера рассмотрим макроэкономическую модель Клейна [2]:

$$CN_t = \alpha_0 + \alpha_1(W_{1t} + W_{2t}) + \alpha_2 P_t + \alpha_3 P_{t-1} + \varepsilon_{1t}, \quad (4.2)$$

$$I_t = \beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + \beta_3 K_{t-1} + \varepsilon_{2t}, \quad (4.3)$$

$$W_{1t} = \gamma_0 + \gamma_1 E_t + \gamma_2 E_{t-1} + \gamma_3 T + \varepsilon_{3t}, \quad (4.4)$$

$$Y_t + TX_t \equiv CN_t + I_t + G_t, \quad (4.5)$$

$$Y_t \equiv P_t + W_t, \quad (4.6)$$

$$K_t \equiv I_t + K_{t-1}, \quad (4.7)$$

$$W_t = W_{1t} + W_{2t}, \quad (4.8)$$

$$E_t \equiv Y_t + TX_t - W_{2t}. \quad (4.9)$$

Первое уравнение называется функцией потребления. Оно соотносит потребление CN и совокупный фонд заработной платы W , равный сумме заработных плат работников занятых в частном секторе W_1 , и государственном секторе W_2 , а также текущий и лаговый незарплатный доход (прибыль) P .

Второе уравнение называется функцией инвестиций. Оно соотносит чистые инвестиции I с текущими и лаговыми прибылями P и запасом капитала K в начале года:

держащие параметров и определяющие фиксированные отношения между переменными, например, соотношения (4.4) – (4.9).

Наличие взаимозависимости между эндогенными переменными в системе одновременных уравнений (4.10) приводит к нарушению предпосылки о независимости объясняющих переменных и случайных членов, в результате чего обычный метод наименьших квадратов будет давать несостоятельные и смещенные оценки параметров.

Если с помощью преобразований исключить зависимые переменные из правых частей уравнений (4.10), то полученная система уравнений называется *приведенной формой модели (ПФМ)*

$$\begin{aligned} \hat{y}_1 &= \delta_{11} \cdot x_1 + \delta_{12} \cdot x_2 + \dots + \delta_{1m} \cdot x_m; \\ \hat{y}_2 &= \delta_{21} \cdot x_1 + \delta_{22} \cdot x_2 + \dots + \delta_{2m} \cdot x_m; \\ &\dots\dots\dots \\ \hat{y}_n &= \delta_{n1} \cdot x_1 + \delta_{n2} \cdot x_2 + \dots + \delta_{nm} \cdot x_m, \end{aligned} \tag{4.15}$$

параметры которой δ_{ij} являются алгебраическими функциями от структурных параметров и называются *приведенными коэффициентами*.

Например, для конъюнктивной модели, определяемой соотношениями:

$$\begin{aligned} C_t &= a_1 + b_{11} \cdot Y_t + b_{12} \cdot C_{t-1} + u_1 && (\text{функция потребления}); \\ I_t &= a_2 + b_{21} \cdot r_t + b_{22} \cdot I_{t-1} + u_2 && (\text{функция инвестиций}); \\ r_t &= a_3 + b_{31} \cdot Y_t + b_{32} \cdot M_t + u_3 && (\text{функция денежного рынка}); \\ Y_t &= C_t + I_t + G_t && (\text{тождество дохода}), \end{aligned} \tag{4.16}$$

где C – расходы на потребление, Y – ВВП, I – инвестиции, r – процентная ставка, M – денежная масса, G – государственные расходы, t и $t-1$ обозначают текущий и предыдущий периоды, u_1, u_2, u_3 – случайные ошибки, приведенная форма модели будет иметь следующий вид:

$$\begin{aligned} C_t &= \delta_{11} \cdot M_t + \delta_{12} \cdot G_t + \delta_{13} \cdot C_{t-1} + \delta_{14} \cdot I_{t-1} + v_1 \\ I_t &= \delta_{21} \cdot M_t + \delta_{22} \cdot G_t + \delta_{23} \cdot C_{t-1} + \delta_{24} \cdot I_{t-1} + v_2 \\ r_t &= \delta_{31} \cdot M_t + \delta_{32} \cdot G_t + \delta_{33} \cdot C_{t-1} + \delta_{34} \cdot I_{t-1} + v_3 \\ Y_t &= \delta_{41} \cdot M_t + \delta_{42} \cdot G_t + \delta_{43} \cdot C_{t-1} + \delta_{44} \cdot I_{t-1} + v_4 \end{aligned} \tag{4.17}$$

По своей структуре приведенная форма модели представляет собой систему независимых уравнений, поэтому ее параметры δ_{ij} можно оценивать с помощью обычного метода наименьших квадратов. Полученные численные значения параметров δ_{ij} позволяют вычислять модельные значения эндогенных переменных через predetermined переменные. На этом процесс построения модели не заканчивается, так как для исследователя наибольший интерес представляют значения именно структурных коэффициентов a_{ij} и b_{ij} , характеризующих внутренние взаимосвязи в системе и допускающих экономическую интерпретацию.

4.2. Оценка параметров структурной формы модели

Получение оценок параметров приведенной формы модели, как уже отмечалось, затруднений не представляет. Следующим этапом должно быть определение оценок параметров структурной формы модели по оценкам приведенной формы модели. Здесь возникает проблема идентифицируемости, заключающаяся в том, что не всегда возможно по приведенным коэффициентам модели однозначно определить ее структурные коэффициенты.

Это связано с тем, что в общем случае структурная и приведенная формы модели содержат разное число параметров $n \cdot (n-1) + n \cdot m$ и $n \cdot m$. Чтобы уравнивать число параметров, необходимо предположить равенство нулю некоторых структурных коэффициентов модели либо наличие между ними определенных соотношений, например, $a_{11} + b_{12} = 0$.

С позиции идентифицируемости можно выделить три вида структурных моделей:

- *идентифицируемые* системы, в которых число параметров структурной и приведенной форм модели совпадает, и структурные коэффициенты модели однозначно оцениваются через параметры приведенной формы модели;
- *неидентифицируемые* системы, в которых число структурных параметров превышает число приведенных, и структурные коэффициенты не могут быть получены из коэффициентов приведенной формы модели;
- *сверхидентифицируемые* системы с числом приведенных параметров превышающих число структурных. В этом случае возможно неоднозначное определение значений структурных коэффициентов при полученных значениях приведенных коэффициентах.

При исследовании структурной модели на идентифицируемость необходимо проверять каждое уравнение. Модель считается идентифицируемой, если каждое уравнение системы идентифицируемо, и неидентифицируемой, если хотя бы одно из уравнений системы неидентифицируемо. Сверхидентифицируемая модель содержит только идентифицируемые и сверхидентифицируемые уравнения.

Необходимое условие идентифицируемости. Обозначим через H число эндогенных переменных в уравнении, а через D – число предопределенных переменных, отсутствующих в уравнении, но присутствующих в системе. Необходимое условие идентифицируемости формулируется следующим образом:

- уравнение идентифицируемо, если $D+1 = H$;
- уравнение неидентифицируемо, если $D+1 < H$;
- уравнение сверхидентифицируемо, если $D+1 > H$.

Иными словами, для того, чтобы уравнение было идентифицируемо, необходимо, чтобы число предопределенных переменных модели, отсутствующих в данном уравнении, было на единицу меньше, чем число эндогенных переменных, входящих в данное уравнение.

Например, для первого уравнения системы (4.16) выполняются соотношения $H = 2$, $D = 3$. Следовательно, $D+1 > H$, и первое уравнение системы (4.16) сверхидентифицируемо.

Достаточное условие идентифицируемости. Уравнение, соответствующее переменной y_i , идентифицируемо, если ранг матрицы, составленной из коэффициентов при переменных модели, отсутствующих в исследуемом уравнении, но входящих в остальные уравнения системы, равен числу эндогенных переменных системы без единицы

$$\text{Rank}([B A]_i) = n - 1,$$

где $[B A]$ – блочная матрица коэффициентов, составленная из матриц B и A ;

$[B A]_i$ – матрица, полученная из матрицы $[B A]$ в результате удаления i -строки и столбцов, соответствующих объясняющим переменным входящим в i -уравнение.

Проверим достаточное условие для первого уравнения системы (4.16) .

Эндогенные переменные модели: C_t, I_t, r_t, Y_t .

Предопределенные переменные модели: $M_t, G_t, C_{t-1}, I_{t-1}$.

Общая матрица $[B A]$ коэффициентов уравнений системы (4.16), столбцы которой соответствуют переменным $C_t, I_t, r_t, Y_t, M_t, G_t, C_{t-1}, I_{t-1}$ имеет вид

$$[B A] = \begin{array}{|c|c|c|c|c|c|c|c|} \hline -1 & & & b_{11} & & & b_{12} & \\ \hline & -1 & b_{21} & & & & & b_{22} \\ \hline & & -1 & b_{31} & b_{32} & & & \\ \hline 1 & 1 & & -1 & & 1 & & \\ \hline \end{array}$$

Первое уравнение содержит переменные C_t, Y_t, C_{t-1} . Запишем матрицу $[B A]_1$, полученную вычеркиванием из матрицы $[B A]$ первой строки и столбцов, соответствующих переменным C_t, Y_t, C_{t-1}

$$[B A]_1 = \begin{array}{|c|c|c|c|c|} \hline -1 & b_{21} & & & b_{22} \\ \hline & -1 & b_{32} & & \\ \hline 1 & & & 1 & \\ \hline \end{array}$$

Ранг матрицы равен трем, т. к.

$$\text{Det} \begin{bmatrix} -1 & 0 & b_{22} \\ b_{32} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} = b_{22} \cdot b_{32} \neq 0$$

Следовательно, достаточное условие идентифицируемости для первого уравнения системы (4.16) выполняется.

4.3. Косвенный метод наименьших квадратов

Наиболее часто для оценки параметров системы одновременных уравнений применяются косвенный, двухшаговый и трехшаговый методы наименьших квадратов (КМНК, ДМНК и ТМНК). Первый из них используется только в случае идентифицируемых уравнений. Реже применяется универсальный, но очень сложный в вычислительном отношении метод максимального правдоподобия.

Косвенный МНК используется в случае идентифицируемой системы уравнений и заключается в следующем:

1) исходная система уравнений преобразуется в приведенную форму модели и определяются численные значения параметров δ_{ij} для каждого ее уравнения в отдельности с помощью традиционного МНК;

2) путем алгебраических преобразований осуществляется переход от приведенной формы к уравнениям структурной формы модели, что автоматически дает численные оценки структурных параметров.

Например, требуется найти структурные параметры модели

$$\begin{aligned} y_1 &= b_{12}y_2 + a_{11}x_1 + \varepsilon_1; \\ y_2 &= b_{21}y_1 + a_{22}x_2 + \varepsilon_2, \end{aligned} \quad (4.18)$$

при условии, что полученная приведенная форма модели описывается уравнениями

$$\begin{aligned} y_1 &= 2 \cdot x_1 + 4 \cdot x_2; \\ y_2 &= x_1 - x_2. \end{aligned}$$

Проверим идентифицируемость уравнений. В модели имеется две эндогенные переменные y_1, y_2 и две экзогенные переменные x_1, x_2 . В первое уравнение входят две эндогенные переменные y_1, y_2 и одна экзогенная переменная x_2 . Следовательно, $H = 2, D = 1$ и $H = D + 1$, и первое уравнение – идентифицируемо. Идентифицируемость второго уравнения доказывается аналогично. Для нахождения структурных коэффициентов можно применить косвенный МНК, т. е. получить их с помощью преобразования приведенных уравнений.

Для этого из 2-го уравнения приведенной формы выразим переменную $x_2 = x_1 - y_2$ и подставим в 1-е уравнение приведенной формы модели

$$y_1 = 2 \cdot x_1 + 4 \cdot (x_1 - y_2) \quad \text{или} \quad y_1 = -4 \cdot y_2 + 6 \cdot x_1.$$

Сравнивая это уравнение с 1-м уравнением структурной формы (4.18) $y_1 = b_{12}y_2 + a_{11}x_1$, определим значения структурных параметров

$$b_{12} = -4; \quad a_{11} = 6.$$

Далее из первого уравнения приведенной формы выразим переменную $x_1 = \frac{1}{2}y_1 - 2 \cdot x_2$ и подставим во 2-е уравнение приведенной формы модели

$$y_2 = \left(\frac{1}{2}y_1 - 2 \cdot x_2\right) - x_2 \quad \text{или} \quad y_2 = \frac{1}{2}y_1 - 3 \cdot x_2.$$

Сравнивая последнее уравнение с 2-м структурной формы (4.16) $y_2 = b_{21}y_1 + a_{22}x_2$, получим

$$b_{21} = \frac{1}{2}; \quad a_{22} = 3.$$

Таким образом, структурная форма модели определяется уравнениями

$$\begin{aligned} y_1 &= -4 \cdot y_2 + 6 \cdot x_1 + \varepsilon_1 \\ y_2 &= \frac{1}{2}y_1 - 3 \cdot x_2 + \varepsilon_2 \end{aligned} \quad (4.19)$$

4.4. Двухшаговый метод наименьших квадратов

Двухшаговый МНК основан на использовании, так называемых, «инструментальных» переменных и является универсальным методом. Как уже отмечалось, в системе одновременных уравнений нарушаются предпосылки о независимости факторов (выражаемых эндогенными переменными) и ошибок уравнений. Для преодоления этой трудности можно использовать замену эндогенных переменных y_i в правых частях уравнений модели на вспомогательные «инструментальные» переменные \hat{y}_i , которые были бы близки к исходным эндогенным переменным и при этом не зависели бы от ошибок уравнений. В качестве таких переменных предлагается использовать переменные, определяемые уравнениями приведенной формы модели (4.15).

Согласно *двухшаговому МНК*, численные значения структурных параметров определяются в следующей последовательности:

1) Исходная система уравнений преобразуется в приведенную форму модели и определяются численные значения параметров δ_{ij} для каждого ее уравнения в отдельности с помощью традиционного МНК;

2) По полученным уравнениям приведенной формы находятся расчетные значения инструментальных переменных \hat{y}_i , соответствующих эндогенным переменным y_i для каждого наблюдения;

3) С помощью обычного МНК определяются параметры каждого структурного уравнения в отдельности, используя в качестве факторов фактические значения предопределенных переменных и полученные расчетные значения инструментальных переменных \hat{y}_i .

Рассмотрим в качестве примера модифицированную модель Кейнса [5]

$$\begin{aligned} C_t &= a_1 + b_{11} \cdot Y_t + \varepsilon_1; \\ I_t &= a_2 + b_{21} \cdot Y_t + b_{22} \cdot Y_{t-1} + \varepsilon_2; \\ Y_t &= C_t + I_t + G_t, \end{aligned} \quad (4.20)$$

где Y – валовой национальный доход; C – личное потребление; I – инвестиции; G – государственные расходы; t и $t-1$ обозначают текущий и предыдущий периоды; ε_1 и ε_2 – случайные ошибки.

Информация об уровнях всех показателей за двенадцать лет дана в табл. 4.1.

Таблица 4.1

Данные для макроэкономической модели Кейнса

Год наблюдения	C_t	I_t	Y_t	Y_{t-1}	G_t	\hat{Y}_t
1	1016,6	267,0	1412,7	–	486,1	–
2	1435,9	376,0	1978,9	1412,7	652,7	2243,7
3	1776,1	408,8	2292,0	1978,9	839,0	2899,5
4	2003,8	407,1	2514,4	2292,0	842,1	3158,6
5	3265,7	670,4	4632,0	2514,4	1258,0	3771,6
6	4476,9	1165,2	7116,6	4632,0	1960,1	6230,0
7	5886,9	1504,7	8819,9	7116,6	2419,4	8736,4
8	7443,2	1762,4	10627,5	8819,9	3422,3	11168,2
9	9024,8	2186,4	12886,1	10627,5	3964,9	13207,8
10	11401,4	2865,0	16679,9	12886,1	4669,7	15784,2
11	14363,5	3611,1	21079,5	16679,9	6820,6	21114,7
12	17742,6	4580,5	26009,7	21079,5	8375,2	26321,7

В модели имеются три эндогенные переменные Y_t , C_t , I_t и две предопределенные переменные Y_{t-1} и G_t .

Первое уравнение свержидентифицируемо, т. к. $H = 2$, $D = 2$ и $H < D + 1$.

Второе уравнение идентифицируемо, т. к. $H = 2$, $D = 1$ и $H = D + 1$.

Применяя обычный МНК получим систему приведенных уравнений

$$\begin{aligned} C_t &= 377,5 + 0,582 \cdot Y_{t-1} + 0,632 \cdot G_t; \\ I_t &= 19,3 + 0,154 \cdot Y_{t-1} + 0,155 \cdot G_t; \\ Y_t &= 412,5 + 0,817 \cdot Y_{t-1} + 1,037 \cdot G_t. \end{aligned} \quad (4.21)$$

Подставляя данные наблюдений из таблицы 4.1 в 3-е уравнение приведенной формы (4.21) определим расчетные значения инструментальной переменной \hat{Y}_t (табл. 4.1), соответствующей эндогенной переменной Y .

Применяя МНК последовательно к уравнениям структурной формы модели

$$\begin{aligned} C_t &= a_1 + b_{11} \cdot \hat{Y}_t + \varepsilon_1; \\ I_t &= a_2 + b_{21} \cdot \hat{Y}_t + b_{22} \cdot Y_{t-1} + \varepsilon_2 \end{aligned}$$

получим окончательный вид структурной формы модели

$$\begin{aligned} C_t &= 97,66 + 0,678 \cdot Y_t + \varepsilon_1; \\ I_t &= -42,47 + 0,150 \cdot Y_t + 0,031 \cdot Y_{t-1} + \varepsilon_2. \end{aligned} \quad (4.22)$$

Из уравнений (4.22) следует, что 67,8 % прироста национального дохода идет на увеличение потребления. На увеличение инвестиций направляется соответственно 15 % и 3,1 % прироста национального дохода текущего года и предыдущего года.

4.5. Трехшаговый метод наименьших квадратов

Более эффективным, но требующим существенно больших вычислительных затрат, является трехшаговый метод наименьших квадратов (ТМНК) [3]. Он заключается в том, что двухшаговый метод наименьших квадратов применяется не к исходным уравнениям модели, а к уравнениям, преобразованным согласно обобщенному методу наименьших квадратов. Трехшаговый МНК является итерационной процедурой:

- 1) Параметры модели определяются обычным или двухшаговым МНК.
- 2) Вычисляются ошибки модели и определяется оценка корреляционной матрицы ошибок.
- 3) Уравнения преобразуются согласно обобщенному МНК.
- 4) Применяется двухшаговый МНК к преобразованным уравнениям и получается улучшенная модель (с улучшенными параметрами).
- 5) Процесс повторяется, начиная со второго шага, пока не будет достигнута заданная точность (либо превышено заданное количество итераций).

Если случайные члены структурной модели не коррелируют, то трехшаговый метод сводится к двухшаговому.

Контрольные вопросы

1. В каких случаях модель строится в виде систем эконометрических уравнений?
2. Какие проблемы возникают при оценке параметров систем эконометрических уравнений?
3. Какие переменные называются эндогенными и предопределенными?
4. Что представляет собой структурная форма модели?
5. Что представляет собой приведенная форма модели?
6. В чем заключается проблема идентифицируемости модели?
7. Как проверяется идентифицируемость уравнений модели?
8. Какие методы применяются для нахождения структурных коэффициентов модели для различных видов систем уравнений?
9. Что представляет собой косвенный МНК?
10. Что представляет собой двухшаговый МНК?
11. Какие требования предъявляются к инструментальным переменным в двухшаговом МНК?

5. Моделирование одномерных временных рядов и прогнозирование

5.1. Составляющие временного ряда

Временным рядом (рядом динамики, динамическим рядом) называется упорядоченная во времени последовательность численных показателей $\{(y_i, t_i), i=1, 2, \dots, n\}$, характеризующих уровни развития изучаемого явления в последовательные моменты или периоды времени (табл. 5.1).

Таблица 5.1

Динамика ВВП Российской Федерации

	2000 г.	2001 г.	2002 г.	2003 г.	2004 г.
ВВП, млрд руб.	7305,6	8943,6	10834,2	13285,2	17048,1

Величины y_i называются уровнями ряда, а t_i – временными метками (моментами или интервалами наблюдения). Обычно рассматриваются временные ряды с равными интервалами между наблюдениями, в качестве значений t_i берутся порядковые номера наблюдений и временной ряд представляется в виде последовательности y_1, y_2, \dots, y_n , где n – количество наблюдений.

Целью исследования временного ряда является выявление закономерностей в изменении уровней ряда и построении его модели в целях прогнозирования и исследования взаимосвязей между явлениями.

При исследовании экономического временного ряда его обычно представляют в виде совокупности трех составляющих:

- долговременной тенденции (Т), т. е. устойчивого увеличения или уменьшения значений уровней ряда (тренда);
- периодических колебаний (S);
- случайных колебаний (E).

На рис. 5.1 показан график временного ряда, на котором прослеживаются все три составляющие.

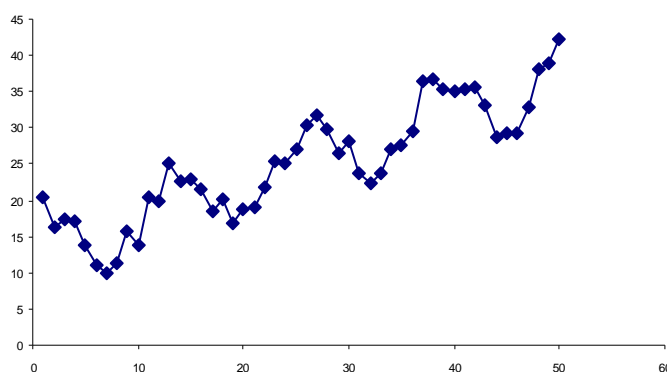


Рис. 5.1. Временной ряд

Различным образом объединяя эти компоненты, можно получить различные модели временного ряда (Y):

- аддитивную

$$Y_t = T_t + S_t + E_t; \quad (5.1)$$

– мультипликативную

$$Y_t = T_t \cdot S_t \cdot E_t; \quad (5.2)$$

– смешанную

$$Y_t = T_t \cdot S_t + E_t. \quad (5.3)$$

В экономике периодические колебания принято подразделять на *сезонные*, у которых период колебаний не превышает одного года (цены на сельскохозяйственную продукцию), вызванные климатическими или социально-экономическими причинами, и *циклические* с периодом колебаний несколько лет, связанные с циклами деловой активности.

Основная задача эконометрического исследования временного ряда заключается в выявлении и придании количественного выражения составляющим его отдельным компонентам.

Как правило, наличие той или иной составляющей можно определить с помощью визуального анализа графика временного ряда (рис. 5.1).

Перед построением модели исходные данные проверяются на сопоставимость (применение одинаковой методики получения или расчета данных), однородность (отсутствие случайных выбросов), устойчивость (наличие закономерности в изменении уровней ряда) и достаточность (число наблюдений должно в 7–10 превосходить число параметров модели).

5.2. Автокорреляция уровней временного ряда

Важной особенностью временных рядов по сравнению с данными наблюдений, относящихся к одному периоду времени, является, как правило, наличие связи между последовательными уровнями ряда, вызванное действием каких-либо долговременных причин, что приводит к наличию таких составляющих ряда, как долговременная тенденция и периодическая составляющая.

Корреляционная зависимость между последовательными уровнями временного ряда называется *автокорреляцией уровней* временного ряда.

Степень тесноты автокорреляционной связи между уровнями ряда может быть определена с помощью коэффициентов автокорреляции, т. е. коэффициентов линейной корреляции между уровнями исходного временного ряда и уровнями ряда, сдвинутыми на несколько шагов назад во времени.

$$r_\tau = \frac{\sum_{t=\tau+1}^n (y_t - \bar{y}_{1\tau}) \cdot (y_{t-\tau} - \bar{y}_{2\tau})}{\sqrt{\sum_{t=\tau+1}^n (y_t - \bar{y}_{1\tau})^2 \cdot \sum_{t=\tau+1}^n (y_{t-\tau} - \bar{y}_{2\tau})^2}}, \quad (5.4)$$

где τ – величина сдвига, называемая *лагом*, определяет порядок коэффициента автокорреляции,

$$\bar{y}_{1\tau} = \frac{\sum_{t=\tau+1}^n y_t}{n - \tau}; \quad \bar{y}_{2\tau} = \frac{\sum_{t=\tau+1}^n y_{t-\tau}}{n - \tau}.$$

Функцию $r(\tau) = r_\tau$ называют *автокорреляционной функцией* временного ряда, а ее график – *коррелограммой*.

Анализ автокорреляционной функции и коррелограммы позволяет выявить структуру ряда, т. е. определить присутствие в ряде той или иной компоненты.

Так, если наиболее высоким оказался коэффициент автокорреляции первого порядка, то исследуемый ряд содержит только тенденцию.

Если наиболее высоким оказался коэффициент автокорреляции порядка m , то ряд содержит циклические колебания с периодичностью в m моментов времени.

Если ни один из коэффициентов автокорреляции не является значимым, то ряд не содержит тенденции и циклических колебаний.

Необходимо подчеркнуть, что линейные коэффициенты автокорреляции характеризуют тесноту только линейной связи текущего и предыдущих уровней ряда. Поэтому по коэффициентам автокорреляции можно судить только о наличии или отсутствии линейной (или близкой к линейной) зависимости. Для проверки ряда на наличие нелинейной тенденции рекомендуется вычислить линейные коэффициенты автокорреляции для временного ряда, состоящего из логарифмов исходных уровней. Отличные от нуля значения коэффициентов автокорреляции будут свидетельствовать о наличии нелинейной тенденции.

5.3. Моделирование тенденции временного ряда

Моделирование тенденции временного ряда является важнейшей классической задачей анализа экономических временных рядов. Решение этой задачи начинается, как правило, с проверки наличия тенденции и формулирования предложений о характере долговременной тенденции, после чего уже строится модель тенденции как функции времени.

5.3.1. Методы определения наличия тенденции

Для диагностирования наличия тенденции наиболее широко применяются метод сравнения средних и метод Фостера-Стюарта.

Метод сравнения средних. Метод сравнения средних применим для выявления монотонной тенденции.

Временной ряд разбивается на две примерно равные части y_1, y_2, \dots, y_{n_1} и $y_{n_1+1}, y_{n_1+2}, \dots, y_{n_1+n_2}$ с количеством уровней n_1 и n_2 и для каждой части вычисляются средние (\bar{y}_1, \bar{y}_2) и выборочные дисперсии (s_1^2, s_2^2) соответственно.

Далее рассчитывается значение критерия Стьюдента по формуле

$$\tau = \frac{|\bar{y}_1 - \bar{y}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}, \quad (5.5)$$

если предполагается, что значения дисперсий на этих участках не равны между собой, т. е. $\sigma_1^2 \neq \sigma_2^2$, и по формуле

$$\tau = \frac{|\bar{y}_1 - \bar{y}_2|}{s^2} \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}, \quad (5.6)$$

где s^2 – общая выборочная дисперсия ряда, если предполагается, что дисперсии одинаковы $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

Нулевая гипотеза о равенстве средних (об отсутствии тенденции) отвергается, если выполняется условие

$$\tau > t_{1-\alpha, m}, \quad (5.7)$$

где $t_{1-\alpha, m}$ – табличное значение t -критерия Стьюдента при уровне значимости α и числе степеней свободы $m = n_1 + n_2 - 2$.

Метод Фостера-Стюарта. Является более универсальным и дает более надежные результаты. Каждому уровню ряда y_i , начиная со второго, ставится в соответствие два значения p_i q_i по следующим правилам:

$p_i = 1$, если уровень y_i меньше всех предыдущих уровней, т. е.

$$y_i < y_1, y_2, \dots, y_{i-1},$$

и $p_i = 0$ в противном случае;

$q_i = 1$, если уровень y_i больше всех предыдущих уровней, т. е.

$$y_i > y_1, y_2, \dots, y_{i-1},$$

и $q_i = 0$ в противном случае.

Вычисляется статистика

$$t_p = \frac{\sum_{i=2}^n (p_i - q_i)}{2 \sum_{i=2}^n \frac{1}{i}}. \quad (5.8)$$

Гипотеза об отсутствии тенденции отвергается, если выполняется условие

$$t_p > t_{1-\alpha, n-1},$$

где $t_{1-\alpha, n-1}$ – табличное значение t -критерия Стьюдента при уровне значимости α и числе степеней свободы $n - 1$.

5.3.2. Сглаживание временного ряда по методу скользящей средней

Цель сглаживания временного ряда заключается в получении ряда с меньшим разбросом уровней, что в ряде случаев позволяет на основе визуального анализа сделать вывод о наличии тенденции и ее характерных особенностях.

Сглаживание временного ряда по методу скользящей средней заключается в замене исходных уровней ряда y_t сглаженными значениями y'_t , которые получаются как среднее значение определенного числа уровней исходного ряда, симметрично окружающего значение y_t . В результате получается временной ряд y'_t , меньше подверженный колебаниям. Если индивидуальный разброс значений временного ряда около своего среднего значения a характеризуется дисперсией σ^2 , то средняя из m членов ряда $(y_1 + y_2 + \dots + y_m)/m$ будет иметь в m раз меньшую дисперсию (σ^2/m).

Для вычисления сглаженных значений y'_t по методу простой скользящей средней используются следующие формулы:

1) *Нечетный интервал сглаживания* $g = 2p+1$ (интервал сглаживания – количество исходных уровней ряда (y_t), используемых для сглаживания):

$$y'_t = \frac{\sum_{i=t-p}^{t+p} y_i}{2p+1} = \frac{y_{t-p} + y_{t-p+1} + \dots + y_{t+p-1} + y_{t+p}}{2p+1}, \quad (5.9)$$

где y_t – фактическое значение уровня исходного ряда в момент t ; y'_t – значение скользящей средней в момент t ; $2p+1$ – длина интервала сглаживания.

Формула (5.9) при интервалах сглаживания $g = 3$ и $g = 5$ принимает вид

$$y'_t = \frac{y_{t-1} + y_t + y_{t+1}}{3}; \quad y'_t = \frac{y_{t-2} + y_{t-1} + y_t + y_{t+1} + y_{t+2}}{5}.$$

2) Четный интервал сглаживания $g = 2p$:

$$\begin{aligned} y'_t &= \frac{\frac{1}{2}y_{t-p} + y_{t-p+1} + \dots + y_{t-1} + y_t + y_{t+1} + \dots + y_{t+p-1} + \frac{1}{2}y_{t+p}}{2p} = \\ &= \frac{\frac{1}{2}y_{t-p} + \sum_{i=t-p+1}^{t+p-1} y_i + \frac{1}{2}y_{t+p}}{2p}. \end{aligned} \quad (5.10)$$

Формула (5.10) при интервалах сглаживания $g = 2$ и $g = 4$ принимает вид

$$y'_t = \frac{\frac{1}{2}y_{t-1} + y_t + \frac{1}{2}y_{t+1}}{2}; \quad y'_t = \frac{\frac{1}{2}y_{t-2} + y_{t-1} + y_t + y_{t+1} + \frac{1}{2}y_{t+2}}{4}.$$

При использовании скользящей средней с длиной активного участка $g = 2p+1$ первые и последние p уровней ряда сгладить нельзя, их значения теряются. Очевидно, что потеря значений последних точек является существенным недостатком, т. к. для исследователя последние «свежие» данные обладают наибольшей информационной ценностью.

Для восстановления потерянных значений временного ряда можно использовать следующий прием:

а) Вычисляется средний прирост Δ_y на последнем активном участке (y_{n-g}, \dots, y_n)

$$\Delta_y = \frac{y_n - y_{n-g}}{g-1},$$

где g – длина активного участка.

б) Определяются значения последних $p = (g-1)/2$ уровней сглаженного временного ряда с помощью последовательного прибавления среднего абсолютного прироста Δ_y к последнему сглаженному значению y'_{n-p}

$$y'_{n-p+1} = y'_{n-p} + \Delta_y, \quad y'_{n-p+2} = y'_{n-p+1} + \Delta_y, \quad \dots, \quad y'_n = y'_{n-1} + \Delta_y.$$

Аналогичная процедура применяется для восстановления первых p уровней временного ряда.

Отметим, что важным свойством процедуры сглаживания является *полное устранение* периодических колебаний из временного ряда, если длина интервала сглаживания берется равной или кратной периоду колебаний. Это обстоятельство используется при выделении периодической составляющей временного ряда (п. 5.)

5.3.3. Метод аналитического выравнивания

Аналитическим выравниванием временного ряда называют нахождение аналитической функции $\hat{y} = f(t)$, характеризующей основную тенденцию изменения уровней ряда с течением времени. Сама функция $f(t)$ носит название *кривой роста*.

При аналитическом выравнивании (нахождении аналитической функции $\hat{y} = f(t)$) исходят из предположения, что аддитивная модель временного ряда может быть представлена как сумма двух компонент

$$y(t) = f(t) + \varepsilon_t, \quad (5.11)$$

где ε_t – случайная компонента с нулевой средней и постоянной дисперсией выражает ошибку модели из-за действия случайных факторов.

Чаще всего в качестве кривой роста применяются следующие функции:

– линейная $y_t = a_0 + a_1 t$; (5.12)

– парабола второго и более высоких порядков

$$y_t = a_0 + a_1 t^1 + a_2 t^2 + \dots + a_k t^k; \quad (5.13)$$

– гиперболическая $y_t = a_0 + a_1 / t$; (5.14)

– экспонента $y_t = e^{a_0 + a_1 t}$; (5.15)

– потенциальная $y_t = a_0 \times a_1^t$; (5.16)

– модифицированная экспонента $y_t = K + a_0 \times a_1^t$; (5.17)

– степенная $y_t = a_0 \times t^{a_1}$; (5.18)

– логистическая кривая $y_t = \frac{K}{1 + a_0 e^{-a_1 t}}$; (5.19)

– кривая Гомперца $y_t = K \times a_0^{a_1^t}$. (5.20)

Построение таких функций ничем не отличается от построения уравнений парной регрессии (линейной или нелинейной) с учетом того, что в качестве зависимой переменной используются фактические уровни временного ряда y_t , а в качестве независимой переменной моменты времени $t = 1, 2, \dots, n$. Для построения кривой роста необходимо выбрать вид аналитической зависимости и затем оценить значения ее параметров.

Для определения вида тенденции (аналитической зависимости) применяются такие методы, как

- качественный анализ изучаемого процесса;
- построение и визуальный анализ графика зависимости уровней ряда от времени;
- расчет и анализ показателей динамики временного ряда (абсолютные приросты, темпы роста и др.);
- анализ автокорреляционной функции исходного и преобразованного временного ряда;

– метод перебора, при котором строятся кривые роста различного вида с последующим выбором наилучшей на основании значения скорректированного коэффициента детерминации \bar{R}^2 (3.28).

5.3.4. Выбор вида тенденции

Выбор вида тенденции на основе качественного анализа. Социально-экономические процессы в зависимости от характера их протекания можно разделить на три класса (рис. 5.2):

I) Процессы с монотонным характером развития и отсутствием пределов роста (рис. 5.2, а). Эти условия справедливы для поведения многих экономических показателей, например, для большинства натуральных показателей промышленного производства. В этом случае для моделирования тенденции могут использоваться: линейная $y_t = a_0 + a_1 t$, параболическая $y_t = a_0 + a_1 t^1 + a_2 t^2 + \dots + a_k t^k$, экспоненциальная $y_t = e^{a_0 + a_1 t}$, степенная $y_t = a_0 \times t^{a_1}$ функции.

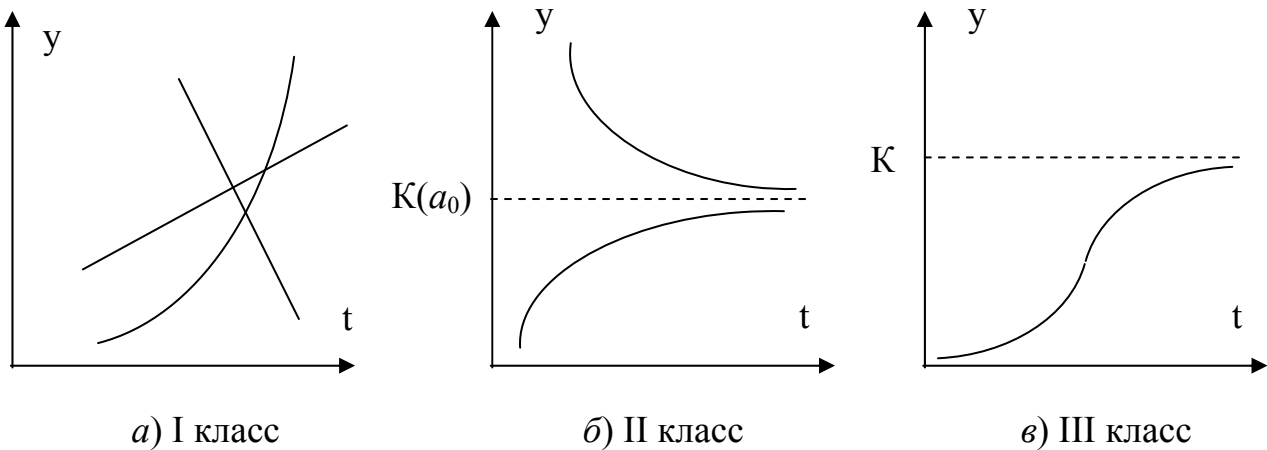


Рис. 5.2. Схемы протекания процессов

II) Процессы, которые имеют предел роста (падения) в исследуемом периоде, так называемые процессы с «насыщением» (рис. 5.2, б). Развитие процесса происходит под влиянием некоторых ограничивающих факторов, величина воздействия которых растет вместе с ростом достигнутого уровня. С такими процессами часто сталкиваются в демографии, при изучении потребностей в товарах и услугах (в расчете на душу населения), при исследовании эффективности использования ресурсов и т. д. Примерами показателей, для которых могут быть указаны пределы роста, являются среднедушевое потребление определенных продуктов питания, расход удобрений на единицу площади и т. п.

В этом случае для моделирования тенденции используются гиперболическая функция $y_t = a_0 + \frac{a_1}{t}$ либо модифицированная экспонента $y_t = K + a_0 \cdot a_1^t$ с параметром a_1 , удовлетворяющим условию $0 < a_1 < 1$.

В случае гиперболы параметр a_0 равен пределу роста, к которому значение уровня процесса приближается (при росте t) снизу в случае $a_1 < 0$, либо сверху при $a_1 > 0$ (рис. 5.2 б).

В случае модифицированной экспоненты параметр K равен пределу роста, к которому значение уровня процесса приближается (при росте t) снизу в случае $a_0 < 0$, либо сверху при $a_0 > 0$ (рис. 5.2, б).

При решении экономических задач часто можно определить значение предела роста исходя из свойств прогнозируемого процесса (например, коэффициент использования оборудования не может превышать 1 и т. п.). Иногда значение предела роста задается экспертным путем.

III) Так называемые S-образные процессы (рис. 5.2, в), представляющие как бы два последовательных лавинообразных процесса (когда прирост зависит от уже достигнутого уровня): один с ускорением развития, а другой – с замедлением. С такими процессами часто сталкиваются в демографических исследованиях, в страховых расчетах, при решении задач прогнозирования научно-технического прогресса, при определении спроса на новый вид продукции.

К S-образным процессам можно отнести процесс развитие новой отрасли (нового производства). Вначале производство развивается очень медленно вследствие того, что технические методы производства еще недостаточно разработаны, издержки производства высоки и спрос на рынке на данный товар еще очень мал, поэтому производство развивается медленно. В дальнейшем, благодаря усовершенствованию технических методов изготовления, переходу к массовому производству и увеличению емкости рынка для данного товара производство растет быстрее. Затем наступает период насыщения рынка, рост производства все более замедляется, и, наконец, почти прекращается. Наступает стабилизация производства на определенном уровне. Следует отметить, что выявленные закономерности развития следует обобщать с определенной осторожностью и, причем, только для достаточно коротких периодов, так как выявленная тенденция развития производства может быть нарушена вследствие внешних факторов, например, технического переворота в данной отрасли или связанных с нею.

Для моделирования тенденции S-образных процессов следует использовать либо логистическую функцию $y_t = \frac{K}{1 + a_0 e^{-a_1 t}}$ (с параметром $a_1 < 1$), либо

кривую Гомперца $y_t = K \times a_0 a_1^t$ с параметрами, удовлетворяющими условиям $0 < a_0, a_1 < 1$. Предел роста в обоих случаях равен параметру K .

Выбор вида тенденции на основе анализа показателей динамики временного ряда. Исследуя последовательные разности уровней ряда y_t первого, второго и следующих порядков

$\Delta y_t = y_t - y_{t-1}$ – последовательные разности первого порядка,

$\Delta^2 y_t = \Delta y_t - \Delta y_{t-1}$ – последовательные разности второго порядка и т. д.,

можно сделать вывод о наличии тенденции, описываемой полиномиальной функцией от времени t .

Если исходный временной ряд содержит тенденцию, а временной ряд последовательных разностей первого порядка не содержит тенденцию, то можно сделать вывод, что тенденция линейно зависит от времени $y_t = a_0 + a_1 t$. Коэффициент a_1 в данном случае численно равен среднему абсолютному приросту уровня явления за единицу измерения временного параметра t (за сутки, неделю, месяц, год и т. д.).

Если исходный временной ряд и временной ряд последовательных разностей первого порядка содержат тенденцию, а временной ряд последовательных разностей второго порядка не содержит тенденцию, то можно сделать вывод, что тенденция задается полиномом второго порядка от времени $y_t = a_0 + a_1 t + a_2 t^2$. Относительно тенденции в виде полинома от t более высокой степени вывод делается аналогично.

Исследуя отношения последовательных уровней ряда (цепные коэффициенты роста) $k_t = y_t/y_{t-1}$, можно сделать вывод о наличии тенденции, задаваемой экспоненциальной функцией от времени t .

Если исходный временной ряд содержит тенденцию, а временной ряд коэффициентов роста не содержит тенденцию, то можно сделать вывод, что тенденция экспоненциально зависит от времени. Иными словами, тенденция имеет вид экспоненциальной функции $y_t = a_0 \times a_1^t$. Величина a_1 в данном случае численно равна среднему коэффициенту роста уровня явления за единицу измерения временного параметра t (за сутки, неделю, месяц, год и т. д.).

Аналогичный результат можно получить, анализируя первые последовательные разности временного ряда, составленного из логарифмов от исходных уровней.

Если наблюдается линейная зависимость между логарифмами уровней ряда $\ln y_t$ и соответствующих промежутков времени $\ln t$, то рекомендуется использовать степенную функцию $y_t = a_0 \times t^{a_1}$.

5.3.5. Оценка адекватности и точности модели тенденции

После построения модели тенденции осуществляется проверка ее качества по характеристикам адекватности (соответствия данным наблюдения) и точности.

Проверка адекватности модели основывается на анализе ряда остатков

$$e_t = y_t - \hat{y}_t. \quad (i = 1, 2, \dots, n) \quad (5.21)$$

Модель считается адекватной, если остатки:

- являются случайными;
- распределены по нормальному закону;
- имеют равное нулю среднее значение $\bar{e} = 0$;
- независимы между собой.

1) Проверка случайности остатков заключается в установлении факта отсутствия или наличия тенденции остатков. Для этой цели может использоваться критерий серий. Предварительно определяется медиана e_m упорядоченного ряда остатков. Каждому элементу ряда остатков e_t ставится в соответствие знак

«+», если $e_t > e_m$, и знак «-», если $e_t < e_m$. Непрерывно идущую последовательность одинаковых знаков принято называть серией. Определяется максимальная длина серии L_{max} и число серий V . Остатки считаются случайными на уровне значимости 0,05, если одновременно выполняются два условия

$$L_{max} < [3,3(\lg n + 1)], \quad (5.22)$$

$$V > \left[\frac{1}{2}(n + 1 - 1,96\sqrt{n-1}) \right]. \quad (5.23)$$

2) Нормальность распределения остатков считается установленной (приближенно) если одновременно выполняются следующие неравенства:

$$|A| < 1,5\delta_A; \quad \left| \dot{Y} + \frac{6}{n+1} \right| < 1,5\delta_{\dot{Y}}, \quad (5.24)$$

где A – выборочная характеристика асимметрии, \mathcal{E} – выборочная характеристика эксцесса, δ_A , $\delta_{\mathcal{E}}$ – среднеквадратические ошибки выборочных характеристик асимметрии и эксцесса, определяемые соотношениями

$$A = \frac{\frac{1}{n} \sum_{t=1}^n e_t^3}{\sqrt{\left(\frac{1}{n} \sum_{t=1}^n e_t^2\right)^3}}; \quad \dot{Y} = \frac{\frac{1}{n} \sum_{t=1}^n e_t^3}{\sqrt{\left(\frac{1}{n} \sum_{t=1}^n e_t^2\right)}} - 3; \quad (5.25)$$

$$\delta_A = \sqrt{\frac{6(n-2)}{(n+1)(n+3)}}; \quad \delta_{\dot{Y}} = \sqrt{\frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}}, \quad (5.26)$$

Если же выполняется хотя бы одно из неравенств

$$|A| \geq 2\delta_A; \quad \left| \mathcal{E} + \frac{6}{n+1} \right| \geq 2\delta_{\mathcal{E}}, \quad (5.27)$$

то гипотеза о нормальном характере распределения остатков отвергается. Другие случаи требуют дополнительной проверки с помощью более мощных критериев.

3) Проверка равенства нулю среднего значения ряда остатков $\bar{e} = 0$ осуществляется с помощью критерия Стьюдента. Гипотеза о равенстве нулю $\bar{e} = 0$ отвергается, если выполняется условие

$$t_p = \frac{|\bar{e}|}{s_e} \sqrt{n} > t_{1-\alpha, n-1}, \quad (5.28)$$

где

$$s_e = \sqrt{\frac{1}{n-1} \sum_{t=1}^n (e_t - \bar{e})^2}.$$

4) Под независимостью ряда остатков понимается отсутствие в нем автокорреляции, т. е. отсутствует зависимость каждого значения ряда от предыдущих значений. Если вид функции, описывающей систематическую составляющую, выбран неудачно, то последовательные значения ряда остатков могут коррелировать между собой.

Для проверки ряда остатков на отсутствие автокорреляции уровней остатков используется критерий Дарбина-Уотсона. Этот критерий основан на расчете величины

$$d = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2}, \quad (5.29)$$

представляющей собой отношение суммы квадратов разностей последовательных значений остатков к остаточной сумме квадратов по модели регрессии.

Между критерием Дарбина-Уотсона d и коэффициентом автокорреляции остатков первого порядка r_1^e имеет место следующее соотношение:

$$d \approx 2 \cdot (1 - r_1^e). \quad (5.30)$$

Таким образом, если в остатках существует полная положительная автокорреляция и $r_1^e = 1$, то $d = 0$. Если в остатках полная отрицательная автокорреляция, то $r_1^e = -1$ и, следовательно, $d = 4$. Если автокорреляция остатков отсутствует, то $r_1^e = 0$ и $d = 2$. Величина d изменяется в диапазоне $0 \leq d \leq 4$.

Применение критерия Дарбина-Уотсона для выявления автокорреляции остатков осуществляется в следующей последовательности.

а) Выдвигается нулевая гипотеза H_0 об отсутствии автокорреляции остатков. Альтернативные гипотезы H_1 и H_1^* состоят, соответственно, в наличии положительной или отрицательной автокорреляции в остатках.

б) По таблицам критерия Дарбина-Уотсона (см. приложение) определяются критические значения критерия d_L и d_U для заданного числа наблюдений n , числа факторов модели k и уровня значимости α . Этими значениями числовой промежутки $[0;4]$ разбивается на пять отрезков $(0, d_L)$, (d_L, d_U) , $(d_U, 4-d_U)$, $(4-d_U, 4-d_L)$, $(4-d_L, 4)$ (рис. 5.3).

в) Выдвинутые гипотезы принимаются или отклоняются с вероятностью $(1-\alpha)$ в зависимости от того, в какой отрезок попадет значение критерия d :

$(0, d_L)$ – принимается H_1 , остатки имеют положительную корреляцию;

(d_L, d_U) – зона неопределенности

$(d_U, 4-d_U)$ – принимается H_0 , автокорреляция остатков отсутствует;

$(4-d_U, 4-d_L)$ – зона неопределенности

$(4-d_L, 4)$ – принимается H_1^* , остатки имеют отрицательную корреляцию.

Применение критерия иллюстрирует рис. 5.2.

Есть положительная автокорреляция остатков. H_0 отклоняется. С вероятностью $P = (1-\alpha)$ принимается H_1	Зона неопределенности	Нет оснований отклонять H_0 (автокорреляция остатков отсутствует)	Зона неопределенности	Есть отрицательная автокорреляция остатков. H_0 отклоняется. С вероятностью $P = (1-\alpha)$ принимается H_1^*		
0	d_L	d_U	2	$4-d_U$	$4-d_L$	4

Рис. 5.3. Алгоритм проверки гипотезы о наличии автокорреляции остатков

К недостаткам критерия Дарбина-Уотсона относится наличие области неопределенности и то, что осуществляется проверка зависимости между ближайшими уровнями ряда.

Другим методом проверки наличия автокорреляции остатков является **тест серий (Бреуша-Годфри)**, основанный на оценке значимости коэффициентов авторегрессионного уравнения

$$e_t = \rho_1 e_{t-1} + \rho_2 e_{t-2} + \dots + \rho_k e_{t-k}, \quad (5.31)$$

полученных методом наименьших квадратов. Наличие значимых коэффициентов говорит об имеющейся автокорреляции остатков и ее характере.

Оценка точности модели тенденции заключается в оценке близости модельных значений тенденции к фактическим уровням ряда и осуществляется с помощью вычисления таких показателей, как:

- дисперсия остатков $\sigma_{\hat{m}\hat{o}}^2$;
- средняя ошибка аппроксимации $\bar{A} = \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right|$;
- коэффициент детерминации R^2 .

5.4. Моделирование периодических колебаний

5.4.1. Выделение периодической компоненты по методу скользящей средней

Простейшим приемом выделения периодической компоненты основано на использовании сглаживания временного ряда по методу простой скользящей средней. Предварительно следует определиться с видом модели временного ряда – аддитивной или мультипликативной. Это можно сделать на основе анализа график временного ряда. Если амплитуда периодических колебаний примерно постоянна, то следует выбрать аддитивную модель

$$Y = T + S + E,$$

в которой амплитуда колебаний периодической компоненты предполагается постоянной, не зависящей от времени. Если амплитуда периодических колебаний возрастает с ростом уровней ряда, то следует выбрать мультипликативную модель временного ряда

$$Y = T \cdot S \cdot E.$$

Выделение периодической компоненты основывается на том, что если исходный временной ряд содержит периодическую компоненту с периодом g , то сглаженный по методу простой скользящей средней временной с интервалом сглаживания g такой компоненты уже не содержит. Таким образом, в случае аддитивной модели периодическая компонента выделяется путем нахождения разности между соответствующими уровнями исходного и сглаженного ряда. В случае мультипликативной модели периодическая компонента выделяется путем нахождения отношения между соответствующими уровнями исходного и

сглаженного ряда. Затем вычисляются средние значения, соответствующие наблюдениям внутри одного периода колебаний.

5.4.2. Моделирование сезонных колебаний с помощью фиктивных переменных

Рассмотрим метод моделирования временного ряда, содержащего сезонные колебания, основанный на включении в модель фиктивных переменных. Количество фиктивных переменных принимается равным числу наблюдений в пределах одного цикла колебаний без единицы. Например, при моделировании поквартальных данных необходимо ввести три дополнительные переменные

$$z_1 = \begin{cases} 1, & \text{весна,} \\ 0, & \text{не весна,} \end{cases} \quad z_2 = \begin{cases} 1, & \text{лето,} \\ 0, & \text{не лето,} \end{cases} \quad z_3 = \begin{cases} 1, & \text{осень,} \\ 0, & \text{не осень.} \end{cases} \quad (5.32)$$

Зиме в этом случае соответствуют нулевые значения всех фиктивных переменных. Уравнение регрессии с учетом фиктивных переменных принимает вид

$$y = a + b \cdot t + c_1 \cdot z_1 + c_2 \cdot z_2 + c_3 \cdot z_3 + \varepsilon. \quad (5.33)$$

Коэффициенты c_i характеризуют отклонение уровней первых трех сезонов по отношению к последнему. Поэтому модель с фиктивными переменными может рассматриваться как частный случай аддитивной модели временного ряда.

5.4.3 Моделирование сезонных колебаний с помощью гармонического анализа

Согласно гармоническому анализу, временной ряд представляется как совокупность гармонических колебательных процессов. Для каждой точки этого ряда справедливо выражение

$$y_t = f(t) + \sum_{k=1} (a_k \cos(kt \frac{2\pi}{n}) + b_k \sin(kt \frac{2\pi}{n})). \quad (t = 1, 2, \dots, n) \quad (5.34)$$

Здесь y_t – фактический уровень ряда в момент (интервал) времени t ; $f(t)$ – выравненный уровень ряда в тот же момент времени, a_k, b_k – параметры колебательного процесса (гармоники) с номером k , в совокупности оценивающие размах (амплитуду) отклонения от общей тенденции и сдвиг колебаний относительно начальной точки.

Общее число колебательных процессов, которые можно выделить для ряда, состоящего из n уровней, равно $n/2$. Обычно ограничиваются меньшим числом наиболее важных гармоник. Параметры гармоники с номером k определяются по формулам:

$$\begin{aligned}
\varepsilon_t &= y_t - f(t); \\
a_k &= \frac{2}{n} \sum_{t=1}^n \varepsilon_t \cdot \cos\left(kt \frac{2\pi}{n}\right); \quad (k = 1, 2, \dots, n/k - 1) \\
b_k &= \frac{2}{n} \sum_{t=1}^n \varepsilon_t \cdot \sin\left(kt \frac{2\pi}{n}\right); \quad (k = 1, 2, \dots, n/k - 1) \\
a_{n/2} &= \frac{1}{n} \sum_{t=1}^n \varepsilon_t \cdot \cos(\pi t); \quad b_{n/2} = 0.
\end{aligned} \tag{5.35}$$

Этот метод хорошо подходит для аналитического выражения сезонных колебаний, имеющих синусоидальную форму. Если ограничиться первой гармоникой, то модель периодических колебаний принимает вид

$$y_t = a_0 + a_1 \cos\left(t \frac{2\pi}{n}\right) + b_1 \sin\left(t \frac{2\pi}{n}\right), \tag{5.36}$$

где

$$a_0 = \bar{y}; \quad a_1 = \frac{2}{n} \sum_{t=1}^n y_t \cdot \cos\left(t \frac{2\pi}{n}\right); \quad b_1 = \frac{2}{n} \sum_{t=1}^n y_t \cdot \sin\left(t \frac{2\pi}{n}\right). \tag{5.37}$$

5.5. Прогнозирование уровней временного ряда на основе кривых роста

5.5.1. Метод аналитического выравнивания

Построенная модель тенденции (кривая роста) может использоваться для прогнозирования. Кривая роста позволяет получить выровненные или теоретические значения уровней динамического ряда. Это те уровни, которые наблюдались бы в случае полного совпадения динамики явления с кривой.

Процедура разработки прогноза с использованием кривых роста включает в себя следующие этапы:

- 1) на основе качественного анализа выбор одной или нескольких кривых, форма которых соответствует характеру изменения временного ряда (п. 5.3);
- 2) оценка параметров выбранных кривых;
- 3) оценка точности и проверка адекватности выбранных кривых прогнозируемому процессу и окончательный выбор кривой роста;
- 4) расчет точечного (по формуле (5.38)) и интервального прогнозов.

Чтобы по имеющемуся временному ряду y_1, y_2, \dots, y_n осуществить прогноз на L шагов вперед, необходимо в построенную модель тенденции (кривую роста) $\hat{y} = f(t)$ подставить значение аргумента, соответствующее интервалу прогноза

$$\hat{y}_{n(+L)} = f(t_{n+L}). \tag{5.38}$$

Полученное значение $\hat{y}_{n(+L)}$ называется точечным прогнозом.

Следующим этапом является определение доверительного интервала прогноза, т. е. пределов, в которых лежит точное значение уровня явления с заданной вероятностью (степенью уверенности). Эта процедура называется вычисле-

нием интервального прогноза. Интервальный прогноз задает границы возможного изменения прогнозируемого показателя.

Несовпадение фактических данных с точечным прогнозом, полученным путем экстраполяции тенденции по кривым роста, может быть вызвано:

- 1) субъективной ошибочностью выбора вида кривой;
- 2) погрешностью оценивания параметров кривых;
- 3) погрешностью, связанной с отклонением отдельных наблюдений от тренда, характеризующего некоторый средний уровень ряда на каждый момент времени.

Погрешность, связанная со вторым и третьим источником, может быть отражена в виде доверительного интервала прогноза.

Доверительный интервал для линейной тенденции по аналогии с парной регрессией вычисляется по формуле

$$\hat{y}_{n(+L)} \pm t_{1-\alpha, n-2} \cdot s_{\hat{y}} \sqrt{1 + \frac{1}{n} + \frac{(t_{n+L} - \bar{t})^2}{\sum_{t=1}^n (t - \bar{t})^2}}, \quad (5.39)$$

где n – длина временного ряда; L – период упреждения; $\hat{y}_{n(+L)}$ – точечный прогноз на момент $n+L$; $t_{1-\alpha, n-2}$ – значение t -статистики Стьюдента при уровне значимости α и числе степеней свободы $n-2$; $s_{\hat{y}}$ – средняя квадратическая ошибка оценки прогнозируемого показателя

$$s_{\hat{y}} = \sqrt{\frac{1}{n-m} \sum_{t=1}^n (\hat{y}_t - y_t)^2}; \quad (5.40)$$

m – число параметров модели кривой роста (для линейной модели $m = 2$).

Для линейной модели формулу (5.39) можно записать следующим образом

$$\hat{y}_{n(+L)} \pm t_{1-\alpha, n-2} \cdot s_{\hat{y}} \sqrt{1 + \frac{1}{n} + \frac{3(n+2L-1)^2}{n(n-1)}}. \quad (5.41)$$

Доверительный интервал для кривой роста в виде полинома второго или третьего порядка вычисляется по формуле

$$\hat{y}_{n(+L)} \pm t_{1-\alpha, n-m} \cdot s_{\hat{y}} \sqrt{1 + \frac{1}{n} + \frac{t_{n+L}^2}{\sum_{t=1}^n t^2} + \frac{\sum_{t=1}^n t^4 - 2t_{n+L}^2 \sum_{t=1}^n t^2 - nt_{n+L}^4}{n \sum_{t=1}^n t^4 - \left(\sum_{t=1}^n t^2\right)^2}}, \quad (5.42)$$

где m – число параметров модели кривой роста. Для полинома второго порядка $m = 3$, для полинома третьего порядка $m = 4$.

Ширина доверительного интервала зависит от уровня значимости, периода упреждения, среднего квадратического отклонения временного ряда от тренда и степени полинома (рис. 5.4).

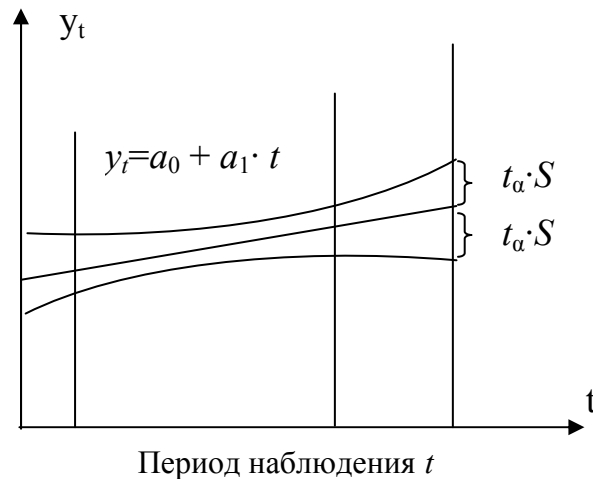


Рис. 5.4. Доверительные интервалы прогноза для линейного тренда

Чем выше степень полинома, тем шире доверительный интервал при одном и том же значении S_y , так как дисперсия уравнения тренда вычисляется как взвешенная сумма дисперсий соответствующих параметров уравнения.

Доверительные интервалы прогнозов, полученных с использованием уравнения экспоненты, определяют аналогичным образом. Отличие состоит в том, что как при вычислении параметров кривой, так и при вычислении средней квадратической ошибки используют не сами значения уровней временного ряда, а их логарифмы.

5.6. Адаптивные модели прогнозирования

5.6.1. Понятие адаптивных методов прогнозирования

При анализе временных рядов часто более важной бывает текущая тенденция (тенденция в данный момент времени, определяемая несколькими последними наблюдениями), а не тенденция, сложившаяся на длительном интервале времени. Соответственно, наиболее ценной является информация последнего периода. Исходя из этого в последнее время важное значение получили, так называемые, адаптивные методы прогнозирования.

Адаптивными называются методы прогнозирования, позволяющие строить самокорректирующиеся (самонастраивающиеся) экономико-математические модели, которые способны оперативно реагировать на изменение условий путем учета результата прогноза, сделанного на предыдущем шаге, и учета различной информационной ценности уровней ряда.

Особенностями адаптивных методов прогнозирования является:

- способность учитывать информационную ценность уровней временного ряда (с помощью системы весов, придаваемых этим уровням);
- использование рекуррентных процедур уточнения параметров модели по мере поступления новых данных наблюдений и тем самым адаптация модели применительно к новым условиям развития явления.

Скорость (быстроту) реакции модели на изменения в динамике процесса характеризует, так называемый, параметр адаптации. Параметр адаптации должен быть выбран таким образом, чтобы обеспечивалось адекватное отображе-

ние тенденции при одновременной фильтрации случайных отклонений. Значение параметра адаптации может быть определено на основе эмпирических данных, выведено аналитическим способом или получено на основе метода проб.

В качестве критерия оптимальности при выборе параметра адаптации обычно принимают минимум среднего квадрата ошибок прогнозирования.

Благодаря указанным свойствам адаптивные методы особенно удачно используются при краткосрочном прогнозировании (при прогнозировании на один или несколько шагов вперед).

Адаптивные методы, как правило, основаны на использовании процедуры экспоненциального сглаживания.

5.6.2. Экспоненциальное сглаживание

Для экспоненциального сглаживания временного ряда y_t используется рекуррентная формула

$$S_t = \alpha y_t + \beta S_{t-1}, \quad (5.43)$$

где S_t – значение экспоненциальной средней в момент t ; y_t – значение временного ряда в момент t ; α – параметр сглаживания, $\alpha = \text{const}$, $0 < \alpha < 1$; $\beta = 1 - \alpha$.

Совокупность значений S_t образует сглаженный временной ряд.

Соотношение (5.43) позволяет выразить экспоненциальную среднюю S_t через предшествующие значения уровней временного ряда y_t . При $n \rightarrow \infty$

$$S_t = \alpha \sum_{i=0}^n \beta^i y_{t-i}. \quad (5.44)$$

Таким образом, величина S_t оказывается взвешенной суммой всех членов ряда. Причем веса отдельных уровней ряда $\alpha \beta^i$ убывают по мере их удаления в прошлое соответственно экспоненциальной функции (в зависимости от «возраста» наблюдений).

Например, при $\alpha = 0,4$ вес текущего наблюдения y_t будет равен $\alpha = 0,4$, вес предыдущего уровня y_{t-1} будет соответствовать $\alpha \cdot \beta = 0,4 \cdot 0,6 = 0,24$; для уровня y_{t-2} вес составит $\alpha \cdot \beta^2 = 0,144$; для y_{t-3} – $\alpha \cdot \beta^3 = 0,0864$ и т. д.

Доказано, что математические ожидания исходного ряда и экспоненциальной средней совпадают. В то же время дисперсия экспоненциальной средней $D(S_t)$ меньше дисперсии временного ряда σ^2 . Чем меньше α , тем это отличие больше.

Таким образом, с одной стороны, желательно увеличивать вес более свежих наблюдений, что может быть достигнуто повышением α (согласно (5.43)), с другой стороны, для сглаживания случайных отклонений величину α нужно уменьшить. Выбор параметра сглаживания α с учетом этих двух противоречивых требований составляет задачу оптимизации модели.

В качестве начального значения S_0 используется среднее арифметическое значение из всех имеющихся уровней временного ряда или из какой-то их части. Из выражения (5.44) следует, что вес, приписываемый этому значению, уменьшается по экспоненциальной зависимости по мере удаления от первого уровня. Поэтому для длинных временных рядов влияние неудачного выбора S_0 погашается.

5.6.3. Использование экспоненциальной средней для краткосрочного прогнозирования

При использовании экспоненциальной средней для краткосрочного прогнозирования предполагается, что модель ряда имеет вид

$$y_t = \alpha_{1,t} + e_t, \quad (5.45)$$

где $\alpha_{1,t}$ – варьирующий во времени средний уровень ряда, e_t – случайные неавтокоррелированные отклонения с нулевым математическим ожиданием и дисперсией σ^2 .

Прогнозная модель определяется соотношением

$$\hat{y}_\tau(t) = \hat{\alpha}_{1,t}, \quad (5.46)$$

где $\hat{y}_\tau(t)$ – прогноз, сделанный в момент t на τ единиц времени (шагов) вперед; $\hat{\alpha}_{1,t}$ – оценка $\alpha_{1,t}$.

Величина параметра модели $\hat{\alpha}_{1,t}$ принимается равной экспоненциальной средней S_t в момент t :

$$\begin{aligned} \hat{a}_{1,t} &= S_t; \\ \hat{a}_{1,0} &= S_0. \end{aligned} \quad (5.47)$$

Прогнозирование предполагает следующую последовательность действий:

– на основании исходного временного ряда y_1, y_2, \dots, y_n вычисление по формуле (5.43) сглаженных уровней ряда S_1, S_2, \dots, S_n ;

– вычисление $\hat{\alpha}_{1,n} = S_n$;

– осуществление прогноза на τ шагов вперед $\hat{y}_\tau(n) = \hat{\alpha}_{1,n}$.

Перегруппировав члены выражение (5.43) можно записать по-другому:

$$S_t = S_{t-1} + \alpha(y_t - S_{t-1}).$$

Если величину $(y_t - S_{t-1})$ рассматривать как погрешность прогноза, то новый прогноз S_t получается как результат корректировки предыдущего прогноза с учетом его ошибки. В этом и состоит адаптация модели.

Экспоненциальное сглаживание является примером простейшей самообучающейся модели. Вычисления выполняются итеративно, причем вся прошлая информация заключена в единственном значении S_{t-1} .

5.6.4. Адаптивные полиномиальные модели

Если для прогнозирования временного ряда, имеющего ярко выраженную линейную тенденцию, использовать подход (5.46) опирающийся на модель экспоненциального сглаживания, то модель, как правило, будет давать смещенные прогнозы, т. е. иметь систематическую ошибку. Для таких временных рядов целесообразно использовать модели линейного роста, в которых процедуре экспоненциального сглаживания подвергаются оценки коэффициентов адаптивной модели.

В этих моделях прогноз может быть получен с помощью следующего выражения:

$$\hat{y}_\tau(t) = \hat{\alpha}_{1,t} + \hat{\alpha}_{2,t}\tau, \quad (5.48)$$

где $\hat{\alpha}_{1,t}$ и $\hat{\alpha}_{2,t}$ – текущие оценки коэффициентов; τ – время упреждения прогноза.

Наиболее часто применяются три модели данного типа, отличающиеся рекуррентными выражениями для пересчета текущих оценок коэффициентов (параметры адаптации или параметры экспоненциального сглаживания $0 < \alpha_1, \alpha_2, \alpha_3, \beta < 1$):

– *двухпараметрическая модель Ч. Хольта*

$$\begin{aligned}\hat{\alpha}_{1,t} &= \alpha_1 y_t + (1 - \alpha_1)(\hat{\alpha}_{1,t-1} + \hat{\alpha}_{2,t-1}), \\ \hat{\alpha}_{2,t} &= \alpha_2 (\hat{\alpha}_{1,t} - \hat{\alpha}_{1,t-1}) + (1 - \alpha_2) \hat{\alpha}_{2,t-1};\end{aligned}\tag{5.49}$$

– *однопараметрическая модель Р. Брауна*

$$\begin{aligned}\hat{\alpha}_{1,t} &= \hat{\alpha}_{1,t-1} + \hat{\alpha}_{2,t-1} + (1 - \beta^2) e_t, \\ \hat{\alpha}_{2,t} &= \hat{\alpha}_{2,t-1} + (1 - \beta^2) e_t;\end{aligned}\tag{5.50}$$

– *трехпараметрическая модель Дж. Бокса и Г. Дженкинса*

$$\begin{aligned}\hat{\alpha}_{1,t} &= \alpha_1 y_t + (1 - \alpha_1)(\hat{\alpha}_{1,t-1} + \hat{\alpha}_{2,t-1}) + \alpha_3 (e_t - e_{t-1}), \\ \hat{\alpha}_{2,t} &= \alpha_2 (\hat{\alpha}_{1,t} - \hat{\alpha}_{1,t-1}) + (1 - \alpha_2) \hat{\alpha}_{2,t-1}.\end{aligned}\tag{5.51}$$

Начальные значения коэффициентов $\hat{\alpha}_{1,t}$ и $\hat{\alpha}_{2,t}$ принимаются равными коэффициентам уравнения регрессии, построенного по начальным уровням ряда.

В эконометрических пакетах чаще представлена модель Ч. Хольта с возможностью выбора оптимальных параметров по критерию минимума среднеквадратической ошибки путем перебора на сетке возможных значений. Рекуррентные формулы для оценки коэффициентов по этой модели могут быть записаны в виде, явно показывающем зависимость «корректирующего воздействия» от величины ошибки:

$$\begin{aligned}\hat{\alpha}_{1,t} &= \hat{\alpha}_{1,t-1} + \hat{\alpha}_{2,t-1} + \alpha_1 e_t, \\ \hat{\alpha}_{2,t} &= \hat{\alpha}_{2,t-1} + \alpha_1 \alpha_2 e_t,\end{aligned}\tag{5.52}$$

где $e_t = y_t - \hat{y}_1(t-1)$ – ошибка прогноза.

Из последних выражений видно, что модель Р. Брауна можно считать частным случаем модели Ч. Хольта. При этом единственный параметр β играет роль коэффициента дисконтирования наблюдений.

5.7. Исследование взаимосвязи двух временных рядов

Модели, построенные на основе данных, характеризующих какой-либо объект за ряд последовательных моментов (периодов) времени, называются *моделями временных рядов*. Исследование взаимосвязи между переменными, заданными при помощи временных рядов имеет существенные особенности.

Наличие в составе временных рядов тенденций и периодических компонент может при применении обычных методов корреляционного или регрессионного анализа привести к явлениям «ложной корреляции» или «ложной регрессии». В этом случае абсолютная величина коэффициента корреляции между переменными x и y , абсолютно не влияющими друг на друга, имеет высокое значение вследствие зависимости каждой из них от времени, либо коэффициент

детерминации свидетельствует о высоком качестве полученной между ними регрессии. Чтобы избежать этого, перед изучением взаимосвязи между переменными x и y необходимо предварительно исключить из уровней временных рядов влияние тенденции и периодической компоненты.

Устранение периодической компоненты из уровней временного ряда можно проводить в соответствии с методикой параграфа 5.4.

Для исключения тенденции применяются такие методы, как метод последовательных разностей, метод отклонений от тренда, метод явного включения в модель регрессии по временным рядам фактора времени.

Метод отклонений от тренда. Рассмотрим два временных ряда x_t и y_t , каждый из которых содержит трендовую компоненту T и случайную компоненту ε . Предположим, что проведено аналитическое выравнивание этих рядов и найдены параметры соответствующих уравнений тенденций $\hat{x}_t = f_1(t)$ и $\hat{y}_t = f_2(t)$. Вычитание расчетных значений уровней ряда \hat{x}_t и \hat{y}_t из фактических x_t и y_t позволяет устранить влияние тенденции в обоих рядах. Дальнейший анализ взаимосвязи рядов проводят с использованием отклонений от тренда $(x_t - \hat{x}_t)$ и $(y_t - \hat{y}_t)$, т. е. уравнение регрессии строится в виде

$$y_t - \hat{y}_t = a + b \cdot (x_t - \hat{x}_t). \quad (5.53)$$

Метод последовательных разностей. Если временной ряд содержит ярко выраженную полиномиальную тенденцию (имеющую вид полинома от времени t), то с целью устранения тенденции можно применить метод последовательных разностей, заключающийся в замене исходных уровней ряда последовательными разностями соответствующих порядков (порядок разности равен порядку полинома).

Последовательными разностями первого порядка называются величины

$$\Delta y_t = y_t - y_{t-1}.$$

Последовательными разностями второго порядка называются величины

$$\Delta^2 y_t = \Delta y_t - \Delta y_{t-1}, \text{ и т. д.}$$

Замена исходных уровней ряда последовательными разностями первого порядка позволяет устранить линейную тенденцию, задаваемую уравнением $y = a + b \cdot t$.

Замена исходных уровней ряда последовательными разностями второго порядка позволяет устранить параболическую тенденцию, задаваемую уравнением в виде полинома второго порядка $y = a + b \cdot t + c \cdot t^2$, и т. д.

Если тенденция временного ряда характеризуется экспоненциальной зависимостью, то временной ряд из логарифмов исходных уровней будет иметь линейную тенденцию, что позволяет применить метод последовательных разностей к ряду логарифмов.

С использованием первых разностей Δy_t , Δx_t уравнение регрессии находится в виде

$$\Delta y_t = a + b \cdot \Delta x_t \quad \text{или} \quad y_t - y_{t-1} = a + b \cdot (x_t - x_{t-1}). \quad (5.54)$$

Включение в модель регрессии фактора времени. Включение фактора времени в модель в качестве независимой переменной позволяет зафиксировать тенденцию с целью исключения ее влияния на параметры модели.

Уравнение парной регрессии в этом случае принимает следующий вид

$$y_t = a + b_1 \cdot x_t + b_2 \cdot t + \varepsilon_t. \quad (5.55)$$

Этот же прием может быть использован, если число факторов больше единицы. Параметры a , b_1 , b_2 модели (5.55) с включением времени в качестве фактора определяются обычным МНК.

Параметры уравнения регрессии (5.55) могут быть проинтерпретированы следующим образом:

– параметр b_1 показывает, насколько в среднем изменится значение результативного признака y_t при увеличении фактора x_t на единицу при неизменной величине других факторов;

– параметр b_2 показывает, насколько в среднем за период наблюдения изменится значение результативного признака y_t за счет воздействия всех факторов, кроме фактора x_t .

5.8. Коинтеграция временных рядов

Не всегда наличие тенденции во временных рядах x_t и y_t приводит к недостоверности оценок параметров регрессии

$$y_t = a + b \cdot x_t + \varepsilon_t, \quad (5.56)$$

полученных с помощью обычного МНК, так как наличие тенденции во временном ряде y_t может являться следствием наличия тенденции во временном ряде x_t . Если нестационарные временные ряды x_t и y_t являются коинтегрируемыми, то оценки параметров регрессии (5.45) оказываются состоятельными.

Нестационарные временные ряды x_t и y_t называются *коинтегрируемыми*, если существует линейная комбинация этих рядов, представляющая собой стационарный временной ряд, т. е. существуют такие числа λ_1 и λ_2 , что временной ряд $\lambda_1 y_t + \lambda_2 x_t$ является стационарным.

Для тестирования временных рядов на коинтеграцию применяется критерий Энгеля-Грэнджера. Согласно этому критерию, исследуются остатки e_t уравнения регрессии (5.56), полученного обычным МНК, для которых рассчитываются параметры уравнения регрессии

$$\Delta e_t = a_e + b_e \cdot e_{t-1}, \quad (5.57)$$

Δe_t – где первые разности остатков. Фактическое значение t -статистики для параметра a_e сравнивается с критическим значением критерия τ . Если фактическое значение меньше критического, то нулевая гипотеза об отсутствии коинтеграции отклоняется. Критические значения критерия τ для уровней значимости 0,01; 0,05 и 0,1 составляют соответственно 2,5899; 1,9439 и 1,6177.

Таким образом, наличие коинтеграции нестационарных временных рядов позволяет при построении регрессионной модели использовать их исходные уровни x_t и y_t .

Через коинтеграцию, к примеру, подтверждаются зависимости между уровнем инфляции, ВВП и денежной массой, ценами на акции и их доходностью, потреблением и уровнем дохода и многие другие экономические зависимости с шумящими переменными.

Но следует отметить, что такой подход применим только к временным рядам, охватывающим достаточно длительные промежутки времени.

Контрольные вопросы

1. Что называют временным рядом?
2. Какие компоненты выделяют в составе экономического временного ряда?
3. В чем заключается основная задача эконометрического исследования временного ряда?
4. Охарактеризуйте понятие автокорреляции уровней временного ряда.
5. Какие методы применяются для проверки наличия тенденции временного ряда?
6. Как осуществляется сглаживание временного ряда по методу скользящей средней?
7. Что понимается под аналитическим выравниванием временного ряда?
8. Какие методы применяются для определения вида тенденции временного ряда?
9. Как осуществляется выбор вида тенденции на основе качественного анализа?
10. Как осуществляется оценка адекватности модели тенденции временного ряда?
11. Как осуществляется оценка точности модели тенденции временного ряда?
12. Для чего применяется критерий Дарбина–Уотсона?
13. Как осуществляется выделение периодической компоненты по методу скользящей средней?
14. Как осуществляется моделирование сезонных колебаний с помощью фиктивных переменных?
15. Как осуществляется прогнозирование уровней временного ряда на основе кривых роста?
16. Что понимается под точечным и интервальным прогнозом?
17. В чем заключаются особенности адаптивных методов прогнозирования?
18. В чем состоит процедура экспоненциального сглаживания временного ряда?
19. Какие сложности возникают при изучении взаимосвязи двух временных рядов?
20. Какие методы применяются для исключения тенденции из временного ряда?
21. Что понимается под коинтеграцией временных рядов?
22. Как проверяется наличие коинтеграции временных рядов?

6. Линейные модели стохастических процессов

6.1. Стационарные стохастические процессы

6.1.1. Основные понятия

Уровни временного ряда x_1, x_2, \dots, x_n при наличии случайной составляющей могут рассматриваться как конкретные значения случайных величин X_1, X_2, \dots, X_n , соответствующих моментам времени t_1, t_2, \dots, t_n , т. е. как отдельная реализация дискретного стохастического процесса.

Стохастическим процессом называется случайная функция $X(t)$ вещественного аргумента t , принадлежащего некоторому подмножеству \check{T} множества действительных чисел. Иными словами, если каждому значению аргумента $t \in \check{T} \subseteq R$ поставлена в соответствие случайная величина $X_t = X(t)$, то совокупность случайных величин $\{X_t\}$ представляет собой стохастический процесс.

Если множество определения \check{T} случайной функции $X(t)$ дискретно, т. е. $\check{T} = \{t_i\}$, то стохастический процесс называется *дискретным*.

Дискретный стохастический процесс представляет собой последовательность случайных величин X_t , соответствующих моментам времени $t_1, t_2, \dots, t_T, \dots$.

Характеристики случайного процесса $X(t)$ в общем случае являются функциями от времени t :

математическое ожидание

$$\mu_t = E[X_t] = \mu(t); \quad (6.1)$$

дисперсия

$$\sigma_t^2 = D[X_t] = E[(X_t - \mu_t)^2] = \sigma^2(t), \quad (6.2)$$

а автоковариация

$$\gamma_{t_1 t_2} = \text{cov}(X_{t_1}, X_{t_2}) = E[(X_{t_1} - \mu_{t_1})(X_{t_2} - \mu_{t_2})] = \gamma(t_1, t_2) \quad (6.3)$$

зависит от t_1 и t_2 .

Стохастический процесс называется *стационарным процессом в узком (сильном) смысле*, если совместное распределение вероятностей случайных величин $X_{t_1}, X_{t_2}, \dots, X_{t_n}$ такое же, как и у случайных величин $X_{t_1+\tau}, X_{t_2+\tau}, \dots, X_{t_n+\tau}$ при любых n, t и τ .

Стохастический процесс называется *стационарным процессом в широком (слабом) смысле*, если математическое ожидание μ_t и дисперсия σ_t^2 не зависят от времени (одинаковы для всех X_t), а автоковариация $\gamma_{t_1 t_2}$ зависит только от величины лага $\tau = t_2 - t_1$, т. е.

$$\begin{aligned} \mu_t &= \mu = \text{const}; \\ \sigma_t^2 &= \sigma^2 = \text{const}; \\ \gamma_{t_1 t_2} &= \text{cov}(X_{t_1}, X_{t_2}) = E[(X_{t_1} - \mu)(X_{t_1+\tau} - \mu)] = \gamma(\tau). \end{aligned} \quad (6.4)$$

Процесс называется *нормальным*, если совместное распределение случайных величин X, X_2, \dots, X_t является n -мерным нормальным распределением.

«Белым шумом» называется последовательность независимых, одинаково распределенных случайных величин a_t . Из определения «белого шума» следует, что

$$\mu_t = \text{const} = \mu; \quad D_t = \sigma_t^2 = \text{const} = \sigma^2; \quad \gamma_{t_1 t_2} = 0, \quad \text{если } t_1 \neq t_2. \quad (6.5)$$

«Белый шум» является стационарным стохастическим процессом и играет важную роль при моделировании остатков стохастического процесса в уравнениях регрессии.

Зависимость автоковариации $\gamma_\tau = \gamma(\tau)$ от длины лага τ называется *автоковариационной функцией*. При $\tau = 0$ ее значение равно дисперсии, т. е. $\gamma_0 = \gamma(0) = \sigma^2$.

Отношение автоковариации $\gamma_\tau = \gamma(\tau)$ к дисперсии $\sigma^2 = \gamma_0$ называется *автокорреляционной функцией* стационарного стохастического процесса:

$$\rho_\tau = \frac{\gamma_\tau}{\gamma_0}, \quad (6.6)$$

причем $-1 \leq \rho_\tau \leq 1$.

Стационарному стохастическому процессу X_t соответствует стационарный временной ряд x_1, x_2, \dots, x_n .

Признаками стационарности временного ряда являются отсутствие тенденции и периодической составляющей, а также систематических изменений размаха колебаний и систематически изменяющихся взаимозависимостей между элементами временного ряда.

Для распознавания стационарности временных рядов могут использоваться следующие подходы:

- визуальный анализ графического представления временного ряда на наличие тенденции и периодической составляющей, на постоянство дисперсии и т. п.;
- анализ временного ряда на наличие автокорреляции;
- тесты на присутствие детерминистического тренда;
- тесты на постоянство статистических характеристик;
- тесты на наличие стохастического тренда, например, тесты на единичный корень.

6.1.2. Параметрические тесты стационарности

Параметрические тесты применяются при относительно строгих предположениях относительно законов распределения временного ряда, его параметров. Они, как правило, оценивают меру близости между эмпирическими характеристиками распределения временного ряда и их теоретическими аналогами, на основании чего делается вывод о целесообразности принятия или отвержения гипотезы о соответствии свойств рассматриваемого ряда стационарному процессу.

Для проверки гипотез о постоянстве математического ожидания и дисперсии на рассматриваемом интервале $t=1, 2, \dots, n$ могут быть использованы критерии Стьюдента и Фишера. Эти критерии применяются в предположении о нормальном законе распределения как значений временного ряда, так и его выборочных параметров, что является достаточно справедливым для многих реальных процессов.

Тестирование математического ожидания. Интервал времени $(1, n)$ (и, соответственно, временной ряд $y_t, t=1, 2, \dots, n$) разбивается на две части, не обязательно одинаковые по количеству содержащихся в них значений y_t с количеством наблюдений n_1 ($n=1, 2, \dots, n_1$) и n_2 ($n=n_1+1, \dots, n$), $n_2 = n - n_1$.

Для каждой из частей определяются оценки \bar{y}_1 и s_1^2 , \bar{y}_2 и s_2^2 – выборочных математического ожидания и дисперсии переменной y_t соответственно. Далее рассчитывается значение критерия Стьюдента по формуле

$$\tau = \frac{|\bar{y}_1 - \bar{y}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}, \quad (6.7)$$

если предполагается, что значения дисперсий на этих участках не равны между собой, т. е. $\sigma_1^2 \neq \sigma_2^2$, и по формуле

$$\tau = \frac{|\bar{y}_1 - \bar{y}_2|}{s^2} \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}, \quad (6.8)$$

если $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

Если оказывается справедливым неравенство

$$\tau < \tau_{кр}(\alpha, k), \quad (6.9)$$

где α – заданный уровень значимости ($\alpha=0,05; 0,01$); $k = n_1+n_2-2$ – число степеней свободы; $\tau_{кр}(\alpha, \nu)$ – критическое значение критерия Стьюдента, соответствующее значениям α и k , то нулевая гипотеза о постоянстве математического ожидания процесса y_t принимается. Вероятность ошибки такого решения при этом составляет α . В противном случае, эта гипотеза отвергается.

Можно интервал наблюдений разделить на несколько частей и проверять гипотезу о равенстве оценок средних значений ряда, рассчитанных на этих частях. Для этих целей используется критерий Фишера. Его расчетное значение в данном тесте определяется как отношение взвешенной суммы квадратов отклонений этих оценок от средней временного ряда в целом к средней дисперсии временного ряда:

$$F = \frac{\frac{1}{k-1} \cdot \sum_{j=1}^k n_j (\bar{y}_j - \bar{y})^2}{\bar{s}^2(k)}, \quad (6.10)$$

где k – число частей разбиения интервала $(1, n)$; n_j – число измерений переменной на j -й части ($j=1, 2, \dots, k$); \bar{y} – среднее значение временного ряда в целом; $\bar{s}^2(k)$ – средняя дисперсия, значение которой рассчитывается на основании следующей формулы

$$\bar{s}^2(k) = \frac{1}{t-k} \cdot \sum_{j=1}^k (n_j - 1) \cdot \bar{s}_j^2,$$

где \bar{s}_j^2 – дисперсия, рассчитанная на j -й части интервала $(1, n)$.

Если оказывается справедливым соотношение

$$F < F(\alpha, \nu_1, \nu_2), \quad (6.11)$$

где $F(\alpha, \nu_1, \nu_2)$ – табличное значение критерия Фишера для уровня значимости α и при числе степеней свободы $\nu_1 = k-1$, $\nu_2 = n_1+n_2+\dots+n_k-k$, то гипотеза о постоянстве математического ожидания временного ряда на всем интервале $(1, n)$ принимается с вероятностью $1-\alpha$. В противном случае она отвергается.

Тестирование дисперсии. Проверка гипотезы о постоянстве дисперсии временного ряда y_t , $t = 1, 2, \dots, n$ в случае разбиения исходного интервала на две части обычно осуществляется с использованием двухстороннего критерия Фишера. Обязательным условием при этом также является нормальный закон распределения значений y_n .

Расчетное значение критерия Фишера определяется по следующей формуле:

$$F = \frac{s_1^2}{s_2^2}, \quad (6.12)$$

где s_1^2 и s_2^2 – оценки дисперсии ряда на первой и второй частях соответственно с числом измерений n_1 и n_2 .

Если для заданного уровня доверительной значимости α оказывается, что значение F удовлетворяет неравенству

$$F(\alpha/2, v_1, v_2) \leq F \leq F(1-\alpha/2, v_1, v_2), \quad (6.13)$$

то гипотеза о постоянстве дисперсии временного ряда может быть принята, т. е. предположение о том, что $s_1^2 = s_2^2 = \sigma^2$ является обоснованным с вероятностью $1-\alpha$.

В выражении (6.13) значения $F(\alpha/2, v_1, v_2)$ и $F(1-\alpha/2, v_1, v_2)$ являются табличными (левосторонним и правосторонним) значениями критерия Фишера, соответствующими вероятности ошибки $\alpha/2$ с числом степеней свободы $v_1 = n_1-1$ и $v_2 = n_2-1$. Заметим, что эти значения удовлетворяют соотношению

$$F(\alpha/2, v_1, v_2) = 1 / F(1-\alpha/2, v_1, v_2). \quad (6.14)$$

Поэтому на практике обычно проверяют только соотношение

$$F \leq F(1-\alpha/2, v_1, v_2), \quad (6.15)$$

при условии, что $s_1 > s_2$.

6.1.3. Непараметрические тесты стационарности

Непараметрические тесты не выдвигают заранее каких-либо предположений о законе распределения тестируемого временного ряда, его параметрах. Они исследуют взаимосвязи между порядком следования образующих его значений, выявляют наличие или отсутствие закономерностей в продолжительности и (или) чередовании их серий, образованных, например, последовательностями единиц совокупности с одинаковыми знаками, сменой знаков у этих единиц и т. п. [10].

Тест Манна-Уитни применяется для тестирования постоянства математического ожидания. Рассмотрим две совокупности значений одного и того же временного ряда $y_1^1, y_2^1, \dots, y_{n_1}^1$ и $y_1^2, y_2^2, \dots, y_{n_2}^2$ и объединим их в один ряд длиной ($n = n_1 + n_2$) в порядке возрастания. Обозначим через u^* число элементов первой совокупности (y^1), предшествующих первому элементу из второй совокупности, плюс число элементов y^1 первой совокупности, предшествующих второму элементу из второй совокупности, включая уже учтенные, плюс и т. д. до последнего элемента второй совокупности.

Величину u^* можно рассчитать через число сумму рангов элементов первой или второй совокупности R_1 и R_2 , определяемых по общей совокупности:

$$u^* = R_1 - \frac{n_1(n_1 + 1)}{2}, \quad (6.16)$$

$$u^* = n_1 \cdot n_2 - \frac{n_2(n_2 + 1)}{2} - R_2. \quad (6.17)$$

Величина z

$$z = \frac{u^* - n_1 \cdot n_2 \pm \frac{1}{2}}{\sqrt{\frac{n_1 \cdot n_2 (n_1 + n_2 + 1)}{12}}}, \quad (6.18)$$

имеет стандартизированное нормальное распределение. В формуле (6.18) поправка $1/2$ прибавляется, если $z < 0$, и отнимается, если $z > 0$. Гипотеза о стационарности процесса отвергается при уровне значимости α , если $|z| > t_{1-\alpha/2}$, где $t_{1-\alpha/2}$ – квантиль стандартизированного нормального распределения порядка $1-\alpha/2$.

Тест Сиджела-Тьюки применяется для тестирования постоянства дисперсии временного ряда и основан на сопоставлении рангов элементов двух совокупностей из рассматриваемого интервала. Согласно тесту Сиджела-Тьюки исходный временной ряд y_1, y_2, \dots, y_n центрируется относительно среднего значения ряда \bar{y} (т. е. определяются значения $\tilde{y}_t = y_t - \bar{y}$) и разделяется на две части (желательно равные) $y_1^1, y_2^1, \dots, y_{n_1}^1$ и $y_1^2, y_2^2, \dots, y_{n_2}^2$, которые затем объединяются в один ряд длиной ($n = n_1 + n_2$) в порядке возрастания.

Элементы полученного ряда ранжируются по следующему правилу. Ранг 1 приписывается наименьшему отрицательному значению. Ранг 2 приписывается наибольшему положительному значению. Ранг 3 приписывается наименьшему значению из еще неранжированных (значению, следующему за наименьшим). Ранг 4 приписывается наибольшему значению из еще неранжированных (значению, следующему за наибольшим), и т. д.

Величина z

$$z = \frac{R_1 - \frac{n_1 \cdot (n_1 + n_2 + 1)}{2} \pm \frac{1}{2}}{\sqrt{\frac{n_1 \cdot n_2 (n_1 + n_2 + 1)}{12}}} \quad (6.19)$$

имеет стандартизированное нормальное распределение. В формуле (6.19) поправка $1/2$ прибавляется, если $z < 0$, и отнимается, если $z > 0$. Гипотеза о стационарности процесса отвергается при уровне значимости α , если $|z| > t_{1-\alpha/2}$, где $t_{1-\alpha/2}$ – квантиль стандартизированного нормального распределения порядка $1-\alpha/2$.

6.2. Линейные модели стационарных временных рядов. Процессы ARMA

6.2.1. Модели авторегрессии (AR)

Авторегрессионным процессом порядка p (обозначается $AR(p)$) называется стохастический процесс X_t , определяемый соотношением

$$X_t = \alpha_0 + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \dots + \alpha_p X_{t-p} + \varepsilon_t, \quad (6.20)$$

где ε_t – процесс типа «белый шум» с $\mu_\varepsilon = 0$. Свободный член α_0 часто приравнивается нулю (т. е. рассматриваются центрированные процессы, средний уровень которых равен нулю).

Авторегрессионная модель временного ряда основана на предположении, что поведение какого-либо экономического явления в будущем определяется только его текущим и предыдущими состояниями.

AR-процесс является *стационарным* тогда и только тогда, когда комплексные решения (корни) его характеристического уравнения

$$1 - \alpha_1 z - \alpha_2 z^2 - \dots - \alpha_p z^p = 0 \quad (6.21)$$

лежат вне единичного круга, т. е. $|z| > 1$ (z — комплексное число).

Процессы, у которых $|z| = 1$, называются *процессами единичного корня* и являются *нестационарными*.

Для процесса AR(1)

$$X_t = \alpha_0 + \alpha_1 X_{t-1} + \varepsilon_t$$

характеристическое уравнение имеет вид

$$1 - \alpha_1 z = 0.$$

Неравенство $|z| > 1$ выполняется, если $|\alpha_1| < 1$. Следовательно, соотношение $|\alpha_1| < 1$ есть условие стационарности процесса AR(1).

Коэффициенты α_i уравнения (6.20) могут быть выражены через коэффициенты автокорреляции r_i . Умножим уравнение (6.20) последовательно на X_{t-k} ($k = 1, \dots, p$) и применим к его правой и левой частям операцию вычисления математического ожидания. В результате получим систему соотношений

$$\begin{aligned} r_1 &= \alpha_1 + \alpha_2 r_1 + \alpha_3 r_2 \dots + \alpha_p r_{p-1}, \\ r_2 &= \alpha_1 r_1 + \alpha_2 + \alpha_3 r_1 \dots + \alpha_p r_{p-2}, \\ &\dots\dots\dots \\ r_p &= \alpha_1 r_{p-1} + \alpha_2 r_{p-2} + \alpha_3 r_{p-3} \dots + \alpha_p, \end{aligned} \quad (6.22)$$

называемых *уравнениями Юла-Уокера*.

В частности, для $p = 1$ имеет место соотношение $\alpha_1 = r_1$.

6.2.2. Модели скользящего среднего (MA)

В моделях скользящего среднего порядка среднее текущее значение стационарного стохастического процесса представляется в виде линейной комбинации текущего и прошедших значений ошибки $\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-p}$, обладающей свойствами «белого шума».

Процессом скользящего среднего порядка q (обозначается MA(q)) называется стохастический процесс X_t , определяемый соотношением

$$X_t = \varepsilon_t - \beta_1 \varepsilon_{t-1} - \beta_2 \varepsilon_{t-2} - \dots - \beta_q \varepsilon_{t-q}, \quad (6.23)$$

где ε_t – процесс типа «белый шум» с $\mu_\varepsilon = 0, \sigma_\varepsilon^2 = \sigma^2$.

Процесс MA(q) обладает следующими свойствами:

$$E[X_t] = 0; \quad D[X_t] = \sigma^2 \sum_{i=0}^q \beta_i^2;$$

$$\gamma_{t,t+\tau} = \begin{cases} 0, & \tau > q; \\ \sigma^2 \sum_{i=0}^{q-\tau} \beta_i \beta_{i+\tau}, & \tau = 0, 1, \dots, q. \end{cases} \quad (6.24)$$

Согласно (6.24) среднее значение, дисперсия и ковариация не зависят от времени, поэтому процесс МА стационарен в широком смысле.

6.2.3. Модели авторегрессии-скользящего среднего (ARMA)

Комбинация процессов авторегрессии и скользящего среднего порядков p и q соответственно называется **авторегрессионным процессом скользящего среднего (ARMA(p, q))**

$$X_t = \alpha_0 + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \dots + \alpha_p X_{t-p} + \varepsilon_t - \beta_1 \varepsilon_{t-1} - \beta_2 \varepsilon_{t-2} - \dots - \beta_q \varepsilon_{t-q}, \quad (6.25)$$

При очень общих условиях стационарный ARMA-процесс может быть представлен как бесконечный AR-процесс или как бесконечный МА-процесс:

$$X_t = \alpha_0 + \varepsilon_t - \beta_1 \varepsilon_{t-1} - \beta_2 \varepsilon_{t-2} - \dots$$

Использование ARMA-процессов позволяет строить более компактные модели реальных временных рядов по сравнению со схожими по поведению AR- или МА-процессами.

6.3. Автокорреляционные функции

6.3.1. Автокорреляционная функция

Автокорреляционная функция (ACF) процесса X_t , определяющая зависимость коэффициентов автокорреляции ρ_τ от величины лага τ , определяется с помощью соотношения (см. (6.3))

$$\rho_\tau = \rho(\tau) = \frac{\gamma_\tau}{\gamma_0} = \frac{1}{\gamma_0} E[(X_t - \mu)(X_{t+\tau} - \mu)]. \quad (6.26)$$

График ρ_τ называется *коррелограммой*.

Для идентификации модели стационарного временного ряда, т. е. для определения типа и порядка процесса могут быть использованы следующие свойства автокорреляционной функции:

а) Для процесса AR(p) коррелограмма представляет собой смесь экспоненциальной кривой и синусоиды.

б) Для процесса МА(q) только первые q автокорреляционных коэффициентов значимо отличны от нуля.

В качестве примера рассмотрим автокорреляционные функции процессов AR(1) и МА(1).

Для процесса AR(1) без свободного члена и с $\alpha_1 < 1$

$$X_t = \alpha_1 X_{t-1} + \varepsilon_t \quad (6.27)$$

автокорреляционная функция определяется соотношениями $\rho_1 = \alpha_1$, и $\rho_k = \alpha_1^k$ (рис. 6.1, а, б).

Для процесса МА(1)

$$X_t = \varepsilon_t - \beta_1 \cdot \varepsilon_{t-1} \quad (6.28)$$

автокорреляционная функция определяется соотношениями $\rho_1 = \frac{-\beta_1}{(1+\beta_1^2)}$, $\rho_2 = 0$, $\rho_3 = 0, \dots$ (рис. 6.2, а, б).

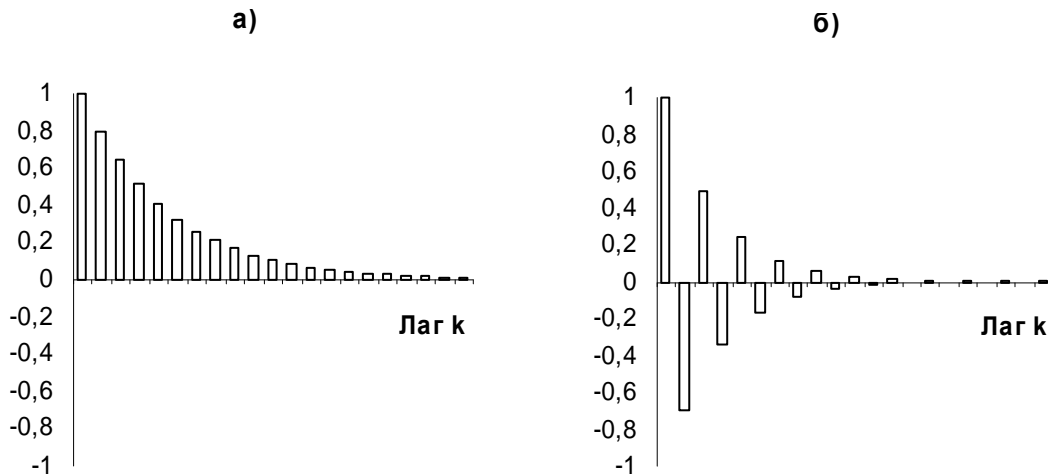


Рис. 6.1 Кореллограмма процесса AR(1)

а) $\alpha_1 > 0$; б) $\alpha_1 < 0$

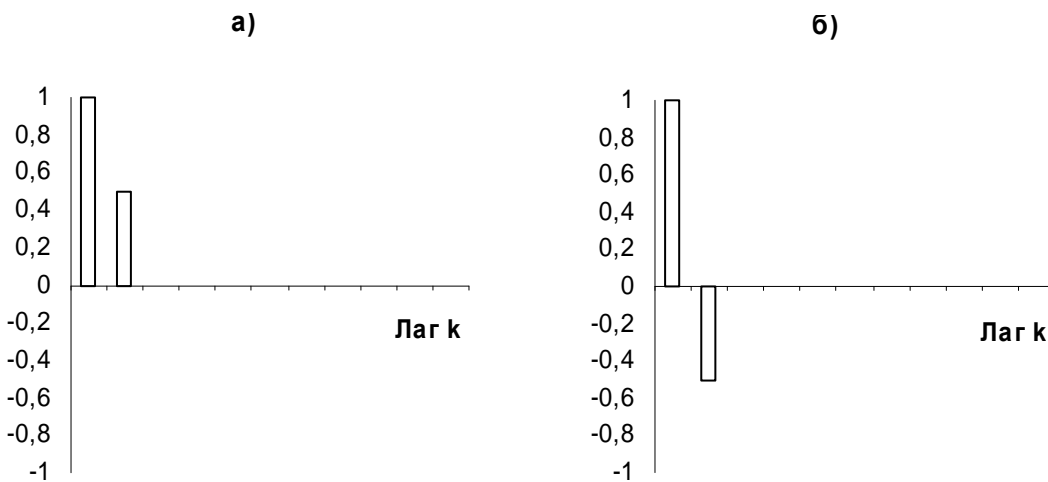


Рис. 6.2 Кореллограмма процесса МА(1)

а) $\beta_1 < 0$; б) $\beta_1 > 0$

6.3.2. Частная автокорреляционная функция

Важную информацию о структуре модели стационарного стохастического процесса можно получить, используя частную автокорреляционную функцию.

Рассмотрим аппроксимацию $AR(k)$ стационарного стохастического процесса X_t

$$X_t^{(k)} = \alpha_{0k} + \alpha_{1k}X_{t-1}^{(k)} + \alpha_{2k}X_{t-2}^{(k)} + \dots + \alpha_{kk}X_{t-k}^{(k)} \quad (6.29)$$

Коэффициент α_{kk} называется *коэффициентом частной автокорреляции* X_t для величины лага k .

Ряд $\rho_{\text{part}}(k) = \alpha_{kk}$ с различными k называется *частной автокорреляционной функцией* (PACF).

Для процесса $AR(p)$ значения частной автокорреляционной функции $\rho_{\text{part}}(\tau)$ равны нулю для величины лага $\tau > p$.

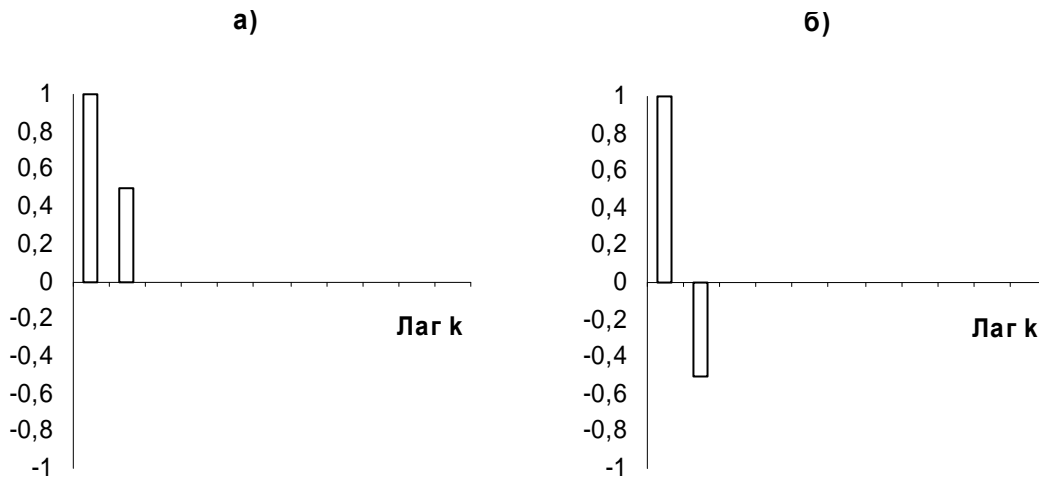


Рис. 6.3 Частная автокорреляционная функция процесса $AR(1)$
а) $\alpha_1 > 0$; б) $\alpha_1 < 0$

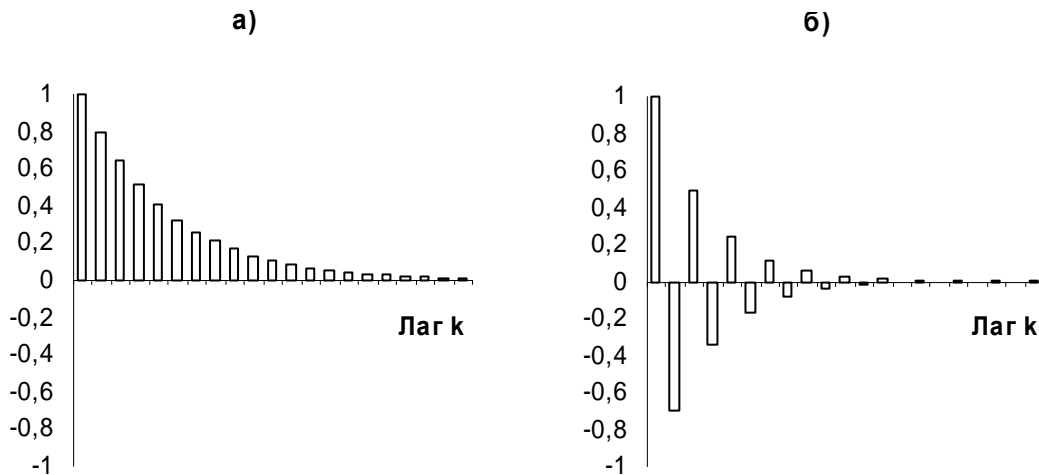


Рис. 6.4 Частная автокорреляционная функция процесса $MA(1)$
а) $\beta_1 < 0$; б) $\beta_1 > 0$

Для процессов $MA(q)$ значения частной автокорреляционной функции экспоненциально убывают с величиной лага q .

В качестве значения частной автокорреляционной функции $\rho_{\text{part}}(k)$ при заданной величине лага k может быть использована оценка коэффициента $\hat{\alpha}_{kk}$ модели $AR(k)$ (6.29), полученная с помощью МНК-оценивания.

6.4. Прогнозирование ARMA-процессов

6.4.1. AR-процессы

Рассмотрим стационарную AR-модель

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots + \alpha_p Y_{t-p} + \varepsilon_t. \quad (6.30)$$

Предположим, что прогноз $\hat{Y}_T(h)$ строится на h шагов вперед, начиная с момента времени T . Запишем уравнение (6.30) для момента времени $T+h$

$$Y_{T+h} = \alpha_0 + \alpha_1 Y_{T+h-1} + \alpha_2 Y_{T+h-2} + \dots + \alpha_p Y_{T+h-p} + \varepsilon_{T+h}. \quad (6.31)$$

При расчете прогнозного значения $\hat{Y}_T(h)$ в правую часть (6.31) вместо Y_{T+i} ($i > 0$) следует подставлять вычисленное ранее прогнозное значение $\hat{Y}_T(i)$. Тогда точечный прогноз будет определяться соотношениями:

$$\begin{aligned} \hat{Y}_T(1) &= \alpha_0 + \alpha_1 Y_T + \alpha_2 Y_{T-1} + \dots + \alpha_p Y_{T-p+1}, \\ \hat{Y}_T(2) &= \alpha_0 + \alpha_1 \hat{Y}_T(1) + \alpha_2 Y_T + \dots + \alpha_p Y_{T-p+2}, \\ &\dots\dots\dots \\ \hat{Y}_T(p) &= \alpha_0 + \alpha_1 \hat{Y}_T(p-1) + \alpha_2 \hat{Y}_T(p-2) + \dots + \alpha_{p-1} \hat{Y}_T(1) + \alpha_p Y_T, \\ \hat{Y}_T(h) &= \alpha_0 + \alpha_1 \hat{Y}_T(h-1) + \alpha_2 \hat{Y}_T(h-2) + \dots + \alpha_p \hat{Y}_T(h-p+1) \quad \text{при } h > p. \end{aligned} \quad (6.32)$$

Доказано, что в бесконечном периоде математическое ожидание прогнозного значения \hat{Y}_T асимптотически сходится к математическому ожиданию процесса Y_t , т. е. условное математическое ожидание ошибки прогноза равно нулю и оценка $\hat{Y}_T(h)$ является несмещенной, а дисперсия прогноза сходится к дисперсии процесса Y_t , т. е. к σ_Y^2 .

Для модели AR(2)

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \varepsilon_t$$

формулы прогнозирования имеют вид:

$$\begin{aligned} \hat{Y}_T(1) &= \alpha_0 + \alpha_1 Y_T + \alpha_2 Y_{T-1}, \\ \hat{Y}_T(2) &= \alpha_0 + \alpha_1 \hat{Y}_T(1) + \alpha_2 Y_T, \\ \hat{Y}_T(h) &= \alpha_0 + \alpha_1 \hat{Y}_T(h-1) + \alpha_2 \hat{Y}_T(h-2) \quad \text{при } h \geq 3. \end{aligned} \quad (6.33)$$

6.4.2. МА-процессы

Рассмотрим теперь стационарную МА-модель

$$Y_t = \varepsilon_t - \beta_1 \varepsilon_{t-1} - \beta_2 \varepsilon_{t-2} - \dots - \beta_q \varepsilon_{t-q}. \quad (6.34)$$

С учетом того, что величина ε_t для прогнозируемых моментов времени не известна точечный прогноз согласно модели (6.34) будет определяться соотношениями:

$$\begin{aligned} \hat{Y}_T(1) &= -\beta_1 \varepsilon_T - \beta_2 \varepsilon_{T-1} - \dots - \beta_q \varepsilon_{T-q+1}, \\ \hat{Y}_T(2) &= -\beta_2 \varepsilon_T - \dots - \beta_q \varepsilon_{T-q+2}, \\ &\dots\dots\dots \\ \hat{Y}_T(q) &= -\beta_q \varepsilon_T, \\ \hat{Y}_T(h) &= 0 \quad \text{при } h > q. \end{aligned} \quad (6.35)$$

Дисперсия ошибки прогноза определяется соотношениями

$$\text{var}(e_T(1)) = \sigma_\varepsilon^2;$$

$$\begin{aligned} \text{var}(e_T(2)) &= \sigma_\varepsilon^2 (1 + \beta_1^2); \\ &\dots\dots\dots \\ \text{var}(e_T(q-1)) &= \sigma_\varepsilon^2 (1 + \beta_1^2 + \dots + \beta_{q-1}^2); \\ \text{var}(e_T(q)) &= \sigma_\varepsilon^2 (1 + \beta_1^2 + \dots + \beta_q^2) = \sigma_Y^2 \quad \text{для } h > q. \end{aligned} \quad (6.36)$$

Для процесса MA(2)

$$Y_t = \varepsilon_t - \beta_1 \varepsilon_{t-1} - \beta_2 \varepsilon_{t-2}$$

формулы для прогнозирования имеют вид

$$\begin{aligned} \hat{Y}_T(1) &= -\beta_1 \cdot \varepsilon_T - \beta_2 \cdot \varepsilon_{T-1} \\ \hat{Y}_T(2) &= -\beta_2 \cdot \varepsilon_T \\ \hat{Y}_T(h) &= 0 \quad \text{при } h \geq 3, \end{aligned} \quad (6.37)$$

а дисперсии ошибки прогноза:

$$\begin{aligned} \text{var}(e_T(1)) &= \sigma_\varepsilon^2; \quad \text{var}(e_T(2)) = \sigma_\varepsilon^2 (1 + \beta_1^2); \\ \text{var}(e_T(h)) &= \sigma_\varepsilon^2 (1 + \beta_1^2 + \beta_2^2) = \sigma_Y^2 \quad \text{для } h \geq 3. \end{aligned}$$

6.4.3. ARMA-процессы

Формулы прогнозирования для процессов ARMA(p, q) получаются объединением формул (6.32) и (6.35).

Для модели ARMA (1,1)

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} - \beta_1 \cdot \varepsilon_{t-1}$$

формулы для прогнозирования имеют вид:

$$\begin{aligned} \hat{Y}_{T(+1)} &= \alpha_0 + \alpha_1 Y_T - \beta_1 \cdot \varepsilon_T \\ \hat{Y}_{T(+h)} &= \alpha_0 + \alpha_1 \hat{Y}_{T(+h-1)} \quad \text{при } h \geq 2. \end{aligned} \quad (6.38)$$

При прогнозировании на практике реальные параметры ARMA-процесса α_k и β_j заменяются их оценками $\hat{\alpha}_k$ и $\hat{\beta}_j$, а случайные воздействия ε_t заменяются на остатки $\hat{\varepsilon}_t$, полученные при оценивании модели, или на ошибки e_{T+h-i} предыдущих прогнозов.

Отметим, что ошибка прогноза данных ARMA-моделей ограничена на бесконечности дисперсией процесса σ_x .

6.5. Нестационарные интегрируемые процессы

6.5.1. Нестационарные стохастические процессы. Нестационарные временные ряды

Признаком нестационарного стохастического процесса является нарушение одного из условий стационарности (6.4). Конкретная реализация нестационарного стохастического процесса представляет собой нестационарный временной ряд. Признаками нестационарности временного ряда могут служить наличие тенденции, систематических изменений дисперсии, периодической составляющей, систематически изменяющихся взаимозависимостей между элементами временного ряда.

Заметим, что, как правило, значения, характеризующие изменение экономических показателей во времени, образуют нестационарные временные ряды.

Рассмотрим авторегрессионный процесс первого порядка, определяемый моделью

$$Y_t = \alpha_0 + \alpha_1 \cdot Y_{t-1} + \varepsilon_t, \quad (6.39)$$

где ε_t – процесс типа «белый шум» с $\mu_\varepsilon = 0$. При $|\alpha_1| < 1$ случайный процесс Y_t будет стационарным. Процесс, определяемый соотношением (6.39) при $\alpha_1 = 1$

$$Y_t = Y_{t-1} + \varepsilon_t. \quad (6.40)$$

является нестационарным и называется «случайным блужданием». Такие нестационарные процессы называют *процессами единичного корня*.

Среднее процесса Y_t постоянно $E(Y_t) = E(Y_{t-1}) + E(\varepsilon_t) = \mu = \text{const}$, а дисперсия $\text{var}(Y_t) = t\sigma^2$ неограниченно возрастает с течением времени. Первые разности Y_t являются «белым шумом» ε_t и стационарны:

$$\Delta Y_t = Y_t - Y_{t-1} = \varepsilon_t.$$

Как показывает практика, рассматриваемые в эконометрических исследованиях нестационарные временные ряды чаще всего относятся именно к этому типу и проблема выявления нестационарности временного ряда сводится к проверке $\alpha_1 = 1$ в модели (6.39). Соответствующие тесты называются «тестами единичного корня».

6.5.2. Тесты Дики-Фуллера

Тест Дики-Фуллера (Dickey-Fuller test, DF-тест) основан на оценке параметра $\lambda = \alpha_1 - 1$ уравнения

$$\Delta Y_t = \lambda \cdot Y_{t-1} + \varepsilon_t, \quad (6.41)$$

эквивалентного уравнению авторегрессии (6.39). Его называют также *тестом на единичный корень*.

Нулевая H_0 и ей альтернативная H_1 гипотезы определяются соотношениями: $H_0: \lambda = 0$; $H_1: \lambda < 0$.

Если значение t -статистики Стьюдента для параметра λ меньше нижнего порогового значения DF -статистики, то нулевую гипотезу $\lambda = 0$ (о наличии единичного корня $\alpha_1 = 1$) следует отклонить и принять альтернативную о стационарности процесса Y_t .

Таблицы теста Дики-Фуллера (DF -теста) рассчитаны для уровней значимости в 1, 5, 10 %. Указанные в таблице значения DF -теста – отрицательные.

DF -тест применим также для тестирования на единичный корень *случайных процессов со смещением и со смещением и линейным детерминистическим трендом* определяемых уравнениями:

$$\Delta Y_t = \alpha_0 + \alpha_1 \cdot Y_{t-1} + \varepsilon_t, \quad (6.42)$$

$$\Delta Y_t = \alpha_0 + \alpha_1 \cdot Y_{t-1} + \alpha_2 \cdot t + \varepsilon_t, \quad (6.43)$$

где α_0 – константа, называемая смещением. При этом используются соответствующие таблицы критических значений DF -теста.

Отметим, что на практике трудно различить ситуации, когда следует применять DF -тест, а когда – DF -тест со смещением.

6.5.3. Модификации теста Дики-Фуллера для случая автокорреляции

При наличии автокорреляции в остатках ε_t используется *обобщенный тест Дики-Фуллера (ADF-тест)*, согласно которому в правую часть уравнения рег-

рессии в качестве дополнительных факторов включаются лаговые значения переменной из левой части Δy_{t-i}

$$\Delta y_t = a_1 y_{t-1} + \sum_{i=1}^k a_{i+1} \Delta y_{t-i} + \varepsilon_t. \quad (6.44)$$

Процедура тестирования, как и ранее, сводится к оценке значения t -критерия Стьюдента для параметра a_1 и сравнении его с критическими значениями для ADF-теста, которые совпадают с критическими значениями обычного DF -теста.

Такой же подход, может быть применен и в случаях тестирования на единичный корень *случайного процесса со смещением* и *случайного процесса со смещением и линейным детерминистическим трендом*:

$$\Delta y_t = a_0 + a_1 y_{t-1} + \sum_{i=1}^k a_{i+1} \Delta y_{t-i} + \varepsilon_t; \quad (6.45)$$

$$\Delta y_t = a_0 + a_1 y_{t-1} + \sum_{i=1}^k a_{i+1} \Delta y_{t-i} + a_{k+2} \cdot t + \varepsilon_t. \quad (6.46)$$

Как и ранее, критические значения для ADF-теста те же самые, что и для обычного DF -теста.

6.5.4. Метод разностей и интегрируемость

Для практики большой интерес представляют, так называемые, интегрируемые нестационарные процессы. Это процессы, для которых с помощью последовательного применения операции взятия последовательных разностей из нестационарных временных рядов можно получить стационарные ряды.

Последовательные разности стохастического процесса определяются соотношениями:

$$\begin{aligned} \Delta Y_t &= Y_t - Y_{t-1} && \text{— первые последовательные разности} \\ \Delta^2 Y_t &= \Delta Y_t - \Delta Y_{t-1} && \text{— вторые последовательные разности и т. д.} \end{aligned}$$

Если первые разности нестационарного ряда Y_t стационарны, то ряд Y_t называется интегрируемым первого порядка. Стационарный временной ряд называется интегрируемым нулевого порядка.

Если первые разности нестационарного ряда нестационарны, а вторые разности стационарны, то ряд Y_t называется интегрируемым второго порядка. Если первый стационарный ряд получается после k -кратного взятия разностей, то ряд Y_t называется интегрируемым k -го порядка.

6.6. Модели ARIMA

6.6.1. Определение и идентификация модели

Рассмотрим интегрируемый порядка d нестационарный процесс X_t . Если при этом процесс $Y_t = \Delta^d X_t$, составленный из первых разностей d -порядка исходного процесса, является процессом $ARMA(p, q)$, т. е.

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots + \alpha_p Y_{t-p} + \varepsilon_t - \beta_1 \varepsilon_{t-1} - \beta_2 \varepsilon_{t-2} - \dots - \beta_q \varepsilon_{t-q}, \quad (6.47)$$

тогда X_t называется процессом $ARIMA(p,d,q)$. На практике свободный член α_0 часто опускается (приравнивается к нулю).

Можно считать, что большинство эмпирических временных рядов является реализациями процессов $ARIMA$.

Основная проблема в анализе временных рядов заключается в определении порядка модели $ARIMA(p,d,q)$.

Необходимо оценить три основных параметра: d – порядок интегрируемости, порядок p компоненты AR и порядок q компоненты MA . Для экономических временных рядов параметр d обычно равен 1, возможны также значения 0 или 2. При определении параметров p и q используются характеристики автокорреляционной функции (ACF) и частной автокорреляционной функции (PACF). При этом предпочтение отдается моделям с наименьшим числом параметров.

6.6.2. Прогнозирование $ARIMA$ -процессов

Для прогнозирования $ARIMA$ -процессов X_t могут быть применены два подхода:

1) Получение прогнозных значений $\hat{Y}_T(h)$ $ARMA$ -процесса $Y_t = \Delta^d X_t$ по методике прогнозирования $ARMA$ -процессов (см. разд. 6.4) с последующим последовательным вычислением прогнозных значений $\Delta^{d-1} \hat{X}_T(h)$, $\Delta^{d-2} \hat{X}_T(h)$ и т. д., пока не будут получены $\hat{X}_T(h)$.

2) Построение прогнозной формулы с помощью модификации уравнения (6.47) путем подстановки разностей $\Delta^d X_t$ вместо Y_t и последующего разрешения полученного уравнения относительно X_t . В результате, будет получена $ARMA$ -модель нестационарного процесса, которая может быть преобразована в формулу для прогнозирования на h шагов вперед величин $\hat{X}_T(h)$ с началом отсчета в момент времени T по методике, описанной в разделе 6.4.

Рассмотрим $ARIMA(0,1,0)$ -модель случайного блуждания $Y_t = \Delta X_t = \varepsilon_t$ или в преобразованном виде $X_t = X_{t-1} + \varepsilon_t$.

Формула экстраполяции имеет вид

$$X_{T+h} = X_{T+h-1} + \varepsilon_t, \quad (6.48)$$

а формула прогноза дается соотношением

$$\hat{X}_T(h) = X_T, \quad \text{для } h \geq 1. \quad (6.49)$$

Дисперсия ошибки прогноза $\text{var}(e_T(h)) = h \cdot \sigma_\varepsilon^2$ увеличивается с ростом h . Ширина доверительного интервала прогноза возрастает пропорционально \sqrt{h} .

Если X_t – случайное блуждание со сдвигом

$$X_t = X_{t-1} + \alpha_0 + \varepsilon_t, \quad (6.50)$$

тогда формула для прогнозирования имеет вид

$$\hat{X}_{T(+h)} = X_T + h\alpha_0, \quad (6.51)$$

что соответствует простому линейному тренду. Дисперсия ошибки прогноза такая же, как и в предыдущем случае с $\alpha_0 = 0$.

Рассмотрим $ARIMA(1,1,1)$ -модель

$$\Delta X_t - \alpha_1 \Delta X_{t-1} = X_t - X_{t-1} - \alpha_1 (X_{t-1} - X_{t-2}) = \alpha_0 + \varepsilon_t - \beta_1 \cdot \varepsilon_{t-1},$$

которая после преобразования принимает вид

$$X_t = \alpha_0 + (1 + \alpha_1) X_{t-1} - \alpha_1 X_{t-2} + \varepsilon_t - \beta_1 \cdot \varepsilon_{t-1}. \quad (6.52)$$

Формулы для прогнозирования в момент $t = T + h$ определяются соотношениями

$$\begin{aligned} \hat{X}_T(1) &= \alpha_0 + (1 + \alpha_1) X_T - \alpha_1 X_{T-1} - \beta_1 \varepsilon_T, \\ \hat{X}_T(2) &= \alpha_0 + (1 + \alpha_1) \hat{X}_T(1) - \alpha_1 X_T, \\ \hat{X}_T(h) &= \alpha_0 + (1 + \alpha_1) \hat{X}_T(h-1) - \alpha_1 X_T(h-2) \quad \text{для } h \geq 3. \end{aligned} \quad (6.53)$$

Контрольные вопросы

1. Дайте определение стохастического процесса.
2. Дайте определение стационарного стохастического процесса в слабом (широком) смысле.
3. Какой стохастический процесс называется нормальным?
4. Какой стохастический процесс называется «белый шумом»?
5. Какими параметрами характеризуется стационарный процесс?
6. Дайте определение автоковариационной функции.
7. Какие методы применяются для распознавания стационарности временных рядов?
8. Приведите примеры параметрических тестов проверки временных рядов на стационарность?
9. Приведите примеры непараметрических тестов проверки временных рядов на стационарность?
10. Охарактеризуйте процессы AR.
11. В каких случаях процессы AR являются стационарными?
12. Охарактеризуйте процессы MA.
13. Охарактеризуйте процессы ARMA.
14. Опишите модель ARMA(3,2).
15. Как используется автокорреляционная функция для идентификации модели стационарного стохастического процесса?
16. Как используется частная автокорреляционная функция для идентификации модели стационарного стохастического процесса?
17. Как осуществляется прогнозирование ARMA-процессов?
18. Что может служить признаком нестационарности временного ряда?
19. Для чего применяются Тесты Дики-Фуллера?
20. Охарактеризуйте процессы ARIMA.
21. Как осуществляется прогнозирование ARMA-процессов?

7. Динамические эконометрические модели

7.1. Общая характеристика динамических моделей

При изучении поведения экономических процессов на достаточно длительном промежутке времени есть все основания предполагать о наличии определенных взаимосвязей между их последовательными состояниями. Т. е. состояние экономического явления в данный момент или период времени определяется, в том числе, и его состояниями, а также состояниями окружающей среды в предшествующие моменты или периоды времени. Данное обстоятельство является следствием наличия запаздывания в действии факторов либо инерционностью изучаемых процессов.

Модели, связывающие состояния экономических явлений в последовательные моменты (периоды) времени, принято называть *динамическими*. Такие модели позволяют изучать явления в динамике, в развитии. Аналитическое представление динамических моделей включает значения переменных, относящиеся как к текущему, так и к предыдущим моментам (периодам) времени.

Эконометрические модели, включающие в качестве факторов значения факторных переменных в предыдущие моменты времени, называются *моделями с распределенным лагом*.

$$y_t = a + b_0 \cdot x_t + b_1 \cdot x_{t-1} + b_2 \cdot x_{t-2} + \dots + b_p \cdot x_{t-p} + \varepsilon_t \quad (7.1)$$

Моделями этого типа описываются ситуации, когда влияние причины (независимых факторов) на следствие (зависимую переменную) проявляется с некоторым запаздыванием. Например, при изучении зависимости объемов выпуска продукции от величины инвестиций, выручки от расходов на рекламу и т. п.

Эконометрические модели, включающие в качестве факторов значения результативной переменной в предыдущие моменты времени. Эти модели называются *моделями авторегрессии*.

$$y_t = a + b_0 \cdot x_t + c_1 \cdot y_{t-1} + c_2 \cdot y_{t-2} + \dots + c_q \cdot y_{t-q} + \varepsilon_t. \quad (7.2)$$

Моделями такого типа предполагают наличие определенной инерционности в изменении рассматриваемого явления, когда уровень изучаемого явления существенно зависит от его уровней, достигнутых в предыдущих периодах. Например, уровень спроса на товар либо уровень ВВП в данном периоде во многом определяется уровнями, достигнутыми в предшествующем периоде.

Применение находят также и различные комбинации упомянутых выше моделей.

Отдельную группу динамических моделей составляют модели, учитывающие ожидаемые уровни переменных, которые определяются экономическими субъектами на основе информации, которой они располагают в текущий и предыдущий момент времени. Например, модели адаптивных ожиданий или частичной корректировки.

Включенные в модель в качестве факторов значения переменных в предыдущие моменты времени называются *лаговыми переменными*. Значениями лаговых переменных являются временные ряды исходных уровней, сдвинутые назад на один или более моментов времени. Величина этого сдвига называется *лагом*.

Включение в эконометрическую модель лаговых значений зависимой пе-

ременной осложняет проблему получения несмещенных и эффективных оценок ее параметров.

Во-первых, наличие нескольких лаговых переменных y_{t-1}, y_{t-2}, \dots либо x_{t-1}, x_{t-2}, \dots , зачастую сильной коррелирующих между собой, ведет к потере качества модели вследствие ухудшения точности оценок ее параметров, снижению их эффективности и устойчивости к незначительным колебаниям исходной информации, ошибкам округления.

Во-вторых, как правило, существует сильная корреляционная зависимость между переменными y_{t-1}, y_{t-2}, \dots и ошибкой ε_t , ведущая к появлению смещения в оценках параметров при использовании МНК.

В-третьих, временной ряд ошибки модели ε_t часто характеризуется наличием автокорреляционной связи, вследствие чего оценки параметров модели, полученные непосредственно на основе МНК, являются неэффективными.

Отметим, что важным этапом при построении моделей с распределенным лагом и моделей авторегрессии является выбор оптимальной величины лага и определение его структуры.

7.2. Модели с распределенным лагом

Рассмотрим модель с распределенным лагом порядка p (7.1)

$$y_t = a + b_0 \cdot x_t + b_1 \cdot x_{t-1} + b_2 \cdot x_{t-2} + \dots + b_p \cdot x_{t-p} + \varepsilon_t.$$

Основную проблему при оценке параметров составляет, как правило, сильная корреляция между факторами $x_t, x_{t-1}, x_{t-2}, \dots$. Для ее преодоления применяется преобразование лаговых переменных, либо делаются определенные предположения о характере коэффициентов регрессии.

7.2.1. Оценка параметров модели с распределенным лагом методом Койка

В методе Койка предполагается, что коэффициенты при лаговых переменных убывают в геометрической прогрессии

$$y_t = a + b \cdot x_t + b \cdot \gamma \cdot x_{t-1} + b \cdot \gamma^2 \cdot x_{t-2} + b \cdot \gamma^3 \cdot x_{t-3} + \dots + \varepsilon_t. \quad (7.3)$$

Модель (7.3) включает три параметра a, b и γ , для определения которых применяется нелинейный метод наименьших квадратов. Согласно этому методу:

а) Задаются границы изменения параметра γ (в простейшем случае 0 и 1) и определяются всевозможные значения параметра γ с достаточно малым шагом (например, 0,01). Для каждого значения параметра γ вычисляются значения новой переменной z

$$z_t = x_t + \gamma \cdot x_{t-1} + \gamma^2 \cdot x_{t-2} + \dots + \gamma^p \cdot x_{t-p}, \quad (7.4)$$

где величина p выбирается такой, чтобы воздействием последующих лаговых значений x_{t-p+i} можно пренебречь;

б) Затем оценивается уравнение регрессии

$$y_t = a + b \cdot z_t + u_t. \quad (7.5)$$

в) Далее выбирается такое значение параметра γ , которому соответствует наибольший коэффициент детерминации R^2 при оценке уравнения (7.5). Полу-

ченные при этом оценки параметров a и b принимаются в качестве оценок параметров исходного уравнения (7.3).

Другой подход к определению параметров уравнения (7.3) основан на так называемом преобразовании Койка. Запишем модель (7.3) для периода $t - 1$

$$y_{t-1} = a + b \cdot x_{t-1} + b \cdot \gamma \cdot x_{t-2} + b \cdot \gamma^2 \cdot x_{t-3} + b \cdot \gamma^3 \cdot x_{t-3} + \dots + \varepsilon_{t-1}.$$

Умножив обе части этого уравнения на γ и вычтя их из уравнения (7.3) после некоторого преобразования получим следующее соотношение

$$y_t = a(1 - \gamma) + b \cdot x_{t-1} + \gamma \cdot y_{t-1} + \varepsilon_t - \gamma \cdot \varepsilon_{t-1}. \quad (7.6)$$

Полученное уравнение представляет собой авторегрессионную модель первого порядка. Оценив параметры этого уравнения, можно получить оценки параметров a , b и γ исходного уравнения (7.3). Заметим, что применение в данном случае для оценки параметров обычного метода наименьших квадратов даст смещенные и несостоятельные оценки вследствие зависимости фактора y_{t-1} от одной из составляющих случайного члена ε_{t-1} .

7.2.2. Оценка параметров модели с распределенным лагом методом Алмон.

В методе Алмон для преодоления сильной корреляции между факторами $x_t, x_{t-1}, x_{t-2}, \dots$ используется переход к $k+1$ новым переменным z_j с меньшей корреляционной зависимостью по формулам

$$z_{jt} = a_{j0} \cdot x_t + a_{j1} \cdot x_{t-1} + \dots + a_{jp} \cdot x_{t-p}, \quad (j = 0, 1, 2, \dots, k) \quad (7.7)$$

где коэффициенты подобраны соответствующим образом.

Согласно методу Алмон? коэффициенты представляют в виде полиномов заданной степени k от величины лага j

$$b_j = c_0 + c_1 j + c_2 j^2 + \dots + c_k j^k. \quad (7.8)$$

В частности:

для полинома первой степени (при $k = 1$): $b_j = c_0 + c_1 j$;

для полинома второй степени (при $k = 2$): $b_j = c_0 + c_1 j + c_2 j^2$ и т. д.

Выражения для коэффициентов b_j модели (7.1) принимают вид:

$$\begin{aligned} b_0 &= c_0; \\ b_1 &= c_0 + c_1 + \dots + c_k; \\ b_2 &= c_0 + 2c_1 + 4c_2 + \dots + 2^k c_k; \\ &\dots \\ b_p &= c_0 + p c_1 + p^2 c_2 + \dots + p^k c_k; \end{aligned} \quad (7.9)$$

Подставив в (7.1) найденные соотношения для b_j , и вводя новые переменные z_j с помощью соотношений (7.11), представим исходное уравнение (7.1) в следующем виде

$$y_t = a + c_0 \cdot z_0 + c_1 \cdot z_1 + c_2 \cdot z_2 + \dots + c_k \cdot z_k + \varepsilon_t, \quad (7.10)$$

где

$$\begin{aligned}
z_0 &= x_t + x_{t-1} + \dots + x_{t-p} = \sum_{j=0}^p x_{t-j}; \\
z_1 &= x_{t-1} + 2x_{t-2} + \dots + p \cdot x_{t-p} = \sum_{j=1}^p j \cdot x_{t-j}; \\
z_2 &= x_{t-1} + 4x_{t-2} + \dots + p^2 \cdot x_{t-p} = \sum_{j=1}^p j^2 \cdot x_{t-j}; \\
&\dots\dots\dots \\
z_k &= x_{t-1} + 2^k x_{t-1} + \dots + p^k \cdot x_{t-p} = \sum_{j=1}^p j^k \cdot x_{t-j}.
\end{aligned} \tag{7.11}$$

После определения численных значений параметров c_j модели (7.10) коэффициенты исходной модели b_j находятся из соотношений (7.9).

Применение метода Алмон для расчета параметров модели с распределенным лагом предполагает предварительное определение максимальной величины лага p и степени полинома k . Оптимальную величину лага можно приближенно определить на основе априорной информации экономической теории или проведенных ранее эмпирических исследований. Приближенно в качестве величины лага можно взять значение максимального лага, для которого парный коэффициент корреляции между y и лаговыми переменными $x_t, x_{t-1}, x_{t-2}, \dots$ остается значимым.

Можно также построить несколько уравнений регрессии с разной величиной лага и выбрать наилучшее.

Что касается степени полинома k , то на практике обычно ограничиваются рассмотрением полиномов второй и третьей степени. Величину k также можно определять путем сравнения моделей, построенных для различных значений k .

Следует отметить, что при наличии сильной корреляционной связи между исходными лаговыми переменными $x_t, x_{t-1}, x_{t-2}, \dots$ переменные z_j , представляющие собой их линейные комбинации, также будут коррелировать между собой. Однако коэффициенты в формулах (7.11) подобраны таким образом, что такая зависимость будет существенно меньше.

Метод Алмон имеет следующие достоинства: он достаточно универсален и с помощью введения небольшого количества вспомогательных переменных z_j в уравнении (7.10) ($k = 2, 3$) позволяет построить модели с распределенным лагом любой длины.

7.2.3. Интерпретация параметров

Из соотношения (7.1) следует, что изменение независимой переменной x в каком-либо периоде времени t влияет на значение переменной y в данном периоде и в течение p следующих периодов времени. В последующие периоды это влияние проявляться не будет. Таким образом, временной интервал влияния конечен и ограничен $p+1$ периодом.

Коэффициент регрессии b_0 при переменной x_t называют *краткосрочным мультипликатором*. Он характеризует среднее абсолютное изменение y_t при

изменении x_t на одну единицу своего измерения в некотором периоде времени t , без учета воздействия лаговых значений фактора x .

Величины $(b_0 + b_1)$, $(b_0 + b_1 + b_2)$ и т. д. называются *промежуточными мультипликаторами*. Они характеризуют изменение y_t в течение двух, трех и т. д. периодов после изменения x_t на одну единицу.

Величина

$$b = b_0 + b_1 + \dots + b_l. \quad (7.12)$$

показывает максимальное суммарное изменение результирующей переменной y , которое будет достигнуто (по окончании текущего и p следующих периодов) под влиянием изменения фактора x на единицу в каком-либо периоде, и называется *долгосрчным мультипликатором*.

Например, для модели

$$y_t = 100 + 70x_t + 25x_{t-1} + 5x_{t-2}$$

краткосрочный мультипликатор равен 70, т. е. увеличение x_t на 1 единицу ведет в среднем к росту показателя y_t на 70 единиц в том же периоде. В течение двух периодов показатель y_t возрастет на $70 + 25 = 95$ единиц, а долгосрочный мультипликатор равен

$$b = (b_0 + b_1 + b_2) = 70 + 25 + 5 = 100,$$

и, следовательно, суммарное изменение показателя y_t составит 100 единиц.

7.3. Модели авторегрессии

7.3.1. Интерпретация параметров

Рассмотрим модель авторегрессии первого порядка

$$y_t = a + b_0 \cdot x_t + c_1 \cdot y_{t-1} + \varepsilon_t. \quad (7.13)$$

Коэффициент b_0 , как и ранее, характеризует краткосрочное изменение y_t под воздействием изменения x_t на единицу в том же периоде. Изменение y_t на b_0 в данном периоде в силу соотношения (7.13) повлечет в следующем периоде изменение y_{t+1} на величину $b_0 \cdot c_1$. В периоде $t + 2$ изменение y_{t+2} составит $b_0 \cdot c_1^2$ и т. д. Долгосрчный мультипликатор в модели авторегрессии рассчитывается как бесконечная сумма

$$b = b_0 + b_0 c_1 + b_0 c_1^2 + b_0 c_1^3 \dots \quad (7.14)$$

Если выполняется условие $|c_1| < 1$, то сумма в правой части (7.14), т. е. величина долгосрочного мультипликатора, будет конечная

$$b = b_0 \cdot (1 + c_1 + c_1^2 + c_1^3 \dots) = \frac{b_0}{1 - c_1}, \quad \text{где } |c_1| < 1. \quad (7.15)$$

Напомним, что неравенство $|c_1| < 1$ является условием стационарности авторегрессионного процесса первого порядка AR(1), определяемого уравнением (7.13).

В модельном примере

$$y_t = 200 + 50x_t + 0,6 y_{t-1},$$

краткосрочный мультипликатор равен 50, следовательно, увеличение x_t на 1 единицу приводит к росту y_t в том же периоде в среднем на 50 единиц. Долгосрчное изменение y_t составит $b = 50 / (1 - 0,6) = 125$ единиц, т. е. изменение x_t на

1 единицу в каком-либо периоде приведет к изменению y_t в долгосрочной перспективе в среднем на 125 единиц.

7.3.2. Оценка параметров моделей авторегрессии

Рассмотрим модель авторегрессии первого порядка

$$y_t = a + b_0 \cdot x_t + c_1 \cdot y_{t-1} + \varepsilon_t. \quad (7.16)$$

Одна из основных проблем при построении моделей авторегрессии (при оценке параметров) связана с наличием корреляционной зависимости между переменной y_{t-1} и остатками ε_t в уравнении регрессии, что приводит при применении обычного МНК к получению смещенной оценки параметра при переменной y_{t-1} .

Для преодоления этой проблемы обычно используется метод *инструментальных переменных*, согласно которому переменная y_{t-1} из правой части модели заменяется на новую переменную \hat{y}_{t-1} , которая, во-первых, должна тесно коррелировать с y_{t-1} , и, во-вторых, не коррелировать с ошибкой модели ε_t .

В качестве такой переменной можно взять регрессию переменной y_{t-1} на переменную x_{t-1} , определяемую соотношением

$$\hat{y}_{t-1} = d_0 + d_1 \cdot x_{t-1}, \quad (7.17)$$

где константы d_1 , d_2 являются коэффициентами уравнения регрессии

$$y_{t-1} = d_0 + d_1 \cdot x_{t-1} + u_t, \quad (7.18)$$

полученными с помощью обычного МНК.

Формула (7.17) получена в предположении о наличии зависимости y_{t-1} от x_{t-1} , как следствия предполагаемой зависимости y_t от x_t . Переменная \hat{y}_{t-1} , во-первых, тесно коррелирует с y_{t-1} , во-вторых, она не будет коррелировать с ошибкой ε_t , так как она линейно зависит от x_{t-1} , некоррелирующей с ε_t по предположению.

В результате, для оценки параметров уравнения (7.16) используется уравнение

$$y_t = a + b_0 \cdot x_t + c_1 \cdot \hat{y}_{t-1} + v_t, \quad (7.19)$$

где значения переменной \hat{y}_{t-1} рассчитаны по формуле (7.17).

Заметим, что функциональная связь между переменными \hat{y}_{t-1} и x_{t-1} (7.13) приводит к появлению высокой корреляционной связи между переменными \hat{y}_{t-1} и x_t . Для преодоления этой проблем в модель (7.16) и, соответственно, в модель (7.19) можно включить фактор времени в качестве независимой переменной. Модель при этом примет вид

$$y_t = a + b_0 \cdot x_t + c_1 \cdot \hat{y}_{t-1} + c_2 \cdot t + v_t. \quad (7.20)$$

Для проверки гипотезы об автокорреляции остатков в моделях авторегрессии (7.16) используется критерий h Дарбина. Фактическое значение критерия вычисляется по формуле

$$h = \left(1 - \frac{d}{2}\right) \cdot \sqrt{\frac{n}{1 - n \cdot V}}, \quad (7.21)$$

где d – фактическое значение критерия Дарбина–Уотсона для данной модели;

n – число наблюдений в модели;

V – квадрат стандартной ошибки при лаговой результативной переменной.

В качестве критических значений критерия при уровне значимости α берутся значения $t_{\alpha/2}$ и $t_{1-\alpha/2}$ квантилей порядка $\alpha/2$ и $1-\alpha/2$ для стандартизованного нормального распределения. Нулевая гипотезы об отсутствии автокорреляции не отвергается, если выполняется условие

$$t_{\alpha/2} < h < t_{1-\alpha/2}. \quad (7.22)$$

Заметим, что этот критерий применим, если $n \cdot V < 1$.

7.4. Модель частичной корректировки

В модели частичной корректировки предполагается, что под воздействием объясняющей переменной x_t формируется не фактическое значение переменной y_t , а ее «желаемый» уровень y^*_t [4]

$$y^*_t = a + b \cdot x_t. \quad (7.23)$$

А фактическое приращение зависимой переменной $y_t - y_{t-1}$ пропорционально разнице $y^*_t - y_{t-1}$ между ее желаемым уровнем и значением в предыдущий период

$$y_t - y_{t-1} = \lambda(y^*_t - y_{t-1}) + v_t, \quad (0 \leq \lambda \leq 1) \quad (7.24)$$

где v_t – случайный член. Записав выражение (7.24) как

$$y_t = \lambda y^*_t + (1 - \lambda)y_{t-1} + v_t, \quad (0 \leq \lambda \leq 1) \quad (7.25)$$

получим, что фактическое значение переменной y_t представляет собой взвешенное среднее желаемого уровня и фактического значения переменной y_t в предыдущем периоде. Чем больше значение λ , тем быстрее происходит корректировка. При $\lambda=1$ корректировка происходит за один период.

Модель, задаваемая соотношениями (7.19) и (7.21), называется *моделью частичной (неполной) корректировки*.

Объединяя соотношения (7.19) и (7.21), получим соотношение

$$y_t = a\lambda + b\lambda x_t + (1 - \lambda)y_{t-1} + v_t, \quad (7.26)$$

представляющее собой уравнение авторегрессии первого порядка, удовлетворяющее предположению Койка о характере изменения коэффициентов модели с распределенным лагом с параметром $\gamma = 1 - \lambda$. Так как, в уравнении (7.26) переменная y_{t-1} не коррелирует со случайным членом v_t , то обычный МНК позволяет получить состоятельные оценки его параметров. Используя эти оценки, можно получить и все параметры модели частичной корректировки a , b и λ .

Другой подход основан на применении обратного преобразования Койка и использовании нелинейного МНК.

Примером использования модели частичной корректировки является модель выплаты дивидендов подробно рассмотренная Дж. Линтнером [4]. Предполагается, что у фирм имеется целевая долгосрочная доля выплат γ и что желаемый объем дивидендов D^*_t соотносится с текущей прибылью Π_t как

$$D^*_t = \gamma \Pi_t. \quad (7.27)$$

Реальный объем дивидендов D^*_t частично корректируется

$$\Delta D_t = \lambda(D^*_t - D_{t-1}) + v_t. \quad (7.28)$$

Соотношение (7.26) принимает вид

$$D_t = \gamma \lambda \Pi_t + (1 - \lambda)D_{t-1} + v_t. \quad (7.29)$$

По данным о деятельности корпоративного сектора США за период 1918–1941 гг. Дж. Линтнером было получено уравнение регрессии

$$\hat{D}_t = 352,3 + 0,15\dot{I}_t + 0,70D_{t-1}. \quad (7.30)$$

Откуда следует, что коэффициент скорости корректировки $\lambda = 0,3$; а оценка для доли выплат $\gamma = 0,5$.

7.5. Модель адаптивных ожиданий

В модели адаптивных ожиданий предполагается, что фактическое значение переменной y_t формируется под воздействием ожидаемого значения объясняющей переменной x_{t+1}^e в следующий момент времени [4]

$$y_t = a + b \cdot x_{t+1}^e + \varepsilon_t, \quad (7.31)$$

а размер корректировки ожидаемого значения пропорционален разности между фактическим и ожидаемым значениями объясняющей переменной

$$x_{t+1}^e - x_t^e = \lambda(x_t - x_t^e). \quad (0 \leq \lambda \leq 1) \quad (7.32)$$

Соотношение (7.28) можно переписать в виде

$$x_{t+1}^e = \lambda x_t + (1 - \lambda)x_t^e, \quad (0 \leq \lambda \leq 1) \quad (7.33)$$

т. е. ожидаемый уровень объясняющей переменной представляет собой взвешенную сумму фактического и ожидаемого уровня в момент t .

Для исключения ожидаемого значения x_t^e из модели запишем уравнение (7.31) для момента $t-1$

$$y_{t-1} = a + b \cdot x_t^e + \varepsilon_{t-1} \quad (7.34)$$

и уравнение (7.27) с учетом соотношения (7.29)

$$y_t = a + b \cdot \lambda x_t + b \cdot (1 - \lambda)x_t^e + \varepsilon_t. \quad (7.35)$$

Исключая x_t^e из последних двух уравнений получим следующее уравнение модели

$$y_t = a \cdot \lambda + b \cdot \lambda x_t + (1 - \lambda)y_{t-1} + u_t, \quad (7.36)$$

где

$$u_t = \varepsilon_t - (1 - \lambda)\varepsilon_{t-1}. \quad (7.37)$$

Модель (7.36) представляет собой модель авторегрессии первого порядка, в которой динамика случайного члена подчинена закону скользящей средней первого порядка MA(1). В этой модели факторная переменная y_{t-1} коррелирует со случайным членом u_t , поэтому для определения параметров следует использовать нелинейный МНК после предварительного выполнения обратного преобразования Койка.

Первым примером использования модели адаптивных ожиданий считается модель гиперинфляции Ф. Кейгана. Объясняемой (зависимой) переменной в этой модели является спрос на реальные денежные остатки $y_t = \log(M_t / P_t)$, где M_t – номинальное количество денег в обращении, P_t – уровень цен. Величина M_t / P_t называется *денежными остатками*. Ф. Кейган предположил, что спрос на реальные денежные остатки в период гиперинфляции определяется ожидаемым уровнем инфляции x_{t+1}^e в момент времени $t+1$

$$y_t = a + b \cdot x_{t+1}^e + \varepsilon_t, \quad (7.38)$$

а корректировка ожидаемого значения инфляции x_{t+1}^e осуществляется по схеме

$$x_{t+1}^e - x_t^e = \lambda(x_t - x_t^e). \quad (0 \leq \lambda \leq 1) \quad (7.39)$$

По данным для семи периодов гиперинфляции, имевших место между 1921 и 1956 гг., были получены следующие значения параметров $b = 4,86$; $\lambda = 0,20$.

Другим известным примером модели адаптивных ожиданий является модель потребления, основанная на гипотезе Фридмена о постоянном доходе [4].

Контрольные вопросы

1. Какие эконометрические модели называются динамическими?
2. Что представляют из себя модели авторегрессии?
3. Что представляют из себя модели с распределенным лагом?
4. Что является значениями лаговых переменных?
5. Как интерпретируются параметры модели с распределенным лагом?
6. Как интерпретируются параметры модели авторегрессии?
7. В чем заключается метод Койка?
8. В чем заключается метод Алмон?
9. Как осуществляется оценка параметров моделей авторегрессии?
10. В чем заключается модель частичной корректировки?
11. В чем заключается модель адаптивных ожиданий?
12. Приведите пример модели частичной корректировки.
13. Приведите пример модели адаптивных ожиданий.

8. Информационные технологии эконометрических исследований

На современном этапе невозможно представить эконометрическое исследование без применения компьютеров. В настоящее время исследователю доступно большое количество разнообразных программных продуктов, которые могут быть использованы для решения эконометрических задач. Сюда относятся, естественно, и все статистические программные пакеты. Практика их использования позволила сформулировать следующие общие требования, предъявляемые к программному обеспечению, применяемому в эконометрических исследованиях:

- наличие удобных средств для работы с исходными данными;
- расчет статистических характеристик;
- поддержка методов построения моделей взаимосвязей;
- поддержка методов оценки адекватности моделей;
- реализация методов анализа и моделирования временных рядов;
- реализация методов прогнозирования;
- реализация статистических критериев;
- обеспечение возможности создания и сохранения сценария исследования, представляющего описание последовательно применяемых процедур;
- визуализация промежуточных и конечных результатов исследования.

Наиболее важными для исследователя являются средства автоматизации процесса моделирования и оценка адекватности полученных моделей.

С точки зрения эффективности использования рабочего времени важное значение имеют такие возможности по работе с исходными данными, как

- удобный ввод данных;
- накопление и хранение эконометрических данных;
- фильтрация и поиск информации;
- предварительная обработка данных.

Удобный ввод данных подразумевает наличие средств копирования данных из других приложений (созданных другими программными продуктами) в табличном виде либо возможности загрузки данных из файлов стандартных форматов (.xml, .xls, .txt).

Важное значение имеет также возможность графического представления исходных данных и результатов средствами 2D и 3D-графики.

Применяемое в эконометрических исследованиях программное обеспечение можно разделить на следующие группы:

- Программы, реализующие технологию электронных таблиц MS Excel, OpenOffice.org Calc и др. Используют представление данных в табличном виде и позволяют решать простейшие эконометрические задачи.

- Статистические пакеты общего назначения: SPSS, STATISTICA, STATGRAPHICS и др.

- Программы, ориентированные на решение эконометрических задач: Econometric Views, STADIA, Matrixer 3.4 и др.

- Специализированные статистические пакеты, предназначенные для решения ограниченного круга задач ЭВРИСТА, МЕЗОЗАВР, ОЛИМП, Forecast Expert и др.;
- Математические пакеты общего назначения Mathcad, Matlab, Maple, Mathematica и др.

8.1. Электронные таблицы Excel

К достоинствам электронных таблиц Excel относятся:

- доступность, так как они входят в состав наиболее используемого пакета Microsoft Office;
- наличие обширного методического обеспечения с примерами использования в эконометрике;
- широкая известность, так как, как правило, Excel изучается в курсе информатики;
- эффективность на этапе обучения эконометрическим методам;
- электронные таблицы согласуются с большинством статистических пакетов прикладных программ.

Электронные таблицы Excel предоставляют исследователю следующие возможности:

- проведение расчетов с использованием большого количества математических и статистических функций: вычисление распределений случайных величин; вычисление статистических характеристик; вычисление критических значений статистических гипотез. С помощью статистической функции «Линейн» можно получить расчеты основных характеристик всех эконометрических моделей;
- использование пакета «Анализ данных», реализующего некоторые статистические методы:
 - однофакторный дисперсионный анализ;
 - двухфакторный дисперсионный анализ;
 - корреляция;
 - ковариация;
 - описательная статистика;
 - экспоненциальное сглаживание;
 - двухвыборочный f -тест для дисперсии;
 - анализ Фурье;
 - скользящее среднее;
 - регрессия;
 - парный двухвыборочный t -тест для средних;
 - двухвыборочный t -тест с одинаковыми дисперсиями;
 - двухвыборочный t -тест с различными дисперсиями;
 - двухвыборочный z -тест для средних;
- имеется также возможность составления программ для реализации различных эконометрических моделей.

8.2. Статистический пакет общего назначения STATISTICA

Статистический пакет общего назначения STATISTICA разработан фирмой StatSoft Incorporation. В данный момент в продаже имеется русскоязычная версия STATISTICA 6 и англоязычная версия STATISTICA 9.

Стандартную комплектацию пакета STATISTICA составляют три модуля.

Базовый пакет (STATISTICA Base) предоставляет обширный выбор основных статистик в едином пакете в сочетании с мощностью, производительностью и простотой использования технологии STATISTICA.

Линейные и нелинейные модели (STATISTICA Advanced Linear/Non-Linear Models) включает большой набор самых современных инструментов для моделирования и прогнозирования, включающий возможность автоматического выбора модели и расширенные интерактивные средства визуализации.

Многомерные разведочные технологии анализа (STATISTICA Multivariate Exploratory Techniques) содержит широкий выбор разведочных технологий анализа различных типов данных в сочетании с богатыми интерактивными средствами визуализации.

Иерархическое меню характеризует возможности программы:

Основные статистики и таблицы.

Множественная регрессия.

Дисперсионный анализ (ДА).

Непараметрическая статистика.

Подгонка распределений.

Углубленные методы анализа:

- общие линейные модели;
- обобщенные линейные и нелинейные модели;
- общие регрессионные модели;
- общие модели частных наименьших квадратов;
- компоненты дисперсии;
- анализ выживаемости;
- нелинейное оценивание;
- множественная нелинейная регрессия;
- логлинейный анализ таблиц частот;
- временные ряды и прогнозирование;
- моделирование структурными уравнениями.

Многомерный разведочный анализ:

- кластерный анализ;
- факторный анализ;
- анализ главных компонент и классификация;
- канонический анализ;
- надежность и позиционный анализ;
- деревья классификации;
- анализ соответствий;
- многомерное шкалирование;

- дискриминантный анализ;
- общие модели дискриминантного анализа.

Анализ мощности, оценка объема выборки.

Нейронные сети.

Добыча данных:

- добытчик данных – все процедуры;
- добытчик данных – чистка и фильтрация;
- добытчик данных – бурение и расслоение данных;
- добытчик данных – углубленные методы классификации;
- добытчик данных – разведчик общих многомерных;
- добытчик данных – обобщенное прогнозирование;
- добытчик данных – нейросетевой разведчик;
- нейронные сети;
- анализ независимых компонент;
- обобщенные методы кластерного анализа;
- правила связи;
- общие деревья классификации;
- общие CHAID модели;
- интерактивные деревья (C&RT, CHAID);
- растущие деревья классификации и регрессии;
- обобщенные аддитивные модели;
- многомерные адаптивные сплайны (NAP-сплайны);
- процедура обучения (Байесовские, ближайшего соседа ...);
- быстрое построение прогноза;
- качество подгонки, классификации, прогноза;
- отсеивание признаков;
- объединение групп (классов).

Добыча данных в управлении качеством.

Текстовая добыча и сканирование Web.

Блочные статистики.

Вероятностный калькулятор.

Широкому распространению пакета STATISTICA способствуют удобный понятный интерфейс, развитые методы визуализации результатов, возможности обмена данными с электронными таблицами Excel и вызова непосредственно из Excel функций пакета STATISTICA. Для поддержки русскоязычных пользователей имеется вэб-страница «Статистический Портал StatSoft» [28].

8.3. Эконометрические программные пакеты. Matrixer 5.1

В настоящее время через Интернет доступен отечественный специализированный эконометрический пакет Matrixer (автор А. Циплаков). Программа предназначена для анализа и обработки данных, проведения эконометрических и статистических расчетов.

Matrixer может использоваться для преподавания эконометрики и проведения эконометрических исследований среднего объема.

В программе реализованы многие классические и современные эконометрические методы и модели:

- Линейная регрессия;
- Биномиальный логит и пробит;
- Упорядоченный пробит;
- Цензурированная (тобит) и усеченная регрессия;
- Пуассоновская и отрицательно-биномиальная регрессия;
- Регрессия с мультипликативной гетероскедастичностью;
- Регрессия с ARMA-процессами в ошибке;
- Модель Бокса-Дженкинса (ARIMA) с динамическими прогнозами;
- GARCH-регрессия;
- ARFIMA-FIGARCH (модели с долгосрочной памятью);
- (Обобщенный) метод инструментальных переменных;
- Непараметрическая регрессия;
- Непараметрическое оценивание плотности;
- Квантильная регрессия;
- Системы одновременных уравнений – 2SLS, 3SLS и FIML;
- Векторная авторегрессия;
- Нелинейная регрессия;
- Нелинейный метод инструментальных переменных;
- Метод максимального правдоподобия.

Программа предоставляет следующие возможности, помогающие обрабатывать и анализировать данные:

- Описательные статистики (среднее, дисперсия и т. п.);
- Различные графики данных;
- Графическое описание данных (гистограмма, оценка спектральной плотности, автокорреляционная функция);
- Таблица корреляций переменных;
- Калькулятор функций распределения (нормальное, Стьюдента, хи-квадрат, Фишера);
- Тест Дики-Фуллера;
- Векторные и матричные вычисления (откуда и название);
- Небольшой встроенный язык программирования для написания макросов;
- Табличный редактор;
- Быстрое построение графиков функций по формулам;
- 3-мерные графики;
- Универсальный модуль импорта данных из текстовых;
- Экспорт данных в нескольких популярных форматах;
- История команд.

Matrixer работает со следующими типами объектов: матрицами, переменными (столбцами матриц), скалярами, строками, моделями. Имеется возможность обмена данными с электронными таблицами Excel.

Управлять программой Matrixer можно с помощью: меню, «горячих клавиш», команд, запускаемых из командных окон.

Программа Matrixer проста в обращении и эффективна. Большое достоинство заключается в том, что русскоязычная версия бесплатная.

8.4. Анализ временных рядов в системе ЭВРИСТА

Система ЭВРИСТА является специализированным статистическим пакетом по исследованию временных рядов, который реализован для операционной системы WINDOWS. Разработчиком системы является ООО «Центр Статистических Исследований». Название системы ЭВРИСТА расшифровывается как «Экспериментальные Временные Ряды Интерактивный Статистический Анализ». В системе ЭВРИСТА реализовано более 100 различных алгоритмов статистического анализа. Система может работать одновременно со 100 рядами данных, размер которых ограничен только вместимостью жесткого диска. Для удобства пользователя в системе также предусмотрены возможности построения графиков, работа с таблицей, экспорт и импорт данных, построение проектов и т.д.

Данные можно импортировать из текстовых файлов, файлов формата DBASE, а также из любых приложений WINDOWS, которые имеют функцию экспорта в папку обмена.

Ниже описаны возможности системы.

Работа с данными

Сравнение двух выборок критериями: Уилкоксона, Клотца, Колмогорова-Смирнова, Хи-квадрат, Стьюдента, Фишера.

Нормализация данных.

Применение к заданным временным рядам преобразования Бокса-Кокса.

Взятие сезонных и несезонных разностей.

Вычисление автокорреляционной функции.

Вычисление аддитивной и мультипликативной сезонной компоненты.

Заполнение пропусков методом скользящего среднего.

Анализ тренда

Оценка тренда методом простого скользящего среднего.

Оценка тренда методом полиномиального скользящего среднего.

Оценка тренда по формулам Спенсера.

Оценка полиномиального тренда и построение доверительных интервалов.

Оценка 12 специальных нелинейных моделей тренда и построение доверительных интервалов.

Закон распределения

Моделирование ряда с заданным законом распределения (всего реализовано 15 законов).

Построение гистограммы и эмпирической функции распределения выборок.

Тестирование выборок на соответствие заданному закону распределения по критерию Хи-квадрат Пирсона.

Тестирование выборок на соответствие заданному закону распределения по критерию Колмогорова-Смирнова.

Автоматическое определение закона распределения и его параметров, наиболее подходящих к данной выборке.

Прогнозирование

Построение прогнозов временных рядов по методу Брауна.

Построение прогнозов сезонных временных рядов с помощью метода Хольта-Уинтерса.

Построение прогнозов временных рядов и доверительных интервалов по 12 специальным нелинейным моделям тренда.

Построение прогнозов временных рядов и доверительных интервалов по модели авторегрессии-скользящего среднего.

Построение прогнозов временных рядов и доверительных интервалов по сезонной модели авторегрессии-скользящего среднего.

Построение прогнозов временных рядов и доверительных интервалов по сезонным и несезонным ARCH моделям.

Спектральный анализ

Построение сглаженных оценок автокорреляционной функции временных рядов

Построение периодограммы временных рядов по частотной или временной шкале

Построение сглаженной периодограммы временных рядов по частотной или временной шкале

ARMA модели

Вычисление автокорреляционной и частной автокорреляционной функции для визуальной идентификации порядков модели.

Автоматическая оценка порядков AR и ARMA модели для данного ряда.

Оценка параметров модели авторегрессии методами Левинсона-Дурбина, Бурга, псевдо-наименьших квадратов и вычисление ее теоретических характеристик.

Оценка параметров ARСС модели методом наименьших квадратов, вычисление стандартных отклонений параметров, проверка гипотез согласия для каждого из них и всей модели в целом, вычисление теоретических характеристик оцененной модели.

Регрессионные модели

Оценка линейной регрессионной модели ряда с некоррелированными ошибками, выдача регрессионной таблицы.

Оценка линейной регрессионной модели ряда с некоррелированными ошибками с автоматическим подбором оптимального набора независимых переменных методом пошаговой регрессии, выдача регрессионной таблицы.

Анализ интервенций

Оценка параметров динамической модели интервенции методом наименьших квадратов, вычисление стандартных отклонений параметров, проверка гипотез согласия для каждого из них и всей модели в целом, вычисление теоретических характеристик оцененной модели. Порядки и вид интервенции могут за-

даваться вручную или вычисляться автоматически. Момент начала интервенции задается пользователем.

Удаление интервенции из ряда.

Гармонические модели

Оценка параметров гармонической модели для заданного ряда методом Байеса, вычисление стандартных отклонений параметров, вычисление теоретических характеристик оцененной модели. Для оценки необходимо задать только размерность модели, все остальные параметры вычисляются автоматически.

В настоящее время в экономической практике большое внимание уделяется анализу рыночных рисков. Программная система ЭВРИСТА предлагает широкий набор методов для применения в этих целях.

Контрольные вопросы

1. Какие требования предъявляются к программному обеспечению эконометрических исследований?
2. Как можно классифицировать программное обеспечение, применяемое в эконометрических исследованиях?
3. Назовите статистические пакеты общего назначения.
4. Каковы основные возможности электронных таблиц MS Excel?
5. Каковы основные возможности статистического пакета STATISTICA?
6. Каковы основные возможности статистического пакета Matrixer 5.1?
7. Каковы основные возможности статистического пакета ЭВРИСТА?

Глоссарий

Автокорреляция – корреляция между временной переменной и лаговой переменной, составленной от той же переменной.

Авторегрессия – регрессия зависимой временной переменной от лаговой переменной, составленной от той же переменной.

Адаптивные полиномиальные модели – адаптивные модели временного ряда, основанные на представлении уровней ряда в виде полиномов от интервала прогнозирования

Белый шум – стохастический процесс, представляющий собой последовательность независимых, одинаково распределенных случайных величин.

Временной ряд (ряд динамики, динамический ряд) – упорядоченная во времени последовательность численных показателей $\{(y_i, t_i), i = 1, 2, \dots, n\}$, характеризующих уровни развития изучаемого явления в последовательные моменты или периоды времени.

Гетероскедастичность – неоднородность относительно дисперсии.

Гомоскедастичность – однородность относительно дисперсии.

Двухшаговый метод наименьших квадратов – применяется для оценки параметров сверхидентифицируемой системы уравнений.

Интервальный прогноз – интервал, в котором с определенной вероятностью находится фактическое значение прогнозируемой переменной.

Коинтеграция нестационарных временных рядов – возможность получить стационарную линейную комбинацию исходных нестационарных временных рядов.

Корреляционная связь – зависимость, при которой изменение среднего значения результативного признака обусловлено изменением факторных признаков.

Корреляционный анализ – раздел математической статистики, изучающий наличие и тесноту связи между переменными без разделения переменных на зависимые и объясняющие.

Косвенный метод наименьших квадратов – применяется для оценки параметров идентифицируемой системы уравнений.

Коэффициент детерминации R^2 – представляет собой квадрат коэффициента корреляции, показывает, какая часть дисперсии результативного признака объяснена уравнением регрессии.

Коэффициент эластичности ε – для зависимости $y = f(x)$ вычисляется по формуле $\dot{Y} = f'(x) \frac{x}{y}$, показывает, на сколько процентов в среднем изменится результативный признак y при изменении фактора x на 1 % от своего номинального значения.

Критерий Стьюдента (t -критерий) – применяется для оценки статистической значимости коэффициентов полученного уравнения регрессии.

Критерий Фишера (F-критерий) – применяется для оценки статистической значимости полученного уравнения регрессии в целом.

Критерий Энгеля-Грэнджера – применяется для тестирования временных рядов на коинтеграцию.

Лаг – задержка. Обычно лаг рассматривают между причиной и следствием.

Линейная аналитическая зависимость – зависимость, определяемая соотношением $y = a + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_p \cdot x_p + \varepsilon$.

Линейный коэффициент корреляции r_{yx} – характеризует тесноту ли-

нейной связи между переменными $r_{xy} = \frac{\frac{1}{n} \sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sigma_x \sigma_y} = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y}$.

Логистическая аналитическая зависимость – зависимость, определяемая соотношением $\hat{y} = \frac{K}{1 + a \cdot e^{-bt}}$.

Ложная корреляция – высокое значение коэффициентов корреляции при отсутствии истинной зависимости между переменными.

Метод Алмон – применяется для оценки коэффициентов модели с распределенным лагом. Согласно методу Алмон, коэффициенты представляются в виде полиномов заданной степени k от величины лага j : $b_j = c_0 + c_1 j + c_2 j^2 + \dots + c_k j^k$.

Метод аналитического выравнивания – заключается в нахождении аналитической функции $\hat{y} = f(t)$, характеризующей основную тенденцию (тренд) изменения уровней временного ряда с течением времени.

Метод включения – применяется при отборе факторов множественной регрессии. Согласно методу включения, сначала строится уравнение регрессии с одним наиболее влияющим фактором (фактор, которому соответствует большее значение парного коэффициента корреляции с результативным признаком r_{yx_i}), затем уравнение регрессии с двумя наиболее влияющими факторами и т. д.

Метод исключения – применяется при отборе факторов множественной регрессии. Согласно методу исключения сначала строится уравнение регрессии с полным набором факторов, из числа которых затем последовательно исключаются наименее значимые факторы.

Метод Койка – применяется для оценки параметров модели с распределенным лагом. Основан на предположении, что коэффициенты при лаговых переменных убывают в геометрической прогрессии.

Метод наименьших квадратов (МНК) – метод оценки параметров уравнения регрессии, основанный на минимизации суммы квадратов отклонений расчетных значений (по уравнению регрессии) зависимой переменной от ее наблюдаемых значений.

Метод сравнения средних – применяется для выявления наличия монотонной тенденции на основе сравнения средних уровней двух отрезков временного ряда.

Метод Фостера-Стюарта – непараметрический метод, применяемый для выявления наличия монотонной тенденции временного ряда.

Модели авторегрессии – динамические модели, определяемые соотношением $y_t = a + b_0 \cdot x_t + c_1 \cdot y_{t-1} + c_2 \cdot y_{t-2} + \dots + c_q \cdot y_{t-q} + \varepsilon_t$.

Модели с распределенным лагом – динамические модели, определяемые соотношением $y_t = a + b_0 \cdot x_t + b_1 \cdot x_{t-1} + \dots + b_p \cdot x_{t-p} + \varepsilon_t$.

Модель адаптивных ожиданий – динамическая модель, основанная на предположении, что фактическое значение переменной y_t формируется под воздействием ожидаемого значения объясняющей переменной x_{t+1}^e в следующий момент времени.

Модель частичной корректировки – динамическая модель, основанная на предположении, что под воздействием объясняющей переменной x_t формируется не фактическое значение переменной y_t , а ее «желаемый» уровень y^*_t .

Мультиколлинеарность – множественная линейная связь.

Непараметрические тесты стационарности – тесты на стационарность, не выдвигающие каких-либо предположений о законе распределения тестируемого временного ряда, его параметрах.

Обобщенный метод наименьших квадратов – метод оценки параметров уравнения регрессии, в котором МНК применяется к уравнению регрессии, преобразованному таким образом, чтобы исключить гетероскедастичность остатков.

Остаточная дисперсия – величина, вычисляемая по формуле $D_{\text{инд}} = \frac{1}{n} \sum (\hat{y}_i - y_i)^2$, где \hat{y}_i, y_i – расчетное и наблюдаемое значение зависимой переменной.

Параметрические тесты стационарности – тесты на стационарность, основанные на относительно строгих предположениях относительно законов распределения временного ряда, его параметров.

Показательная аналитическая зависимость – зависимость, определяемая соотношением $\hat{y} = a \cdot b^x$.

Поле корреляций – совокупность точек на координатной плоскости, изображающих наблюдения.

Полулогарифмическая аналитическая зависимость – зависимость, определяемая соотношением $y = a + b_1 \ln x_1 + b_2 \ln x_2 + \dots + b_p \cdot \ln x_p$.

Предмет эконометрики – количественная оценка взаимосвязи между случайными событиями, признаками, показателями, факторами переменных экономических объектов.

Приведенная форма модели – система одновременных уравнений, которая отражает зависимость эндогенных переменных только от predetermined переменных.

Регрессионный анализ – раздел математической статистики, изучающий форму зависимости характеристик стохастического процесса от одного или нескольких факторов.

Регрессия – зависимость среднего значения какой-либо величины (y) от некоторой другой величины или от нескольких величин (x_i).

Регрессия множественная – зависимость с несколькими зависимыми переменными.

Регрессия парная – зависимость между двумя переменными.

Результативный признак – признак, изменяющийся под действием факторных признаков.

Случайное блуждание – стохастический процесс, определяемый соотношением $Y_t = Y_{t-1} + \varepsilon_t$.

Средняя квадратическая ошибка модели – величина, вычисляемая по формуле $\bar{\varepsilon}_{\hat{a}} = \sqrt{\frac{1}{n} \sum (\hat{y}_i - y_i)^2}$, где \hat{y}_i, y_i – расчетное и наблюдаемое значение зависимой переменной.

Стационарный стохастический процесс в широком (слабом) смысле – стохастический процесс, у которого математическое ожидание μ_t и дисперсия σ_t^2 не зависят от времени (одинаковы для всех X_t), а автоковариация $\gamma_{t_1 t_2}$ зависит только от величины лага $\tau = t_2 - t_1$.

Стационарный стохастический процесс в узком (сильном) смысле – стохастический процесс, для которого совместное распределение вероятностей случайных величин $X_{t_1}, X_{t_2}, \dots, X_{t_n}$ такое же, как у случайных величин $X_{t_1+\tau}, X_{t_2+\tau}, \dots, X_{t_n+\tau}$ при любых n, t и τ .

Степенная аналитическая зависимость – зависимость, определяемая соотношением $y = a \cdot x_1^{b_1} \cdot x_2^{b_2} \cdot \dots \cdot x_p^{b_p} \cdot \varepsilon$.

Стохастическая связь – причинная зависимость, которая проявляется не в каждом отдельном случае, а в общем, среднем при большом числе наблюдений.

Стохастический – случайный, вероятностный.

Стохастический процесс – случайная функция $X(t)$ вещественного аргумента t .

Структурная форма модели – исходная система одновременных уравнений, содержащая в качестве факторов зависимые переменные.

Тенденция – изменение, определяющее общее направление развития, основную тенденцию временного ряда.

Теорема Гаусса-Маркова – теорема, определяющая условия, при соблюдении которых оценки параметров линейной регрессии, полученные методом наименьших квадратов, будут несмещенными и эффективными в классе линейных несмещенных оценок

Тест Глейзера – применяется для проверки остатков регрессии на гетероскедастичность. Проверяет существование функциональной зависимости следующего вида $\sigma_i = \alpha + \beta x_i^j$.

Тест Гольдфельда-Квандта – применяется для проверки остатков регрессии на гетероскедастичность. Проверяет равенство дисперсий остатков на двух интервалах изменения фактора.

Тест Дики-Фуллера – применяется для тестирования временного ряда на нестационарность, основан на оценке параметра $\lambda = \alpha_1 - 1$ уравнения $\Delta Y_t = \lambda \cdot Y_{t-1} + \varepsilon_t$.

Тест Манна-Уитни – непараметрический тест стационарности применяется для тестирования постоянства математического ожидания и основан на сопоставлении рангов элементов двух совокупностей из рассматриваемого интервала.

Тест на единичный корень (тест Дики-Фуллера) – применяется для тестирования временного ряда на нестационарность, основан на оценке параметра $\lambda = \alpha_1 - 1$ уравнения $\Delta Y_t = \lambda \cdot Y_{t-1} + \varepsilon_t$.

Тест ранговой корреляции Спирмена – применяется для проверки остатков регрессии на гетероскедастичность. Проверяет наличие монотонной зависимости между дисперсией ошибки и величиной фактора.

Тест Сиджела-Тьюки – применяется для тестирования постоянства дисперсии временного ряда и основан на сопоставлении рангов элементов двух совокупностей из рассматриваемого интервала.

Точечный прогноз – среднее прогнозное значение изучаемой переменной экономического объекта.

Тренд – изменение, определяющее общее направление развития, основную тенденцию временного ряда.

Трехшаговый метод наименьших квадратов – применяется для оценки параметров системы уравнений при нарушении предпосылок о независимости и гомоскедастичности остатков системы одновременных уравнений.

Условия Гаусса-Маркова – условия, обеспечивающие состоятельность, несмещенность и эффективность оценок параметров уравнения регрессии с помощью МНК.

Факторный признак – признак, обуславливающий изменение другого, связанного с ним признака.

Фиктивные переменные – применяются для учета в модели влияния качественного фактора.

Функциональная связь – связь, при которой определенному значению факторного признака соответствует одно и только одно значение результативного признака.

Частные уравнения регрессии – характеризующие изолированное влияние одного из факторов x_i на результативную переменную y при исключении влияния остальных факторов, входящих в общее уравнение регрессии.

Экзогенные (внешние, независимые) переменные – переменные, значения которых определяются вне рассматриваемой модели, для которой они являются заданными.

Эконометрика – раздел науки, изучающий конкретные количественные и качественные взаимосвязи экономических объектов и процессов с помощью математических и статистических методов и моделей.

Экспоненциальная аналитическая зависимость – зависимость, определяемая соотношением $y = e^{a+b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_p \cdot x_p + \varepsilon}$.

Экспоненциальное сглаживание – сглаживание временного ряда y_t на основе рекуррентной формулы $S_t = \alpha y_t + \beta S_{t-1}$, где S_t – значение экспоненциальной средней в момент t ; y_t – значение временного ряда в момент t ; α – параметр сглаживания, $\alpha = \text{const}$, $0 < \alpha < 1$; $\beta = 1 - \alpha$.

Экстраполяция – прогноз, получение расчетных значений при условии, что значения аргумента выходят за пределы области определения функции.

Эндогенные (внутренние, зависимые) переменные – переменные, значения которых определяются внутри рассматриваемой модели.

Приложение

1. Нормированная функция Лапласа

$$\Phi(e) = \frac{1}{\sqrt{2\pi}} \int_0^e e^{-\frac{t^2}{2}} \cdot dt = \frac{1}{\sqrt{2\pi}} \int_{-e}^0 e^{-\frac{t^2}{2}} \cdot dt$$

t	0	1	2	3	4	5	6	7	8	9
0	0	40	80	120	160	199	239	279	319	359
0,1	398	438	478	517	557	596	636	675	714	753
0,2	793	832	871	910	948	987	1026	1064	1103	1141
0,3	1179	1217	1255	1293	1331	1368	1406	1443	1480	1517
0,4	1554	1591	1628	1664	1700	1736	1772	1808	1844	1879
0,5	1915	1950	1985	2019	2054	2088	2123	2157	2190	2224
0,6	2257	2291	2324	2357	2389	2422	2454	2486	2517	2549
0,7	2580	2611	2642	2673	2704	2734	2764	2794	2823	2852
0,8	2881	2910	2939	2967	2995	3023	3051	3078	3106	3133
0,9	3159	3186	3212	3238	3264	3289	3315	3340	3365	3389
1	3413	3438	3461	3485	3508	3531	3554	3577	3599	3621
1,1	3643	3665	3686	3708	3729	3749	3770	3790	3810	3830
1,2	3849	3869	3888	3907	3925	3944	3962	3980	3997	4015
1,3	4032	4049	4066	4082	4099	4115	4131	4147	4162	4177
1,4	4192	4207	4222	4236	4251	4265	4279	4292	4306	4319
1,5	4332	4345	4357	4370	4382	4394	4406	4418	4429	4441
1,6	4452	4463	4474	4484	4495	4505	4515	4525	4535	4545
1,7	4554	4564	4573	4582	4591	4599	4608	4616	4625	4633
1,8	4641	4649	4656	4664	4671	4678	4686	4693	4699	4706
1,9	4713	4719	4726	4732	4738	4744	4750	4756	4761	4767
2	4772	4778	4783	4788	4793	4798	4803	4808	4812	4817
2,1	4821	4826	4830	4834	4838	4842	4846	4850	4854	4857
2,2	4861	4864	4868	4871	4875	4878	4881	4884	4887	4890
2,3	4893	4896	4898	4901	4904	4906	4909	4911	4913	4916
2,4	4918	4920	4922	4925	4927	4929	4931	4932	4934	4936
2,5	4938	4940	4941	4943	4945	4946	4948	4949	4951	4952
2,6	4953	4955	4956	4957	4959	4960	4961	4962	4963	4964
2,7	4965	4966	4967	4968	4969	4970	4971	4972	4973	4974
2,8	4974	4975	4976	4977	4977	4978	4979	4979	4980	4981
2,9	4981	4982	4982	4983	4984	4984	4985	4985	4986	4986
3	4987	4987	4987	4988	4988	4989	4989	4989	4990	4990
3,1	4990	4991	4991	4991	4992	4992	4992	4992	4993	4993
3,2	4993	4993	4994	4994	4994	4994	4994	4995	4995	4995
3,3	4995	4995	4995	4996	4996	4996	4996	4996	4996	4997
3,4	4997	4997	4997	4997	4997	4997	4997	4997	4997	4998
3,5	4998	4998	4998	4998	4998	4998	4998	4998	4998	4998

*Значения ординат увеличены в 10 000 раз

2. Значения критических уровней $t_{\alpha,k}$ в зависимости от k степеней свободы и заданного уровня значимости α для распределения Стьюдента

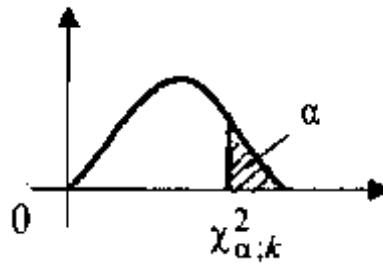
$k \backslash \alpha$	0,1	0,05	0,025	0,02	0,01	0,005	0,001
1	6,31	12,71	25,45	31,82	63,66	127,32	636,62
2	2,92	4,30	6,21	6,96	9,92	14,09	31,60
3	2,35	3,18	4,18	4,54	5,84	7,45	12,92
4	2,13	2,78	3,50	3,75	4,60	5,60	8,61
5	2,02	2,57	3,16	3,36	4,03	4,77	6,87
6	1,94	2,45	2,97	3,14	3,71	4,32	5,96
7	1,89	2,36	2,84	3,00	3,50	4,03	5,41
8	1,86	2,31	2,75	2,90	3,36	3,83	5,04
9	1,83	2,26	2,69	2,82	3,25	3,69	4,78
10	1,81	2,23	2,63	2,76	3,17	3,58	4,59
12	1,78	2,18	2,56	2,68	3,05	3,43	4,32
14	1,76	2,14	2,51	2,62	2,98	3,33	4,14
16	1,75	2,12	2,47	2,58	2,92	3,25	4,01
18	1,73	2,10	2,45	2,55	2,88	3,20	3,92
20	1,72	2,09	2,42	2,53	2,85	3,15	3,85
22	1,72	2,07	2,41	2,51	2,82	3,12	3,79
24	1,71	2,06	2,39	2,49	2,80	3,09	3,75
26	1,71	2,06	2,38	2,48	2,78	3,07	3,71
28	1,70	2,05	2,37	2,47	2,76	3,05	3,67
30	1,70	2,04	2,36	2,46	2,75	3,03	3,65
32	1,69	2,04	2,35	2,45	2,74	3,01	3,62
34	1,69	2,03	2,35	2,44	2,73	3,00	3,60
36	1,69	2,03	2,34	2,43	2,72	2,99	3,58
38	1,69	2,02	2,33	2,43	2,71	2,98	3,57
40	1,68	2,02	2,33	2,42	2,70	2,97	3,55
45	1,68	2,01	2,32	2,41	2,69	2,95	3,52
50	1,68	2,01	2,31	2,40	2,68	2,94	3,50
∞	1,64	1,96	2,24	2,33	2,58	2,81	3,29

3. Значения F -критерия Фишера на уровне значимости $\alpha = 0,05$

$k_2 \backslash k_1$	1	2	3	4	5	6	8	12	24	∞
1	161,45	199,50	215,71	224,58	230,16	233,99	238,88	243,91	249,05	254,31
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,45	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,85	8,74	8,64	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91	5,77	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,84	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,41	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,90	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,61	2,40
12	4,75	3,89	3,49	3,26	3,11	3,00	2,85	2,69	2,51	2,30
13	4,67	3,81	3,41	3,18	3,03	2,92	2,77	2,60	2,42	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,35	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,31	2,11	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,08	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,05	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,03	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,37	2,20	2,01	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	1,98	1,73
25	4,24	3,39	2,99	2,76	2,60	2,49	2,34	2,16	1,96	1,71
26	4,23	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,95	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,31	2,13	1,93	1,67
28	4,20	3,34	2,95	2,71	2,56	2,45	2,29	2,12	1,91	1,65
29	4,18	3,33	2,93	2,70	2,55	2,43	2,28	2,10	1,90	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62
35	4,12	3,27	2,87	2,64	2,49	2,37	2,22	2,04	1,83	1,56
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,79	1,51
45	4,06	3,20	2,81	2,58	2,42	2,31	2,15	1,97	1,76	1,47
50	4,03	3,18	2,79	2,56	2,40	2,29	2,13	1,95	1,74	1,44
60	4,00	3,15	2,76	2,53	2,37	2,25	2,10	1,92	1,70	1,39
70	3,98	3,13	2,74	2,50	2,35	2,23	2,07	1,89	1,67	1,35
80	3,96	3,11	2,72	2,49	2,33	2,21	2,06	1,88	1,65	1,32
90	3,95	3,10	2,71	2,47	2,32	2,20	2,04	1,86	1,64	1,30
100	3,94	3,09	2,70	2,46	2,31	2,19	2,03	1,85	1,63	1,28
∞	3,84	3,00	2,60	2,37	2,21	2,10	1,94	1,75	1,52	1,00

4. Значения F -критерия Фишера на уровне значимости $\alpha = 0,01$

$k_2 \backslash k_1$	1	2	3	4	5	6	8	12	24	∞
1	4052,18	4999,50	5403,35	5624,58	5763,65	5858,99	5981,07	6106,32	6234,63	6365,86
2	98,50	99,00	99,17	99,25	99,30	99,33	99,37	99,42	99,46	99,50
3	34,12	30,82	29,46	28,71	28,24	27,91	27,49	27,05	26,60	26,13
4	21,20	18,00	16,69	15,98	15,52	15,21	14,80	14,37	13,93	13,46
5	16,26	13,27	12,06	11,39	10,97	10,67	10,29	9,89	9,47	9,02
6	13,75	10,92	9,78	9,15	8,75	8,47	8,10	7,72	7,31	6,88
7	12,25	9,55	8,45	7,85	7,46	7,19	6,84	6,47	6,07	5,65
8	11,26	8,65	7,59	7,01	6,63	6,37	6,03	5,67	5,28	4,86
9	10,56	8,02	6,99	6,42	6,06	5,80	5,47	5,11	4,73	4,31
10	10,04	7,56	6,55	5,99	5,64	5,39	5,06	4,71	4,33	3,91
11	9,65	7,21	6,22	5,67	5,32	5,07	4,74	4,40	4,02	3,60
12	9,33	6,93	5,95	5,41	5,06	4,82	4,50	4,16	3,78	3,36
13	9,07	6,70	5,74	5,21	4,86	4,62	4,30	3,96	3,59	3,17
14	8,86	6,51	5,56	5,04	4,69	4,46	4,14	3,80	3,43	3,00
15	8,68	6,36	5,42	4,89	4,56	4,32	4,00	3,67	3,29	2,87
16	8,53	6,23	5,29	4,77	4,44	4,20	3,89	3,55	3,18	2,75
17	8,40	6,11	5,18	4,67	4,34	4,10	3,79	3,46	3,08	2,65
18	8,29	6,01	5,09	4,58	4,25	4,01	3,71	3,37	3,00	2,57
19	8,18	5,93	5,01	4,50	4,17	3,94	3,63	3,30	2,92	2,49
20	8,10	5,85	4,94	4,43	4,10	3,87	3,56	3,23	2,86	2,42
21	8,02	5,78	4,87	4,37	4,04	3,81	3,51	3,17	2,80	2,36
22	7,95	5,72	4,82	4,31	3,99	3,76	3,45	3,12	2,75	2,31
23	7,88	5,66	4,76	4,26	3,94	3,71	3,41	3,07	2,70	2,26
24	7,82	5,61	4,72	4,22	3,90	3,67	3,36	3,03	2,66	2,21
25	7,77	5,57	4,68	4,18	3,85	3,63	3,32	2,99	2,62	2,17
26	7,72	5,53	4,64	4,14	3,82	3,59	3,29	2,96	2,58	2,13
27	7,68	5,49	4,60	4,11	3,78	3,56	3,26	2,93	2,55	2,10
28	7,64	5,45	4,57	4,07	3,75	3,53	3,23	2,90	2,52	2,06
29	7,60	5,42	4,54	4,04	3,73	3,50	3,20	2,87	2,49	2,03
30	7,56	5,39	4,51	4,02	3,70	3,47	3,17	2,84	2,47	2,01
35	7,42	5,27	4,40	3,91	3,59	3,37	3,07	2,74	2,36	1,89
40	7,31	5,18	4,31	3,83	3,51	3,29	2,99	2,66	2,29	1,80
45	7,23	5,11	4,25	3,77	3,45	3,23	2,94	2,61	2,23	1,74
50	7,17	5,06	4,20	3,72	3,41	3,19	2,89	2,56	2,18	1,68
60	7,08	4,98	4,13	3,65	3,34	3,12	2,82	2,50	2,12	1,60
70	7,01	4,92	4,07	3,60	3,29	3,07	2,78	2,45	2,07	1,54
80	6,96	4,88	4,04	3,56	3,26	3,04	2,74	2,42	2,03	1,49
90	6,93	4,85	4,01	3,53	3,23	3,01	2,72	2,39	2,00	1,46
100	6,90	4,82	3,98	3,51	3,21	2,99	2,69	2,37	1,98	1,43
∞	6,63	4,61	3,78	3,32	3,02	2,80	2,51	2,18	1,79	1,00

5. Значения $\chi^2_{\alpha;k}$ критерия Пирсона

Число степеней свободы k	Вероятность α												
	0,99	0,98	0,95	0,90	0,80	0,70	0,50	0,30	0,20	0,10	0,05	0,02	0,01
1	0,00	0,00	0,00	0,02	0,06	0,15	0,45	1,07	1,64	2,71	3,84	5,41	6,64
2	0,02	0,04	0,10	0,21	0,45	0,71	1,39	2,41	3,22	4,60	5,99	7,82	9,21
3	0,11	0,18	0,35	0,58	1,00	1,42	2,37	3,66	4,64	6,25	7,82	9,84	11,3
4	0,30	0,43	0,71	1,06	1,65	2,20	3,36	4,88	5,99	7,78	9,49	11,7	13,3
5	0,55	0,75	1,14	1,61	2,34	3,00	4,35	6,06	7,29	9,24	11,1	13,4	15,1
6	0,87	1,13	1,63	2,20	3,07	3,83	5,35	7,23	8,56	10,6	12,6	15,0	16,8
7	1,24	1,56	2,17	2,83	3,82	4,67	6,35	8,38	9,80	12,0	14,1	16,6	18,5
8	1,65	2,03	2,73	3,49	4,59	5,53	7,34	9,52	11,0	13,4	15,5	18,2	20,1
9	2,09	2,53	3,32	4,17	5,38	6,39	8,34	10,7	12,2	14,7	16,9	19,7	21,7
10	2,56	3,06	3,94	4,86	6,18	7,27	9,34	11,8	13,4	16,0	18,3	21,2	23,2
11	3,05	3,61	4,58	5,58	6,99	8,15	10,3	12,9	14,6	17,3	19,7	22,6	24,7
12	3,57	4,18	5,23	6,30	7,81	9,03	11,3	14,0	15,8	18,5	21,0	24,1	26,2
13	4,11	4,76	5,89	7,04	8,63	9,93	12,3	15,1	17,0	19,8	22,4	25,5	27,7
14	4,66	5,37	6,57	7,79	9,47	10,8	13,3	16,2	18,1	21,1	23,7	26,9	29,1
15	5,23	5,98	7,26	8,55	10,3	11,7	14,3	17,3	19,3	22,3	25,0	28,3	30,6
16	5,81	6,61	7,96	9,31	11,1	12,6	15,3	18,4	20,5	23,5	26,3	29,6	32,0
17	6,41	7,26	8,67	10,1	12,0	13,5	16,3	19,5	21,6	24,8	27,6	31,0	33,4
18	7,02	7,91	9,39	10,9	12,9	14,4	17,3	20,6	22,8	26,0	28,9	32,3	34,8
19	7,63	8,57	10,1	11,6	13,7	15,3	18,3	21,7	23,9	27,2	30,1	33,7	36,2
20	8,26	9,24	10,8	12,4	14,6	16,3	19,3	22,8	25,0	28,4	31,4	35,0	37,6
21	8,90	9,92	11,6	13,2	15,4	17,2	20,3	23,9	26,2	29,6	32,7	36,3	38,9
22	9,54	10,6	12,3	14,0	16,3	18,1	21,3	24,9	27,3	30,8	33,9	37,7	40,3
23	10,2	11,3	13,1	14,8	17,2	19,0	22,3	26,0	28,4	32,0	35,2	39,0	41,6
24	10,9	12,0	13,8	15,7	18,1	19,9	23,3	27,1	29,6	33,2	36,4	40,3	43,0
25	11,5	12,7	14,6	16,5	18,9	20,9	24,3	28,2	30,7	34,4	37,7	41,7	44,3
26	12,2	13,4	15,4	17,3	19,8	21,8	25,3	29,2	31,8	35,6	38,9	42,9	45,6
27	12,9	14,1	16,1	18,1	20,7	22,7	26,3	30,3	32,9	36,7	40,1	44,1	47,0
28	13,6	14,8	16,9	18,9	21,6	23,6	27,3	31,4	34,0	37,9	41,3	45,4	48,3
29	14,3	15,6	17,7	19,8	22,5	24,6	28,3	32,5	35,1	39,1	42,6	46,7	49,6
30	14,9	16,3	18,5	20,6	23,4	25,5	29,3	33,5	36,2	40,3	43,8	48,0	50,9

6. Значения статистик Дарбина-Уотсона d_L d_U при уровне значимости $\alpha = 0,05$

(n – число наблюдений; m – число объясняющих переменных)

n	m=1		m=2		m=3		m=4		m=5	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
	0,61	1,40	–	–	–	–				
7	0,70	1,36	0,47	1,90	–	–				
8	0,76	1,33	0,56	1,78	0,37	2,29				
9	0,82	1,32	0,63	1,70	0,46	2,13				
10	0,88	1,32	0,70	1,64	0,53	2,02				
11	0,93	1,32	0,66	1,60	0,60	1,93				
12	0,97	1,33	0,81	1,58	0,66	1,86				
13	1,01	1,34	0,86	1,56	0,72	1,82				
14	1,05	1,35	0,91	1,55	0,77	1,78				
15	1,08	1,36	0,95	1,54	0,82	1,75	0,69	1,97	0,56	2,21
16	1,10	1,37	0,98	1,54	0,86	1,73	0,74	1,93	0,62	2,15
17	1,13	1,38	1,02	1,54	0,90	1,71	0,78	1,90	0,67	2,10
18	1,16	1,39	1,05	1,53	0,93	1,69	0,82	1,87	0,71	2,06
19	1,18	1,40	1,08	1,53	0,97	1,68	0,86	1,85	0,75	2,02
20	1,20	1,41	1,10	1,54	1,00	1,68	0,90	1,83	0,79	1,99
21	1,22	1,42	1,13	1,54	1,03	1,67	0,93	1,81	0,83	1,96
22	1,24	1,43	1,15	1,54	1,05	1,66	0,96	1,80	0,86	1,94
23	1,26	1,44	1,17	1,54	1,08	1,66	0,99	1,79	0,90	1,92
24	1,27	1,45	1,19	1,55	1,10	1,66	1,01	1,78	0,93	1,90
25	1,29	1,45	1,21	1,55	1,12	1,66	1,04	1,77	0,95	1,89
26	1,30	1,46	1,22	1,55	1,14	1,65	1,06	1,76	0,98	1,88
27	1,32	1,47	1,24	1,56	1,16	1,65	1,08	1,76	1,01	1,86
28	1,33	1,48	1,26	1,56	1,18	1,65	1,10	1,75	1,03	1,85
29	1,34	1,48	1,27	1,56	1,20	1,65	1,12	1,74	1,05	1,84
30	1,35	1,49	1,28	1,57	1,21	1,65	1,14	1,74	1,07	1,83
31	1,36	1,50	1,30	1,57	1,23	1,65	1,16	1,74	1,09	1,83
32	1,37	1,50	1,31	1,57	1,24	1,65	1,18	1,73	1,11	1,82
33	1,38	1,51	1,32	1,58	1,26	1,65	1,19	1,73	1,13	1,18
34	1,39	1,51	1,33	1,58	1,27	1,65	1,21	1,73	1,15	1,81
35	1,40	1,52	1,34	1,58	1,28	1,65	1,22	1,73	1,16	1,80
36	1,41	1,52	1,35	1,59	1,29	1,65	1,24	1,73	1,18	1,80
37	1,42	1,53	1,36	1,59	1,31	1,66	1,25	1,72	1,19	1,80
38	1,43	1,54	1,37	1,59	1,32	1,66	1,26	1,72	1,21	1,79
39	1,43	1,54	1,38	1,60	1,33	1,66	1,27	1,72	1,22	1,79
40	1,44	1,54	1,39	1,60	1,34	1,66	1,28	1,72	1,23	1,78
100	1,65	1,69	1,63	1,72	1,61	1,74	1,59	1,76	1,57	1,78

7. Критические значения f -критерия для DF-, ADF- и PP-тестов, рассчитанные по Маккиннону [5]

Уровень значимости	Тип функции		
	без параметра смещения и тренда	включая параметр смещения	включая параметр смещения и тренд
T=25			
0,01	-2,6756	-3,7667	-4,4415
0,05	-1,9574	-3,0038	-3,6330
0,01	-1,6238	-2,6417	-3,2535
T=23			
0,01	-2,6700	-3,7497	-4,4167
0,05	-1,9566	-2,9969	-3,6219
0,1	-1,6235	-2,6381	-3,2474
T=88			
0,01	-2,5894	-3,5055	-4,0648
0,05	-1,9438	-2,8943	-3,4608
0,1	-1,6177	-2,5840	-3,1564
T= 118			
0,01	-2,5831	-3,4865	-4,0380
0,05	-1,9427	-2,8859	-3,4481
0,1	-1,6171	-2,5796	-3,1489
T= 150			
0,01	-2,5793	-3,4749	-4,0216
0,05	-1,9420	-2,8807	-3,4403
0,1	-1,6168	-2,5769	-3,1444

8. Критические значения коинтеграционного ADF-критерия для уровней значимости 0,01; 0,05; 0,1; m=1 [5]

Уровень значимости	Тип функции			
	Без параметра смещения		С параметром смещения	
	ADF_l	ADF_u	ADF_l	ADF_u
T=20				
0,01	-4,43	-3,94	-4,73	-4,61
0,05	-3,63	-3,52	-3,85	-3,78
0,1	-3,25	-3,17	-3,44	-3,38
T=25				
0,01	-4,66	-4,17	-4,57	-4,46
0,05	-3,74	-3,63	-3,76	-3,69
0,1	-3,30	-3,23	-3,38	-3,32

Библиографический список

1. Айвазян С. А. Прикладная статистика. Основы эконометрики : учебник для вузов : В 2 т. / Айвазян С. А., Мхитарян В. С. ; С. А. Айвазян, В. С. Мхитарян. – 2-е изд., испр. – М. : ЮНИТИ, 2001. – Т. 1: Теория вероятностей и прикладная статистика. – 656 с.
2. Берндт, Э. Р. Практика эконометрики: классика и современность : учебник / Э. Р. Берндт. – М. : ЮНИТИ-ДАНА, 2005. – 863 с.
3. Джонстон, Дж. Эконометрические методы / Дж. Джонстон. – М. : Статистика, 1980. – 350 с.
4. Доугерти, К. Введение в эконометрику / К. Доугерти. – М. : ИНФРА-М, 1997. – 402 с.
5. Елисеева, И. И. Эконометрика : учебное пособие / И. И. Елисеева, С. В. Курышева, Д. М. Гордиенко и др. – М.: Финансы и статистика, 2001.
6. Замков, О.О. Эконометрические методы в макроэкономическом анализе : курс лекций / О. О. Замков. – М. : ГУ ВШЭ, 2001. – 122 с.
7. Кремер, Н. Ш. Эконометрика / Н. Ш. Кремер, Б. А. Путко. – М. : ЮНИТИ, 2005. – 311 с.
8. Магнус, Я. Р. Эконометрика. Начальный курс / Я. Р. Магнус, П. К. Катышев, А. А. Пересецкий. – М. : Дело, 1997. – 247с.
9. Практикум по эконометрике : учебное пособие / под ред. И. И. Елисеевой. – М. : Финансы и статистика, 2002. – 191с.
10. Тихомиров, Н. П. Эконометрика : учебник для вузов / Н. П. Тихомиров, Е. Ю. Дорохина ; Рос. экон. акад. им. Г. В. Плеханова. – М. : Экзамен, 2003. – 510 с.

Интернет-ресурсы

11. Айвазян С. А., Степанов Ц. С. Программное обеспечение по статистическому анализу данных: Методология сравнительного анализа и выборочный обзор рынка. – Режим доступа : <http://pubhealth.spb.ru/SAS/STatProg.htm>
12. Демоверсия пакета STADIA. – Режим доступа : <http://www.protein.bio.msu.ru/akula/index.htm>.
13. КВАНТИЛЬ международный эконометрический журнал на русском языке. http://uisrussia.msu.ru/docs/nov/quantile/2/quantile_2_index.htm.
14. Колеников С. О. Прикладной эконометрический анализ в статистическом пакете Stata6 в формате PDF. – Режим доступа : <http://www.komkon.org/~tacik/Stata6Ec.pdf>.
15. Математическое Бюро. Учебники по эконометрике и статистике. – Режим доступа : http://www.matburo.ru/st_subject.php?p=ec.
16. Образовательный математический сайт. – Режим доступа : <http://www.exponenta.ru>.
17. Программа для эконометрики Eviews, сайт разработчика. – Режим доступа : <http://www.eviews.com>.
18. Профессиональный статистический пакет Stata. – Режим доступа : <http://www.stata.com>.

19. Рассылка «Эконометрика». – Режим доступа : <http://subscribe.ru/archive/science.humanity.econometrika/>
20. Ресурсы по статистике и эконометрике. – Режим доступа : <http://dist-economics.eu.spb.ru/HTML/predmet/econometrics.htm>.
21. Ресурсы по статистике и эконометрике. – Режим доступа : <http://research.by/rus/links/>.
22. Ресурсы по статистике и эконометрике. – Режим доступа : <http://www.ecsocman.edu.ru/db/msg/163749.html>.
23. Сайт «Высокие статистические технологии». – Режим доступа : <http://orlovs.pp.ru>.
24. Сайт компании Manugistics Group – поставщика пакета. – Режим доступа : STATGRAPHICS PLUS. <http://www.manugistics.com>.
25. Сайт компании SPSS Rus. – Режим доступа : <http://www.spss.ru/>
26. Сайт компании разработчика пакета STATGRAPHICS. – Режим доступа : <http://www.statgraphics.com/>
27. Сайт фирмы Statsoft разработчика пакета STATISTICA. – Режим доступа : <http://www.statsoft.ru/>
28. Статистический Портал StatSoft. – Режим доступа : <http://www.statistica.ru/>
29. Упражнения по эконометрики в соответствии с учебником [8]. – Режим доступа : <http://econometrics.nes.ru/mkp/>
30. Эконометрика. Библиотека. Единое окно доступа к образовательным ресурсам. – Режим доступа : http://window.edu.ru/window/library?p_rubr=2.2.76.4.8
31. Эконометрическая программа Matrixer 5.1. А. Цыплаков (tsy@land4.nsu.ru). – Режим доступа : <http://www.nsu.ru/ef/tsy/ecmr/mtx/index.htm>.
32. Эконометрическая страничка НГУ. – Режим доступа : <http://www.nsu.ru/ef/tsy/ecmr/index.htm>.
33. Экономическая библиотека онлайн. – Режим доступа : <http://www.elobook.com/ekonometr/index.htm>
34. Электронные библиотеки России. Полнотекстовые pdf-учебники. – Режим доступа : http://www.gaudeamus.omskcity.com/PDF_library_economic_7.html
35. Электронный учебник по статистике. – М.: StatSoft, Inc, 2001. <http://www.statsoft.ru/home/download/textbook/default.htm>.

Учебное издание

ШАНЧЕНКО Николай Иванович

Лекции по эконометрике

Учебное пособие
для студентов высших учебных заведений, обучающихся
по специальности «Прикладная информатика (в экономике)»

Редактор Штаева М.

Компьютерная верстка Н. И. Шанченко

ЛР №020640 от 22.10.97.

Подписано в печать 28.12.2008. Формат 60×84/16.

Усл. печ. л. 8,14. Тираж 100 экз.

Ульяновский государственный технический университет
432027, г. Ульяновск, Северный Венец, 32.

Типография УлГТУ, 432027, Ульяновск, Северный Венец, 32.