

**Министерство науки и высшего образования Российской Федерации**  
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ  
**«Санкт-Петербургский государственный университет  
промышленных технологий и дизайна»**  
**Высшая школа технологии и энергетики**  
**Кафедра финансов и учета**

# **ЭКОНОМЕТРИКА**

Текст лекций для студентов всех форм обучения  
по направлению подготовки  
38.03.01 — Экономика

Составители:  
М. П. Власов  
О. А. Морозов

Санкт-Петербург  
2025

Утверждено  
на заседании кафедры ФиУ  
28.03.2025 г., протокол № 8

Рецензенты:  
М. Г. Трейман, Е. М. Фрейдкина

Текст лекций соответствует программам и учебным планам дисциплины «Эконометрика» для студентов, обучающихся по направлению подготовки 38.03.01 «Экономика». Текст лекций охватывает весь объем материалов по дисциплине. Издание предназначено для самостоятельной работы студентов. Текст лекций предназначен для бакалавров всех форм обучения.

Утверждено Редакционно-издательским советом ВШТЭ СПбГУПТД  
в качестве текстов лекций

Режим доступа: [http://publish.sutd.ru/tp\\_get\\_file.php?id=202016](http://publish.sutd.ru/tp_get_file.php?id=202016), по паролю.  
- Загл. с экрана.

Дата подписания к использованию 15.05.2025 г. Рег. № 5123/25

Высшая школа технологии и энергетики СПбГУПТД  
198095, СПб., ул. Ивана Черных, 4.

## ОГЛАВЛЕНИЕ

|                                                            |    |
|------------------------------------------------------------|----|
| ВВЕДЕНИЕ .....                                             | 5  |
| 1. ПРЕДМЕТ ЭКОНОМЕТРИКИ И ЕЕ МЕТОДОЛОГИЯ .....             | 7  |
| 1.1. Определение эконометрики .....                        | 7  |
| 1.2. Предмет эконометрики .....                            | 10 |
| 1.3. Метод эконометрики .....                              | 11 |
| 1.4. Измерения в эконометрике .....                        | 18 |
| 1.5. Определение и необходимость спецификации модели ..... | 20 |
| 2. МЕТОДЫ СТАТИСТИЧЕСКОГО АНАЛИЗА ДАННЫХ .....             | 25 |
| 2.1. Статистические методы .....                           | 25 |
| 2.2. Частотный анализ .....                                | 27 |
| 2.3. Анализ соответствий .....                             | 30 |
| 2.4. Методы сравнения средних .....                        | 33 |
| 2.5. Канонический анализ .....                             | 36 |
| 3. ЭКОНОМЕТРИЧЕСКИЕ МОДЕЛИ .....                           | 39 |
| 3.1. Определение эконометрической модели .....             | 39 |
| 3.2. Классификация эконометрических моделей .....          | 41 |
| 3.3. Процесс формирования эконометрической модели .....    | 46 |
| 3.4. Эконометрические модели дискретного выбора .....      | 49 |
| 3.4.1. Логистическая регрессия .....                       | 50 |
| 3.4.2. Пробит-регрессия .....                              | 52 |
| 3.4.3. Тобит-регрессия .....                               | 54 |
| 3.5. Линейные модели бинарного выбора .....                | 55 |
| 4. КЛАССИФИКАЦИЯ В ЭКОНОМЕТРИКЕ .....                      | 57 |
| 4.1. Процесс классификации .....                           | 57 |
| 4.2. Классификация многомерных наблюдений .....            | 60 |
| 4.3. Автоматическая классификация .....                    | 62 |
| 4.4. Дискриминантный анализ .....                          | 65 |
| 4.5. Таблица сопряженности и анализ соответствия .....     | 67 |
| 5. СПЕЦИФИКАЦИЯ ЭКОНОМЕТРИЧЕСКОЙ МОДЕЛИ .....              | 70 |
| 5.1. Функциональные и статистические зависимости .....     | 70 |
| 5.2. Выявление зависимости между показателями .....        | 75 |

|                                                                  |     |
|------------------------------------------------------------------|-----|
| 5.3. Спецификация эконометрических моделей .....                 | 82  |
| 6. МЕТОДЫ ПРОГНОЗИРОВАНИЯ В ЭКОНОМЕТРИКЕ .....                   | 92  |
| 6.1. Механизм прогнозирования.....                               | 92  |
| 6.2. Проблемы прогнозирования .....                              | 95  |
| 6.3. Классификация методов прогнозирования .....                 | 100 |
| 6.4. Регрессия в прогнозировании и планировании .....            | 105 |
| 6.5. Формы многофакторной корреляционной зависимости.....        | 110 |
| 6.6. Мультиколлинеарность данных .....                           | 112 |
| 6.7. Выбор наилучшей функции регрессии.....                      | 116 |
| 6.8. Временные ряды в прогнозировании.....                       | 117 |
| 7. СИСТЕМЫ ОДНОВРЕМЕННЫХ УРАВНЕНИЙ.....                          | 126 |
| 7.1. Определение системы одновременных уравнений .....           | 126 |
| 7.2. Классификация систем одновременных уравнений.....           | 132 |
| 7.3. Проблема идентификации систем одновременных уравнений ..... | 135 |
| 7.4. Метод максимального правдоподобия.....                      | 142 |
| 8. РЯДЫ ДИНАМИКИ, ОПРЕДЕЛЕНИЯ И ПОКАЗАТЕЛИ.....                  | 144 |
| 8.1. Временные ряды (ряды динамики), определения .....           | 144 |
| 8.2. Сопоставимость и смыкание рядов динамики .....              | 149 |
| 8.3. Показатели временного ряда .....                            | 153 |
| 8.4. Средние показатели рядов динамики .....                     | 158 |
| 8.5. Выявление основной тенденции ряда динамики.....             | 160 |
| 8.6. Выбор формы уравнения тренда .....                          | 167 |
| 8.7. Анализ периодических колебаний .....                        | 174 |
| БИБЛИОГРАФИЧЕСКИЙ СПИСОК .....                                   | 183 |

## ВВЕДЕНИЕ

Слово «эконометрика» является комбинацией двух терминов:

- «экономика»;
- «метрика» (от греч. «метрон» – правило определения расстояния между двумя точками в пространстве).

Наука «эконометрика» как «измерения в экономике» предназначена для построения математических моделей реальных экономических явлений и их анализа. Эконометрика призвана подтвердить или опровергнуть тот или иной экономический закон либо гипотезу. Эконометрика дает количественное выражение взаимосвязей экономических явлений и процессов. Эконометрика представляет синтез:

- экономической теории;
- статистики;
- математики.

Эконометрика заимствовала большинство методов и приемов математической статистики. В эконометрике используются статистические данные, которые обладают рядом особенностей:

- 1) данные не являются результатом контролируемого эксперимента;
- 2) отсутствует возможность проводить многократные эксперименты (из-за изменения условий, в которых проводятся измерения);
- 3) экономические данные часто содержат ошибки измерения.

В эконометрике применяются методы, позволяющие, если не устранить, то, по крайней мере, снизить влияние этих ошибок на полученные результаты. Эти особенности рожают ряд специфических проблем, решение которых не являются предметом математической статистики. Таким образом, эконометрика, предлагая математические модели, связывает экономическую теорию и экономическую статистику, придавая количественное выражение закономерностям экономической теории.

Предмет эконометрики – экономические явления, характеризующиеся экономической статистикой.

Объект эконометрики – количественное выражение экономических явлений в виде эконометрических моделей.

Основные задачи эконометрики:

1. Построение эконометрических моделей, т. е. представление экономических моделей в математической форме, удобной для проведения эмпирического анализа.
2. Оценка параметров построенной модели, делающих выбранную модель наиболее адекватной реальным данным.
3. Проверка качества найденных параметров модели и самой модели в целом.
4. Использование построенных моделей для объяснения поведения исследуемых экономических показателей, прогнозирования и предсказания, а также для осмысленного проведения экономической политики.

Таким образом, стандартная схема анализа зависимостей состоит в осуществлении ряда последовательных процедур:

1. Подбор начальной модели (этап спецификации). Он осуществляется на основе экономической теории, предыдущих знаний об объекте исследования, опыта исследователя и его интуиции.

2. Оценка параметров модели на основе имеющихся статистических данных (этап параметризации).

3. Осуществление проверки качества модели (этап верификации).

4. При наличии хотя бы одного неудовлетворительного ответа по какому-либо критерию модель совершенствуется с целью устранения выявленного недостатка.

5. При положительных ответах по всем критериям модель считается качественной.

Она используется для анализа и прогноза объясняемой переменной. Однако необходимо предостеречь от абсолютизации полученного результата, поскольку даже качественная модель является подгонкой спецификации модели под имеющийся набор данных. Поэтому вполне реальна картина, когда исследователи, обладающие разными наборами данных, строят разные модели для объяснения одного и того же явления. Проблематичным является и использование модели для прогнозирования значений объясняемой переменной. Иногда хорошие с точки зрения критериев проверки модели обладают весьма низкими прогнозными качествами.

# 1. ПРЕДМЕТ ЭКОНОМЕТРИКИ И ЕЕ МЕТОДОЛОГИЯ

## 1.1. Определение эконометрики

Понятие эконометрика было впервые использовано австро-венгерским бухгалтером П. Цемпой, затем в 1930 году было основано эконометрическое общество в ходе заседания Американской ассоциации развития науки, а норвежский исследователь Р. Фриш назвал новую науку «эконометрика». Эконометрика – это комбинация двух слов «экономики» и «метрики». Таким образом, термин выражает специфику и содержание эконометрики: количественное выражение соотношений и связей, раскрытых и обоснованных экономической теорией. Эконометрика – это область научного знания, выражающая количественно взаимосвязи экономических процессов и явлений. Цель эконометрики – придание количественной меры экономическим отношениям.

**Эконометрика** (эконометрия) (от экономики и греч. *metreo* – измеряю), научная дисциплина, позволяющая на базе положений экономической теории и результатов экономических измерений придавать конкретное количественное выражение общим (качественным) закономерностям. При этом основную роль в математическом оснащении этой дисциплины играют методы математической статистики, и в первую очередь, – многомерного статистического анализа.

Суть эконометрики в синтезе экономической теории, экономической статистики и прикладной математики. Использование эконометрики в экономической теории позволяет выявить объективно существующих (на качественном уровне) экономических законов и связей между экономическими показателями, но и подходами к их формализации при спецификации и идентификации соответствующих моделей с учетом их идентифицируемости. При рассмотрении экономической статистики как составной части эконометрики, прежде всего, интересует тот аспект, который непосредственно связан с информационным обеспечением анализируемой эконометрической модели, хотя в этих рамках зачастую приходится решать спектр соответствующих задач:

- выбор необходимых экономических показателей и обоснование способа их измерения;
- определение плана статистического исследования и т. п.

В качестве основной составляющей математический инструментарий эконометрики содержит ряд специальных разделов многомерного статистического анализа:

- линейные (классическая и обобщенная) и некоторые специальные модели регрессии, методы и модели анализа временных рядов;
- обобщенный метод моментов;
- так называемые системы одновременных уравнений;
- статистические методы классификации и снижения размерности анализируемого признакового пространства.

Однако эконометрика использует понятия, постановки и методы решения задач и из многих других разделов математики:

- теории вероятностей;
- математического программирования;
- численных методов решения задач линейной алгебры;
- систем нелинейных уравнений;
- теории нахождения неподвижных точек отображений.

При всей своей условности и неполноте приведенная схема эконометрического исследования дает общее представление об эконометрике и ее месте в ряду других экономических и статистических дисциплин. Именно использование экономической теории на базе экономической статистики для определения вполне определенных количественных взаимосвязей является ключевым моментом в понимании сущности эконометрики. Это позволяет разграничить эконометрику с такими дисциплинами, как математическая экономия, описательная экономическая статистика и математическая статистика. Так, математическая экономия, которая часто определяется как математически сформулированная экономическая теория, изучает взаимосвязи между экономическими переменными на общем (неколичественном) уровне. Она преобразуется в эконометрику, когда символически представленные в этих взаимосвязях коэффициенты заменяются конкретными численными оценками, полученными на базе соответствующих экономических данных.

Таким образом, призвание эконометрики состоит в создании моделей оценки (прогнозирования и планирования) состояния и поведения экономических объектов и процессов на основе статистических методов.

Эконометрические модели могут характеризоваться с точки зрения одного, двух и более показателей (статические модели), или изменения во времени (динамические модели, модели временных рядов).

В широком смысле эконометрика представляет совокупность экономических исследований, проводимых с использованием математических методов (рис. 1.1).

В узком смысле эконометрика – это использование статистических методов в экономических исследованиях, более конкретно – это построение математико-статистических моделей экономических процессов с оценкой параметров этих моделей.

Эконометрика является комбинацией трех научных областей: экономической статистики, экономической теории и математической статистики.

Таким образом, суть эконометрики – именно в синтезе экономической теории, экономической статистики и прикладного математического инструментария. Эконометрика интересуется не просто выявлением объективно существующих (на качественном уровне) экономических законов и связей между экономическими показателями, но и подходами к их формализации, включающими методы спецификации и идентификации соответствующих моделей с учетом решения проблемы их идентифицируемости.

При рассмотрении экономической статистики интересен аспект, который непосредственно связан с информационным обеспечением формируемой эконометрической модели, в этих рамках зачастую приходится решать полный



спектр вопросов: выбор необходимых экономических показателей и обоснование способа их измерения, определение плана статистического обследования и т. п.

Наконец, математической инструментарий эконометрики в качестве своей основной составляющей содержит ряд специальных разделов многомерного статистического анализа:

- линейные (классическая и обобщенная) и некоторые специальные модели регрессии;
- методы и модели анализа временных рядов;
- обобщенный метод моментов;
- системы одновременных уравнений;
- статистические методы классификации и снижения размерности анализируемого признакового пространства.



Рисунок 1.1 – Структура эконометрического исследования

Представленная схема (рис. 1.1) при всей своей условности и неполноте в целом дает общее наглядное представление об эконометрике и ее месте в ряду других экономических и статистических дисциплин.

Ключевыми моментами в понимании сущности эконометрики является моделирование закономерностей экономической теории на базе конкретной экономической статистики и извлечение с помощью подходящего математического аппарата количественных взаимосвязей являются.

## **1.2. Предмет эконометрики**

Из определения эконометрики следует, что предметом этой дисциплины являются экономические и социально-экономические приложения, а именно модельное описание конкретных количественных взаимосвязей, существующих между анализируемыми показателями.

К числу типовых экономических моделей, конструируемых и изучаемых с помощью эконометрических методов, относятся:

- прогнозирование и планирование величины спроса, цен, объема выпуска товаров;
- производственные функции, выражающие взаимосвязи между затратами и результатами производственной деятельности экономических систем различных уровней;
- прогнозирование и планирования доходов социальных групп;
- модели функционирования национальной экономики;
- типологизация объектов и поведения агентов (стран, регионов, предприятий, потребителей);
- целевые функции потребительского предпочтения и функции спроса;
- модели распределительных отношений в обществе;
- модели рынка и экономического равновесия;
- модели интернационализации национальных экономик;
- модели межстранового и межрегионального анализа и др.

При всем разнообразии спектра решаемых с помощью эконометрики задач их тем не менее было бы удобно расклассифицировать по трем направлениям:

- по конечным прикладным целям;
- по уровню иерархии;
- по профилю анализируемой экономической системы.

По конечным прикладным целям выделим:

а) прогнозирование и планирование экономических и социально-экономических показателей (переменных), характеризующих состояние и развитие анализируемой системы;

б) имитация различных возможных сценариев социально-экономического развития анализируемой системы, когда статистически выявленные взаимосвязи между характеристиками производства, потребления, социальной и финансовой политики.

Сценарный или ситуационный анализ проводится для отслеживания возможных изменений параметров производства или распределения на значениях «выходных» характеристик».

По уровню иерархии анализируемой экономической системы выделяются макроуровень (т. е. страны в целом), мезоуровень (регионы, отрасли, корпорации) и микроуровень (домашние хозяйства, предприятия, фирмы).

В некоторых случаях профиль эконометрического моделирования может быть сконцентрирован на проблемах рынка, инвестиционной, финансовой или социальной политики, ценообразования, распределительных отношений, спроса и потребления, или на определенном комплексе проблем. Однако, чем шире охват анализируемых проблем эконометрическое исследование, тем меньше шансов провести его достаточно эффективно.

Большое внимание в эконометрике уделяется проблеме данных – специальным методам работы при наличии данных с пропусками, влиянию агрегирования данных на эконометрические измерения. Информация может отсутствовать по единицам совокупности и быть только на уровне более крупных единиц (агрегатов). Например, не по отдельным организациям, а по организациям в пределах административного района, т. е. по районам, и т. д. При агрегировании данных во времени опасность искажения результатов измерений (скажем, корреляции между временными рядами) гораздо больше, чем при агрегировании пространственных данных.

Эконометрическая модель, как правило, основана на предположении о взаимосвязи переменных и характере связи между ними. При всем стремлении к «наилучшему» описанию связей приоритет отдается качественному анализу. Поэтому в качестве этапов эконометрического исследования можно указать:

- формулировку проблемы;
- анализ возможности получения данных и их получение, анализ их качества;
- спецификацию модели;
- оценку параметров модели;
- интерпретацию результатов расчетов по модели.

Этот список менее подробен, чем предыдущий, и включает те стадии, которые проходит любое исследование, независимо от того, на использование каких данных оно ориентировано: пространственных или временных.

### **1.3. Метод эконометрики**

Эконометрика как система специфических методов начала развиваться с осознания своих задач – отражения особенностей экономических переменных и связей между ними. Эконометрический метод может быть описан следующим образом:

- постулируется, что анализируемые переменные (экономические показатели)  $x^{(1)}, x^{(2)}, \dots, x^{(p)}$  являются случайными величинами, совместный закон распределения вероятностей которых не известен исследователю, но принадлежит некоторому семейству функций;

- в процессе функционирования анализируемой экономической системы генерируются наблюдаемые значения  $x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(p)}$  ( $i = 1, 2, \dots, n$ ) интересующих переменных;

- идентификация модели (анализируемой системы) заключается в выборе из упомянутого семейства конкретного закона распределения вероятностей, наиболее хорошо (в определенном смысле) согласующегося с имеющимися в распоряжении исследователя сгенерированными системой данными;

- различные спецификации (конкретизации, основанные на дополнительных исходных допущениях) такой постановки проблемы приводят к широкому спектру методов и моделей эконометрического анализа: регрессии, временным рядам, системам одновременных уравнений и другим методам, используемым при решении задач экономического прогноза, ситуационного анализа, оценивания важных экономических характеристик.

Все эконометрические модели, независимо от того, относятся они ко всему хозяйству или к его элементам (т. е. к макроэкономике, отрасли, предприятию или рынку), имеют некоторые общие особенности:

1) основаны на предположении, что поведение экономических переменных определяется с помощью совместных и одновременных операций с некоторым числом экономических соотношений.

2) принимается гипотеза, в силу которой модель, допуская упрощение сложной действительности, тем не менее, улавливает главные характеристики изучаемого объекта.

3) создатель модели полагает, что на основе достигнутого с ее помощью понимания реальной системы удастся предсказать ее будущее движение и, возможно, управлять им в целях улучшения экономического благосостояния.

*Пример.* Предположим, что экономическая теория позволяет сформулировать следующие положения:

- потребление есть возрастающая функция от имеющегося в наличии дохода, но возрастающая, видимо, медленнее, чем рост дохода;

- объем инвестиций есть возрастающая функция национального дохода и убывающая функция некоторых характеристик государственного регулирования (например, нормы процента);

- национальный доход есть сумма потребительских, инвестиционных и государственных закупок товаров и услуг.

Первая задача – перевести эти положения на математический язык. Здесь открывается многообразие возможных решений, удовлетворяющих сформулированным априорным требованиям теории. Какие соотношения выбрать между переменными – линейные или нелинейные? Если остановиться на нелинейных, то какими они должны быть – логарифмическими, полиномиальными или какими-либо еще? Даже после определения формы конкретного соотношения, остается еще нерешенной проблема выбора для различных уравнений запаздываний по времени. Будут ли, например, инвестиции текущего периода реагировать только на национальный доход, произведенный в последнем периоде, или же на них скажется динамика

нескольких предыдущих периодов? Обычный выход из этих трудностей состоит в выборе при первоначальном анализе наиболее простой из возможных форм этих соотношений. Тогда появляется возможность записать на основе указанных выше положений следующую линейную относительно анализируемых переменных и аддитивную относительно случайных составляющих модель:

$$y_t^{(1)} = \alpha_0 + \alpha_1(y_t^{(3)} - x_t^{(1)}) + \varepsilon_t^{(1)}, \quad (1.1)$$

$$y_t^{(2)} = \beta_1 y_{t-1}^{(3)} + \beta_2 x_t^{(2)} + \varepsilon_t^{(2)}, \quad (1.2)$$

$$y_t^{(3)} = y_t^{(1)} + y_t^{(2)} + x_t^{(3)}, \quad (1.3)$$

где априорные ограничения выражены неравенствами:

$$0 < \alpha_1 < 1; \beta_1 > 0; \beta_2 < 0.$$

Эти три соотношения вместе с ограничениями образуют модель:

$y_t^{(1)}$  – потребление;

$y_t^{(2)}$  – инвестиции;

$y_t^{(3)}$  – национальный доход;

$x_t^{(1)}$  – подоходный налог;

$x_t^{(2)}$  – норма процента как инструмент государственного регулирования;

$x_t^{(3)}$  – государственные закупки товаров и услуг, измеренные в «момент времени»  $t$ .

Присутствие в уравнениях (1.1) и (1.2) «остаточных» случайных составляющих  $\varepsilon_t^{(1)}$  и  $\varepsilon_t^{(2)}$  обусловлено необходимостью учесть влияние соответственно на  $(y_t^{(1)})$  и  $(y_t^{(2)})$  ряда неучтенных факторов. Действительно, нереалистично ожидать, что величина потребления  $(y_t^{(1)})$  будет однозначно определяться уровнями национального дохода  $(y_t^{(3)})$  и подоходного налога  $(x_t^{(1)})$ ; аналогично величина инвестиций  $(y_t^{(2)})$  зависит, очевидно, не только от достигнутого в предыдущий год уровня национального дохода  $(y_{t-1}^{(3)})$  и от величины нормы процента  $(x_t^{(2)})$ , но и от ряда не учтенных в уравнении (1.3) факторов.

Полученная модель содержит два уравнения, объясняющих поведение потребителей и инвесторов, и одно тождество. Сформулировали ее для дискретных периодов времени и выбрали запаздывание (лаг) в один период для отражения воздействия национального дохода на инвестиции.

В дальнейшем этот пример используется для пояснения ряда основных понятий эконометрического моделирования.

**Основные понятия** эконометрического моделирования. В любой эконометрической модели в зависимости от конечных прикладных целей ее использования все участвующие в ней переменные подразделяются на:

- **экзогенные**, т. е. задаваемые как бы «извне», автономно, в определенной степени управляемые (планируемые);
- **эндогенные**, т. е. такие переменные, значения которых формируются в процессе и внутри функционирования анализируемой социально-

экономической системы в существенной мере под воздействием экзогенных переменных и, конечно, во взаимодействии друг с другом; в эконометрической модели они являются предметом объяснения;

- **предопределенные**, т. е. выступающие в системе в роли факторов-аргументов, или объясняющих переменных.

Множество предопределенных переменных формируется из всех экзогенных переменных (которые могут быть «привязаны» к прошлым, текущему или будущим моментам времени) и так называемых лаговых эндогенных переменных, т. е. таких эндогенных переменных, значения которых входят в уравнения анализируемой эконометрической системы измеренными в прошлые (по отношению к текущему) моменты времени, а, следовательно, являются уже известными, заданными.

Набор взаимосвязанных регрессионных уравнений, в которых одни и те же переменные могут одновременно играть роль (в различных уравнениях системы) результирующих показателей и объясняющих переменных (предикторов) называют системой одновременных уравнений (СОУ). Модель (1.1)–(1.3) представляет пример СОУ. В данном примере потребление ( $y_t^{(1)}$ ), инвестиции ( $y_t^{(2)}$ ) и национальный доход ( $y_t^{(3)}$ ) в текущий момент времени  $t$  являются эндогенными переменными; подоходный налог ( $x_t^{(1)}$ ), норма процента как инструмент государственного регулирования ( $x_t^{(2)}$ ) и государственные закупки товаров и услуг ( $x_t^{(3)}$ ) – экзогенные переменные, которые вместе с национальным доходом в предшествующий момент времени ( $y_{t-1}^{(3)}$ ) образуют множество предопределенных переменных.

Таким образом, можно сказать, что эконометрическая модель служит для объяснения поведения эндогенных переменных в зависимости от значений экзогенных и лаговых эндогенных переменных.

При построении и анализе эконометрической модели следует различать ее структурную и приведенную формы. Для пояснения этих понятий условимся в дальнейшем обозначать латинской буквой  $X$  вектор-столбец всех предопределенных переменных (он включает в себя все экзогенные переменные и все участвующие в модели лаговые эндогенные переменные). Пусть общее число эндогенных переменных равно  $m$ , а общее число предопределенных переменных –  $p$ . Общее число уравнений и тождеств в эконометрической модели равно числу эндогенных переменных, т. е. равно  $m$ . И пусть из общего числа соотношений модели имеется  $m_1$  уравнений, включающих случайные остаточные компоненты, и  $m_2$  тождеств ( $m_1 + m_2 = m$ ). Разобьем вектор эндогенных переменных  $Y_t = (y_t^{(1)}, y_t^{(2)}, \dots, y_t^{(m)})^T$  на два подвектора  $Y_t^1 = (y_t^{(1)}, y_t^{(2)}, \dots, y_t^{(m_1)})^T$  и  $Y_t^2 = (y_t^{(m_1+1)}, y_t^{(m_1+2)}, \dots, y_t^{(m_1+m_2)})^T$ , при этом порядок, в котором перенумерованы эндогенные переменные, не имеет значения.

Тогда общий вид линейной эконометрической модели может быть представлен в форме:

$$\begin{cases} B_1 Y_t^{(1)} + B_2 Y_t^{(2)} + C_1 X_t = \Delta_t, \\ B_3 Y_t^{(1)} + B_4 Y_t^{(2)} + C_2 X_t = 0, \\ t = 1, 2, \dots, n, \end{cases} \quad (1.4)$$

где  $B_1 = (\beta_{i,j})_{i=\overline{1,m_1}, j=\overline{1,m_1}}$  – матрица размерности  $(m_1 \times m_1)$  из коэффициентов при  $y_t^{(1)}, y_t^{(2)}, \dots, y_t^{(m_1)}$  в  $m_1$  первых уравнениях;

$B_2 = (\beta_{i,j})_{i=\overline{1,m_1}, j=m_1+1, m_1+m_2}$  – матрица из коэффициентов при  $y_t^{(m_1+1)}, \dots, y_t^{(m_1+m_2)}$  в  $m_1$  первых уравнениях;

$X_t = (x_t^{(0)}, x_t^{(1)}, x_t^{(2)}, \dots, x_t^{(p)})^T$  – вектор-столбец предопределенных переменных (в нем  $x_t^{(0)} \equiv 1$ );

$C_1 = (c_{i,j})_{i=\overline{1,m_1}, j=0,p}$  – матрица размерности  $m_1 \times (p+1)$  из коэффициентов при предопределенных переменных в первых  $m_1$  уравнениях (очевидно, коэффициенты  $c_{i,0}$  играют роль свободных членов уравнений);

$B_3 = (\beta_{i,j})_{i=m_1+1, m_1+m_2, j=\overline{1,m_1}}$  – матрица размерности  $m_2 \times m_1$  из коэффициентов при  $y_t^{(1)}, l, y_t^{(m_1)}$  в  $m_2$  тождествах системы;

$B_4 = (\beta_{i,j})_{i=\overline{m_1+1, m_1+m_2}, j=m_1+1, m_1+m_2}$  – матрица размерности  $m_2 \times m_1$  из коэффициентов при  $y_t^{(m_1+1)}, l, y_t^{(m_1+m_2)}$  в  $m_2$  тождествах системы;

$C_2 = (c_{i,j})_{i=m_1+1, m_1+m_2, j=0,p}$  – матрица размерности  $m_2 \times (p+1)$  из коэффициентов при предопределенных переменных в  $m_2$  тождествах системы;

$\Delta_t = (\varepsilon_t^{(1)}, \varepsilon_t^{(2)}, \dots, \varepsilon_t^{(m_1)})^T$  – вектор-столбец размерности  $m_1$  случайных остаточных составляющих  $m_1$  первых уравнений системы;

$0_{m_2} = (0, 0, \dots, 0)^T$  – вектор-столбец размерности  $m_2$  состоящий из нулей.

Заметим, что исходными статистическими данными, необходимыми для проведения статистического анализа системы (1.4) (а именно, для оценки неизвестных коэффициентов  $\beta_{i,j}$  и  $c_{i,j}$  проверки статистических гипотез, например, о линейном характере исследуемых зависимостей и т. п.), являются матрицы:

$$Y = \begin{pmatrix} Y_1^T \\ \dots \\ Y_n^T \end{pmatrix} \text{ и } X = \begin{pmatrix} X_1^T \\ \dots \\ X_n^T \end{pmatrix} \quad (1.5)$$

соответственно размерностей  $n \times m$  и  $n \times (p+1)$ , а все элементы матриц  $B_3$ ,  $B_4$  и  $C_2$  являются известными (их числовые значения определяются содержательным смыслом соответствующих тождеств системы).

Система (1.4) может быть записана также в виде:

$$BY_t + CX_t = \bar{\Delta}_t, \quad t = 1, 2, \dots, n, \quad (1.6')$$

или в виде

$$YB^T + XC^T = \bar{\Delta}, \quad (1.6'')$$

где

$$Y_t = \begin{pmatrix} Y_t^{(1)} \\ Y_t^{(2)} \end{pmatrix}, B = \begin{pmatrix} B_1 & B_2 \\ B_3 & B_4 \end{pmatrix}, C = \begin{pmatrix} C_1 \\ C_2 \end{pmatrix}, \bar{\Delta}_t = \begin{pmatrix} \Delta_t \\ 0_{m_2} \end{pmatrix}, \bar{\Delta} = \begin{pmatrix} \bar{\Delta}_1^T \\ \vdots \\ \bar{\Delta}_n^T \end{pmatrix},$$

а матрицы  $U$  и  $X$  определены в (1.5).

Система уравнений и тождеств вида (1.4) (или эквивалентных ей записей (1.6') или (1.6'')) называется структурной формой линейной эконометрической модели. При этом предполагается, что коэффициент при эндогенной переменной  $i$  в структурном стохастическом уравнении  $i$  ( $i = 1, 2, \dots, m$ ) равен единице (правило нормировки системы), а матрицы  $B_4$  и  $B$  не вырождены (*Примечание: Невырожденная матрица (иначе неособенная матрица) – квадратная матрица, определитель которой отличен от нуля*) (допускаются и другие способы нормировки системы).

Поскольку при реализации конечных прикладных целей эконометрического моделирования (т. е. при прогнозе значений эндогенных переменных и при различных имитационных расчетах) главный интерес представляют соотношения, позволяющие явно выразить все эндогенные переменные  $Y_t$  через predetermined  $X_t$ , то одновременно со структурной формой имеет смысл рассмотреть так называемую приведенную (редуцированную) форму линейной эконометрической модели. Требуемый результат получим, домножив слева обе части соотношений (1.6') на матрицу  $B^{-1}$  и уединив затем  $Y_t$ :

$$Y_t = -B^{-1}CX_t + B^{-1}\bar{\Delta}_t, t = 1, 2, \dots, n, \quad (1.7)$$

или

$$Y_t = \Pi X_t + \tilde{\varepsilon}_t, t = 1, 2, \dots, n, \quad (1.7')$$

где  $m(p+1)$  матрица  $\Pi$  и вектор остаточных случайных составляющих  $\tilde{\varepsilon}_t$  определяются соотношениями:

$$\Pi = -B^{-1}C, \quad (1.8)$$

$$\tilde{\varepsilon}_t = B^{-1}\bar{\Delta}_t, \quad (1.9)$$

Система соотношений (1.7), в которой все эндогенные переменные эконометрической модели явно линейно выражены через predetermined переменные и случайные остаточные компоненты, называется приведенной формой линейной эконометрической модели.

Проиллюстрируем введенные понятия на примере (1.1) – это (1.3).

В этом примере число эндогенных переменных, так же, как и общее число всех соотношений модели, равно трем ( $m = 3$ ). Среди этих соотношений мы имеем одно тождество (следовательно,  $m_1 = 2$ ,  $m_2 = 1$ ). Общее число predetermined переменных  $p = 4$ , в том числе три экзогенные переменные ( $x_t^{(1)}, x_t^{(2)}, x_t^{(3)}$ ) и одна лаговая эндогенная переменная ( $y_{t-1}^{(3)}$ ), которую в соответствии с принятой договоренностью кодируем как  $x_t^{(4)}$  (т. е.  $y_{t-1}^{(3)} \equiv x_t^{(4)}$ ).

Структурная форма модели в данном примере задается соотношениями (1.1)–(1.3). В общих матричных обозначениях, использованных в (1.4), имеем:



$$B_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, B_2 = \begin{pmatrix} -\alpha_1 \\ 0 \end{pmatrix}, C_1 = \begin{pmatrix} -\alpha_0 & -\alpha_1 & 0 & 0 & 0 \\ 0 & 0 & -\beta_2 & 0 & 0 \end{pmatrix},$$

$$B_3 = (-1; -1), B_4 = (1), C_2 = (0, 0, 0, -1, 0)$$

и

$$\Delta_t = (\varepsilon_t^{(1)}, \varepsilon_t^{(2)})^T.$$

Если же структурная форма записана в виде (1.6'), то в данном примере участвующие в этой записи матрицы конкретизируются в виде:

$$B = \begin{pmatrix} 1 & 0 & -\alpha \\ 0 & 1 & 0 \\ -1 & -1 & 1 \end{pmatrix}; C = \begin{pmatrix} \alpha_0 & -\alpha_1 & 0 & 0 & 0 \\ 0 & 0 & -\beta_2 & 0 & -\beta_1 \\ 0 & 0 & 0 & -1 & 0 \end{pmatrix}.$$

и

$$\bar{\Delta}_t = (\varepsilon_t^{(1)}, \varepsilon_t^{(2)}, 0)^T.$$

Отметим, что, во-первых, выполнено условие нормировки ( $y_t^{(i)}$  входит в уравнение  $i$  системы,  $i = 1, 2$ , с коэффициентом единица); во-вторых, значения элементов матриц  $B_3$ ,  $B_4$  и  $C_2$  известны, они определяются содержательным смыслом тождества; в третьих, требование невырожденности матриц  $B_4$  и  $B$  соблюдено; и, наконец, в четвертых, матрицы  $B_1, B_2$  и  $C_1$  относительно «слабо заполнены» неизвестными (подлежащими статистическому оцениванию) коэффициентами: их всего четыре  $-\alpha_0, \alpha_1, \beta_1$  и  $\beta_2$ . Последняя особенность рассматриваемой эконометрической модели является достаточно общей отличительной чертой систем эконометрических уравнений. Если бы это было не так, т. е. если бы мы были вынуждены иметь дело с системами, «сильно заполненными» неизвестными коэффициентами, то задача статистического анализа таких систем оказывалась бы принципиально неразрешимой: имеющихся исходных статистических данных просто не хватало бы для корректного проведения такого анализа. Ведь при построении и анализе систем эконометрических уравнений, описывающих макроэкономические модели, исследователю зачастую приходится иметь дело с десятками и сотнями эндогенных и экзогенных переменных.

Приведенная форма модели (1.1)–(1.3) в данном примере имеет вид:

$$\varepsilon_t^{(1)} = \frac{1}{1-\alpha_1} (\alpha_0 - \alpha_1 x_t^{(1)} + \alpha_1 \beta_2 x_t^{(2)} + \alpha_1 x_t^{(3)} + \alpha_1 \beta_1 x_t^{(4)}) + \left( \varepsilon_t^{(1)} + \frac{\alpha_1}{1-\alpha_1} \varepsilon_t^{(2)} \right),$$

$$y_t^{(2)} = \beta_2 x_t^{(2)} + \beta_1 x_t^{(4)} + \varepsilon_t^{(2)},$$

$$y_t^{(3)} = \frac{1}{1-\alpha_1} (\alpha_0 - \alpha_1 x_t^{(1)} + \beta_2 x_t^{(2)} + \alpha_1 x_t^{(3)} + \alpha_1 \beta_1 x_t^{(4)}) + \left( \varepsilon_t^{(1)} + \frac{1}{1-\alpha_1} \varepsilon_t^{(2)} \right).$$

#### 1.4. Измерения в эконометрике

Понятие «эконометрика» включает экономические измерения, то есть получение, сравнение и упорядочение информации. Это определение исходит из того, что измерение предполагает выделение некоторого свойства, по которому производится сравнение объектов в определенном отношении.

Эконометрика использует конкретные экономические данные и количественно описывает конкретные взаимосвязи, другими словами, заменяет коэффициенты в общем виде конкретными значениями.

Многие эконометрические методы и приемы заимствованы из математической статистики. Между тем, математико-статистические методы являются универсальными и учитывают специфику экономических данных.

К особенностям измерений в экономике можно отнести следующее:

- субъективный фактор участвует в каждом отдельном экономическом акте;
- в распоряжении исследователя не всегда имеется полная информация;
- нельзя провести контролируемый эксперимент в отличие от химии, физики или биологии. В экономической науке не представляется возможным проведение многократных экспериментов, поскольку постоянно изменяются внешние условия. Эта особенность порождает другие специфические проблемы, решение которых не связано с математической статистикой;
- отсутствует единая мера, эталон соизмерения экономических величин.

Измерение понимается по-разному. Прежде всего признаками измерения называют получение, сравнение и упорядочение информации. Это определение измерения в широком смысле. В нем подчеркивается, что измерение предполагает выделение некоторого свойства, по которому проводится сравнение объектов в определенном отношении.

Все понятия измерения могут быть объединены на базе определения шкалы измерения. Тип шкалы определяется допустимым преобразованием.

Допустимое преобразование – это преобразование, при котором отношения между элементами системы сохраняются неизменными: истинные утверждения не становятся ложными, а ложные – истинными.

Поскольку понятие «эконометрика» включает экономические измерения, остановимся подробнее на этом вопросе.

Другое понимание измерения исходит из числового выражения результата, т. е. измерение трактуется как операция, в результате которой получается численное значение величины, причем числа должны соответствовать наблюдаемым свойствам, фактам, качествам, законам науки и т. д.

Первый, низший уровень измерения предполагает сравнение объектов по наличию или по отсутствию исследуемого свойства. На этом уровне измерения используются термины «номинация», «классификация», «нумерация».

Второй уровень предполагает сравнение объектов по интенсивности проявляемых свойств. На этом уровне используются термины «шкалирование», «топология», «упорядочение».

Третий, высший уровень измерения предполагает сравнение объектов с эталоном (в контексте физического измерения). На этом уровне используются

термины «измерение», «квантификация». Такой подход к измерению связан с обязательным наличием единицы измерения (эталоны). Это определение измерения в узком смысле. Для определения любой шкалы измерения необходимо дать название объекта, отождествить объект с некоторым свойством или группой свойств (предприятие промышленное, станок токарный, девушка сероглазая, автомобиль легковой и т. д.). Если это требование оказывается единственным, то шкала называется шкалой наименований или номинальной шкалой.

Измерением в номинальной шкале можно считать любую классификацию, по которой класс получает числовое наименование (например, номер научной или учебной специальности и т. д.).

Шкала, в которой порядок элементов по уровню проявления некоторого свойства существенен, а количественное выражение различия несущественно или плохо осуществимо, называется порядковой, или ранговой. Шкала порядка, или ординальная шкала, допускает операции «равенство-неравенство», «больше-меньше».

Кроме номинальной и порядковой шкал для определения измерения используются интервальные шкалы.

Измерения в интервальных шкалах в известном смысле более совершенны, чем в порядковых. Применение этих шкал дает возможность не только упорядочить объекты по количеству свойства, но и сравнить между собой разности количеств.

Формально интервальная шкала определяется как единственная до линейного преобразования шкала вида:

$$y = ax + b,$$

где  $a$  и  $b$  – числа, для которых определены операции сложения и умножения, соответственно  $a > 0$ ,  $b > 0$ .

Параметр  $a$  называется масштабом, а параметр  $b$  – началом отсчета.

В случаях, когда на шкале можно указать абсолютный нуль, мы имеем несколько более высокий уровень измерения, а именно шкалу отношений (или пропорциональную шкалу). При измерении на такой шкале можно, например, сделать вывод, что  $x_4$  вдвое больше  $x_2$ , если, например,  $x_4 = 40k$ , а  $x_2 = 20k$ .

Шкала отношений – это единственная с точностью до линейных преобразований шкала вида:

$$y = ax,$$

при  $a > 0$ , где  $a$  – масштаб.

Шкала разностей допускает операции «равенство-неравенство», «больше-меньше», «равенство-неравенство интервалов» и операцию вычитания, на основе которой устанавливается величина интервала в фиксированном масштабе. К шкале разностей относятся логарифмические шкалы, а также процентные и аналогичные им шкалы измерений, задающие безразмерные величины. Например, указание года рождения – это представление возраста в шкале разностей.

Шкала разностей существенна с точностью до линейного преобразования вида:

$$y = x + b.$$

Таким образом, в определении шкал участвуют понятия равенства, порядка, дистанции между пунктами шкалы (интервалы), начала отсчета и единицы измерения. В зависимости от наличия или отсутствия этих элементов возникают различные типы шкал.

Кроме классификации информации по типам шкал различают также перекрестные (пространственные) данные и временные ряды. В первом случае имеют дело с информацией, собранной для одного такта времени по нескольким объектам, охарактеризованным по нескольким признакам (например, работники фирмы, охарактеризованные по их профессиональным качествам). В случае же временных рядов имеем информацию по одному показателю в течение нескольких периодов времени (например, зарплата работника фирмы за год по неделям).

При работе с пространственной информацией, как правило, имеют дело со случайной выборкой, взятой из генеральной совокупности, поэтому здесь возможно применение теории оценивания, разработанной в курсе математической статистики.

В дальнейшем числовые характеристики генеральной совокупности будем называть параметрами, а их аналоги, вычисленные на основе выборки – выборочными характеристиками.

Основное свойство выборки – это ее случайность, поэтому все выборочные характеристики – случайные числа, подчиняющиеся тому или иному закону распределения.

### **1.5. Определение и необходимость спецификации модели**

Методы эконометрики позволяют выявлять новые, ранее не известные связи между экономическими показателями, а также уточнять или отвергать гипотезы о существовании таких связей.

Эконометрическая модель, таким образом, представляет формализованную математическую модель взаимосвязи между экономическими показателями, построенную на основе использования статистических методов.

Можно сказать, что она представляет частный случай экономико-математической модели, т. е. экономическую модель, представленную в математической форме.

Первый вопрос, который решают при построении эконометрической модели, – это вопрос спецификации.

Спецификация модели – это математическая форма записи уравнения зависимости результирующей переменной от одного или нескольких факторов. По сути, она представляет отбор факторов, включаемых в модель, и выбор вида уравнений.

В зависимости от спецификации эконометрических моделей классифицируют на несколько типов. Соответственно, основных классификационных признаков, связанных с классификацией, два: количество включенных факторов (т. е. экономических показателей) и тип зависимости между показателями в уравнении регрессии (корреляции). Кроме того, эконометрические модели классифицируют по типу исходных данных.

Таким образом, спецификация модели [specification of a model] – один из этапов построения экономико-математической модели, на котором, на основании предварительного анализа рассматриваемого экономического объекта или процесса, в математической форме выражаются обнаруженные связи и соотношения, а значит параметры и переменные, которые на данном этапе представляются существенными для цели исследования.

Эта проблема включает:

- а) определение конечных целей моделирования (прогноз, имитация различных сценариев социально-экономического развития анализируемой системы, оценка определенных экономических характеристик);
- б) определение списка экзогенных и эндогенных переменных;
- в) определение состава анализируемой системы уравнений и тождеств, их структуры и соответственно списка предопределенных переменных;
- г) способ параметризации модели, т. е. определение общего вида искомых функциональных зависимостей, связывающих между собой анализируемые переменные;
- д) формулировку исходных предпосылок и априорных ограничений относительно:

- стохастической природы остатков  $\Delta_t$  (в классических вариантах моделей постулируются их взаимная статистическая независимость или некоррелированность, нулевые значения их средних величин и, иногда, сохранение постоянными в процессе наблюдения значений их дисперсий – гомоскедастичность);
- числовых значений отдельных параметров модели.

Итак, спецификация модели – это первый и, быть может, важнейший шаг эконометрического исследования. От того насколько удачно решена проблема спецификации и, в частности, насколько реалистичны наши решения и предположения относительно состава эндогенных, экзогенных и предопределенных переменных, структуры и общего вида самой системы уравнений и тождеств, стохастической природы случайных остатков и конкретных числовых значений части неизвестных параметров модели, решающим образом зависит успех всего эконометрического исследования.

Эконометрический анализ как правило проводится в шесть этапов:

1. На первом этапе (постановочном) формируются цели исследования, определяются участвующие в модели экономические переменные. Прежде чем выбрать экономические переменные, необходимо каждую из них теоретически обосновать. Объясняющие переменные не должны находиться в тесной корреляционной или функциональной зависимости, поскольку это может вызвать мультиколлинеарность (*Примечание: Мультиколлинеарность (multicollinearity)*

– в эконометрике – наличие линейной зависимости между объясняющими переменными (факторами) регрессионной модели. При этом различают полную коллинеарность, которая означает наличие функциональной (тождественной) линейной зависимости и частичную или просто мультиколлинеарность – наличие сильной корреляции между факторами).

2. На втором этапе (априорном) анализируется сущность изучаемого объекта, формируется и формализуется априорная, т. е. известная до начала моделирования, информация.

3. На третьем этапе (параметризации) осуществляется моделирование, т. е. определяется общий вид модели, выявляются входящие в нее связи. Основной задачей данного этапа является выбор типа функции  $f(x)$  для эконометрической модели. Правильное решение проблемы спецификации модели позволяет достигать наиболее успешного эконометрического моделирования.

4. На четвертом этапе (информационном) собирается необходимая статистическая информация – наблюдаемые значения экономических переменных.

5. На пятом этапе (идентификации модели) проводится статистический анализ исследуемой модели и оцениваются ее параметры.

На заключительном, 6 этапе верификации модели проверяется ее истинность и адекватность. Определяется точность расчетов для данной модели, степень соответствия построенной модели моделируемому экономическому процессу или объекту. При наличии статистических данных, характеризующих моделируемый экономический процесс в предшествующий или данный момент времени, чтобы верифицировать модель, построенную для прогноза, необходимо провести сравнение реальных значений переменных в последующий момент времени со значениями, которые получены на основании рассматриваемой модели.

Спецификация опирается как на имеющиеся экономические теории, специальные знания или интуитивные представления исследователя об анализируемой экономической системе, так и на специальные методы и приемы (в том числе, математико-статистические) так называемого разведочного анализа.

Уравнение модели имеет структурную форму, если оно содержит более одной эндогенной переменной. Уравнение модели имеет приведенную форму, если оно содержит только одну эндогенную переменную.

Форма модели в виде системы нескольких уравнений считается структурной, если хотя бы одно из уравнений представлено в структурном виде. На этапе спецификации модели из нескольких уравнений, как правило, имеют структурную форму.

Модели в виде изолированного уравнения всегда имеет приведенную форму. Если исследователя интересуют только параметры приведенной формы и задача прогноза эндогенных переменных, то он может ограничиться применением обычного метода наименьших квадратов к каждому отдельному уравнению приведенной формы (с последующей оценкой, если это необходимо, идентифицируемых параметров структурной формы).

Структурная и приведенная формы модели – это две различные формы записи одной модели.

Уравнение структурной формы эконометрической модели называется точно идентифицируемым, если все участвующие в нем неизвестные (т. е. априори не заданные) коэффициенты однозначно восстанавливаются по коэффициентам приведенной формы без каких-либо ограничений на значения последних.

Различают два способа исследования массовых экономических явлений в зависимости от полноты охвата объекта: сплошное и несплошное. Разновидностью несплошного наблюдения является выборочное.

Выборочное наблюдение (выборочное исследование) заключается в обследовании определенного числа единиц совокупности, отобранного, как правило, случайным образом. При выборочном методе обследованию подлежит сравнительно небольшая часть всей изучаемой совокупности (обычно до 5–10 %, реже до 15–20 %). Выборочное наблюдение ставит перед собой задачу: по обследуемой части дать характеристику всей совокупности единиц. Отбор единиц из генеральной совокупности производится таким образом, чтобы выборочная совокупность была представительна (репрезентативна) и характеризовала генеральную совокупность. Степень представительности выборки зависит от способа организации выборки и от ее объема. Полной репрезентативности выборки достичь не удастся. Поэтому необходима оценка надежности результатов выборки и возможности их распространения на генеральную совокупность.

В зависимости от характеристик выборочных совокупностей выборки могут быть представительными, расслоенными, засоренными и цензурированными:

а) Представительная выборка – выборка наблюдений из генеральной совокупности, наиболее полно и адекватно представляющая ее свойства.

б) Расслоенная выборка – выборка, включающая ряд выборочных совокупностей, взятых из соответствующих слоев генеральной совокупности. Широко используется при выборочном обследовании в экономике, демографии и социологии.

в) Засоренная выборка – выборка наблюдений, содержащая «грубые» ошибки. Основная масса элементов засоренной выборки является реализацией случайной величины  $X$ , закон распределения которой известен. Такие элементы – «типичные» появляются в совокупности с вероятностью  $1 - \xi$ , где  $\xi$  ( $0 < \xi < 1$ ). Элементы совокупности оказываются реализацией другой случайной величины  $Y$ , закон распределения которой в общем случае неизвестен. Такие элементы называются «грубыми» ошибками. Обычные оценки, например, средняя арифметическая выборочная, на засоренной выборке теряют свои оптимальные свойства (эффективность, несмещенность) с ростом интенсивности засорения.

г) Цензурированная выборка – выборка, полученная из вариационного ряда наблюдений путем отбрасывания некоторого числа экстремальных наблюдений. Если отбрасывание производится по признаку выхода наблюдений

за пределы заданного интервала, то такой прием называется цензурирование первого типа. В этом случае число оставшихся наблюдений является случайной величиной. Если отбрасывается фиксированная доля крайних малых значений и фиксированная доля крайних больших значений, то это называется цензурированием второго типа уровня  $(\delta, \lambda)$ . При этом число оставшихся в рассмотрении наблюдений является величиной заранее заданной.

Цензурирование или урезание результатов выборочного обследования возникает при исследовании:

- а) «длительности жизни» какого-либо процесса или элемента;
- б) времени нахождения системы (элемента) в определенном состоянии: время жизни индивида, период безотказной работы прибора, время поиска работы безработным, длительность забастовки.

Совокупность измерений подобного рода показателей в течение некоторого периода времени и представляет *временной ряд*. Временной ряд – это совокупность наблюдений какого-либо показателя  $x(t_1), x(t_2), \dots, x(t_N)$  за несколько последовательных моментов или периодов времени. Включает как динамические, так и статические последовательности уровней какого-либо показателя. Временной ряд  $x_t$  называется интегрируемым порядка  $k$ , если он становится впервые стационарным после  $k$ -кратного применения к нему разностного оператора  $\Delta$ .



## 2. МЕТОДЫ СТАТИСТИЧЕСКОГО АНАЛИЗА ДАННЫХ

### 2.1. Статистические методы

Статистические методы – научные методы описания и изучения массовых явлений, допускающих количественное (численное) выражение. Слово «статистика» (от итал. stato – государство) имеет общий корень со словом «государство». Первоначально оно относилось к науке управления и означало сбор данных о некоторых параметрах жизнедеятельности государства. Со временем статистика стала охватывать сбор, обработку и анализ данных о массовых явлениях вообще; ныне статистические методы охватывают собою практически все области знаний и жизнедеятельности общества.

Статистические методы включают и экспериментальное, и теоретическое начало. Статистика исходит прежде всего из опыта; недаром ее зачастую определяют как науку об общих способах обработки результатов эксперимента. Обработка массовых опытных данных представляет самостоятельную задачу. Иногда простая регистрация некоторых рядов наблюдений приводит к тому или иному значимому выводу. Так, если в некоторой стране из года в год растет объем валового внутреннего продукта, то это говорит об ее устойчивом развитии. Однако в большинстве случаев для обработки опытного статистического материала используются математические модели исследуемого явления, основу которых составляют идеи и методы теории вероятностей.

Теория вероятностей есть наука о массовых случайных явлениях. Массовость означает, что исследуются огромные количества однородных явлений (объектов, процессов). Случайность же означает, что значение рассматриваемого параметра отдельного явления (объекта) в своей основе не зависит и не определяется значениями этого параметра у других явлений, входящих в ту же совокупность. Основной характеристикой массового случайного явления является распределение вероятностей. Теорию вероятностей можно определить как науку о вероятностных распределениях – их свойствах, видах, законах взаимосвязей, распределении величин, характеризующих исследуемый объект, и законах изменения распределений во времени. Так, говорят о распределении молекул газа по скоростям, о распределениях доходов граждан в некотором обществе и т. д.

Эмпирически задаваемые распределения соотносятся с т. н. генеральной совокупностью, т. е. с наиболее полным теоретическим описанием распределений соответствующих массовых явлений. При этом во многих случаях бывает нецелесообразно «перебирать» все элементы рассматриваемых совокупностей либо в силу чрезвычайно большого их числа, либо в силу того, что при наличии некоторого числа «перебранных» элементов учет новых не внесет существенных изменений в общие результаты. Для этих случаев разработан специальный выборочный метод исследования общих свойств статистических систем на основе изучения лишь части соответствующих элементов, взятых на выборку. Так, при оценке политических симпатий граждан некоторого региона или страны перед предстоящими выборами невозможно

проводить сплошной опрос граждан. В этих случаях и прибегают к выборочному методу. Чтобы выборочное распределение достаточно надежно характеризовало исследуемую систему, оно должно удовлетворять специальным условиям репрезентативности. Репрезентативность требует случайного выбора элементов и учета макроструктуры всего массового явления.

Распределения представляют наиболее общую характеристику массовых случайных явлений. Задание исходного распределения нередко предполагает построение математической модели соответствующих областей действительности. Построение и анализ таких моделей и составляет основную направленность статистических методов. Построенная математическая модель, в свою очередь, указывает, какие переменные следует измерять и какие из них имеют основное значение. Но главное в построении математической модели состоит в объяснении исследуемых явлений и процессов. Если модель достаточно полна, то она описывает зависимости между основными параметрами этих явлений.

Следует, однако, отметить, что во многих случаях исходные вероятностные распределения задаются не путем непосредственной обработки массового материала. Вероятностная гипотеза чаще всего вводится гипотетически, косвенно, на основе теоретических предпосылок. Возможность подобного задания вероятностных распределений и проверки их справедливости обусловлена характером и природой самих распределений, математическое выражение которых обладает самостоятельными характеристиками, достаточно независимыми от конкретных значений элементов.

Особые сложности возникают при применении статистических методов в изучении социальных явлений. Анализ общих направлений социальных процессов и внутренних механизмов, вызывающих конкретные статистические результаты, необычайно трудоемок. Так, благосостояние людей характеризуется весьма многими параметрами и соответствующими распределениями – уровнем доходов, участием в общественно-полезном труде, уровнем образования и здравоохранения и др. показателями жизнедеятельности человека. Выявление взаимосвязи этих распределений и тенденций их изменения требует решения многих сложных задач. Состояние общества можно определить через такие параметры, как внутренний валовый продукт, потребление энергии на душу населения, расслоение общества по доходам и т. п. Вместе с тем общество представляет необычайно сложную систему, а познание сложных систем основывается на разработке многих моделей, выражающих различные аспекты их структуры и функционирования. Соответственно, для более полной характеристики состояния общества требуется оперировать весьма многими параметрами и их распределениями. Так, говорят об экономической, производственной, сельскохозяйственной, социальной и многих других статистиках. Для объединения данных этих статистик в единую целостную картину необходимо выявление субординации, иерархии параметров, характеризующих состояние общества.

## 2.2. Частотный анализ

Анализ результатов экономической деятельности всегда включает частотный анализ поведения объектов, попавших случайно или специально в поле интересов, т. е. всей той информации, которую предоставляет деятельность, независимо от того в какой форме информация была передана и получена.

Частотный анализ распределения признака в поле информации является той единственной возможностью, которая позволяет исследователю самоопределяться и определять позиционирование другого в непосредственной ему среде, определить траекторию движения другого объекта относительно самого себя и пр. Это и условие выживания и условие развития. Текст только одна из форм представления информации о социальной или природной среде, так же осознанно или нет, подвергающийся частотному анализу интересующего человека признака.

Поэтому анализ информации всегда был статистическим. Это означало, что интересующий признак всегда рассматривался в некотором варианте частотного выражения. Не важно был ли задан кем-то, например, самим человеком, или же частотный вариант возникал спонтанно. Другими словами, статистический анализ информации означает как часто в некотором континууме организованной и целенаправленной информации, проявлялся интересующий человека или заданный им признак.

Так, если частота проявления знака опасности достигала некоторой установленной эмпирически предельной нормы, то надо принимать экстренные меры. Но если частота проявления признака опасности не достигала предельной нормы опасности, то можно постоять и попытаться проанализировать от куда, от кого и для кого проистекает опасность. Но каждый раз это именно частотный анализ события, но различного уровня.

Понятие частотный (синонимы: частота, густота, плотность, частый) есть количественное выражение проявления признака наблюдаемых объектов за установленный промежуток времени и при их перемещении в пространстве. Иначе говоря, интересуется в частотном распределении, сколько раз проявлялся данный признак в статистическом поле и в каких местах.

Мы существуем в мире постоянного частотного распределения многих или точнее всех наблюдаемых признаков и чаще всего не обращаем на них равным счетом никакого внимания, но до тех пор, пока частота их распределения находится в некой установленной норме. В этом случае данная частота полностью вливается в норму жизнеобеспечения и не является, по меньшей мере, опасной. Но как только частота проявления меняется в какую-либо сторону, то это становится предметом внимания и анализа, ибо такое частотное изменение может нести опасность.

Статистический анализ частотного распределения признака в каком-либо событийном поле на так скажем бытовом, повседневном уровне, осуществляется исключительно интуитивно. Отсюда и интуитивное решение возникает спонтанно. Интуитивно означает, что сознание производит статистический частотный анализ без актуального его осознания и понимания и естественно без какой-либо формализации. Оно просто осуществляется где-то в глубинах

сознания. Так, когда принимается решение перепрыгнуть яму, в уме интуитивно, но тщательно просчитывается расстояние, силу отталкивания для прыжка, высоту прыжка и пр. И не дай бог ошибиться. Но ошибаются и довольно часто.

Интуитивное в научной литературе чаще всего воспринимается как нечто если и не негативное, то во всяком случае, как бы несовершенное. Такое взаимодействие с предметным полем внешней действительности не считается обоснованным, так как приводит к ошибкам, неверным решениям и пр. По сравнению с актуализированным и формализованным знанием, которое чаще всего и, как правило, естественниками рассматривается как синоним точного и совершенного метода познания. На самом деле противопоставлять их и тем более в соотношении плохо-хорошо вряд ли возможно и целесообразно.

Частотное распределение, например, опасности по видам можно представить таким образом, когда каждый вид опасности имеет свой признак и свою частоту проявления данного признака опасности. Такое углубление интуитивного статистического анализа, безусловно, имело свое положительное значение.

Но, с другой стороны, усложнение системы статистического анализа по многим параметрам не позволяло в полной мере проявиться именно из-за ограниченности возможностей интуитивного статистического анализа. Возможности мозга запоминать не безграничны. Собственно, поэтому появилась потребность записывать и создавать новый механизм статистического анализа частотного распределения признака в какой-то, не обязательно опасной, ситуации.

Контент-анализ как особое направление оформился окончательно тогда, когда единицы распределения в статистической совокупности можно было сосчитать и представить в цифровом выражении, который был синонимом точности. Однозначно, что такой признак имеет процентное наполнение, т. е. имеет количественное выражение отношения к каким-то другим признакам и ко всей совокупности в целом уже было много. Это означало, что появилась возможность для количественной оценки проявления признаков. Другими словами, появилась возможность делить и отнимать, и каждый раз переставляя выделенный признак в то место, которое интересно исследователю.

Благодаря контент-анализу возможности для глубокого статистического анализа текста оказались таким образом практически безграничными. Они ограничивались лишь только разработкой математического и статистического аппарата и потребностями человека в уровне такого анализа для решения тех или иных возникающих задач.

Тем самым открылась принципиально новая страница в частотном описании бесконечного количества вариаций ситуаций, а значит и обогащения эмпирического анализа и опыта природной и социальной целесообразной деятельности. Фактически произошел отказ от интуитивного анализа статистической совокупности со всеми плюсами и минусами и переход на формализованный частотный анализ структуры совокупности.

Структура частотного наполнения имеет и свою пороговую величину, за пределами которой цифровой анализ теряет смысл. Пороговое значение устанавливает исследователь в зависимости от многих привходящих факторов.

Практика количественного анализа позволяет установить данную величину предельно близкой той величины, которая позволяет в свою очередь решить поставленную задачу. Имеется еще одна особенность частотного распределения. Мы всегда все измеряем, любое событие, попавшее в поле нашего интереса, при решении задачи получает свое измерение в каких-то определенных единицах. Только измеряя, можно оперировать с явлениями. При этом различные явления в силу особенности и сущности самого явления имеют разную систему и единицы измерения. Дать оценку одному событию еще как-то можно, не прибегая к сложному измерению. Но если приходится соединять в одно целое различные сущности, то возникают большие трудности. Понятно, что более сложные комбинации оказываются вообще вне какой-либо системы измерения.

Таким образом, частотный анализ проводят при анализе дискретных данных. Частота (Frequency) – количество наблюдений, в которых признак принимает определенное значение или находится в определенном интервале. Результаты анализа сводятся в таблицы частот, или как еще их называют одновходовые таблицы, которые представляют простейший метод анализа категориальных переменных.

Таблицы частот могут быть с успехом использованы также для исследования количественных переменных, хотя при этом могут возникнуть трудности с интерпретацией результатов. Данный вид статистического исследования часто используют как одну из процедур разведочного анализа, чтобы посмотреть, каким образом различные группы наблюдений распределены в выборке, или как распределено значение признака на интервале от минимального до максимального значения. Как правило, таблицы частот графически иллюстрируются при помощи гистограмм.

**Кросстабуляция (сопряжение)** – процесс объединения двух (или нескольких) таблиц частот так, что каждая ячейка в построенной таблице представляется единственной комбинацией значений или уровней табулированных переменных. Кросстабуляция позволяет совместить частоты появления наблюдений на разных уровнях рассматриваемых факторов. Исследуя эти частоты, можно выявить связи между табулированными переменными и исследовать структуру этой связи. Обычно табулируются категориальные или количественные переменные с относительно небольшим числом значений. Если надо табулировать непрерывную переменную (предположим, уровень сахара в крови), то вначале ее следует перекодировать, разбив диапазон изменения на небольшое число интервалов (например, уровень: низкий, средний, высокий).

### 2.3. Анализ соответствий<sup>1</sup>

Анализ соответствий (correspondence analysis) – это статистический метод анализа двумерных и многомерных таблиц сопряженности, содержащих информацию по взаимосвязанным категориальным переменным.

Анализ соответствий содержит описательные и разведочные методы анализа двухвходовых и многовходовых таблиц. Эти методы по своей природе похожи на методы факторного анализа и позволяют исследовать структуру группирующих переменных, включенных в таблицу. Одной из наиболее общих разновидностей многовходовых таблиц типа являются частотные таблицы сопряженности.

Анализ соответствий – метод многомерного шкалирования качественных данных. Цель анализа соответствий – преобразовать исходный набор категориальных переменных в гораздо меньшее количество агрегированных переменных, чтобы представить исходную информацию графически в пространстве с более низкой размерностью.

Таблица сопряженности, или таблица контингентности, факторная таблица в статистике – средство представления совместного распределения двух переменных, предназначенное для исследования связи между ними. Строки таблицы сопряженности соответствуют значениям одной переменной, столбцы – значениям другой переменной, при этом количественные шкалы предварительно должны быть сгруппированы в интервалы.

На пересечении строки и столбца указывается частота совместного появления  $f_{i,j}$  соответствующих значений двух признаков  $x_i$  и  $y_j$ . Сумма частот по строке  $f_i$  называется маргинальной частотой строки; сумма частот по столбцу  $f_j$  – маргинальной частотой столбца. Сумма маргинальных частот равна объему выборки  $n$ ; их распределение представляет одномерное распределение переменной, образующей строки или столбцы таблицы. В таблице сопряженности могут быть представлены как абсолютные, так и относительные частоты (в долях или процентах). Относительные частоты могут рассчитываться по отношению:

- а) к маргинальной частоте по строке;
- б) к маргинальной частоте по столбцу;
- в) к объему выборки.

Анализ соответствий – метод многомерного шкалирования качественных данных. Цель анализа соответствий – преобразовать исходный набор категориальных переменных в гораздо меньшее количество агрегированных переменных, чтобы представить исходную информацию графически в пространстве с более низкой размерностью.

В классическом анализе соответствий частоты в таблице сопряженности стандартизуются таким образом, чтобы сумма наблюдений во всех ячейках была

---

<sup>1</sup>СтатМод.ру: [сайт] – URL: [https://www.statmod.ru/wiki/\\_media/study](https://www.statmod.ru/wiki/_media/study) (дата обращения: 15.01.2025).

равна 1. Одной из целей анализа соответствий является представление содержимого таблицы относительных частот в виде расстояний между отдельными строками и/или столбцами таблицы в пространстве возможно более низкой размерности. Каким образом это достигается, лучше всего показать на простом примере, который будет рассмотрен далее. Заметим, что имеется некоторое сходство в интерпретации результатов анализа соответствий.

Таблица сопряженности является наиболее универсальным средством изучения статистических связей, так как в ней могут быть представлены переменные с любым уровнем измерения. Таблицы сопряженности часто используются для проверки гипотезы о наличии связи между двумя признаками с использованием точного теста Фишера или критерия согласия Пирсона.

Основная цель анализа соответствий состоит в том, чтобы перейти от исходной матрицы данных к новой, более простой матрице, потеряв при этом как можно меньше информации. Кроме того, анализ соответствий позволяет представить полученные результаты графически, что существенно облегчает интерпретацию решения. Одна из задач анализа соответствий – это описание взаимосвязей между двумя номинальными переменными в таблице соответствий в малоразмерном пространстве при одновременном описании взаимосвязей между категориями для каждой переменной. Для каждой переменной расстояния между обозначающими категории точками на графике отображает взаимосвязь между категориями, при этом сходные категории изображаются близко друг к другу. Точки проекции для одной переменной на векторе из начала координат к точке категории для другой переменной, описывают взаимосвязь между переменными.

Анализ соответствий позволяет изучить структуру категориальных переменных, входящих в таблицу сопряженности.

Анализ соответствий по сравнению с частотным анализом содержит более мощные описательные и разведочные методы анализа двухвходовых и многовходовых таблиц. Метод, так же, как и таблицы сопряженности, позволяет исследовать структуру и взаимосвязь группирующих переменных, включенных в таблицу. В классическом анализе соответствий частоты в таблице сопряженности стандартизуются (нормируются) таким образом, чтобы сумма элементов во всех ячейках была равна 1.

Часто с помощью анализа соответствий анализируются двумерные или двухвходовые таблицы сопряженности. Например, среднюю оценку удовлетворенности для нескольких услуг, можно отобразить наглядно в виде точек в двумерном пространстве или на карте. Это облегчает анализ и интерпретацию сложного набора данных.

Метод анализа соответствий используют при решении маркетинговых задач. Например, на анализе соответствий построен метод построения карт позиционирования брендов (Brand Mapping). Карты позиционирования брендов часто используют для иллюстрации представления клиентов о продуктах и их атрибутах вместе на одной карте. Это позволяет точно интерпретировать восприятие клиентами компании множества продуктов и атрибутов сервиса одновременно.

Одна из целей анализа соответствий – представление содержимого таблицы относительных частот в виде расстояний между отдельными строками и/или столбцами таблицы в пространстве более низкой размерности.

**Шкалирование координат (возможности стандартизации).** Еще одним важным решением, которое принимает аналитик, является решение о шкалировании координат. Решение о выборе той или иной опции зависит от того, собираетесь ли Вы анализировать относительные проценты по рядам, по столбцам или по тем и другим одновременно. В контексте предыдущего примера, проценты по рядам сравнивались для того, чтобы проиллюстрировать наблюдаемое сходство процентов в таблице для близких точек на диаграмме. Другими словами, координаты точек прямо связаны с анализом *матрицы профилей строк*, в которой сумма элементов строки равна 1 (каждый элемент  $r_{ij}$  в матрице профилей строк интерпретируется как условная вероятность того, что элемент строки  $i$  принадлежит столбцу  $j$ ). Таким образом, координаты вычисляются так, чтобы максимизировать расстояние между профилями строк (процентов по строкам). Координаты строк вычисляются по матрице профилей строк, а координаты столбцов вычисляются по матрице профилей столбцов.

Имеется также опция *канонической* стандартизации, эта разновидность стандартизации применяется к столбцам и строкам матрицы относительных частот. Данная стандартизация применяется для шкалирования стандартизаций профилей по строкам и столбцам и не является широко распространенной. Отметим, что существует возможность задавать собственную стандартизацию по выбору пользователя, если имеются значения *собственных значений* и *собственных векторов*.

**Метрика координатной системы.** Термин *расстояние* может использоваться для обозначения различий между строками и столбцами матрицы относительных частот, которые, в свою очередь, представлялись в пространстве меньшей размерности в результате использования методов анализа соответствий. В действительности расстояния, представленные в виде координат в пространстве соответствующей размерности, это не просто евклидовы расстояния, вычисленные по относительным частотам столбцов и строк, а некоторые взвешенные расстояния. Процедура подбора весов устроена таким образом, чтобы в пространстве более низкой размерности метрикой являлась бы метрика *Хи-квадрат*, учитывая, что сравниваете точки-строки и выбираете стандартизацию профилей строк или стандартизацию профилей строк и столбцов или что вы сравниваете точки-столбцы и выбираете стандартизацию профилей столбцов или стандартизацию профилей строк и столбцов.

В этом случае (но *не* в случае канонической стандартизации) возведенное в квадрат евклидово расстояние между, например, двумя точками-строками  $i$  и  $i'$  в некоторой координатной системе соответствующей размерности аппроксимируют взвешенное (например, *Хи-квадрат*) расстояние между относительными частотами.

Другими словами, только расстояния между точками-строками и, соответственно, между точками-столбцами могут интерпретироваться в



подобной манере, однако расстояния между строками и столбцами такой интерпретации не подлежат.

**Замечание о «статистической значимости».** Необходимо отметить, что анализ соответствий является *разведочным методом*. Данный метод был разработан на базе методологии, рассматривающей построение моделей с точки зрения их соответствия данным, а не наоборот («второй принцип» Jean-Paul Benzecri<sup>2</sup> постулирует так: «Модель должна удовлетворять имеющимся данным, а не наоборот»<sup>3</sup>). Следовательно, не существует статистических тестов, которые могли бы быть использованы для проверки результатов анализа соответствий. Главной целью анализа соответствий является представление в упрощенном виде (пространстве меньшей размерности) информации, содержащейся в больших частотных таблицах (или таблицах с аналогичными мерами соответствия).

## 2.4. Методы сравнения средних

Сравнение двух выборок по признаку, измеренному в метрической шкале, обычно предполагает сравнение средних значений с использованием параметрического критерия t-Стьюдента. Следует различать три ситуации по соотношению выборок между собой: случай независимых и зависимых выборок (измерений признака) и дополнительно – случай сравнения одного среднего значения с заданной величиной (критерий t-Стьюдента для одной выборки).

К параметрическим методам относится и сравнение дисперсий двух выборок по критерию F-Фишера. Иногда этот метод приводит к ценным содержательным выводам, а в случае сравнения средних для независимых выборок сравнение дисперсий является обязательной процедурой.

При сравнении средних или дисперсии двух выборок проверяется ненаправленная статистическая гипотеза о равенстве средних (дисперсий) в генеральной совокупности. Соответственно, при ее отклонении допустимо принятие двусторонней альтернативы о конкретном направлении различий в соответствии с соотношением выборочных средних (дисперсий). Для принятия статистического решения в таких случаях применяются двусторонние критерии и, соответственно, критические значения для проверки ненаправленных альтернатив.

Сравнение средних значений различных выборок относится к наиболее часто применяемым методам статистического анализа. При этом всегда должен быть выяснен вопрос, можно ли объяснить имеющееся различие средних значений статистическими колебаниями или нет. Если нет, говорят о статистически значимом различии между сравниваемыми группами.

В прикладных исследованиях часто встречаются случаи, когда средний результат некоторого признака одной серии экспериментов отличается от

---

<sup>2</sup> Benzecri, J. P. L' Analyse des Donnees. Paris, 1973. Vol. 2. 619 p.

<sup>3</sup> Theory and Applications of Correspondence Analysis. Front Cover. Michael J. Greenacre. Academic Press, 1984. Mathematics. 364 p.

среднего результата другой серии. Так как средние – это результаты измерений, то, как правило, они всегда различаются, вопрос в том, можно ли объяснить обнаруженное расхождение средних неизбежными случайными ошибками эксперимента, или оно вызвано определенными причинами. Если идет речь о сравнении двух средних, то можно применять критерий Стьюдента (t-критерий). Это параметрический критерий, так как предполагается, что признак имеет нормальное распределение в каждой серии экспериментов. В настоящее время модным стало применение непараметрических критериев сравнения средних.

Основные термины:

1. Нулевая гипотеза – предположение о том, что между определенными статистическими параметрами генеральной совокупности не существует связи или различия. Ее подтверждение не требует от компании каких-либо действий.

2. Альтернативная гипотеза – утверждение о том, что между определенными статистическими параметрами генеральной совокупности есть связь или различия. Ее подтверждение означает, что руководству компании следует предпринять какие-либо действия или менять свои взгляды на положение дел.

3. Параметрические методы проверки гипотез – применяются для переменных, измеренных с помощью интервальных шкал.

4. Непараметрические методы проверки гипотез – применяются для переменных, измеренных с помощью номинальных или порядковых шкал.

5. Размах вариации  $R$  – это разность между самым большим и самым малым значениями признака у единиц данной совокупности:  $R = X_{max} - X_{min}$ .

6. Дисперсия – это средний квадрат отклонений индивидуальных значений признака от его средней величины.

$$D = \frac{\sum_{i=1}^n (x - \bar{x})^2}{n},$$

где  $D$  – дисперсия;

$x$  – анализируемый показатель;

$\bar{x}$  – среднее значение показателя;

$n$  – количество значений в анализируемой совокупности данных.

7. Стандартное отклонение – корень из дисперсии (среднего квадрата отклонений).

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x - \bar{x})^2}{n}}.$$

8. Коэффициент вариации – наиболее универсальный показатель, отражающий степень разбросанности значений независимо от их масштаба и единиц измерения. Коэффициент вариации измеряется в процентах и может быть использован для сравнения вариации различных процессов и явлений.

$$V = \frac{\sigma}{\bar{x}}.$$

9. F-критерий или критерий Фишера – статистический критерий, с помощью которого проверяется гипотеза о равенстве дисперсий двух

разных совокупностей. F-критерий – это отношение двух выборочных дисперсий.

10. Т-критерий – одномерный метод проверки гипотез, использующий Т-распределение. Т-критерий применяется для небольших выборок, когда стандартное отклонение не известно.
11. Метод Т-теста используется для проверки достоверности различия двух выборок по количественной переменной. Нулевая гипотеза формулируется следующим образом: «Взаимосвязи между исследуемыми переменными нет». Например, формулируя исходную (нулевую) гипотезу, предполагаем равенство среднего чека покупателей разного возраста, это говорит о том, что люди разного возраста в среднем тратят одинаковую сумму на покупки. В результате проведенного исследования данная гипотезе либо подтверждается, либо опровергается.

Сравнивать средние величины возможно при помощи нескольких методов. Сравнение средних результатов один из способов выявления зависимостей между переменными признаками, характеризующими исследуемую совокупность объектов (наблюдений). Если при разбиении объектов исследования на подгруппы при помощи категориальной независимой переменной (предиктора) верна гипотеза о равенстве средних некоторой зависимой переменной в подгруппах, то это означает, что существует стохастическая взаимосвязь между этой зависимой переменной и категориальным предиктором. Так, например, если установлено, что неверна гипотеза о равенстве средних показателей физического и интеллектуального развития детей в группах матерей, куривших и не куривших в период беременности, то это означает, что существует зависимость между курением матери ребенка в период беременности и его интеллектуальным и физическим развитием.

Наиболее общий метод сравнения средних дисперсионный анализ. В терминологии дисперсионного анализа категориальный предиктор называется фактором.

Дисперсионный анализ можно определить как параметрический, статистический метод, предназначенный для оценки влияния различных факторов на результат эксперимента, а также для последующего планирования экспериментов. Поэтому в дисперсионном анализе можно исследовать зависимость количественного признака от одного или нескольких качественных признаков факторов. Если рассматривается один фактор, то применяют однофакторный дисперсионный анализ, в противном случае используют многофакторный дисперсионный анализ.

## 2.5. Канонический анализ

Основная цель канонического анализа – поиск максимальных корреляционных связей между двумя группами исходных переменных.

Канонический анализ предназначен для анализа зависимостей между двумя списками признаков (независимых переменных), характеризующих объекты. Например, можно изучить зависимость между различными неблагоприятными факторами и появлением определенной группы симптомов заболевания, или взаимосвязь между двумя группами клинико-лабораторных показателей (синдромов) больного.

Вместе с тем, канонический анализ может применяться как метод решения классификационных задач. Действительно, способности канонических переменных наилучшим образом разделять совокупности наблюдений, могут также быть истолкованы как свойства этих переменных находить такие направления межгрупповой вариации, вдоль которых различия центральных точек выборок будут максимальными.

Канонические переменные формируются как взвешенные суммы исходных переменных по двум группам.

Собственные значения – это собственные значения корреляционной матрицы. Эти значения равны доле дисперсии, объясняемой корреляцией между соответствующими каноническими переменными. Вычисляются в порядке убывания значений.

Канонические корреляции (канонические корни) – коэффициенты корреляции между каноническими переменными. Вычисляются как квадратные корни из собственных значений. Число канонических корней равно числу переменных в наименьшем множестве.

Канонические веса – коэффициенты во взвешенной сумме, соответствующей каноническому корню (каждый корень в действительности представляет две взвешенные суммы, по одной на каждое множество переменных). Канонические веса аналогичны частным корреляциям переменных, соответствующих каноническому корню, а также коэффициентам факторных нагрузок.

Канонические значения – значения канонических переменных (взвешенные суммы значений исходных переменных). Канонические веса определяются для стандартизированных ( $z$ -преобразованных) переменных.

Факторная структура. Еще одним способом интерпретации канонических корней является рассмотрение обычных корреляций между каноническими переменными (факторами) и переменными из каждого множества. Эти корреляции также называются каноническими нагрузками факторов.

Извлеченная дисперсия. Коэффициенты канонической корреляции соответствуют корреляции между взвешенными суммами по двум множествам

переменных. Квадраты этих корреляций будут отражать долю дисперсии, объясняемую каждой переменной. Для каждого корня можно вычислить среднее значение этих долей. При этом получится средняя доля изменчивости, объясненной в этом множестве на основании соответствующей канонической переменной. Другими словами, это средняя доля дисперсии, извлеченной каждым корнем.

Избыточность. Каноническая корреляция при возведении в квадрат дает долю дисперсии, общей для сумм по каждому множеству (канонической переменной). Если умножить эту долю на долю извлеченной дисперсии, получится мера избыточности множества переменных, т. е. величина, показывающая, насколько избыточно одно множество переменных, если задано другое множество.

Канонический анализ является обобщением множественной корреляции как меры связи между одной переменной и множеством других переменных. Как известно, множественная корреляция есть максимальная корреляция между одной переменной и линейной функцией других переменных. Эта концепция была обобщена на случай связи между множествами переменных – признаков, характеризующих объекты. При этом достаточно ограничиться рассмотрением небольшого числа наиболее коррелированных линейных комбинаций из каждого множества.

Пусть, например, первое множество переменных состоит из признаков  $y_1, \dots, y_p$ , второе множество состоит из  $x_1, \dots, x_q$ , тогда взаимосвязь между данными множествами можно оценить как корреляцию между линейными комбинациями:

$$a_1y_1 + a_2y_2 + \dots + a_py_p, b_1x_1 + b_2x_2 + \dots + b_qx_q,$$

которая называется канонической корреляцией.

Задача канонического анализа в нахождении весовых коэффициентов таким образом, чтобы каноническая корреляция была максимальной.

Каждая каноническая переменная выявляет определенную ось, вдоль которой центральные точки некоторых выборок будут отстоять друг от друга так далеко, как это только возможно. Любая такая ось может считаться переменной, описывающей некоторое информативное направление межгрупповой вариации именно центральных точек выборок. При этом вариация индивидуальных наблюдений нас может специально и не интересовать. Таким образом, канонический анализ может считаться инструментом изучения межгрупповой вариации примерно так же, как компонентный анализ позволяет рассматривать внутригрупповую изменчивость. Коэффициенты исходных признаков у канонических переменных позволяют интерпретировать морфологический смысл этих переменных и выяснять какие комплексы значений исходных признаков характерны для тех или иных групп выборок, локализованных на полюсах осей этих новых признаков.

Вторым важным свойством канонического анализа, которое позволяет его расценивать как метод решения задач классификации, является возможность наглядно представлять расположение центральных точек выборок в координатах канонических переменных. Однако на этой картине можно размещать только центральные точки выборок, не интересуясь при этом индивидуальными наблюдениями. Взаимное расположение одних только центральных точек позволит наглядно представить их пространственные взаимоотношения, наличие определенной структуры и т. д.

В-третьих, канонический анализ позволяет выделить набор информативных признаков, выявляющих различия данного набора выборок наиболее отчетливо. Такое суждение можно сделать по значениям стандартизованных коэффициентов признаков у канонических переменных. Очевидно, вклад каждого признака  $j$  в вариацию канонической переменной  $i$  будет соответствовать его стандартизованному коэффициенту  $b_{ij}$ . Аналогичный вклад этого признака в вариацию всех  $n$  канонических переменных может быть определен как сумма  $n$  его стандартизованных коэффициентов у всех этих переменных:

$$g_j = (b_{1j}^2 + b_{2j}^2 + b_{3j}^2 + \dots + b_{nj}^2)^{1/2}.$$

Суммирование квадратов значений  $b_{ij}$  необходимо для того, чтобы на величину  $g_j$  не влияли различия в их знаках.

### 3. ЭКОНОМЕТРИЧЕСКИЕ МОДЕЛИ

#### 3.1. Определение эконометрической модели

При анализе экономических явлений на основе экономико-математических методов особое место занимают модели, выявляющие количественные связи между изучаемыми показателями и влияющими на них факторами. Научной дисциплиной, предмет которой составляет изучение этой количественной стороны экономических явлений и процессов средствами математического и статистического анализа, является эконометрика, в которой результаты теоретического анализа экономики синтезируются с выводами математики и статистики. Основная задача эконометрики – проверка экономических теорий на фактическом (эмпирическом) материале при помощи методов математической статистики.

Главным инструментом эконометрики служит эконометрическая модель, т. е. экономико-математическая модель факторного анализа, параметры которой оцениваются средствами математической статистики. Эта модель выступает в качестве средства анализа и прогнозирования конкретных экономических процессов на основе реальной статистической информации.

Эконометрическая модель – модель, в которой параметры оцениваются с помощью методов математической статистики. Такие модели используются в качестве средств анализа и прогнозирования конкретных экономических процессов как на макро-, так и на микроэкономическом уровне.

В наиболее общем виде любую эконометрическую модель, построенную в виде системы линейных уравнений, можно записать так:

$$y(t) = Ay(t) + \sum_{I=1}^p Zy(t-I) + Cx(t),$$

где  $y$  – вектор текущих значений эндогенных переменных модели;

$A$  – матрица коэффициентов взаимодействия между текущими значениями эндогенных переменных модели;

$Z$  – матрица коэффициентов влияния запаздывающих (лаговых) переменных модели на текущие значения эндогенных и моделируемых показателей;

$C$  – матрица коэффициентов внешних воздействий;

$x$  – вектор значений экзогенных показателей;

$t$  – индекс временного периода;

$I$  – индекс запаздывания (лага);

$p$  – продолжительность максимального лага.

При построении эконометрических моделей могут использоваться два принципиально различных типа исходных информационных массивов – статический и динамический.

**Статический массив** выражает взаимосвязи между результирующей (зависимой, объясняемой и т. п.) переменной  $y$  и влияющими на нее факторами (независимыми, объясняющими переменными)  $x_i$ , характерными для однородной совокупности объектов в определенный период времени. Примером

таких объектов является некоторая совокупность однотипных промышленных предприятий (заводов одной отраслевой направленности). В качестве  $y$  в практических исследованиях часто рассматриваются показатели производительности труда, объемов выпускаемой продукции и некоторые другие. В качестве  $x_i$  – влияющие на уровень этих показателей факторы – объемы используемых фондов, квалификация рабочей силы и т. п.

Другой пример статической информации характерен для социальных исследований, когда в качестве  $y$  рассматривается заболеваемость (смертность) населения, уровень которых в каждом из регионов страны определяют независимые факторы, отражающие достигнутый материальный уровень жизни, климатические условия, состояние окружающей среды и т. п. В этом случае необходимая для построения эконометрической модели информация собирается по совокупности регионов страны за фиксированный промежуток времени.

Таким образом, необходимая для построения эконометрической модели статическая информация выражается следующими массивами взаимно соответствующих наборов данных:

$$\begin{array}{l} y_1 \leftrightarrow x_{11} \quad x_{21} \dots x_{i1} \dots x_{n1}; \\ y_2 \leftrightarrow x_{12} \quad x_{22} \dots x_{i2} \dots x_{n2}; \\ \text{-----} \\ y_i \leftrightarrow x_{1j} \quad x_{2j} \dots x_{ij} \dots x_{nj}; \\ \text{-----} \\ y_N \leftrightarrow x_{1N} \quad x_{2N} \dots x_{iN} \dots x_{nN}; \end{array}$$

$y_j$  – уровень зависимой переменной на объекте  $j$  совокупности;

$x_{ij}$  – уровень фактора  $i$  на объекте  $j$  совокупности;

$i = 1, 2, \dots, n$ ;

$j = 1, 2, \dots, N$ .

В общем случае эконометрическая модель, использующая динамическую информацию, связывает значения некоторой зависимой переменной  $y_t$  в моменты времени  $t$  со значениями независимых переменных (факторов)  $x_{it}$ , рассматриваемых в те же моменты времени (или в предшествующие). Такая информация может отражать, например, уровни производительности труда на одном из заводов и определяющие ее характеристики факторов в последовательные моменты времени.

Несложно заметить, что принципиального различия между статическим и динамическим массивами не существует. С абстрактных позиций момент времени  $t$  выражает единицу совокупности, так что набор  $y_1, y_2, \dots, y_T$  может рассматриваться как выборка из  $T$  заводов (регионов) и наоборот. Это же относится и к элементам  $x_{ij}$  и  $x_{it}$ .

Вследствие этого в дальнейшем при изложении материала (если это не оговорено специально) для определенности будем использовать динамические обозначения.



Данный массив образован вектором-столбцом значений зависимой переменной  $y = (y_1, y_2, \dots, y_T)'$  и матрицей значений независимых переменных:

$$X = \begin{bmatrix} x_{11} & x_{21} & \dots & x_{n1} \\ x_{12} & x_{22} & \dots & x_{n2} \\ \dots & \dots & \dots & \dots \\ x_{1T} & x_{2T} & \dots & x_{nT} \end{bmatrix}$$

размерностью  $T \times n$ , таким образом, что каждому элементу  $y_t$  вектора  $y$  соответствует строка матрицы  $X$ .

Эконометрическая модель, отражающая взаимосвязь переменных  $y$  и  $x_i$ ,  $i = 1, 2, \dots, n$  в общем виде может быть представлена следующим уравнением:

$$y_t = f_t(a, x) + \varepsilon_t, \quad (3.1)$$

где  $f_t(a, x)$  – функционал, выражающий закономерность взаимосвязи между переменными  $y_t$  и  $x_{it}$ ;

$x = (x_1, x_2, \dots, x_n)$  – вектор независимых переменных (факторов);

$a = (a_0, a_1, \dots, a_n)$  – вектор параметров модели;

$a_i$  – параметр выражает степень влияния фактора  $x_i$  на переменную  $y$ ;

$a_0$  – постоянная модели;

$\varepsilon_t$  – случайная ошибка модели в момент  $t$ , в отношении которой выдвигается предположение о равенстве нулю ее математического ожидания и конечности дисперсии.

Под структурой эконометрической модели понимается совокупность переменных и их взаимосвязей, входящих в правую часть выражения (3.1). Форма эконометрической модели отражает особенности взаимосвязи между переменными  $y$  и  $x_i$ ,  $i = 1, 2, \dots, n$ .

Проблема построения эконометрической модели состоит в определении конкретного состава независимых переменных  $x_i$ , в выборе вида функционала, связывающего их с зависимой переменной  $y$  и в оценке его параметров  $a_i$ ,  $i = 0, 1, 2, \dots, n$ , на основании известных компонент вектора  $y$  и элементов матрицы  $X$ .

Состав переменных  $x_i$  и функционал  $f$  могут отражать либо экономическую концепцию, лежащую в основе взаимосвязи между зависимой и независимыми переменными, либо эмпирические (т. е. выявленные в ходе конкретных исследований) взаимосвязи между ними в период  $(1, T)$ .

### 3.2. Классификация эконометрических моделей

Эконометрические модели можно систематизировать по ряду классификационных признаков. Так, по аналитической форме модели (уравнения) выделяют линейные, нелинейные, степенные модели, модели Брандона и др. Например, модель Брандона имеет вид:

$$\hat{y} = \bar{y} f_1(x_1) f_2(x_2) \dots f_m(x_m),$$

где  $\bar{y}$  – изучаемый показатель (будем называть его **результативным признаком**), черта над ним означает среднюю величину (математическое ожидание);

$x_1, x_2, \dots, x_m$  – влияющие на изучаемый показатель величины (будем называть их **факторными признаками**). Факторы в модели должны быть независимыми друг от друга, т. е. сохраняется требование некоррелированности величин.

$f_i(x_i)$  в модели располагаются по степени влияния фактора на результат. Чем она больше, тем меньше порядковый номер функции.

Одной из основных классификационных рубрик эконометрических моделей является классификация по направлению и сложности причинных связей между показателями, характеризующими экономическую систему. Если пользоваться термином «переменная», то в любой достаточно сложной экономической системе можно выделить внутренние переменные (например, выпуск продукции, численность работников, производительность труда) и внешние переменные (например, поставка ресурсов, климатические условия и др.). Тогда по **направлению** и **сложности** связей между внутренними (**эндогенными**, выходными) переменными и внешними (**экзогенными**, входными) переменными выделяют следующие эконометрические модели: регрессионные модели, взаимозависимые системы, рекурсивные системы.

**Регрессионными** называют модели, основанные на уравнении регрессии, или системе регрессионных уравнений, связывающих величины эндогенных и экзогенных переменных.

Термин регрессия в статистике понимают как:

- а) функцию связи, зависимости;
- б) направление развития явления вспять;
- в) функцию анализа случайных событий во времени;
- г) уравнение линии связи.

Суть регрессионного анализа заключается в нахождении наиболее важных факторов, которые влияют на зависимую переменную. Связь между переменными может быть положительная, отрицательная.

Существует 5 видов регрессии:

- Линейная регрессия.
- Полиномиальная регрессия.
- Гребневая (ридж) регрессия.
- Регрессия по методу «лассо».
- Регрессия «эластичная сеть».

Наиболее часто выделяют уравнения (модели) парной и множественной регрессии. Если для обозначения эндогенных переменных использовать букву  $y$ , а для экзогенных переменных букву  $x$ , то в случае линейной модели уравнение парной регрессии имеет вид  $\hat{y} = a_0$ , а уравнение множественной регрессии:  $\hat{y} = a_0 + a_1x_1 + a_2x_2 + \dots + a_mx_m$ . Подобного типа модели рассматриваются подробно в следующих параграфах. Отметим только, что параметры моделей парной и множественной регрессии находятся на основе метода наименьших квадратов. Одним из часто применяемых видов нелинейных многофакторных регрессионных моделей являются **производственные функции**.

**Взаимозависимые** системы наиболее полно описывают экономическую систему, содержащую, как правило, множество взаимосвязанных эндогенных и экзогенных переменных.

Такие модели задаются системой взаимозависимых уравнений следующего вида ( $n$  – число эндогенных переменных,  $m$  – число экзогенных переменных):

$$\begin{aligned}\hat{y} &= a_0 + a_1x_1 + a_2x_2 + \dots + a_mx_m; \\ \hat{y} &= a_{1,0} + a_{1,1}x_1 + a_{1,2}x_2 + \dots + a_{1,m}x_m + b_{1,2}y_2 + b_{1,3}y_3 + \dots + b_{1,s}y_n; \\ \hat{y} &= a_{2,0} + a_{2,1}x_1 + a_{2,2}x_2 + \dots + a_{2,m}x_m + b_{2,1}y_1 + b_{2,3}y_3 + \dots + b_{2,s}y_n; \\ &\dots\dots\dots \\ \hat{y} &= a_0 + a_1x_1 + a_2x_2 + \dots + a_mx_m + b_{n,1}y_1 + b_{n,2}y_2 + \dots + b_{n,s}y_n.\end{aligned}$$

Для нахождения параметров системы взаимозависимых уравнений используются более сложные методы: двух- и трехшаговый метод наименьших квадратов, методы максимального правдоподобия с полной и неполной информацией и др.

На практике стремятся упростить взаимозависимые системы и привести их к так называемому рекурсивному виду. Для этого сначала выбирают эндогенную переменную (внутренний показатель), зависящую только от экзогенных переменных (внешних факторов), обозначают ее  $y_1$ . Затем выбирается внутренний показатель, который зависит только от внешних факторов и от  $y_1$  и т. д.; таким образом, каждый последующий показатель зависит только от внешних факторов и от внутренних предыдущих. Такие системы называются **рекурсивными**. Параметры первого уравнения рекурсивных систем находят методом наименьших квадратов, их подставляют во второе уравнение, и опять применяется метод наименьших квадратов, и т. д.

В зависимости от характера ограничений и статистической структуры переменных эконометрические модели классифицируются на линейные модели с одной, двумя и большим числом переменных, а также на модели дискретного выбора, непрерывной длительности (выживания), логит-модель, пробит-модель, тобит-модель.

В практике эконометрических исследований используется достаточно широкий круг функциональных зависимостей между переменными. Основные из них следующие:

1. Линейная эконометрическая модель:

$$y_t = a_0 + a_1x_{1,t} + \dots + a_nx_{n,t} + \varepsilon. \quad (3.2)$$

2. Правая полулогарифмическая эконометрическая модель:

$$y_t = a_0 + a_1\ln x_{1,t} + \dots + a_n\ln x_{n,t} + \varepsilon. \quad (3.3)$$

3. Степенная эконометрическая модель:

$$y_t = a_0x_{1,t}^{a_1}x_{2,t}^{a_2} \dots x_{n,t}^{a_n}\varepsilon. \quad \dots\dots\dots (3.4)$$

4. Гиперболическая эконометрическая модель:

$$y_t = a_0 + a_1/x_{1,t} + \dots + a_n/x_{n,t} + \varepsilon. \quad (3.5)$$

5. Логарифмическая гиперболическая эконометрическая модель:

$$\ln y_t = a_0 + \frac{a_1}{x_{1,t}} + \dots + \frac{a_n}{x_{n,t}} + \varepsilon. \quad (3.6)$$

6. Обратная линейная (функция Торнквиста) эконометрическая модель:

$$\frac{1}{y_t} = a_0 + \frac{a_1}{x_{1,t}} + \dots + \frac{a_n}{x_{n,t}} + \varepsilon. \quad (3.7)$$

7. Функция с постоянной эластичностью замены:

$$y_t = [a_1 x_{1,t}^{-\rho} + \dots + a_n x_{n,t}^{-\rho}]^{-\frac{\lambda}{\rho}} \times \varepsilon, \quad (3.8)$$

где  $\lambda$  и  $\rho$  – также параметры функции.

Следует отметить, что в практических исследованиях могут встретиться и комбинации рассмотренных выше зависимостей. Например,

$$y_t = a_0 + \frac{a_1}{x_{1,t}} + a_2 \ln x_{2,t} + a_3 x_{3,t}^{a_4} + \dots + \varepsilon. \quad (3.9)$$

Здесь необходимо отметить, что значительное большинство функций с помощью определенного набора преобразований могут быть приведены к линейной форме (3.2). Например, если  $y$  и  $x_i$  связаны зависимостью  $y \sim 1/x_i$  (выражение (3.5)), то, введя переменные  $y_i = \frac{1}{x_i}$ , получим выражение (3.2) с точностью до преобразования исходных факторов.

Аналогичным образом, используя преобразование  $v_i = \ln x_i$ , получим линейную модель при логарифмической взаимосвязи между переменными  $y$  и  $x_i$ , т. е.  $y \sim \ln x_i$ .

Заметим, что в основе использования степенной функции (3.4) обычно лежит концептуальное допущение о постоянстве частной эластичности выпуска  $y$  по каждому ресурсу (фактору)  $x_i$ . Напомним, что частная эластичность в точке  $t$  показывает, на сколько процентов изменится зависимая переменная  $y_t$  при изменении фактора  $x_{i,t}$  на 1 % при условии постоянства значений остальных факторов в этой точке. Эластичность определяется следующим выражением:

$$\mathcal{E}_{i,t} = \frac{\partial y_t}{\partial x_{i,t}} \times \frac{x_{i,t}}{y_t}. \quad (3.10)$$

Подставим вместо  $y_t$  в правую часть выражения (3.10) функцию:

$$a_0 \times x(1,t)^{a_1} \times \dots \times x(n,t)^{a_n}.$$

Учитывая, что

$$\frac{\partial y_t}{\partial x_{i,t}} = a_i a_0 \times x_{1,t}^{a_1} \times \dots \times x_{i,t}^{a_i-1} \times \dots \times x_{n,t}^{a_n}$$

получим:

$$\mathcal{E}_i = a_i. \quad (3.11)$$

Таким образом, коэффициент модели (3.4)  $a_i$  сразу определяет значение эластичности  $y$  по фактору  $x_i$  на интервале  $(1, T)$ .

Удобство экономической интерпретации параметров модели (3.4), относительная простота ее записи и послужили причиной ее широкого использования, особенно в макроэкономических исследованиях.

Например, двухфакторная функция Кобба Дугласа:

$$y_t = a_0 x_{1,t}^{a_1} x_{2,t}^{1-a_1} \varepsilon_t \quad (3.12)$$

обычно применяется в макроэкономических исследованиях при анализе взаимосвязи между объемом полученного валового внутреннего продукта ( $y$ ) и используемыми ресурсами ( $x_1$  – основные фонды и  $x_2$  – затраты живого труда).

Функция с постоянной эластичностью замены (3.8) обычно используется в предположении о постоянстве эластичности замещения изменения одного фактора соответствующим изменением другого, обеспечивающего постоянство зависимой переменной  $y$ . Иными словами, значение этого коэффициента показывает, на сколько процентов необходимо изменить значение фактора  $j$  при изменении фактора  $i$  на 1 % при условии, что зависимая переменная не изменится. Значения других факторов при этом предполагаются неизменными. Таким образом, эластичность замещения определяется выражением:

$$\varepsilon_{i,j} = \frac{dx_i}{dx_j} \times \frac{x_j}{x_i}. \quad (3.13)$$

Проводя расчеты по формуле (3.13) для функции (3.8), получим, что для всех  $i, j$  и для всех значений  $t = 1, 2, \dots, T$  эластичность замещения прироста одного фактора соответствующим изменением другого является постоянной:

$$\varepsilon_{i,j} = \frac{1}{1+p}.$$

Для многих практических исследований столь строгие теоретические концепции о характере взаимодействия между переменными отступают на второй план. Для них главным является установление взаимосвязи между переменными  $y$  и  $x_i$ ,  $i = 1, 2, \dots, n$ , наиболее адекватной тенденциям изменений этих величин на временном интервале  $(1, T)$ . Правильный выбор формы таких взаимосвязей обеспечит наилучшее приближение теоретических (расчетных) значений:

$$y_t = f_t(a, x)$$

к действительным значениям  $y_t$ . Обычно такой выбор осуществляется на основе графического анализа тенденций развития соответствующих процессов. Например, если переменные  $y$  и  $x_i$  изменялись во времени согласно графикам, представленным на рис. 3.1, то логично предположить, что  $y \sim 1/x_{it}$ .

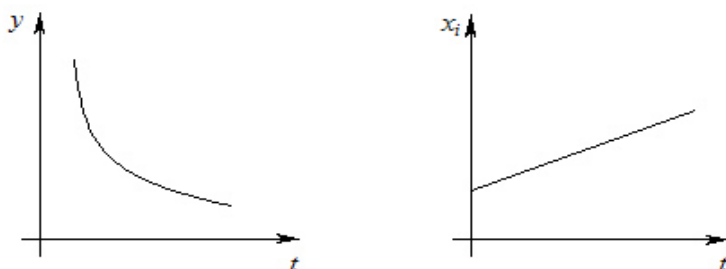


Рисунок 3.1 – Графики гиперболической и линейной функций

Для графиков, представленных на рис. 3.2, характерной является логарифмическая зависимость  $y_t \sim \ln x_{it}$ .

В этих и во многих других случаях, как правило, с учетом замены переменных, в качестве функции  $f(a, x)$  выбирается линейная форма (3.2). Заметим, что значение частичной эластичности  $y$  по фактору  $x_i$ , рассчитанное на основе выражения (3.13) для функции (3.2) равно:

$$\mathfrak{E}_{i,t} = a_i \frac{x_{i,t}}{y_t}.$$

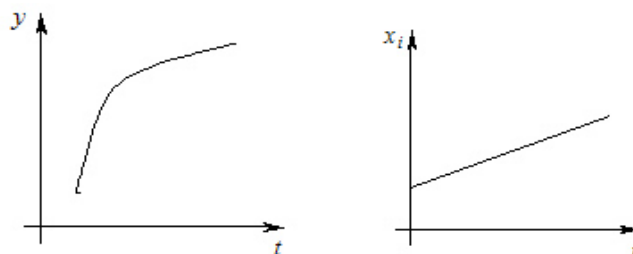


Рисунок 3.2 – Графики логарифмической и линейной зависимости

Таким образом, этот показатель изменяется во времени в соответствии с изменениями  $y$  и  $x_i$ .

Аналогично можно показать, что эластичность замещения факторов  $i$  и  $j$  для функции (3.2) также является переменной величиной:

$$\mathfrak{E}_{i,j,t} = -\frac{a_1 x_{2,t}}{a_2 x_{1,t}},$$

и ее значение также зависит от соотношения уровней рассматриваемых факторов в каждый момент времени.

### 3.3. Процесс формирования эконометрической модели

Процесс построения и использования эконометрических моделей является достаточно сложным и включает следующие основные этапы:

- определение цели исследования;
- построение системы показателей и логический отбор факторов, наиболее влияющих на каждый показатель;
- выбор формы связи изучаемых показателей между собой и отобранными факторами, другими словами, выбор типа эконометрической модели;
- сбор исходных данных и анализ информации;
- построение эконометрической модели, т. е. определение ее параметров;
- проверка качества построенной модели, в первую очередь ее адекватности изучаемому экономическому процессу; использование модели для экономического анализа и прогнозирования.

При практической реализации указанных этапов очень важным является построение системы показателей исследуемого экономического процесса и определение перечня факторов, влияющих на каждый показатель.

Укажем основные требования, предъявляемые к включаемым в эконометрическую модель факторам:

- каждый фактор должен быть обоснован теоретически;

- в модель целесообразно включать только важнейшие факторы, оказывающие существенное воздействие на изучаемые показатели;
- количество включаемых в модель факторов не должно превышать одной трети от числа наблюдений в выборке (длины временного ряда).
- факторы не должны быть линейно зависимы, поскольку эта зависимость означает, что они характеризуют аналогичные свойства изучаемого явления. Например, заработная плата работников зависит, наряду с другими факторами, от роста производительности труда и от объема выпускаемой продукции. Однако эти факторы могут быть тесно взаимосвязаны, коррелированы и, следовательно, в модель целесообразно включать только один из этих факторов. Включение в модель линейно взаимосвязанных факторов приводит к возникновению явления мультиколлинеарности, которое отрицательно сказывается на качестве модели;
- факторы, влияющие на экономический процесс, могут быть количественные и качественные;
- в модель включаются только такие факторы, которые могут быть численно измерены;
- в одну модель нельзя включать совокупный фактор и образующие его частные факторы. Одновременное включение таких факторов приводит к неоправданно увеличенному их влиянию на зависимый показатель, к искажению реальной действительности.

При отборе влияющих факторов используются статистические методы отбора. Выбор метода отбора позволяет задать то, каким образом независимые переменные включаются в анализ. Используя различные методы, можно построить целый ряд эконометрических моделей для одного и того же набора переменных.

- *Ввод (регрессия).* Процедура для выбора переменной, когда все переменные в блоке вводятся на одном шаге.
- *Пошаговый.* На каждом шаге в уравнение включается новая независимая переменная с наименьшей вероятностью  $F$ , при условии, что эта вероятность достаточно мала. Переменные, уже введенные в регрессионное уравнение, исключаются из него, если их вероятность  $F$  становится достаточно большой. Алгоритм останавливается, когда не остается переменных, удовлетворяющих критерию включения или исключения.
- *Удалить.* Процедура отбора переменных, при которой все переменные блока исключаются на одном шаге.
- *Отбор исключением.* Процедура отбора переменных, при которой все переменные вводятся в уравнение, а затем последовательно исключаются из него. Первым кандидатом на удаление считается переменная, имеющая наименьшую частную корреляцию с зависимой переменной. Если она удовлетворяет критерию исключения, ее удаляют. Следующим кандидатом на исключение становится переменная, имеющая наименьшую среди оставшихся переменных частную корреляцию с

зависимой переменной. Процедура останавливается, когда не остается переменных, удовлетворяющих критерию исключения.

- *Отбор включением.* Шаговая процедура отбора переменных, при которой переменные последовательно включаются в модель. Первым кандидатом на ввод служит переменная с наибольшим модулем корреляции с зависимой переменной. Если эта переменная удовлетворяет критерию ввода, она включается в модель. Если первая переменная включена в модель, то следующим кандидатом на включение среди оставшихся вне модели переменных становится переменная, имеющая наибольшую частную корреляцию. Процедура останавливается, когда не остается переменных, удовлетворяющих критерию ввода.

Значения значимостей в выводе результатов основаны на подгонке единственной модели. Поэтому значения значимостей, как правило, некорректны при применении шагового метода (Шаговый отбор, Включение или Исключение).

Вне зависимости от выбранного метода отбора, каждая переменная должна удовлетворять критерию допуска (толерантности) для того, чтобы быть введенной в уравнение. По умолчанию значение уровня толерантности (допуска) равно 0,0001. Кроме того, переменная не будет введена в модель, если это повлечет за собой снижение толерантности переменной, уже введенной в уравнение, до величины меньшей, чем значение критерия допуска.

Существенного сокращения числа влияющих факторов можно достичь с помощью процедур отбора переменных. Но ни одна из этих процедур не гарантирует получения оптимального набора переменных. Однако при практическом применении они позволяют получать достаточно хорошие наборы существенно влияющих факторов; кроме того, их можно сочетать с другими подходами к решению данной проблемы, например, с экспертными оценками значимости факторов. Среди пошаговых процедур отбора факторов наиболее часто используются процедуры пошагового включения и исключения факторов. Обе эти процедуры хорошо формализованы и потому успешно реализованы в различных машинных программах статистического анализа.

Метод исключения предполагает построение уравнения, включающего всю совокупность переменных, с последующим последовательным (пошаговым) сокращением числа переменных в модели до тех пор, пока не выполнится некоторое наперед заданное условие. Суть метода включения – в последовательном включении переменных в модель до тех пор, пока регрессионная модель не будет отвечать заранее установленному критерию качества. Последовательность включения определяется с помощью частных коэффициентов корреляции: переменные, имеющие относительно исследуемого показателя большее значение частного коэффициента корреляции, первыми включаются в регрессионное уравнение.

Выше отмечено, что одной из предпосылок применения методов регрессионного анализа для построения эконометрических моделей является отсутствие среди независимых переменных (факторов) линейно связанных. Если данная предпосылка не выполняется, то возникает, как уже сказано выше,



явление мультиколлинеарности, т. е. наличие сильной корреляции между независимыми переменными (включенными в модель факторами). В математическом аспекте мультиколлинеарность приводит к слабой обусловленности матрицы системы нормальных уравнений, т. е. близости ее определителя к нулю, а в содержательном аспекте – к искажению смысла коэффициентов регрессии и затруднению выявления наиболее существенно влияющих факторов.

Основные причины, вызывающие мультиколлинеарность – независимые переменные, либо характеризующие одно и то же свойство изучаемого явления, либо являющиеся составными частями одного и того же признака.

В настоящее время существует ряд методов, позволяющих оценить наличие мультиколлинеарности в совокупности независимых переменных, измерить ее степень, выявить взаимно коррелированные переменные и устранить или ослабить ее негативное влияние на регрессионную модель. Наиболее распространенным методом выявления мультиколлинеарности является метод корреляции. На практике считают, что две переменные коллинеарны (линейно зависимы), если парный коэффициент корреляции между ними по абсолютной величине превышает 0,8. Устраняют мультиколлинеарность чаще всего путем исключения из модели одного из коррелированных факторов.

### **3.4. Эконометрические модели дискретного выбора**

Модели дискретного выбора [discrete choice models], иначе называемые моделями качественного отклика [qualitative response models], определяют вероятностное распределение дискретных зависимых переменных как функцию независимых переменных и неизвестных параметров. Их применение в эконометрике определяется тем, что решение экономического субъекта часто включает дискретный выбор, например, решение поступить на работу или не поступать, выбор занятия, выбор маршрута перевозки груза и т. п.

Модели дискретного выбора – это экономические (эконометрические) модели, позволяющие описывать, объяснять и прогнозировать выбор между двумя или более альтернативами (то есть когда множество альтернатив не более чем счетно). Модели дискретного выбора позволяют на основе некоторых характеристик (атрибутов) экономического субъекта или ситуации оценить вероятность выбора той или иной альтернативы.

В каком-то смысле эти модели противоположны агрегированным макроэкономическим моделям, которые описывают массовые, а не индивидуальные факты. В разных постановках модели дискретного выбора в качестве математического аппарата применяют:

- а) цепи Маркова;
- б) модели с бинарными переменными;
- в) многомерные модели (совместное распределение вероятностей для двух или большего числа дискретных зависимых переменных);

г) случайные выборки и др.

Логит-модель, пробит-модель, тобит-модель относятся к классу моделей, для анализа которых неприменимы обычные методы регрессионного анализа.

### 3.4.1. Логистическая регрессия

Логит-модель или логистическая регрессия (англ. Logit model) – статистическая модель, используемая для прогнозирования вероятности возникновения некоторого события путем его сравнения с логистической кривой. Эта регрессия выдает ответ в виде вероятности бинарного события (1 или 0).

Отличие логит-модели состоит в том, что зависимая переменная (признак-фактор) может принимать лишь ограниченное число значений, в простейшем случае – либо 0, либо 1.

$$y = \frac{1}{1+e^{-z}}.$$

Задача состоит в том, чтобы определить вероятность принятия зависимой переменной значения из заданного набора. Для этого в качестве аналитического средства применяется логистическая функция, выраженная в логарифмической форме.

Логистическая регрессия применяется для прогнозирования вероятности возникновения некоторого события по значениям множества признаков. Для этого вводится так называемая зависимая переменная  $y$ , принимающая лишь одно из двух значений – как правило, это числа 0 (событие не произошло) и 1 (событие произошло), и множество независимых переменных (также называемых признаками, предикторами или регрессорами) – вещественных  $x_1, x_2, \dots, x_n$ , на основе значений которых требуется вычислить вероятность принятия того или иного значения зависимой переменной. Как и в случае линейной регрессии, для простоты записи вводится фиктивный признак  $x_0 = 1$ .

Делается предположение о том, что вероятность наступления события  $y = 1$  равна:

$$P\{y = 1 | x\} = f(z),$$

где  $z = \theta^T x = \theta_0 + \theta_1 x_1 + \dots + \theta_n x_n$ ,  $x$  и  $\theta$  – векторы-столбцы значений независимых переменных  $1, x_1, \dots, x_n$  и параметров (коэффициентов регрессии) – вещественных чисел  $\theta_0, \dots, \theta_n$ , соответственно;

$f(z)$  – так называемая приведенная выше логистическая функция (иногда также называемая сигмоидом или логит-функцией):

$$f(z) = \frac{1}{1+e^{-z}}.$$

Для подбора параметров  $\theta_0, \dots, \theta_n$  необходимо составить обучающую выборку, состоящую из наборов значений независимых переменных и соответствующих им значений зависимой переменной  $y$ . Формально, это множество пар  $(x^{(1)}, y^{(1)}), \dots, (x^{(m)}, y^{(m)})$ , где  $x^{(i)} \in R^n$  – вектор значений независимых переменных, а  $y^{(i)} \in \{0, 1\}$  – соответствующее им значение  $y$ . Каждая такая пара называется обучающим примером.

Обычно используется метод максимального правдоподобия, согласно которому выбираются параметры  $\theta$ , максимизирующие значение функции правдоподобия на обучающей выборке.

Для улучшения обобщающей способности получающейся модели, то есть уменьшения эффекта переобучения, на практике часто рассматривается логистическая регрессия с регуляризацией.

Регуляризация заключается в том, что вектор параметров  $\theta$  рассматривается как случайный вектор с некоторой заданной априорной плотностью распределения  $p(\theta)$ . Для обучения модели вместо метода наибольшего правдоподобия при этом используется метод максимизации апостериорной оценки, то есть ищутся параметры  $\theta$ , максимизирующие величину:

$$\prod_{i=1}^m P\{y^i|x^i, \theta\} \times p(\theta).$$

В качестве априорного распределения часто выступает многомерное нормальное распределение  $N(0, \sigma^2 I)$  с нулевым средним и матрицей ковариации  $\sigma^2 I$ , соответствующее априорному убеждению о том, что все коэффициенты регрессии должны быть небольшими числами, идеально – многие малозначимые коэффициенты должны быть нулями. Подставив плотность этого априорного распределения в формулу выше, и прологарифмировав, получим следующую оптимизационную задачу:

$$\sum_{i=1}^m \log P\{y^i|x^i, \theta\} - \|\theta\|^2 \rightarrow \max,$$

где  $\lambda = \text{const}/\sigma^2$  – параметр регуляризации. Этот метод известен как  $L2$ -регуляризованная логистическая регрессия, так как в целевую функцию входит  $L2$ -норма вектора параметров для регуляризации.

Эта модель часто применяется для решения задач классификации – объект  $x$  можно отнести к классу  $y = 1$ , если предсказанная моделью вероятность:

$$P\{y = 1|x\} > 0,5,$$

и к классу  $y = 0$  в противном случае. Получающиеся при этом правила классификации являются линейными классификаторами (*Примечание: Линейный классификатор – способ решения задач классификации, при котором решение принимается на основании линейного оператора над входными данными. Линейные классификаторы представляют семейство статистических ранжирования алгоритмов. Роль классификатора состоит в том, чтобы классифицировать по группам (классам) образцы, которые имеют аналогичные свойства, измеренные на основе наблюдений. Линейный классификатор – это особый тип классификатора, который вычисляет решение путем линейной комбинации выборок).*

### 3.4.2. Пробит-регрессия

Пробит-регрессия (пробит-модель, англ. probit) – применяемая в различных областях (эконометрика, токсикология и др.) статистическая (нелинейная) модель и метод анализа зависимости качественных (в первую очередь – бинарных) переменных от множества факторов, основанная на нормальном распределении (в отличие от, например, аналогичной логит-регрессии, основанной на логистическом распределении). В экономике (эконометрике) пробит-модели (наряду с логит-, гомпит- и др.) используются в моделях бинарного выбора или в моделях множественного выбора между различными альтернативами, для моделирования дефолтов компаний, в страховании жизни – для оценки вероятности смерти в зависимости от возраста и пола и т. д. В токсикологии пробит-регрессия используется для оценки влияния дозы или концентрации тех или иных веществ на биологические объекты.

**Пробит-модель** (probit model) – это статистическая модель бинарного выбора, используемая для предсказания вероятности возникновения интересующего события на основе функции стандартного нормального распределения.

Отличие пробит-модели от логит-модели состоит в том, что вместо логистической функции для описания вероятности наступления события используется функция стандартного нормального распределения.

$$P(Y_i = 1) = \Phi(z_i) = \Phi(\beta_1 + \beta_2 \times x_i^2 + \dots + \beta_k \times x_i^k),$$

$$z_i = \beta_1 + \beta_2 \times x_i^2 + \dots + \beta_k \times x_i^k.$$

Здесь  $\Phi(z_i)$  – функция стандартного нормального распределения

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}.$$

Оценивание, тестирование гипотез и интерпретация результатов в рамках пробит-анализа проводится полностью аналогично случаю логит-анализа с поправкой на использование функции стандартного нормального распределения вместо логистической функции. В частности, предельный эффект изменения переменной  $x^{(j)}$  может быть вычислен вот так:

$$\begin{aligned} \frac{\partial P(Y_i = 1)}{\partial x^j} &= \Phi'(\beta_1 + \beta_2 \times x_i^2 + \dots + \beta_k \times x_i^k) \times \beta_j = \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{(\beta_1 + \beta_2 \times x_i^2 + \dots + \beta_k \times x_i^k)^2}{2}} \times \beta_j. \end{aligned}$$

Здесь  $\Phi'(\cdot)$  – производная от функции стандартного нормального распределения, то есть функция плотности вероятности стандартного нормального распределения.

Пробит-модель позволяет оценить вероятность того, что анализируемая (зависимая) переменная примет значение 1 при заданных значениях факторов (то есть это оценка доли «единиц» при данном значении факторов). В пробит-модели пробит-функция от вероятности моделируется как линейная комбинация факторов (включая константу). Пробит-функцией принято называть функцию, обратную к интегральной функции (CDF) стандартного нормального

распределения, то есть функцию, определяющую квантиль стандартного нормального распределения для заданной вероятности  $x_q = \Phi^{-1}(q)$ .

Термин «probit» как производное от англ. *probability unit* предложил (впервые использовал) Честер Блисс (Chester Ittner Bliss [1899–1979]) в своей статье, посвященной количественному анализу смертельного действия ядов на примере действия никотина на щавелевую тлю (*Aphisrumicis* L.). С тех пор метод пробит-анализа особенно популярен в токсикологии. Само использование функции нормального распределения для описания зависимости «доза – эффект» восходит к английскому математику J. W. Trevan, который показал, что интенсивность клеточного ответа на данную дозу лекарственного вещества подчиняется распределению Гаусса.

Пробит-модель является частным случаем модели бинарного выбора, в которой используется нормальное распределение. А именно, пусть зависимая переменная  $Y$  является бинарной, то есть может принимать только два значения, которые для упрощения предполагаются равными 1 и 0. Например,  $Y$  может означать наличие/отсутствие каких-либо условий, успех или провал чего-либо, ответ да/нет в опросе и т. д. Пусть также имеется вектор регрессоров (факторов)  $X$ , которые оказывают влияние на  $Y$ . В пробит-модели предполагается, что вероятность того, что  $Y = 1$  определяется нормальным распределением, таким образом пробит-модель имеет вид:

$$p(x) = P(Y = 1 | X = x) = \Phi(x^T b),$$

где  $\Phi$  – интегральная функция распределения (CDF) стандартного нормального распределения;

$b$  – неизвестные параметры, которые требуется оценить.

Использование именно стандартного нормального распределения не ограничивает общности модели, так как возможное ненулевое среднее учтено в константе, которая обязательно присутствует в числе факторов, а возможная неединичная дисперсия учитывается за счет соответствующего нормирования всех коэффициентов  $b$ .

Как и в общем случае модели бинарного выбора в основе модели лежит предположение о наличии некоторой скрытой (ненаблюдаемой) переменной  $Y^*$ , в зависимости от значений которой наблюдаемая переменная  $Y$  принимает значение 0 или 1:

$$Y = \begin{cases} 1, Y^* > 0 \\ 0, Y^* < 0 \end{cases}$$

Предполагается, что скрытая переменная зависит от факторов  $X$  в смысле обычной линейной регрессии  $y^* = x^T b + \varepsilon$ , где случайная ошибка в данном случае имеет стандартное нормальное распределение  $N(0, 1)$ . Тогда

$$\begin{aligned} p(x) &= P(Y^* > 0 | X = x) = P(x^T b + \varepsilon > 0) = P(\varepsilon > -x^T b) = 1 - \Phi(-x^T b) \\ &= \Phi(x^T b). \end{aligned}$$

Последнее равенство следует из симметричности нормального распределения.

Также модель может быть обоснована через полезность альтернатив – не наблюдаемой функции  $U(y, x)$ , то есть фактически двух функций:

$$U_1(x) = x^T b_1 + \varepsilon_1 \text{ и } U_0(x) = x^T b_0 + \varepsilon_0$$

соответственно для двух альтернатив. Функция разности полезностей альтернатив здесь выполняет роль той самой скрытой переменной.

Пробит-модель содержит дихотомические переменные («все или ничего»). Например – модели принятия решений: владеть или арендовать собственность, модели выбора маршрута или профессии.

Для оценки пробит-модели применяется метод максимального правдоподобия.

### 3.4.3. Тобит-регрессия

Тобит-регрессия (цензурированная регрессия, англ. Censored regression) – регрессия, с зависимой переменной, наблюдаемой с ограничением (цензурированием) возможных значений. При этом модель может быть цензурирована только с одной стороны (справа или слева) или с обеих сторон. Цензурированная регрессия отличается от усеченной регрессии (англ. truncated regression), тем что значения факторов, в отличие от зависимой переменной, наблюдаются без ограничений.

Каноническая цензурированная регрессия, цензурированная снизу нулевым значением, носит название тобит (по аналогии с пробит, логит и т. д.), названная в честь лауреата премии имени Нобеля по экономике Джеймса Тобина. Собственно, исследования цензурированных моделей начались с работы Дж. Тобина в 1958 году, в которой рассматривались расходы семей на автомобили. Для оценки эластичности спроса на автомобили по доходу необходимо оценить модель зависимости логарифма расходов на логарифм доходов. Однако, как показал Тобин, такая оценка будет смещенной и несостоятельной, так как для семей с низким доходом (ниже некоторого порога) величина расходов равна нулю независимо от конкретной величины дохода и других факторов. Тобин впервые и предложил подход к оценке таких моделей, который позволяет получить состоятельные оценки параметров модели.

Тобит-модель включает как непрерывные, так и дискретные переменные. Например, модели рынков, на которых часть цен свободна, а часть – нет.

Модель Тобита – это любой из класса моделей регрессии, в котором наблюдаемый диапазон зависимой переменной каким-то образом подвергнут цензуре. Для оценки моделей применяется метод максимального правдоподобия.

### 3.5. Линейные модели бинарного выбора

Модели бинарного выбора описывают выбор между двумя альтернативами. Это формализуется с помощью переменной  $y$ , принимающей 0 для одной альтернативы и 1 для другой. Для такой переменной *математическое ожидание* равно вероятности  $p$  выбора «единицы». Вероятность выбора «нуля» равна  $1-p$ .

Пусть  $x$  – набор факторов, от которых может зависеть выбор. Предполагается, что выбор определяется некоторым неявным механизмом, связанным с тем, превышает или нет некоторая неявная переменная  $y^*$ , зависящая от факторов  $x$ , некоторое пороговое значение. Без потери общности в качестве порогового значения всегда можно использовать 0. Зависимость неявной переменной от факторов является вероятностной, так как при моделировании выбора выбираются не все возможные факторы, а лишь наиболее существенные, поэтому в модели имеется случайная компонента. Обычно для  $y^*$  предполагается модель линейной регрессии по факторам  $x$ :

$$y^* = x^T b + \varepsilon,$$

где  $b$  – параметры модели (в том числе обязательно константа  $b_0$ );

$\varepsilon$  – случайная компонента.

Выбор осуществляется следующим образом: если  $y^* > 0$ , то выбирается альтернатива  $y = 1$ , иначе –  $y = 0$ . Следовательно искомая вероятность  $p$  равна:

$$P = P(y^* > 0) = P(x^T b + \varepsilon > 0) = P(\varepsilon > -x^T b) = 1 - P(\varepsilon < -x^T b) = 1 - F(-x^T b),$$

где  $F$  – функция распределения случайной компоненты  $\varepsilon$ .

При этом функция распределения может быть определена с точностью до некоторых неизвестных параметров  $\theta$ , которые тоже подлежат оценке на основе статистических данных вместе с параметрами  $b$ . Если распределение симметричное, то есть  $1 - F(-z) = F(z)$ , то модель бинарного выбора принимает вид:

$$p = F(x^T b).$$

В качестве функций распределения чаще всего используются нормальное распределение (пробит-модель) или логистическое распределение (логит-модель).

Модель бинарного выбора – применяемая в эконометрике модель зависимости бинарной переменной (принимающей всего два значения – 0 и 1) от совокупности факторов. Построение обычной линейной модели для таких зависимых переменных теоретически некорректно, так как условное математическое ожидание таких переменных равно вероятности того, что зависимая переменная примет значение 1, а линейная модель допускает с тем числе отрицательные значения и значения выше 1 (притом, что вероятность должна быть от 0 до 1). Поэтому обычно используются некоторые интегральные функции распределения. Чаще всего используются нормальное распределение (пробит), логистическое распределение (логит), распределение Гомперца (гомпит).

Еще одно обоснование заключается в использовании понятия полезности альтернатив – не наблюдаемой функции:

$$U(y, x),$$

то есть фактически двух функций:

$$U_1(x) = x^T b_1 + \varepsilon_1 \text{ и } U_0(x) = x^T b_0 + \varepsilon_0$$

соответственно для двух альтернатив. Логично предположить, что если при заданных значениях факторов полезность одной альтернативы больше полезности другой, то выбирается первая и наоборот. В связи с этим разумно рассмотреть функцию разности полезностей альтернатив:

$$\Delta U(x) = U_1(x) - U_0(x) = x^T(b_1 - b_0) + (\varepsilon_1 - \varepsilon_0) = x^T b + \varepsilon.$$

Если она больше нуля, то выбирается первая альтернатива, если меньше или равна нулю – то вторая. Таким образом, функция разности полезностей альтернатив здесь выполняет роль той самой скрытой переменной. Наличие случайной ошибки в моделях полезностей позволяет учесть не абсолютную детерминированность выбора (по крайней мере не детерминированность данным набором факторов, хотя элемент случайности выбора есть при любом наборе факторов).

В пробит-модели в качестве  $F$  используется интегральная функция стандартного нормального распределения  $\Phi$ :

$$p(x) = 1 - \Phi(-x^T b) = \Phi(x^T b).$$

В логит-модели используется CDF логистического распределения:

$$p(x) = 1 - e^{-x^T b} / (1 + e^{-x^T b}) = e^{x^T b} / (1 + e^{x^T b}).$$

В гомпит-модели используется распределение экстремальных значений – распределение Гомперца:

$$p(x) = 1 - e^{-(e^{(-x^T b)})} = e^{-(1 - e^{(-x^T b)})}.$$

Оценка моделей производится методом максимального правдоподобия.

Во многих случаях приходится иметь дело не с двумя, а с несколькими альтернативами. В таком случае говорят о *моделях множественного выбора*. Задача этих моделей – оценить вероятности  $p_i$  выбора различных альтернатив ( $i = 1, \dots, k$ ). В случае, если эти альтернативы можно упорядочить каким-либо образом, то говорят о *моделях упорядоченного выбора*.



## 4. КЛАССИФИКАЦИЯ В ЭКОНОМЕТРИКЕ

### 4.1. Процесс классификации

Классификация заключается в разделении множества объектов технико-экономической и социальной информации на подмножества по их сходству или различию в соответствии с принятыми методами, обеспечивающими систематизацию объектов классификации по определенным выбранным признакам (свойства, характеристики или параметры объектов). Количество значений признака определяет число образуемых классификационных группировок. Для классификации продукции используются, например, следующие признаки: отраслевая принадлежность, назначение, область применения, принцип действия, конструктивные особенности, используемый для изготовления способ и материал. В частности, в Общероссийском классификаторе продукции (ОКП) трубы стальные классифицируются:

- по способу изготовления – тянутые, сварные, центробежнолитые и др.;
- по назначению – водогазопроводные, нефтепроводные, подшипниковые и др.;
- по виду материала – нержавеющие, углеродистые, эмалированные;
- по размеру – диаметром 1020, 1220, 1420 мм и др.

Для описания процесса классификации информации используются следующие термины и определения понятий:

- Система классификации – совокупность методов и правил классификации и ее результат.
- Объект классификации – элемент классифицируемого множества.
- Признак классификации – свойство или характеристика объекта, по которому производится классификация.
- Степень классификации – этап классификации при иерархическом методе, в результате которого получается совокупность классификационных группировок.
- Глубина классификации – число ступеней классификации, которое зависит от степени конкретизации группировок и числа признаков, необходимых для решения конкретных задач.
- Группировка классификационная – подмножество объектов, полученное в результате классификации.

Методы классификации подразделяются на иерархический и фасетный.

Иерархический метод классификации – последовательное разделение множества объектов на подчиненные классификационные группировки. В этом методе множество объектов делится сначала по некоторому выбранному признаку (основанию деления) на крупные группировки, затем каждая из этих группировок делится на ряд последующих группировок по другому признаку, конкретизируя объект классификации. Таким образом между классификационными группировками устанавливается иерархия (подчиненность).

Фасетный метод классификации – параллельное разделение множества объектов на независимые классификационные группировки. В этом методе классификационное множество объектов информации описывается набором независимых фасетов (списков), не имеющих жесткой взаимосвязи друг с другом, которые можно использовать отдельно для решения различных задач. Например, в Общероссийском классификаторе информации о населении (ОКИН) используются следующие фасеты: пол, возраст, гражданство, национальность, язык, родство и др.

И так классификация заключается в разбиении множества на классы (подмножества) с целью выявления его структуры, изучения связей его элементов (главным образом, на основе выявления их сходства/различия). При этом под классификацией понимают как саму указанную процедуру, так и ее результат, т. е. представление исходного множества через объединение его подмножеств. Эта процедура равносильна заданию на исходном множестве отношения эквивалентности (два элемента множества находятся в отношении эквивалентности тогда и только тогда, когда принадлежат одному классу). В таком случае получается строгая классификация:

- объединение классов совпадает с исходным множеством;
- выделенные классы не пересекаются;
- среди них нет пустых.

Иначе говоря, каждый элемент попадает в какой-либо класс, никакой элемент не попадает хотя бы в два класса и в каждом классе содержится хотя бы один элемент. Формально: пусть  $M$  – исходное множество,  $M_i$  – его классы,  $n$  – число классов, так что  $i = 1, 2, \dots, n$ . Тогда  $M = \bigcup_{i=1}^n M_i$ , причем  $M_i \cap M_j = \emptyset$  при  $i \neq j$  и  $M_i \neq \emptyset$  для каждого  $i$ .

Процесс классификации может быть продолжен, если процедура классификации применяется хотя бы к некоторым классам  $M_i$ . Тогда получается многоуровневая иерархия:

- исходному множеству соответствует 0-й уровень;
- классам, которые получены его разбиением, – 1-й;
- далее образуются классы 2-го уровня и т. д.

Формирование классов производится по какому-либо содержательному признаку, называемому основанием классификации (например, семьи можно классифицировать по душевому доходу). Для классов одного уровня основание классификации должно быть общим, хотя иногда встречаются нарушения этого правила – в плохо структурированных задачах, где только отыскиваются подходы к научной классификации. При переходе к следующему уровню в иерархических классификациях основание обычно меняется, производится замена классификационного признака (например, семьи любого класса, выделенного по доходу, можно разбить на два подкласса в соответствии с тем, где они проживают – в городе или на селе). Однако иногда классы разбивают на подклассы, используя тот же признак, но при его более дробных (или дезагрегированных) значениях (например, для классификации семей по душевому доходу при

формировании классов первого уровня можно воспользоваться интервалами дохода с шагом в 10 тыс. руб. в год, а второго – 2 тыс. руб.).

Если признак, используемый в качестве основания классификации при формировании очередного уровня, может быть представлен как принимающий только два значения (0 или 1), он называется бинарным.

Если классификация построена при использовании только бинарных признаков, она называется бинарной. Например, классификация семей по душевому доходу может быть бинарной, если на первом шаге весь интервал значений дохода  $(0, \infty)$  разбить на два подинтервала –  $(0, d)$  и  $(d, \infty)$ , на втором шаге аналогично можно разбить каждый из этих подинтервалов и т. д.

Именно бинарные классификации во многих отношениях наиболее удобны в использовании. В частности, они обладают самой простой структурой и наилучшим образом приспособлены для формального (логического или математического) описания.

Даже в случаях, когда используемые в качестве оснований классификации признаки дискретны, она может представлять весьма сложную задачу. Так, хотя биологические классификации насчитывают не менее двух с половиной веков, в них почти непрерывно производятся изменения, причем далеко не всегда они обусловлены открытием новых биологических видов, т. е. расширением исходного множества. Тем более сложной может оказаться задача классификации в случаях, когда признаки различия классифицируемых объектов нечетки, очень разнообразны по содержанию, способам идентификации и измерения. Именно такие задачи классификации встречаются в социально-экономических исследованиях. В подобных случаях широко применяются методы автоматической классификации.

Идея классификации пронизывает всю современную науку и является одной из основополагающих для научной методологии. «Объект» научного исследования на самом деле представляет (репрезентирует) совокупность, класс объектов реального мира (например, моря, деревья, органы слуха) либо абстрагированные от них формы, свойства, признаки (например, длина, поверхность, периодичность, твердость, размытость, упорядоченность), которые, опять-таки, могут быть выявлены только через соотнесение какому-либо классу. Тем самым, уже на стадии формирования объекта научного исследования решается, по сути, классификационная задача: вычленяется класс объектов «вещного» мира. Для объединения (мысленного, концептуального) реальных объектов в такой класс в них необходимо усмотреть сходство в том или ином смысле, зависящем, естественно, от цели научного исследования. Такой подход вполне определенно просматривается еще у Аристотеля.

По-видимому, первой целенаправленно и последовательно построенной классификацией, была биологическая систематика, предложенная шведским естествоиспытателем К. Линнеем в середине 18 в. (эта классификация – строго бинарная). Периодическая система химических элементов Д. И. Менделеева также представляет классификацию. Она была построена через упорядочивание химических элементов по атомному весу (точнее, как выяснилось впоследствии,

по заряду атомного ядра), а выявившаяся при этом периодичность элементов по химическим свойствам позволила строго сформировать их классы (щелочные металлы, щелочноземельные металлы и т. д. вплоть до инертных газов). Тот же результат был бы получен, если процедуру классификации изначально основывать на сходстве элементов по их химическим свойствам.

Основные признаки классификации всех задач, рассматриваемых эконометрикой:

1. Классификация задач в зависимости от конечных прикладных целей (прогноз социально-экономических показателей, моделирование возможных вариантов социально-экономического развития системы, выявление факторов, оказывающих наиболее мощное влияние на состояние системы в целом).

2. Классификация задач в зависимости от уровня иерархии (макроуровень, мезоуровень, микроуровень).

3. Классификация задач в зависимости от профиля изучаемой экономической системы (рынок, инвестиции, спрос, потребление, ценообразование и прочее).

Много общих черт существует у эконометрики и статистики – в некоторой степени, эконометрика является частью статистической науки: статистика изучает глобальные явления всех направлений, в том числе и экономических, а эконометрика занимается теми же процессами, но ограничивается исключительно экономическими задачами.

Термин «Классификация» используется также для обозначения любого упорядочения рассматриваемых объектов.

## **4.2. Классификация многомерных наблюдений**

Классификация многомерных наблюдений представляет разделение рассматриваемой совокупности объектов  $O_1, O_2, \dots, O_n$ , каждый из которых ( $O_i$ ) представлен вектором характеризующих его признаков  $X_i = (x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(p)})$ , на некоторое число (заранее заданное или нет) однородных в определенном смысле групп. При этом термин «классификация», в зависимости от контекста, используют как для обозначения самого процесса разделения, так и его результата. Это понятие тесно связано с такими терминами, как группировка, типологизация, систематизация, дискриминация, кластеризация, образование таксонов, и является одним из основополагающих в практической и научной деятельности человека.

Если наряду с объектами, подлежащими классификации, исследователь располагает «представительскими порциями» полностью описанных объектов (т. е. с известными значениями анализируемых признаков и адресами классов, к которым они принадлежат) от каждого из классов, то такую информацию называют «обучающей» (или «обучающими выборками»), а самую задачу – задачей классификации с обучением (задачей распознавания образов, задачей дискриминантного анализа).

Именно к таким ситуациям относятся типичные задачи медицинской диагностики, когда в клинических условиях в качестве исходных данных исследователь располагает как «входами» – результатами инструментальных обследований пациентов, так и «выходами» – уже установленным диагнозом («болен» – «здоров») по каждому из них. Цель исследований такого типа – использование имеющегося «обучения» для отбора из множества результатов обследований небольшого числа наиболее информативных (с точки зрения диагностической силы) показателей и для построения на их основе формального диагностирующего правила.

Однако в задачах социально-экономического профиля исследователь в качестве исходных данных располагает, как правило, лишь данными об объектах, подлежащих классификации (получение обучающей информации здесь связано с организацией трудоемкой и дорогостоящей специальной системы предварительных экспертных оценок). В подобных ситуациях, т. е. в случаях, когда приходится решать задачу классификации, не располагая обучающими выборками, говорят об автоматической классификации (или о классификации без обучения, кластер-анализе, численной таксономии).

Среди типов прикладных задач классификации (конечных прикладных целей) выделяют:

1) комбинационные группировки и их непрерывные обобщения – разбиение совокупности на интервалы (области) группирования;

2) простая типологизация: выявление естественного расслоения анализируемых данных (объектов) на четко выраженные «сгустки» (кластеры), лежащие друг от друга на некотором расстоянии, но не разбивающиеся на столь же удаленные друг от друга части;

3) связная неупорядоченная типологизация: использование реализованной в пространстве результирующих показателей простой типологизации в качестве обучающих выборок при классификации той же совокупности объектов в пространстве описательных признаков;

4) связная упорядоченная типологизация, отличающаяся от связной неупорядоченной возможностью экспертного упорядочения классов, полученных в пространстве результирующих показателей, и использованием этого упорядочения для построения сводного латентного результирующего показателя как функции от описательных переменных;

5) структурная типологизация: дает на «выходе» задачи дополнительно к описанию классов еще и описание существующих между ними и их элементами структурных (в т. ч. иерархических) связей;

6) типологизация динамических траекторий системы: в качестве классифицируемых объектов выступают характеристики динамики исследуемых систем, например, дискретные или непрерывные временные ряды или траектории систем, которые в каждый момент времени могут находиться в одном из заданных состояний.

Хотя авторы этих классификаций и не располагали современным математическим аппаратом многомерного статистического анализа, его

основные идеи и методологические принципы пронизывают логику их конструкций, а подчас и прямо формулируются.

Основные методологические принципы, лежащие в основе большинства конструкций многомерного статистического анализа:

а) необходимость учета эффекта существенной многомерности анализируемых данных (используемые в конструкциях характеристики должны учитывать структуру и характер статистических взаимосвязей исследуемых признаков);

б) возможность лаконичного объяснения природы анализируемых многомерных структур (допущение, в соответствии с которым существует сравнительно небольшое число определяющих, подчас латентных, т. е. непосредственно не наблюдаемых, факторов, с помощью которых могут быть достаточно точно описаны все наблюдаемые исходные данные, структура и характер связей между ними);

в) максимальное использование «обучения» в настройке математических моделей классификации и снижения размерности (под «обучением» понимается та часть исходных данных, в которой представлены «статистические фотографии» соотношений «входов» и «выходов» анализируемой системы);

г) возможность оптимизационной формулировки задач многомерного статистического анализа (в т. ч. задач классификации), т. е. нахождение наилучшей процедуры статистической обработки данных с помощью оптимизации некоторого экзогенного заданного критерия качества метода.

Первые два принципа относятся к природе обрабатываемых данных, а следующие два – к логике построения соответствующих аппаратных средств.

### **4.3. Автоматическая классификация**

Автоматическая классификация (кластер-анализ, кластерный анализ) представляет совокупность многомерных статистических методов, предназначенных для формирования относительно «отдаленных» друг от друга групп «однородных» объектов по информации о расстояниях или связях (меры близости) между ними. Используется для анализа:

- структуры совокупности социально-экономических показателей по заданной матрице коэффициентов корреляции между ними;
- социально-экономических объектов (предприятий, регионов и т. д.), описанных многими априорно равноправными признаками, и т. п.

Выделяются два основных типа методов кластерного анализа в зависимости от того, одновременно (1) или последовательно (2) отыскиваются кластеры. Среди первых широкое распространение получили т. н. вариационные методы, основанные на оптимизации того или иного показателя качества выявленной кластерной структуры, и агломеративные методы, основанные на последовательном объединении пар наиболее близких кластеров. Среди вторых можно отметить методы, основанные на явном определении понятия кластера, как правило, в терминах максимально допустимого «радиуса» или «порога

существенности» связей. Агломеративные процедуры кластерного анализа обычно включают параметры, задаваемые исследователем (число классов, порог значимости и т. п.), что позволяет получать несколько решений, из которых исследователь выбирает наилучшие с точки зрения интерпретации.

Таким образом, автоматическая классификация (кластер-анализ) – это математически-формализованная процедура разбиения анализируемой совокупности объектов  $O_1, O_2, \dots, O_n$  (или любого из последовательно поступающих «на вход» объектов) на некоторое число (заранее известное или нет) однородных в определенном смысле классов в условиях отсутствия обучающих выборок. При этом исходная информация о классифицируемых объектах представлена либо значениями многомерного признака (по каждому объекту в отдельности), либо матрицей попарных расстояний (или близостей) между объектами, а понятие однородности основано на предположении, что геометрическая близость двух или нескольких объектов означает близость их «физических» состояний, их сходство. Полученные в результате разбиения классы часто называют кластерами (таксонами, образами), а методы их нахождения соответственно кластер-анализом, численной таксономией, распознаванием образов с самообучением.

Математическая постановка задачи автоматической классификации требует формализации понятия «качество разбиения». С этой целью вводится понятие критерия (функционала) качества разбиения  $Q(S)$ , который задает способ сопоставления с каждым возможным разбиением  $S$  заданного множества объектов на классы некоторого числа  $Q(S)$ , оценивающего (в определенной шкале) степень оптимальности данного разбиения. Тогда задача поиска наилучшего разбиения  $S^*$  сводится к решению оптимизационной задачи вида:

$$Q(S) \rightarrow \underset{S \in A}{extr},$$

где  $A$  – множество всех допустимых разбиений.

В зависимости от наличия и характера априорных сведений о природе искомых классов и от конечных прикладных целей применяется одна из трех составных частей математического аппарата классификации в условиях отсутствия обучающих выборок:

1) метод расщепления смесей вероятностных распределений (каждый класс интерпретируется как параметрически заданная одномодальная генеральная совокупность при неизвестном значении определяющего ее параметра, а классифицируемые наблюдения – как выборка из смеси таких генеральных совокупностей);

2) метод собственно кластер-анализа (не имеется оснований для параметризации модели, а иногда и для интерпретации последовательности классифицируемых наблюдений в качестве выборки из генеральной совокупности);

3) классификационные процедуры иерархического типа (главная цель – получение наглядного представления о стратификационной структуре всей классифицируемой совокупности, например, в виде дендрограммы).

**Выбор метрики** (или меры близости) между объектами, каждый из которых представлен значениями характеризующего его многомерного признака, является узловым моментом исследования, от которого решающим образом зависит окончательный вариант разбиения объектов на классы при любом используемом для этого алгоритме разбиения. В каждой конкретной задаче этот выбор должен производиться по-своему, в зависимости от главных целей исследования, физической и статистической природы анализируемого многомерного признака, априорных сведений о его вероятностной природе и т. п. В этом смысле схемы, основанные на анализе смесей распределений, а также классификация по исходным данным, уже представленным в виде матрицы попарных расстояний (близостей), находятся в выгодном положении, поскольку не требуют решения вопроса о выборе метрики.

Важное место в построении классификационных процедур, в первую очередь иерархических, занимает проблема выбора способа вычисления расстояния между подмножествами объектов. Изыскное обобщение большинства используемых в статистической практике вариантов вычисления расстояний между двумя группами объектов дает расстояние, подсчитываемое как обобщенное степенное среднее всевозможных попарных расстояний между представителями рассматриваемых двух групп.

В статистической практике выбор функционала качества разбиения  $Q(S)$  обычно осуществляется произвольно, опирается скорее на эмпирические и профессионально-интуитивные соображения, чем на какую-либо точную формализованную схему. Однако ряд распространенных в статистической практике функционалов качества удастся постфактум обосновать и осмыслить в рамках строгих математических моделей. Возможность этого появляется при наличии дополнительных априорных сведений о классах, позволяющих, например, представлять каждый класс в качестве параметрически заданной одномеральной генеральной совокупности.

Еще один подход к осмыслению и обоснованию методов автоматической классификации представлен аппроксимационными моделями, когда искомая классификация характеризуется матрицей определенной структуры (например, ультраметрической матрицей близости или аддитивными кластерами), а задача состоит в том, чтобы оценить параметры этой структуры таким образом, чтобы она минимально отличалась от матрицы исходных данных. В такой постановке проблема классификации сближается с проблемами факторного анализа. Поэтому данный подход в определенной мере интегрирует традиционные методы кластер-анализа («компактность» кластеров в признаковом пространстве) и многомерной статистики (декомпозиция разброса исходных данных на «вклады» отдельных кластеров и других элементов решения).

В теории и практике статистического анализа данных широко известны такие методы автоматической классификации, как ФОРЕЛЬ, К-средних (Мак-Куина), ИСОМАД (IZODATA), «метод динамических сгущений», «ближайшего соседа».



#### 4.4. Дискриминантный анализ

Дискриминантный анализ – это совокупность многомерных статистических методов классификации многомерных наблюдений в ситуации, когда исследователь обладает т. н. обучающими выборками.

Пусть результатом наблюдения над объектом является реализация  $k$ -мерного случайного вектора  $x = (x_1, x_2, \dots, x_k)^T$ . Требуется установить правило, согласно которому по наблюдаемому значению вектора  $X$  объект относят к одной из возможных совокупностей  $\pi_i, i = 1, 2, \dots, l$ .

Правило дискриминации выбирается в соответствии с определенным принципом оптимальности на основе априорной информации о совокупностях  $\pi_i$ , и вероятностях  $p_i$  извлечения объекта из  $\pi_i$ . При этом следует учитывать размер убытка от неправильной дискриминации.

Обычно в задаче различения переходят от вектора признаков, характеризующих объект, к линейной функции от них, дискриминантной функции – гиперплоскости, наилучшим образом разделяющей совокупность выборочных точек.

Наиболее изучен случай нормального распределения векторов признаков в каждой совокупности при отсутствии информации о параметрах этих распределений. Неизвестные параметры распределения в дискриминантной функции заменяются их наилучшими оценками. Правила дискриминации основываются на отношении правдоподобия.

Таким образом, дискриминантный анализ – это раздел статистического анализа, посвященный получению правил классификации, наблюдений (объектов) в один из нескольких описанных некоторым образом классов (групп, категорий, популяций).

Пусть имеются исходные, подлежащие классификации наблюдения в виде матрицы данных  $X$ , причем предполагается, что они представляют выборку из генеральной совокупности, являющейся смесью наблюдений из  $k$  классов  $D_1, \dots, D_k$ . При этом классу  $i$  соответствует закон распределения вероятностей  $P_i(X)$ ,  $i = 1, \dots, k$ . Априорное описание классов заключается в том, что распределение  $P_i(X)$  либо известны, либо могут быть статистически оценены по имеющимся в распоряжении исследователя обучающим выборкам:

$$X_{j,n_j}, j = 1, 2, \dots, k, X_{i,j} \in D_j.$$

Выборка  $X_{i,t} = \{X_{j,l}\}, l = 1, \dots, t$  называется обучающей, если известно, что все ее наблюдения  $X_i$  извлечены из одной и той же генеральной совокупности  $D_j$ . Требуется указать правило классификации  $S$ , относящее наблюдения из  $X$  к одному из классов (законов распределения)  $D_j$  с минимальными, в определенном смысле, потерями. Обучающие выборки могут составлять часть выборки  $X$ .

Общий вид функционала качества классификации в дискриминантном анализе определяется в виде:

$$Q(S) = \sum_{\substack{i,j=1 \\ i \neq j}}^k c_{i,j} q_j P_s(i | j),$$

где  $c_{i,j}, q_j, P_s(i | j)$  – соответственно потери от классификации объекта из класса  $j$  в класс  $i$ , удельный вес (априорная вероятность) объектов класса  $j$  и вероятность ошибочной классификации объектов класса  $j$  в класс  $i$  при использовании классифицирующего правила  $S$ .

Во многих реальных задачах трудно оценить потери  $c_{i,j}$  из-за неправильной классификации объектов, а иногда и априорные вероятности  $q_j$  появления объектов различных классов. В этих случаях можно воспользоваться т. н. байесовским подходом, в соответствии с которым наблюдение  $X$  должно быть отнесено к классу с тем номером  $j$ , для которого соответствующая апостериорная вероятность, а следовательно и логарифмическая функция правдоподобия  $\ln(q_i P_j(X))$  оказываются максимальными по всем  $i = 1, 2, \dots, k$ . Таким образом, решение об отнесении «текущего» наблюдения к одному из классов принимается на основании значения функции  $d_j = \ln(q_i P_j(X))$ .

Функции  $d_j(X)$  и определяемые ими поверхности  $d_{i,j}(X) = d_i(X) - d_j(X)$  принято называть дискриминантными функциями. Так что искомое правило классификации  $S$  определяется соответствующей системой дискриминантных функций:

$$S : X \in D_j, \text{ если } d_j(X) = \arg \max_{1 \leq i \leq k} d_i(X).$$

Правило классификации, основанное на байесовском подходе, минимизирует среднюю величину ошибочной классификации:

$$Q(S) = \sum_{\substack{i,j=1 \\ i \neq j}}^k q_j P_s(i | j).$$

Помимо байесовского подхода к построению дискриминантных функций, практикуется подход, непосредственно не опирающийся на идею отношения логарифмический функций правдоподобия. Вместо этого пытаются подобрать такую систему функций  $d_i(X) = d_j(X; \theta)$  из параметрического семейства  $\{d(X; \theta)\}$ ,  $\theta \in H$ , на которой максимизируется та или иная мера различия между анализируемыми классами:

$$\tilde{Q}_s(\theta) = \sum \omega_{i,j} \frac{[\hat{M}_i\{d_{i,j}(X; \theta_i, \theta_j)\} - \hat{M}_j\{d_{i,j}(X; \theta_i, \theta_j)\}]^2}{\hat{q}_i \hat{D}_i\{d_{i,j}(X; \theta_i, \theta_j)\} - \hat{q}_j \hat{D}_j\{d_{i,j}(X; \theta_i, \theta_j)\}},$$

где  $d_{j,i}(X; \theta_i, \theta) = d_i(X; \theta) - d_j(X; \theta_j)$ ,

$\hat{q}_i$  – оценка априорной вероятности  $q_i$ , например,  $\hat{q}_i = \frac{n_i}{n}$ .

В этом отношении «веса»  $\omega_{i,j}$  – учитывают относительную важность правильности различия классов с номерами  $i$  и  $j$  среди всех возможных пар классов, т. е.  $\omega_{i,j}$  суть аналоги весов  $c_{i,j}$ , а  $\hat{M}_i\{\xi(X)\}$  и  $\hat{D}_i\{\xi(X)\}$  – это соответственно оценки математического ожидания и дисперсии случайной величины  $\xi(X)$ , вычисленные в предположении, что аргумент  $X$  принадлежит классу  $i$ , т. е. оцененные по соответствующей обучающей выборке.

#### 4.5. Таблица сопряженности и анализ соответствия

Таблица сопряженности (перекрестная классификация, парная группировка), задает распределение объектов по категориям двух или нескольких нечисловых признаков.

Пусть имеется ряд предприятий и для каждого из них известны размер (признак  $x$ ) и уровень рентабельности ( $y$ ), разбитые соответственно на  $r$  и  $c = 3$  категорий. Представим эти данные в виде таблицы сопряженности (табл. 4.1).

Здесь  $f_{i,j}$  – число предприятий, попадающих в категорию размера  $i$  и категорию рентабельности  $j$ ;

$f_{i,\bullet}$  – суммарное число предприятий размера  $i$ , т. е. сумма  $f_{i,j}$  в строке  $i$ ;

$f_{\bullet,j}$  – число предприятий, имеющих категорию рентабельности  $j$ , или сумма по столбцу  $j$ ;

$f_{\bullet,\bullet}$  – общее число предприятий в выборке.

Таблица 4.1 – Таблица сопряженности

| Размер          | Рентабельность  |                 |                 |                       |
|-----------------|-----------------|-----------------|-----------------|-----------------------|
|                 | менее 5 %       | 5–25 %          | более 25 %      | S                     |
| до 100 чел.     | $f_{1,1}$       | $f_{1,2}$       | $f_{1,3}$       | $f_{1,\bullet}$       |
| 100-200 чел.    | $f_{2,1}$       | $f_{2,2}$       | $f_{2,3}$       | $f_{2,\bullet}$       |
| ...             | ...             | ...             | ...             | ...                   |
| $x_i$           | $f_{i,1}$       | $f_{i,2}$       | $f_{i,3}$       | $f_{i,\bullet}$       |
| ...             | ...             | ...             | ...             | ...                   |
| более 5000 чел. | $f_{r,1}$       | $f_{r,2}$       | $f_{r,3}$       | $f_{r,\bullet}$       |
| S               | $f_{\bullet,1}$ | $f_{\bullet,2}$ | $f_{\bullet,3}$ | $f_{\bullet,\bullet}$ |

Очевидно, что последний столбец ( $f_{i,\bullet}$ ) и последняя строка ( $f_{\bullet,j}$ ) задают выборочные маргинальные распределения признаков  $x$  и  $y$  соответственно. Вместо целых чисел  $f_{i,j}$  в клетках могут стоять частоты  $\frac{f_{i,j}}{f_{\bullet,\bullet}}$  или проценты.

Применение статистических методов зависит от того, как была получена анализируемая таблица. Рассмотрим следующие три выборочные схемы:

1) частоты одного из признаков фиксированы. Например, берется по 100 предприятий каждого размера, и рассматриваются их распределения по рентабельности;

2) предполагается, что таблица является реализацией полиномиального распределения  $f_{\bullet\bullet}$  объектов по  $r \times c$  клеткам. Общее число объектов фиксировано;

3) третья схема возникает, когда  $f_{\bullet\bullet}$  является случайной величиной.

Таблицу сопряженности можно рассматривать как двумерную гистограмму. Первая выборочная схема позволяет сравнивать ряд одномерных гистограмм или распределения предприятий по рентабельности при всевозможных категориях размера. Имеются соответствующие статистические критерии. Для целей визуализации первая схема обладает тем преимуществом, что суммы по строкам (по столбцам) зафиксированы, обычно берутся одинаковые маргинальные частоты, что позволяет сравнивать числа, содержащиеся в одном столбце (строке). При второй или третьей выборочных схемах  $f_{i,j}$  и  $f_{i,j}$  в общем случае нельзя сравнивать, так как различие может быть обусловлено неодинаковыми  $f_{i\bullet}$  и  $f_{\bullet j}$ . Для нивелирования этих эффектов обычно считают проценты по строкам  $100 \times \frac{f_{i,j}}{f_{i\bullet}}$  или по столбцам  $100 \frac{f_{i,j}}{f_{\bullet j}}$ .

Процедура мостеллеризации при выполнении некоторых условий позволяет так преобразовать таблицу, чтобы суммы по строкам и по столбцам были одновременно равны постоянной величине. По значениям в клетках полученной таблицы можно судить о близости соответствующих категорий признаков  $x$  и  $y$ .

Если в таблице имеется много клеток с нулевыми значениями, то можно попытаться так переставить строки и столбцы, чтобы ненулевые элементы образовали блоки, расположенные вдоль главной диагонали. Тем самым решается задача классификации.

Анализ соответствий позволяет одновременно представить категории обоих признаков и объекты как комбинации категорий в виде точек в евклидовом пространстве (как правило, на плоскости), что позволяет лучше понять структуру данных, выявить аномальные наблюдения, возможные кластеры.

При проверке гипотезы о независимости признаков  $x$  и  $y$  предпочтительнее работать в предположениях второй или третьей выборочных схем. При этом оценка максимального правдоподобия элемента таблицы с номерами  $i, j$  равна  $\frac{f_{i\bullet} \cdot f_{\bullet j}}{f_{\bullet\bullet}}$ . Если число строк равно числу столбцов, особенно когда  $x$  и  $y$  имеют

смысл одного признака различных объектов (мужчин и женщин) или состояния в два различных момента времени (таблица обмена), бывает полезно проверять гипотезы о симметричности таблицы и об однородности маргинальных распределений. Для таблицы обмена имеет смысл гипотеза квазисимметричности: оценка элемента таблицы представляется в виде произведения  $a_i b_j s_{i,j}$ , где  $s_{i,j} = s_{j,i}$  — показатель близости состояний  $i$  и  $j$ ,  $a_i$  — показатель «отталкивания»  $i$ , а  $b_j$  — «привлекательность» состояния  $j$ . Три последние гипотезы проверяются только для второй и третьей выборочных схем.

При проверке всех гипотез полезно определять вклады клеток исходной таблицы в статистику критерия. С их помощью можно понять, какие комбинации значений признаков нарушают гипотезу.

Если гипотеза о независимости отвергается, то можно пытаться измерить тесноту связи рассматриваемых признаков. Для этого существуют многочисленные коэффициенты. Некоторые из них (Пирсона, Чупрова, Крамера) основаны на величине отклонения наблюдаемых частот от значений, ожидаемых при гипотезе о независимости. Другие коэффициенты используют понятие количества информации (мера Кульбака), относительную величину ошибки предсказания значения одного признака при известном другом (Х-меры). Для признаков, измеренных в ординальной шкале, имеются специальные меры (Гудмена–Краскала). Для некоторых мер известны распределения. Нулевое или очень маленькое значение отдельного коэффициента в общем случае не говорит об отсутствии связи. Это свидетельствует лишь об отсутствии конкретного вида связи, который измеряется данным коэффициентом. Универсальная мера не существует.

Возможны многомерные таблицы сопряженности. Например, если предприятия, составляющие приведенную выше таблицу, принадлежат одному министерству и известны аналогичные таблицы для других министерств, то такие данные можно представить в виде трехмерной таблицы сопряженности. Министерство будет определять двумерный слой. При этом у всех величин появится дополнительный индекс. Встречаются таблицы и большей размерности.

Имеются обобщения анализа соответствий и модели квазисимметричности на многомерные таблицы. Структура зависимости в многомерных таблицах значительно сложнее, чем в двумерных, так как уже третий признак  $z$  может зависеть не только от признаков  $x$  и  $y$  в отдельности, но и от их взаимодействия. В общем случае возможны взаимодействия и более высоких порядков. Для формирования и проверки адекватности моделей многомерных таблиц используется аппарат лог-линейного анализа, являющегося наиболее строгим с математической точки зрения методом анализа нечисловых данных. Он обобщает многие известные модели, однако для его успешного применения необходимо большое, по сравнению с числом клеток в таблице, количество наблюдений. Содержательная интерпретация взаимодействий высоких порядков часто вызывает сложности.

Анализ соответствий (метод взаимных средних, оптимальное шкалирование, канонический (компонентный) анализ качественных признаков) – это методология изучения таблиц сопряженности, основанная на представлении их строк и столбцов точками одного и того же многомерного пространства. Используется преимущественно для визуализации данных.

В конце 1980-х гг. предложен ряд обобщений анализа соответствий для ситуаций, когда данные имеют более сложный характер, в частности, для многомерных и динамических данных.

## 5. СПЕЦИФИКАЦИЯ ЭКОНОМЕТРИЧЕСКОЙ МОДЕЛИ

### 5.1. Функциональные и статистические зависимости

Исследование объективно существующих связей между явлениями имеет большое значение в социальной и экономической сфере. Изучение любого явления будет не полным, если не исследованы его связи с другими явлениями и процессами. В основе исследования связей лежит общефилософское положение о всеобщей связи в природе и обществе. В процессе исследования зависимостей вскрываются причинно-следственные отношения между явлениями, что позволяет выявлять факторы (признаки), оказывающие основное влияние на вариацию изучаемых явлений. Целью эконометрического изучения связи социально-экономических процессов является не только выявление наличия и направление связи, но и количественная ее оценка, а также аналитическое выражение.

Классифицировать методы анализа данных можно исходя из ряда соображений.

Во-первых, следует различать описательные и вероятностные методы. Именно в последнем случае понятие вероятностной модели данных, а часто и вероятностные свойства оценок, и законы их изменчивости проявляются в явном виде. В описательной статистике вероятностные аспекты или совсем отсутствуют, или им придается небольшое значение. Если в выводах необходимо отразить в явном виде степень неопределенности заключений, более подходящими представляются вероятностные методы.

Во-вторых, внутри численных методов следует выделить детерминированные и вероятностные методы. Для детерминированных методов характерно признание взаимно-однозначной зависимости между признаками, т. е. конкретному значению одного фактора соответствует одно или несколько конкретных значений другого признака. Для вероятностных (статистических) методов эта ситуация выглядит следующим образом: конкретному значению одного признака соответствует область значений другого признака, имеющая некоторую плотность распределения. Таким образом, если в выводах необходимо отразить вероятностные оценки, то более подходящими представляются статистические методы. Частным случаем статистической связи является корреляционная связь. При корреляционной связи конкретному значению признака-фактора соответствуют определенные средние значения признака-результата. Возможны и другие случаи статистической связи. Например, с изменением значения признака-фактора прослеживается закономерность изменения таких статистических показателей, как коэффициент вариации, показатель асимметрии или эксцесса. Статистическая связь между двумя признаками предполагает, что каждый из них является случайной величиной со своей вариацией относительно средней величины. Если такую вариацию имеет один из признаков, а значения другого являются детерминированными, то такая связь классифицируется как регрессионная.

В-третьих, существует разграничение между графическими и численными

методами. Завершающий этап анализа данных с помощью вероятностных методов почти всегда носит численный, а не графический характер, но даже при использовании этих методов графические методы могут являться ценным подспорьем в представлении выводов. Графические методы особенно ценны при представлении качественных аспектов выводов в наглядной форме. Но при последующем использовании результаты все-таки лучше представлять в численной форме.

Методы изучения связи применимы не только для изучения простых парных связей между двумя процессами, но и множественных (между несколькими явлениями). При изучении связей следует учитывать, что большинство социально-экономических процессов представляет результат воздействия множества факторов. Поэтому основное внимание в статистике уделяется изучению многофакторных связей. Это базируется на предположении множественности причин и следствий, объективно присущих социально-экономическим процессам.

Выбор метода изучения связей зависит от цели исследования, поставленной задачи. Очень часто необходимо лишь определить направление и характер связи, выявить воздействие одних факторов на другие.

Связи между явлениями и признаками ввиду большого их разнообразия классифицируются в статистике по ряду оснований.

Признаки по характеру их роли во взаимосвязи подразделяются на два больших класса. Признаки, обуславливающие изменение других, связанных с ними признаков, называют *факторными* или *признаками-факторами*.

Признаки, изменяющиеся под действием факторных признаков, называют *результативными* (иногда *признаки-результаты*).

Например: квалификация (фактор) – производительность (результат).

Связи между явлениями, их признаками подразделяют, прежде всего, по степени тесноты связи, затем по направлению и аналитическому выражению.

По степени тесноты связи различают полную, или функциональную, связь и связь неполную, статистическую, корреляционную, стохастическую, или регрессионную.

*Функциональными* называют такие связи, в которых определенному значению факторного признака (признаков) соответствует строго определенное значение (или значения) результативного признака.

Однако функциональные зависимости не исчерпывают всех возможных видов взаимосвязи между явлениями. Например, на вопрос «Связан ли вес человека с ростом?» можно уверенно ответить утвердительно, хотя о функциональной связи в этом случае не может быть и речи. Следовательно, хотя здесь и имеет место зависимость, но эта зависимость не является функциональной. Такой же нефункциональный характер имеет большинство зависимостей, с которыми приходится сталкиваться в экономике, финансах, маркетинге и менеджменте. Да и действие экономических законов проявляется в виде некоторых тенденций, которые нельзя описать в виде точных функциональных зависимостей. Такого рода зависимости носят название корреляционных, или стохастических. Основоположниками теории корреляции

считаются английские биометрики Ф. Гальтон (1822–1911) и К. Пирсон (1857–1936). Термин «корреляция» был заимствован из естествознания и обозначает соотношение, соответствие. Представление о корреляции как об отношении взаимозависимости между случайными переменными величинами лежит в основе математико-статистической теории корреляции. Корреляция характеризует связь между признаками, проявляющуюся не в каждом конкретном случае (строго функционально), а лишь в среднем по совокупности. При этом каждому значению фактора соответствует не одно значение признака-результата, а распределение значений, варьирующих около средней величины.

В статистической взаимосвязи взаимно однозначного соответствия между изменениями факторного и результативного признаков нет – одному и тому же значению признака-фактора соответствует область значений результативного признака (при одном и том же уровне квалификации работников их производительность труда может иметь самые различные значения).

Использование корреляционного анализа предполагает выполнение ряда предпосылок. Одним из важнейших условий адекватного применения методов корреляционного анализа является требование однородности единиц совокупности, которые подвергаются изучению. Одним из возможных вариантов количественной оценки однородности совокупности являются значения относительных показателей вариации. Другим условием, обеспечивающим надежность выводов при использовании методов корреляционного анализа, является требование достаточности числа наблюдений. Как отмечалось выше, влияние систематических причин может быть искажено действием случайных факторов, «взаимоисключающее» влияние которых возрастает при увеличении числа наблюдений. Кроме того, все основные положения теории корреляции базировались на гипотезе о нормальном характере распределения исследуемых признаков. В случае применения на практике не все предпосылки выполняются, поэтому получаемые показатели могут оказаться недостаточно точными и достоверными, что необходимо учитывать в окончательных выводах.

*По направлению* выделяют связь прямую и обратную.

*Прямая* – это связь, при которой факторный и результативный признаки изменяются в одном и том же направлении – по мере увеличения или уменьшения значения факторного признака значение результативного соответственно увеличивается или уменьшается.

В случае *обратной* связи значения результативного признака изменяются под действием факторного, но в противоположном направлении по сравнению с изменением факторного признака.

Например:

- 1) по мере роста производительности труда себестоимость продукции, как правило, уменьшается;
- 2) по мере снижения цены объем спроса, как правило, увеличивается.

По аналитическому выражению обычно в наиболее общей классификации выделяют связи прямолинейные (или просто линейные) и нелинейные. Статистическую связь называют *линейной* связью, если она может быть



приближенно выражена математически уравнением прямой линии. А если статистическая связь может быть выражена уравнением какой-либо кривой линии (параболы, гиперболы и т. п.), то ее называют *нелинейной*.

Основными этапами изучения зависимостей между признаками (показателями) являются следующие.

1. Выдвижение гипотезы о наличии связи между признаками. На этом этапе статистического изучения связи основным является выделение основных причинно-следственных связей и отделение их от второстепенных. Следовательно, в основе первого этапа статистического исследования связей лежит качественный анализ изучаемого явления, связанный с анализом сущности социального или экономического явления методами политической экономии, социологии, конкретной экономики.

2. Проверка факта наличия зависимости между признаками, например, методом группировки.

3. Установление направления и выбор формы зависимости между признаками с помощью графических методов.

4. Установление тесноты зависимости между признаками, например, с помощью такого показателя, как эмпирическое корреляционное отношение.

5. Построение модели связано с определением параметров аналитической зависимости между признаками методом наименьших квадратов.

6. Оценка достоверности параметров аналитической зависимости и оценка точности результатов с помощью статистических критериев.

7. Экономическая интерпретация результатов и практическое использование полученной модели процесса, которая вновь связана с качественными особенностями изучаемого явления.

Построение эконометрической модели может базироваться на методах математической и общей теории статистики: законов распределения, группировок, средних величин, показателей вариации и т. д.

Неоднозначный характер проявления корреляционных зависимостей определяется тем обстоятельством, что на изменение изучаемого процесса влияет не только выделенный фактор, но и множество причин, многие из которых могут быть исследователям неизвестны.

Для выявления связи, ее характера, направления в статистике используются методы приведения параллельных данных, балансовый, аналитических группировок, графический.

Суть *метода приведения параллельных данных* состоит в следующем: приводятся два ряда данных о двух явлениях или двух признаках, связь между которыми желают выявить. По характеру изменений делают заключение о наличии (если изменение величин одного ряда следует за изменением величин другого ряда) или об отсутствии (если никакого твердого, устойчивого соответствия в их изменениях нет) связи (рис. 5.1).

На рисунке 5.1 приведено графическое представление рядов двух признаков-факторов, упорядоченных относительно значений признака-результата.

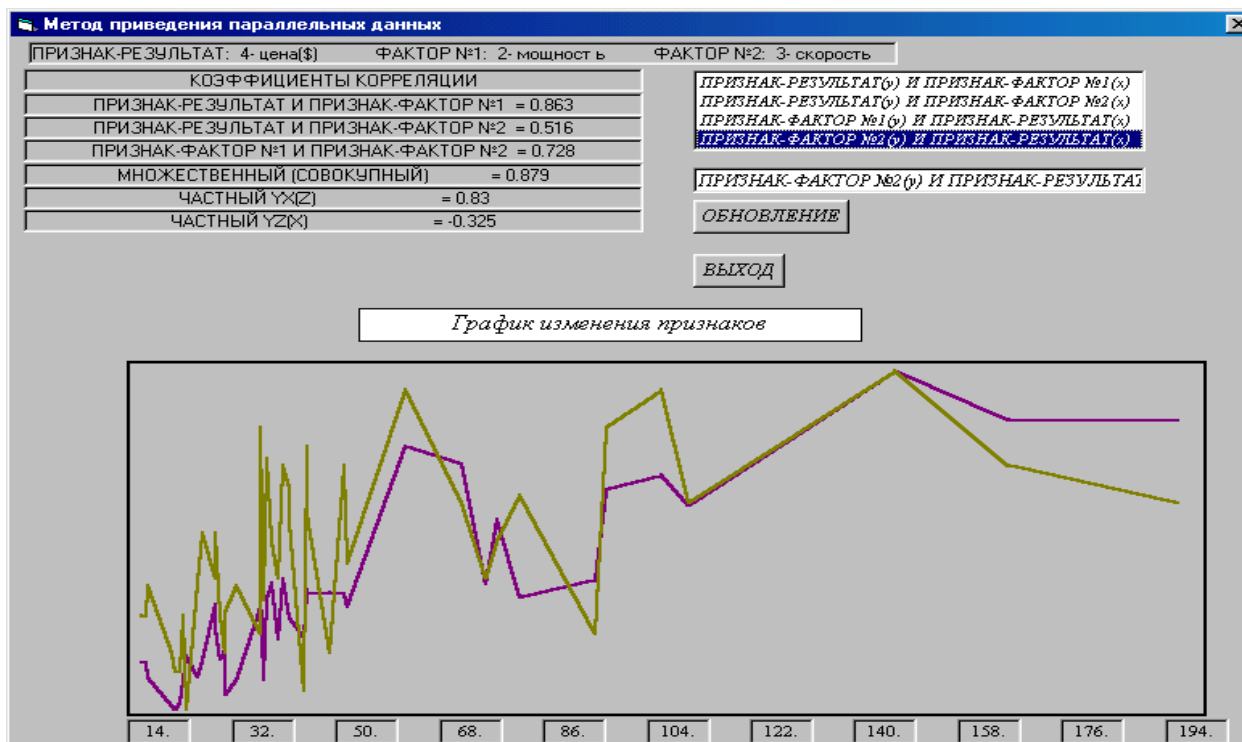


Рисунок 5.1 – Иллюстрация метода приведения параллельных данных

*Балансовый метод* заключается в построении балансов-таблиц, в которых итог одной части равен итогу другой (например, баланс производства какого-либо продукта и его потребления).

Удобной формой изложения данных при проведении корреляционного анализа является корреляционная таблица (табл. 5.1).

Таблица 5.1 – Корреляционная таблица

| Количество работников, чел. | Количество маршрутов, обслуживаемых одной туристической фирмой, $x$ |          |           |           |           |           |           | Итого |
|-----------------------------|---------------------------------------------------------------------|----------|-----------|-----------|-----------|-----------|-----------|-------|
|                             | 5–7                                                                 | 7–9      | 9–11      | 11–13     | 13–15     | 15–17     | 17–19     |       |
| 10–15                       | <u>7</u>                                                            | 4        |           | 2         | 1         |           |           | 14    |
| 15–20                       | 3                                                                   | <u>8</u> | 5         | 4         |           |           |           | 20    |
| 20–25                       |                                                                     | 2        | <u>11</u> | 8         |           | 2         |           | 23    |
| 25–30                       |                                                                     |          | 5         | <u>13</u> | 7         | 1         |           | 26    |
| 30–35                       |                                                                     |          |           | 1         | <u>16</u> | 3         |           | 20    |
| 35–40                       |                                                                     |          |           | 2         | 6         | <u>19</u> | 3         | 30    |
| 40–45                       |                                                                     |          |           |           | 3         | 7         | <u>18</u> | 28    |
| Итого                       | 10                                                                  | 14       | 31        | 30        | 33        | 32        | 21        | 161   |

Эта таблица позволяет сделать предположение о наличии или отсутствии связи между признаками. Например, если таблица показывает, что частоты концентрируются у главной диагонали, то связь между количеством маршрутов и численностью работников прямая (с увеличением числа маршрутов увеличивается численность работников) или близкая к прямой (концентрация

частот идет почти по прямой). Если частоты принимают примерно одинаковое значение и заполняют всю таблицу, то можно предполагать отсутствие связи.

Однако это обстоятельство еще не означает, что корреляционная связь между признаками отсутствует. При этом важно установить, как расположена основная масса наблюдений. Для того чтобы восприятие корреляционной таблицы было более наглядным, целесообразно воспользоваться графическим методом.

## **5.2. Выявление зависимости между показателями**

В любой сфере экономики, маркетинга, менеджмента встречается большое разнообразие взаимосвязей между показателями, являющимися причиной или следствием явления. При спецификации эконометрической модели необходимо выявления наличия зависимости между показателями, что вызывает необходимость выполнения аналитической группировки. В этой группировке участвуют два показателя, один из которых называется факторным, другой – результативным. При построении аналитических группировок важно правильно определить показатель-результат и показатель-фактор. Показатель, влияние которого на другие признаки исследуется, называется показателем-фактором. Показатель, испытывающий влияние факторного, называется показателем-результатом. Взаимосвязи между показателями можно классифицировать следующим образом:

1. факторный показатель является количественным признаком, а результативный показатель – качественный признак (например, стаж работы в фирме и квалификация работников (тарифный разряд));
2. основанием группировки является качественный показатель, а результативный показатель – количественный признак (например, тарифный разряд – величина заработной платы);
3. показатель-фактор и показатель-результат являются качественными признаками (например, квалификация работников и уровень их образования);
4. факторный и результативный показатели являются количественными признаками (например, стаж работы и уровень оплаты труда).

Следствием выполнения аналитической группировки является вывод о наличии и направлении связи между показателями. Логическим продолжением статистического исследования после выполнения аналитической группировки является определение тесноты связи аналитической зависимости между показателями.

Факторный показатель оказывает влияние на результативный показатель. Чтобы установить связь между показателями, аналитическая группировка всегда осуществляется по показателю-фактору. Затем по каждой группе отбираются соответствующие значения показателя-результата и рассчитывается его среднее значение. Сопоставляя изменение средних значений показателя-результата от группы к группе с изменениями показателя-фактора, можно сделать вывод о

наличии или отсутствии взаимосвязи, а также о ее направлении. Различие групповых средних позволяет утверждать, что признаки взаимозависимы. Если изменение величины показателя-фактора в определенном направлении вызывает изменение величины показателя-результата в том же направлении, то связь прямая, в противном случае – связь обратная. Результаты аналитической группировки представляются в табличном виде (табл. 5.2).

Таблица 5.2 – Результаты аналитической группировки

| Наименование показателя-фактора (единица измерения) | Количество элементов совокупности в группе | Среднее значение показателя-результата (единица измерения) |
|-----------------------------------------------------|--------------------------------------------|------------------------------------------------------------|
| ...                                                 | ...                                        | ...                                                        |
| ...                                                 | ...                                        | ...                                                        |
| ...                                                 | ...                                        | ...                                                        |
| Итого                                               | Число элементов совокупности               | —                                                          |

Аналитическая группировка позволяет исключить влияние всех прочих факторов, сохранив влияние лишь того, по которому производится группировка.

Аналитическая группировка позволяет выявить наличие или отсутствие зависимости. Вместе с тем в рамках этого метода не удастся аналитически описать эту зависимость, а также не удастся выяснить «тесноту» или «существенность» этой зависимости.

Целью аналитической группировки является выявление наличия зависимости между факторным и результативным показателями.

В большинстве случаев аналитическая группировка реализуется по количественным показателям. Аналитические группировки могут быть равноинтервальные и неравноинтервальные.

Интервал – разность между верхним и нижним значением факторного показателя в данной группе (табл. 5.3).

Таблица 5.3 – Пример интервалов

| Границы |         | Интервал |
|---------|---------|----------|
| Нижняя  | Верхняя |          |
| 20      | 50      | 30       |
| 50      | 200     | 150      |
| 200     | 500     | 300      |

Если о распределении по группам ничего неизвестно, то рекомендуется рассматривать сначала равноинтервальные группировки. Если в результате выполнения этой группировки явно просматривается тенденция изменения показателя, то процесс выполнения группировки на этом завершается. Если же тенденция просматривается, но требует уточнения, то можно попытаться это сделать за счет изменения числа групп или за счет выполнения

неравноинтервальной группировки. Неравноинтервальная группировка позволяет уточнить распределение по группам и установить закон распределения, т. е. представить тенденцию изменения показателя в «явном» виде.

В роли обобщающей характеристики структурной группировки выступает число единиц наблюдений, попавших в каждую отдельную группу.

Другой итоговой характеристикой является относительный показатель структуры:

$$a_i = \frac{N_i}{N} \cdot 100\%,$$

где  $N_i$  – число единиц наблюдения в группе  $i$ ;

$N$  – объем совокупности.

Для оценки интенсивности изменения структуры используют различные показатели:

1. Линейный коэффициент структурных различий (сдвигов) изменяется в пределах  $0 \leq d \leq 1$ .

$$d = \frac{\sum_{i=1}^n |a_{i1} - a_{i2}|}{n},$$

где  $a_{i1}$  – относительные показатели структуры элемента  $i$  в период 1;

$n$  – число элементов рассматриваемой совокупности или структурных составляющих.

2. Квадратичный коэффициент структурных сдвигов изменяется в пределах  $0 \leq \sigma \leq 1$ .

$$\sigma = \sqrt{\frac{\sum_i (a_{i1} - a_{i2})^2}{n}}.$$

Если относительные показатели структуры заданы в процентах, то  $0 \leq d, \sigma \leq 100$ .

3. Интегральный коэффициент К. Гатева изменяется в пределах  $0 \leq K_v \leq 1$ .

$$K_v = \sqrt{\frac{\sum_i (a_{i1} - a_{i2})^2}{\sum_i a_{i1}^2 + \sum_i a_{i2}^2}}.$$

4. Индекс Салаи изменяется в пределах  $0 \leq I \leq 1$ .

$$I = \sqrt{\frac{1}{n} \sum_i \left( \frac{a_{i1} - a_{i2}}{a_{i1} + a_{i2}} \right)^2}.$$

*Пример.* По данным о распределении работающих и безработных по уровню образования (табл. 5.4) определить показатели структурных различий.

Таблица 5.4 – Распределение работающих и безработных по уровню образования

| Образование               | Работающие, % | Безработные, % |
|---------------------------|---------------|----------------|
| Высшее профессиональное   | 19,6          | 10,9           |
| Среднее профессиональное  | 34,4          | 24,5           |
| Среднее (полное) общее    | 33,3          | 39,9           |
| Основное общее            | 11,1          | 20,5           |
| Не имеют основного общего | 1,6           | 4,2            |

*Решение:*

1) линейный коэффициент структурных различий (сдвигов):

$$d = 0,0744;$$

2) квадратичный коэффициент структурных сдвигов:

$$\sigma = 7,90\%;$$

3) интегральный коэффициент К. Гатева:

$$K_v = 0,237;$$

4) индекс Салаи:

$$I = 0,285.$$

**Выводы:**

- Структуры занятых и безработных существенно различаются, независимо от используемого показателя.
- Расчет показывает, что между показателями существует следующее соотношение:

$$0 \leq d \leq \sigma \leq K_v \leq I \leq 1.$$

Для равноинтервальной группировки границы группы устанавливаются по следующей схеме:

- выбирается показатель группировки;
- исходя из числа объектов (наблюдений) устанавливается первоначальное число групп  $n$ .
- Основные трудности при построении группировок представляют выбор числа групп и выбор длины каждого интервала. При выборе числа групп следует учитывать, что:
- увеличение их числа приводит к меньшей достоверности результатов, так как в каждый интервал попадает меньше наблюдений;
- уменьшение числа групп приводит к меньшей точности, так как любой группировке сопутствует потеря информации о форме распределения.

Первоначальное число интервалов  $n$  составляет величину порядка:

$$n = \sqrt{N} - 1$$

( $N > 15$  – объем совокупности).

На практике, при ограниченном объеме выборки, приходится эту величину корректировать. Однако этот результат справедлив при выборе равных интервалов. В дальнейшем при построении неравноинтервальной группировки при «очень большом» числе наблюдений в отдельных интервалах целесообразно эти интервалы разбить на два. При «малом» числе наблюдений в соседних

интервалах целесообразно их объединить. Для повышения достоверности результатов при заданном числе интервалов следует стремиться к тому, чтобы в каждом интервале относительный показатель структуры был примерно равен  $a_j = \frac{1}{n_j}$ , где  $n_j$  – число наблюдений в группе  $j$ . Если в рассматриваемой группе

концентрируется большая часть наблюдений, то необходимо эту группу разбить на несколько с интервалами меньшей длины, а в местах «малой» интенсивности – большей. В каждом конкретном случае необходимо руководствоваться методом «проб и ошибок», повторяя построение статистического ряда до получения удовлетворительного результата. При этом величину  $n$  следует выбирать не более 10–20 и выполнить следующее:

- определить величину интервала:

$h$  – величина интервала (в равноинтервальной группировке принимается одинаковой для всех групп),

$$h = \frac{x_{\max} - x_{\min}}{n},$$

где  $x_{\max}, x_{\min}$  – максимальное и минимальное значения показателя;

- установить границы групп (табл. 5.5);

Таблица 5.5 – Границы групп

| № группы | Границы         |                 |
|----------|-----------------|-----------------|
|          | нижняя          | верхняя         |
| 1        | $x_{\min}$      | $x_{\min} + h$  |
| 2        | $x_{\min} + h$  | $x_{\min} + 2h$ |
| 3        | $x_{\min} + 2h$ | $x_{\min} + 3h$ |
| ...      | ...             | ...             |
|          | $x_{\max} - h$  | $x_{\max}$      |

При определении границ интервалов целесообразно округлять результат до двух, максимум трех значащих цифр. В ряде случаев в группировке могут быть использованы открытые интервалы, например, число путевок:

- до 200 штук – нижняя граница открыта;
- 200–300;
- 300–400;
- 400 и более – верхняя граница открыта.

В этом случае длина первого интервала условно принимается равной длине второго, а длина последнего – длине предпоследнего;

- распределить единицы наблюдения по установленным группам.

Существует два разных подхода.

1. Последовательный просмотр всех единиц наблюдения с отнесением каждой из них в ту или иную группу (табл. 5.6).

Таблица 5.6 является средством для подсчета числа наблюдений в каждой

группе при последовательном рассмотрении единиц наблюдения. Но при изменении границ интервалов возникает необходимость повторения этой процедуры, что является существенным недостатком, от которого избавлен следующий способ.

Таблица 5.6 – Определение частоты

| № группы | Интервалы | Рабочее поле   |       | Число единиц | Структура, % |
|----------|-----------|----------------|-------|--------------|--------------|
| 1        | 2–5       | 11111          | ☑     | 5            | 20           |
| 2        | 5–8       | 11111 11111 11 | ☑ ☑ ⊥ | 12           | 48           |
| 3        | 8–11      | 11111 111      | ☑ ⊥   | 8            | 32           |
| ...      | ...       | ...            |       | ...          | ...          |

2. Распределение единиц совокупности по группам может быть выполнено с помощью ранжирования, которое представляет собой расположение единиц наблюдения в порядке возрастания или убывания, т. е. предполагается использование ранжированного ряда распределения (табл. 5.7).

Таблица 5.7 – Ранжированный ряд распределения

| Значение показателя | Число единиц |
|---------------------|--------------|
| 4                   | 10           |
| 7                   | 5            |
| 10                  | 7            |
| 15                  | 4            |
| 24                  | 12           |

*Пример.* Выполнить аналитическую группировку. Цель – оценить процесс концентрации случайной величины, т. е. определить распределение марок автомобилей по стоимости. Данные представлены в табл. 5.8.

Факторный показатель: объем двигателя.

Результативный показатель: стоимость (цена).

Первоначальное число групп

$$n = \sqrt{N} - 1, n = 5,$$

тогда длина интервала равна:

$$h = \frac{x_{\max} - x_{\min}}{n} = \frac{6750 - 1600}{5} \cong 1000,$$

а результаты проведения равноинтервальной аналитической группировки представлены в табл. 5.9 и графически на рис. 5.2.



Таблица 5.8 – Исходные данные по автомобилям типа седан

| Наблюдения              | Объем двигателя, см <sup>3</sup> | Цена, долл. |
|-------------------------|----------------------------------|-------------|
| 1. «BMW» 316            | 1600                             | 28 851      |
| 2. «BMW» 3181           | 1800                             | 30 631      |
| 3. «BMW» 325            | 2500                             | 43 000      |
| 4. «BMW» 518            | 1800                             | 35 495      |
| 5. «BMW» 525            | 2500                             | 49 501      |
| 6. «BMW» 730            | 3000                             | 80 000      |
| 7. «BMW» 740            | 4000                             | 95000       |
| 8. «BMW» 850            | 5000                             | 104501      |
| 9. «Bentley»            | 6750                             | 165 000     |
| 10. «Cadilac-1»         | 4900                             | 74 000      |
| 11. «Cadilac-2»         | 5700                             | 76 000      |
| 12. «Cadilac-3»         | 4900                             | 93 000      |
| 13. «Cadilac-4»         | 3800                             | 42 500      |
| 14. «Fiat»              | 1600                             | 20 200      |
| 15. «Ford-1»            | 1800                             | 24 000      |
| 16. «Ford-2»            | 2400                             | 14 000      |
| 17. «Ford-3»            | 2400                             | 15 000      |
| 18. «Ford-4»            | 3800                             | 28 730      |
| 19. Lincoln             | 4600                             | 47 000      |
| 20. «Mersedes-Benz» 190 | 2600                             | 27 000      |
| 21. «Mersedes-Benz» 300 | 3000                             | 50 000      |
| 22. «Mersedes-Benz» 500 | 5000                             | 60 000      |
| 23. «Mersedes-Benz» 560 | 5500                             | 109 227     |
| 24. «Mersedes-Benz» 600 | 6000                             | 145 000     |
| 25. «Mitsubishi-1»      | 1800                             | 21 000      |
| 26. «Mitsubishi-2»      | 1600                             | 21 414      |
| 27. «Mitsubishi-3»      | 3000                             | 38 800      |
| 28. «Nissan»            | 1600                             | 19500       |
| 29. «Opel»              | 2000                             | 38000       |
| 30. «Rolls-Roys»        | 6750                             | 194000      |
| 31. «Saab»              | 2300                             | 43000       |
| 32. «Volvo-1»           | 2000                             | 27000       |
| 33. «Volvo-2»           | 2500                             | 40 000      |
| 34. «Volvo-3»           | 2300                             | 28 000      |
| 35. «Volvo-4»           | 3000                             | 36 950      |

Вывод: можно предполагать наличие зависимости между объемом двигателя и стоимостью автомобиля. Связь прямая.

Таблица 5.9 – Результаты проведения равноинтервальной аналитической группировки

| Интервал значений<br>группировочного признака | Среднее значение результативного<br>показателя |
|-----------------------------------------------|------------------------------------------------|
| 1 600–2 600                                   | 29,33                                          |
| 2 600–3 600                                   | 46,55                                          |
| 3 600–4 600                                   | 55,41                                          |
| 4 600–5 600                                   | 81,29                                          |
| 5 600–6 750                                   | 145,00                                         |

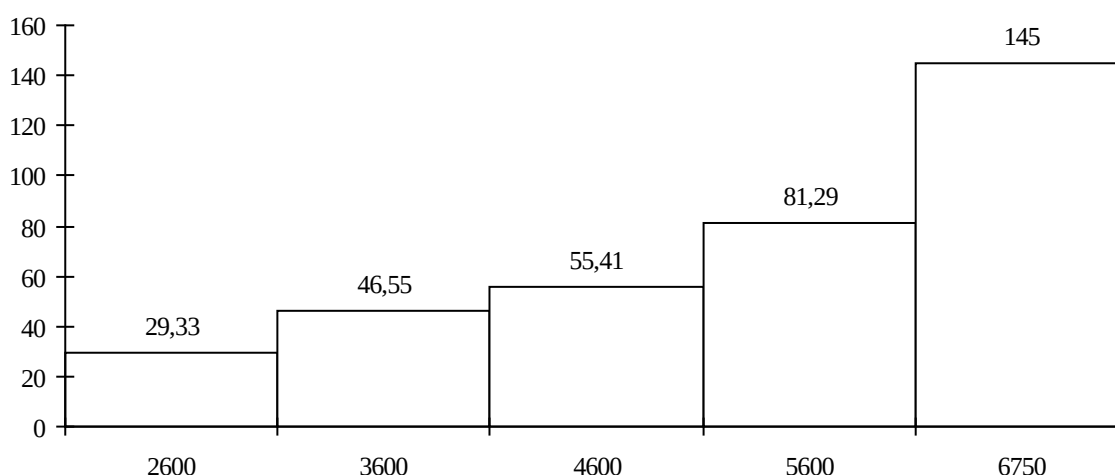


Рисунок 5.2 – Графическое представление результатов равноинтервальной аналитической группировки

В примере полученная равноинтервальная группировка не потребовала уточнения, т. е. отсутствовала необходимость в построении неравноинтервальной группировки.

### 5.3. Спецификация эконометрических моделей

В эконометрике широко используются методы статистики, которые позволяют установить количественные описания взаимосвязей между экономическими переменными. Эконометрика, прежде всего, связана с методами регрессии и корреляции для количественное описание взаимосвязей между экономическими переменными. Целью эконометрики является получение модели, адекватно описывающее поведение объясняемой переменной. Под эконометрической моделью понимается упрощенное представление действительного объекта и/или протекающих в нем процессов с заданной точки зрения, предназначенное для изучения данного объекта.

Модель – это упрощенный объект, сохраняющий с определенной точки зрения важнейшие свойства настоящего существующего объекта или системы. Точка зрения определяется цель проводимого эконометрического исследования,

но ее определение является субъективным, так как формулируется менеджментом или собственником предприятия.

В процессе спецификации эконометрических моделей следует учитывать следующие принципы:

- 1) эконометрические утверждения и закономерности должны быть формализованы (представлены в виде уравнений, формул);
- 2) в эконометрической модели количество эндогенных переменных должно быть равно числу уравнений;
- 3) переменные должны быть привязаны к определенному периоду времени (датированы);
- 4) модель должна включать параметр случайной ошибки для характеристики влияния случайных факторов.

Например, для получения эконометрической модели может быть использован корреляционно-регрессионный аппарат, в котором, в зависимости от количества факторов, включенных в уравнение регрессии, принято различать простую (парную) и множественную регрессии.

*Простая регрессия* представляет модель, где среднее значение зависимой (объясняемой) переменной  $y$  рассматривается как функция одной независимой (объясняющей) переменной  $x$ , т. е. это модель вида:

$$y = f(x).$$

*Множественная регрессия* представляет модель, где среднее значение зависимой (объясняемой) переменной  $y$  рассматривается как функция нескольких независимых (объясняющих) переменных  $x_1, x_2, \dots$ , т. е. это модель вида:

$$y = f(x_1, x_2, \dots, x_k).$$

В регрессионном анализе рассматривается односторонняя зависимость случайной переменной ( $y$ ) от одной или нескольких детерминированных (неслучайных) независимых переменных ( $x$ ).

Поэтому любое эконометрическое исследование начинается со *спецификации модели*, т. е. с формулировки вида модели исходя из соответствующей теории связи между переменными. Иными словами, исследование начинается с теории, устанавливающей тесноту связи между явлениями.

Слово «**спецификация**» (от лат. species – род, вид, разновидность и facio – делаю) термин, обозначающий набор требований и параметров, которым удовлетворяет некоторая сущность.

Спецификацией эконометрической модели поведения объясняемой переменной называется исследование, направленное на:

- отбор факторов, включаемых в модель, с учетом их влияния и тесноты связи с объясняемой переменной;
- выбор математической функции, наилучшим способом фиксирующей связь отобранных факторов с объясняемой переменной;
- определение параметров выбранной функции.

Спецификация эконометрической модели является одним из этапов формирования эконометрической модели:

- 1) определение цели исследования;
- 2) определение факторов, влияющие на зависимый показатель;
- 3) проверка связи между переменными;
- 4) составление перечня факторов, которые могут быть включены в модель;
- 5) спецификация эконометрической модели;
- 6) отбор факторов, включаемых в модель множественной регрессии;
- 7) оценка параметров линейных уравнений регрессии;
- 8) оценка качества подбора уравнения;
- 9) проверка статистической значимости модели и ее параметров;
- 10) получение точечного и интервального прогноза;
- 11) выводы и предложения.

Спецификация эконометрической модели считается произведенной, если выполнены следующие требования:

- 1) должна быть определена цель исследования;
- 2) должны быть определены факторы, влияющие на зависимый показатель;
- 3) должна быть проведена проверка связи между переменными;
- 4) должен быть определен перечень факторов, которые претендуют на включение в модель;
- 5) должен быть выбран вид модели, учитывающей влияние выбранных факторов;
- 6) должно быть проверено условие: объем выборки должен быть в 6–7 раз больше количества факторов, включенных в модель;
- 7) должно быть проверено условие возможности расчета коэффициентов методом наименьших квадратов;
- 8) должна быть изучена возможность однозначной экономической интерпретации коэффициентов (параметров) модели и проведения прогнозирования.

**Первое** требование предполагает определение цели эконометрического исследования, соответствующей целям и задачам предприятия. Любая организация или предприятие имеют утвержденные миссию, видение, стратегические цели и задачи на определенный период. Цель эконометрического исследования должна соответствовать целям и задачам предприятия.

Целью эконометрического исследования может быть: проверка достижения основных плановых значений экономических показателей, таких как обоснование объема выпуска, цен на выпускаемую продукцию в зависимости от стратегической зоны хозяйствования, обоснования издержек, размера торговых запасов и т. д. с помощью использования эконометрических моделей.

Результатом эконометрического моделирования должно быть создание модели с проверкой ее достоверности, точечным и интервальным прогнозом зависимой переменной.

**Второе** требование предполагает определение факторов, влияющих на зависимый показатель (табл. 5.10).

Таблица 5.10 – Классификация основных эконометрических переменных

| Термин                                               | Определение                                                                                                                                                                                                                                                                             |
|------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Переменная                                           | показатель экономической системы, численные значения которого изменяются                                                                                                                                                                                                                |
| Фактор                                               | причина, которая влияет на зависимую переменную                                                                                                                                                                                                                                         |
| Зависимая переменная                                 | следствие, которое испытывает влияние со стороны факторов                                                                                                                                                                                                                               |
| Основные факторы                                     | факторы, без которых не будет существовать зависимая переменная                                                                                                                                                                                                                         |
| Второстепенные факторы                               | факторы, которые влияют на зависимую переменную, но без которой зависимая переменная сможет существовать.                                                                                                                                                                               |
| Числовая переменная                                  | переменная, которая имеет дискретные или непрерывные численные значения                                                                                                                                                                                                                 |
| Качественная переменная                              | переменная, значения которой принадлежат к определенному классу                                                                                                                                                                                                                         |
| Количественная переменная                            | переменная, которая количественным образом описывает качественную переменную                                                                                                                                                                                                            |
| Текущие переменные                                   | переменные, измеренные в текущий момент времени                                                                                                                                                                                                                                         |
| Пространственные переменные                          | переменные, характеризующие предприятия за один интервал времени                                                                                                                                                                                                                        |
| Временные переменные                                 | переменные, которые характеризуют предприятие за несколько интервалов времени                                                                                                                                                                                                           |
| Пространственно-временные переменные                 | переменные, которые характеризуют несколько предприятий за несколько интервалов времени                                                                                                                                                                                                 |
| Лаговые переменные                                   | переменные, численные значения которых измерены в предшествующие моменты времени по отношению к текущим значениям зависимой переменной. Лаговая переменная может влиять, но не может быть зависимой, так как прошлое может влиять на будущее, но прошлое не зависит от текущего времени |
| Эндогенные переменные                                | внутренние переменные, которые принадлежат к экономической системе. При этом эндогенные переменные могут влиять на другие эндогенные переменные и могут от них зависеть                                                                                                                 |
| Экзогенные переменные                                | внешние переменные, которые влияют на переменные экономической системы, но от них не зависят. Экзогенной переменной является лаговая эндогенная переменная                                                                                                                              |
| Среднее динамическое равновесное значение переменной | значение эндогенной переменной, которое остается стабильным в экономической системе                                                                                                                                                                                                     |

В процессе эконометрического исследования на значения переменных накладывают ограничения:

- предполагается, что численные значения переменной должны изменяться. Это условие обусловлено тем, что нельзя определить степень влияния переменной, значения которой не изменяются. Степень синхронности изменения переменных влияет на величину их связи как между собой, так и с объясняемой переменной;

- количество значений в переменной или объем выборки должен быть в 6-7 раза больше, чем количество факторов в модели;

- все переменные в модели должны иметь изменяющиеся численные значения. Если в модели используется качественная переменная, то необходимо каждому классу дать численное значение;

- численные значения переменных, по возможности, не должны содержать ошибок измерений;

- в значениях переменной не должно быть пропусков. Если какие-то значения в переменной отсутствуют, то следует заполнить их средними значениями или интерполяционными значениями.

Эконометрические исследования должны учитывать следующие свойства значений переменных:

- сильно выделяющееся значение зависимой переменной (выброс) может быть результатом влияния сильно изменившегося одного влияющего фактора или является результатом однонаправленного воздействия большинства объясняющих переменных. Вероятность однонаправленного воздействия большинства объясняющих переменных очень мала. Поэтому причиной этого явления могло послужить изменение такого синхронизирующего фактора, который одновременно повлиял на эти объясняющие переменные. В экономических исследованиях такими синхронизирующими факторами могут быть: праздники; стихийные бедствия; предвыборные компании; сезонность; события, которые вызывают ожидание перемен и другие явления, которые одновременно воздействуют на все население или только на ее часть;

- если переменные изменяются синхронно, то можно предполагать, что эти переменные связаны между собой. Если переменные изменяются хаотично и связь между ними не прослеживается, то можно предполагать, что эти переменные не связаны между собой.

Таким образом, определение влияния на зависимый показатель состоит в определении главных и второстепенных факторов, оказывающих влияние на объясняемую (зависимую) переменную.

Зависимая переменная зависит от:

- условий окружающей физической среды: температура, влажность, давление и степень загрязнения воздуха, освещенность рабочего места, уровень шума и радиации и т. д.;
- условий антропогенной среды: правовая и законодательная среда деятельности предприятия, стиль руководства, моральный климат и традиции, существующие на предприятии и т. д.;
- состояния основных средств: оборудование, технология;
- состояния оборотных средств: материалы, измерения, методы;
- характеристики персонала;
- цикла деловой активности.

Далее составляется список факторов, которые могут влиять на зависимую переменную. В эконометрическом моделировании используется следующее правило: в модель надо включать все «основные» факторы и несколько «второстепенных» с определенной точки зрения. Поэтому в список факторов должны войти все основные и некоторые второстепенные, которые являются наиболее существенными.

**Третье** требование предполагает проведение проверки наличия связи между переменными.

После определения предполагаемых факторов, влияющих на зависимую переменную, необходимо проверить эти связи на фактических данных. Для этой цели надо собрать данных для зависимой переменной  $y$  и предполагаемых факторов  $x$ .

Вычислить матрицу парных коэффициентов корреляции между  $y$  и всеми факторами для определения наличия линейной тенденции с использованием программы Excel «Корреляция».

Если зависимость  $y$  от факторов будет нелинейной, то коэффициент корреляции может быть недостоверным и эти факторы могут выйти из анализа, что приведет к ошибкам спецификации модели. Поэтому необходимо произвести визуальный анализ графиков зависимости  $y$  от всех факторов для определения формы зависимости.

Можно определить вид зависимости  $y$  от  $x$  по ряду функций с помощью графических средств Excel. После обнаружения факторов, линейно или нелинейно влияющих на  $y$ , необходимо перейти к спецификации эконометрической модели.

**Четвертое** требование предполагает проведение ранжирования факторов по мере уменьшения их влияния на зависимую переменную с целью выделения главных факторов, существенно влияющих на зависимую переменную.

**Пятое** требование, предполагающее выбор и обоснование вида функции, может выполняться на основе экономической теории изучаемого процесса или имеющейся фактической связи между переменными.

Для выбора необходимой функции надо знать ее свойства: графики функций или какие виды тенденций она может воспроизвести, линейность коэффициентов и переменных, аддитивный или мультипликативный вид соединения коэффициентов для выбора метода расчета ее коэффициента.

Но в программном продукте Excel, который наиболее распространен, используется не более 10 видов функций. Имеются пакеты прикладных программ, которые позволяют использовать более 2000 видов функций для воспроизведения выявленной закономерности между  $y$  и одним или несколькими факторами.

Таким образом программные средства позволяют воспроизвести любую обнаруженную закономерность между  $y$  и  $x$ .

Если выбранная функция не соответствует экономическому процессу, то ее заменяют другой.

**Шестое требование** предполагает сбор данных по каждой переменной в таком количестве, чтобы он превышал количество факторов в модели в 6–7 раз. Это условие выполнить трудно, особенно для временных рядов, когда количество лет анализируемого показателя очень ограничено или имеется ограниченное количество предприятий. Очень часто в многофакторных моделях количество факторов может превышать объем выборки по каждому фактору. При этом если использовать шаговый метод построения модели, то в результате в полученной модели третье условие может быть выполнено.

**Седьмое требование** предполагает проверку возможности проведения расчетов с помощью метода наименьших квадратов (МНК).

Расчеты коэффициентов модели МНК можно произвести, если:

- модель является линейной относительно коэффициентов и переменных;
- коэффициенты модели соединены аддитивно;
- количество факторов в модели меньше объема выборки.

Дадим более подробное пояснение к этим условиям.

Общий вид множественной регрессионной модели для выборочной совокупности имеет следующую структуру:

$$y_i = f(x_{1,i}, x_{2,i}, x_{3,i}, \dots) + e_i = y_{p,i} + e_i,$$

где  $y_i$  – случайная зависимая переменная;

$y_{p,i}$  – расчетные значения зависимой переменной;

$f(x_{1,i}, x_{2,i}, x_{3,i}, \dots)$  – математическая функция отражает детерминированную составляющую  $y_i$ , которая диктуется законодательной, правовой средой общества в сочетании с потребностями членов общества, закономерностями существующими между экономическими показателями;

$x_{1,i}, x_{2,i}, x_{3,i}, \dots$  – факторы, оказывающие влияние на  $y_i$ ;

$e_i = y_i - y_{p,i}$  – случайная составляющая (остатки, возмущения)  $y_i$ , которая содержит ошибки спецификации, ошибки выборки и ошибки измерения;

$i$  – порядковый номер измерения  $i = 1, \dots, n$ ,

$n$  – объем выборки, при этом объем выборки должен быть в 6-7 раз больше количества факторов, включенных в модель.

В зависимости от количества факторов в модели различают однофакторные (парные, простые) и многофакторные модели.

Пример линейной однофакторной (парной, простой) регрессионной модели:

$$y_i = a_0 + a_1 \times x_i + e_i = y_{p,i} + e_i,$$

где  $a_0$  – свободный коэффициент, равный  $y_p$  при  $x = 0$ ;

$a_1$  – коэффициент пропорциональности зависимости  $y_p$  от  $x$ , численно равный приросту  $y_p$  при изменении  $x$  на 1;

$e$  – остатки модели, которые содержат ошибки спецификации, ошибки выборки и ошибки измерения.

Приведем вид линейной трехфакторной регрессионной модели:

$$y_i = a_0 + a_1 \times x_{1,i} + a_2 \times x_{2,i} + a_3 \times x_{3,i} + e_i,$$

где  $x_1, x_2, x_3$  – факторы, независимые переменные, оказывающие влияние на  $y$ .

Приводим классификацию регрессионных моделей по признаку линейности, переменных, виду соединения коэффициентов.

Переменные являются линейными, если они находятся в первой степени.

Коэффициенты регрессии могут соединяться между собой аддитивно (сложением) или мультипликативно (умножением).

Линейность уравнения регрессии определяют по коэффициентам и по переменным.

Например, линейными уравнениями по коэффициентам и по переменным, коэффициенты соединены аддитивно являются следующие регрессии:



- $y = a_0 + a_1 \times x_1 + e$  – линейное уравнение регрессии, так как коэффициенты  $a_0, a_1$  и переменные  $y, x$  имеют степень 1, коэффициенты соединены аддитивно (сложением).
- $y_t = a_0 + a_1 \times x_{1,t} + a_2 \times x_{1,(t-1)} + a_3 \times x_{1,(t-2)} + et$  – модель распределенного лага – зависимость последующих значений зависимой переменной от предыдущих значений объясняемой переменной.
- $y_t = a_0 + a_1 \times x_t + a_2 \times e(t-1)$  – модель, учитывающая автокорреляцию остатков – зависимость последующих значений остатков модели от своих же предыдущих значений.
- $y_t = a_0 + a_1 \times x_t + a_2 \times y(t-1) + a_3 \times y(t-2) + et$  – авторегрессионная модель, учитывающая влияние на  $y$  от их же предыдущих значений.

Модель считается линейной по коэффициентам, нелинейной по переменным и аддитивным остаткам, если выполняются следующие условия:

- коэффициенты имеют первую степень;
- переменные имеют степень, отличную от 1;
- коэффициенты связаны аддитивно;
- остатки включены в модель аддитивно.

Примерами классов функций, наиболее часто используемых на практике, могут служить:

1. Множество всех линейных зависимостей (парабола первого порядка):  
 $y = a_0 + a_1 x$ .
2. Множество всех квадратичных трехчленов (парабола второго порядка):  
 $y = a_0 + a_1 x + a_2 x^2$ .
3. Множество всех полиномов определенной степени (параболы более высоких порядков):  $y = a_0 + a_1 x + a_2 x^2 + \dots + a_k x^k$ .
4. Множество всех гиперболических зависимостей вида

$$y = a_0 + \frac{a_1}{x}.$$

5. Множество всех показательных зависимостей:  $y = a_0 a_1^x$ .
6. Множество всех степенных зависимостей:  $y = a_0 x^{a_1}$ .
7. Множество логарифмических зависимостей:  $y = a_0 + a_1 \log x$ .

Кроме вышеприведенных, часто используются и другие аналитические зависимости:

$$y = a_0 + a_1 \log(x);$$

$$y = \frac{a_0}{a_1 + x};$$

$$y = a_0 \exp\left(\frac{a_1}{x}\right);$$

$$y = a_0 \times a_1^x;$$

$$y = a_0 + a_1 x^n;$$

$$y = \frac{1}{a_0 + a_1 \exp(x)};$$

$$y = a_0 x^{a_1};$$

$$y = \frac{1}{a_0 + a_1 \log(x)};$$

$$y = \frac{a_0 x}{a_1 + x};$$

$$y = a_0 \times a_1^{\frac{1}{x}};$$

$$y = \frac{1}{a_0 + a_1 x};$$

$$y = \frac{x}{a_0 + a_1 x};$$

$$y = a_0 \exp(a_1 x);$$

$$y = \frac{1}{a_0 + a_1 \exp(-x)}.$$

Коэффициенты уравнения регрессии можно рассчитать методом наименьших квадратов, если коэффициенты являются линейными и соединены аддитивно (сложением).

Если к регрессионному уравнению не применять метод наименьших квадратов, то можно:

- преобразовать его к линейному аддитивному виду с помощью логарифмирования функции и заменой преобразованной переменной;
- предложить специально разработанные формулы расчетов коэффициентов уравнения регрессии, ориентируясь на выявленные тенденции;
- использовать приближенные итеративные методы расчетов коэффициентов, например, с помощью Excel программы «Поиск решения», в которой переменными являются определяемые коэффициенты, целевой функцией является сумма квадратов остатков с условием ее минимизации и позволяет рассчитать коэффициенты всех известных функций.

**Восьмое** требование спецификации модели предполагает наличие возможности четкой экономической интерпретации коэффициентов модели и возможности проведения прогнозирования.

Проблемы с экономической интерпретацией коэффициентов модели возникают для преобразованных данных, которые решаются для каждой модели индивидуально.

Целью всех эконометрических моделей является получение прогноза, поэтому:

- вид математической функции для эконометрической модели должен быть как можно проще, чтобы было удобно проводить расчеты;
- использование сложной функции может воспроизводить не только тенденцию, но и часть случайной составляющей, что приводит к увеличению ошибки в расчетах.

Одним из базовых предположений построения качественной модели является адекватная спецификация эконометрической модели. Адекватная спецификация эконометрической модели должна отражать соотношение между экономическими показателями, участвующими в модели. Это является необходимой предпосылкой дальнейшего качественного оценивания использования результатов моделирования. Неправильный выбор функциональной формы или набора объясняющих переменных называется ошибками спецификации. Поэтому к ошибкам спецификации будут относиться неправильный выбор той или иной математической функции, и недоучет в эконометрической модели кого-либо существенного фактора, например, использование парной регрессии вместо множественной. Таким образом можно классифицировать следующие случаи ошибок спецификации модели:

1. Отбрасывание значимой переменной.
2. Добавление незначимой переменной.
3. Выбор неправильной функциональной формы.

Наряду с ошибками спецификации могут быть ошибки выборки, так как при установлении закономерностей чаще всего используются выборочные данные. Ошибки выборки и появляются в силу неоднородности данных в исходной статистической совокупности, что невозможно исключить при изучении экономических процессов. Если исходная совокупность неоднородна, то построение эконометрической модели лишено смысла. Поэтому при построении эконометрической модели необходимо исключать из совокупности единицы с аномальными значениями исследуемых признаков. И в этом случае результаты регрессии представляют собой выборочные характеристики.

Наибольшую опасность в практическом использовании эконометрического моделирования представляют ошибки измерения. Если ошибки спецификации можно уменьшить, изменяя форму модели (вид математической формулы), а ошибки выборки – увеличивая объем исходных данных. Ошибки измерения практически сводят на нет все усилия по количественной оценке связи между признаками.

## 6. МЕТОДЫ ПРОГНОЗИРОВАНИЯ В ЭКОНОМЕТРИКЕ

### 6.1. Механизм прогнозирования

Прогнозирование необходимо для определения наиболее вероятного будущего, которое произойдет независимо от предпринимаемых усилий. При этом часто прогнозирование заключается в определении значений экономических показателей в будущем. Это может касаться цен на товары, объемов их продаж. Прогнозирование всегда предшествует планированию, которое на основе предполагаемого будущего позволяет выбрать реальные цели, оценить возможность, целесообразность, пути и средства их достижения.

В рыночной экономике прогнозирование является основой всей системы управления. Анри Файоль (1841–1925), один из выдающихся представителей менеджмента первой четверти XX века, разработавший основы административного подхода к управлению, уделял прогнозированию и планированию особое значение. Он указал на необходимость краткосрочного и долгосрочного планирования в каждой организации. По его мнению, предвидение составляет наиболее существенную часть управления. «Предвидеть – ... означает исчислять будущее и подготавливать его; предвидеть – это уже почти действовать»<sup>4</sup>. Прогнозы необходимы в финансировании, маркетинге, подборе кадров и различных производственных областях, в коммерческих и государственных организациях. Можно с уверенностью сказать, что любое управляющее решение, последствия которого проявятся в будущем, основывается на том или ином способе прогнозирования. Всякий раз, когда принимается решение относительно будущих действий, учитываются прежде всего прогностические оценки. Не вызывает никаких сомнений то, что научно обоснованные прогнозы являются по сравнению со случайными и чисто «интуитивными» прогнозами более точными и, следовательно, более эффективными.

Аппарат прогнозирования может использоваться для того, чтобы помочь обосновать решения тому, кто его принимает, опираясь на понимание количественного и качественного аппаратов прогнозирования и разумное их использование. Профессиональный менеджер в процессе принятия решений должен выделить те ситуации, которые можно с большей или меньшей точностью описать математически, т. е. для которых можно построить соответствующие модели и, конечно, должен уметь извлечь из этих моделей необходимую для принятия решений информацию.

С развитием аппарата прогнозирования, а также с появлением компьютеров, оснащенных соответствующим программным обеспечением, прогнозированию уделяется все больше и больше внимания. В настоящее время каждый менеджер имеет возможность использовать в целях прогнозирования аппарат анализа данных, необходимо подчеркнуть, что знание этого аппарата является весьма существенным. Специалисты, непосредственно использующие

---

<sup>4</sup> Файоль, А. Общее и промышленное управление / А. Файоль. – М.: Дело, 1991.

прогнозы в своей деятельности, должны понимать опасность выбора неадекватных методов прогнозирования, так как некорректные прогнозы могут привести к принятию неверных решений.

В отличие от планирования, при котором директивным образом задается будущее движение, сущность экономического прогнозирования состоит в описании и анализе будущего развития.

Прогноз – это вероятностное научно обоснованное суждение о перспективах, возможных состояниях того или иного явления в будущем и (или) об альтернативных путях и сроках их осуществления. Процесс разработки прогнозов называется прогнозированием (от греческого *prognosis* – предвидение, предсказание).

Типология прогнозов строится в зависимости от различных критериев и признаков. Конструктивная классификация позволяет изобразить совокупность методов прогнозирования в виде иерархического дерева и охарактеризовать каждый уровень своим классификационным признаком.

1. В зависимости от **объектов прогнозирования** прогнозы подразделяются на: научно-технические, демографические, социальные, экономические, прогнозы природных ресурсов, внешне-политические прогнозы.

- Научно-технические прогнозы рассматривают достижения научно-технического прогресса, развитие фундаментальных и прикладных исследований, новых видов техники и технологии, определяют последствия НТП.

- Демографические прогнозы оценивают вероятную численность населения в более или менее отдаленном будущем при некоторых научно обоснованных предположениях о возможных изменениях рождаемости и смертности.

- Социальные прогнозы позволяют дать научное обоснование предполагаемых изменений социальных отношений и социальных процессов или их элементов.

- Экономические прогнозы – результат научных исследований о возможных направлениях будущего развития экономики и ее отдельных сегментов.

- Естественно-природные прогнозы характеризуют запасы природных ресурсов и возможности их вовлечения в хозяйственный оборот, состояние растительного и животного мира, окружающей среды.

- Внешнеполитические прогнозы рассматривают перспективы сотрудничества с зарубежными странами, проблемы интеграции национальной экономики в мировое хозяйство, вопросы рационализации экспорта и импорта и т. д.

2. В зависимости от **цели прогнозирования** прогнозы делятся на: поисковые и нормативные.

- Поисковый прогноз основан на условном продолжении в будущее тенденции развития исследуемого объекта в прошлом и настоящем, абстрагируясь от условий, способных изменить эти тенденции. Такой прогноз

отвечает на вопрос: что вероятнее всего произойдет при условии сохранения существующих тенденций?

- Нормативный прогноз представляет определение путей и сроков достижения возможных состояний объекта прогнозирования, принимаемых в качестве цели. Такой прогноз отвечает на вопрос: каким образом можно достичь желаемого состояния прогнозируемого объекта?

3. В зависимости от масштабности объекта прогнозирования экономические прогнозы могут охватывать все уровни от глобального уровня, где существующие закономерности рассматриваются в мировом масштабе; макроуровня, анализирующего экономическое развитие в масштабе страны; до микроуровня, рассматривающего прогнозы развития отдельных предприятий, производств, продуктов.

4. По периоду времени, промежутку времени, на который рассчитан прогнозы различаются:

- Оперативный, как правило, рассчитан на перспективу, на протяжении которой не ожидается существенных изменений объекта исследования – ни количественных, ни качественных.

- Краткосрочный – рассчитан на перспективу, на протяжении которой не ожидается существенных количественных изменений объекта исследования.

- Долгосрочный – рассчитан на перспективу. На протяжении которой ожидаются существенные не только количественные, но и существенные качественные изменения объекта исследования.

- Среднесрочный – охватывает перспективу между кратко- и долгосрочными прогнозами с преобладанием количественных изменений над качественными.

- Дальнесрочный (сверхдолгосрочный) – охватывает перспективу, в течение которой ожидается столь значительные качественные изменения, что можно говорить лишь о самых общих перспективах развития исследуемого явления и процесса.

Временная градация прогнозов является в определенной мере условной и зависит от характера и цели данного прогноза. В социально-экономических прогнозах эмпирически установлен следующий временной масштаб: оперативные прогнозы имеют продолжительность от одного месяца, краткосрочные – до одного года, среднесрочные рассчитаны на несколько (до пяти) лет, долгосрочные – на период свыше пяти лет и, примерно, до пятнадцати – двадцати лет, дальнесрочные – за пределами долгосрочных.

5. В зависимости от характера и объема информации, необходимой для их применения, методы и модели экономического прогнозирования делятся на: одномерные, многомерные (многофакторные) и интуитивные.

- Для одномерных методов прогнозирования общим является использование в качестве независимой переменной только одного фактора – времени и, соответственно, использование только одного временного ряда, составленного из наблюдений того показателя, прогноз которого предстоит

осуществить. К одномерным методам прогнозирования относят кривые роста и адаптивные методы.

- Многофакторные методы и модели позволяют учитывать конкретные, количественные взаимосвязи экономических объектов и процессов. При их использовании для прогнозирования экономических показателей необходимо иметь временные ряды по всем показателям и факторам, участвующим в построении многофакторной модели. К многофакторным моделям относят регрессионные и структурные уравнения.

- Методы экспертных оценок используются в случаях, когда велика вероятность качественного изменения исследуемого объекта или когда невозможно математически формализовать исследуемый процесс из-за отсутствия достаточного количества информации. Основой интуитивных методов и моделей прогнозирования служит мнение эксперта, чаще группы экспертов, основанное на их всестороннем знании исследуемого объекта, интуиции и опыте. Интуитивные методы включают, таким образом, методы, базирующиеся на экспертной информации, а также методы аналогий, патентный метод и другие.

## **6.2. Проблемы прогнозирования**

Методологической основой прогнозирования служит теория развития объекта, которая раскрывает существо закономерностей и содержание основных причинно-следственных связей рассматриваемого процесса. Методы прогнозирования позволяют найти меру влияния отдельных закономерностей и причин на общую тенденцию развития. Это позволяет представить объект прогноза как динамическую систему, поведение которой определяется взаимодействием различных явлений и тем самым представить с определенной степенью вероятности поведение этой системы в будущем.

Практически любой объект может быть представлен как система и в этом качестве он характеризуется следующими признаками:

- целостность (принципиальная несводимость свойств системы к сумме составляющих ее элементов и невыводимость из последних свойств целого);
- зависимость каждого элемента, свойства и отношения системы от его места, функций и т. д. внутри целого);
- сложность (существование очень большого числа различных элементов, неизвестным образом связанных друг с другом);
- природа явлений, протекающих в системе, неизвестна;
- для изучения сложных систем требуется решать задачи с непомерно большим объемом вычислений;
- структурность (возможность описания системы через установление ее структуры, т. е. сети связей и отношений; обусловленность поведения системы не столько поведением ее отдельных элементов, сколько свойствами ее структуры);

- иерархичность (важный компонент системы в свою очередь может рассматриваться как система, а исследуемая система является одним из компонентов более широкой системы);
- взаимозависимость системы и внешней среды;
- целесообразное поведение системы и т. д.

Экономическое прогнозирование характеризует будущее развитие, исходя из гипотезы, что основные тенденции, действующие факторы и зависимости прошлого и текущего периода сохранятся на период прогноза, либо можно предопределить и учесть направление их изменений в рассматриваемой перспективе. Развитие прогнозируемых социально-экономических процессов в этом случае должно быть представлено в виде определенной траектории, то есть процессы в той или иной мере должны обладать инерционностью. Использование перечисленных предположений основано на специфике прогноза как одного из способов познания. Это объясняется тем, что при прогнозировании на первый план выдвигается исследование не того, что есть, а того, что будет, то есть необходимо смоделировать процесс и с помощью модели воспроизвести его характеристики в перспективе. В философии такой способ познания называют опережающим отражением. Оно возможно при выполнении следующих предпосылок<sup>5</sup>:

- прогнозируемая система в своем развитии имеет определенные объективные черты, закономерности, проявляющиеся в повторяемости и устойчивости ряда связей и отношений;
- в развитии системы имеют место сходные, циклически повторяющиеся ситуации, которые позволяют накопить достаточно информации о познавательных сходных ситуациях и на этой основе создать условия для выработки запаздывающего отражения;
- запаздывающее отражение не приводит непосредственно к опережающему отражению, поскольку в развитии любой сложной системы не бывает точного повторения одних и тех же ситуаций, и накопленная информация не может непосредственно использоваться для прогнозирования, необходимо наличие активного субъекта, который переработал бы информацию о развитии прогнозируемого объекта, окружающей его среды и в результате своей деятельности воспроизвел бы этот образ в прогнозируемом периоде.

Прогнозы любого типа немыслимы без определенной системы задач и целей. При этом следует учитывать, что все цели взаимосвязаны и взаимообусловлены (рис. 6.1).

---

<sup>5</sup> Шелобаев, С. И. Теоретико-методологические основы экономического прогнозирования / С. И. Шелобаев, И. С. Шелобаева, В. Д. Бушуев // Фундаментальные исследования. – 2008. – № 1 – С. 125–128.



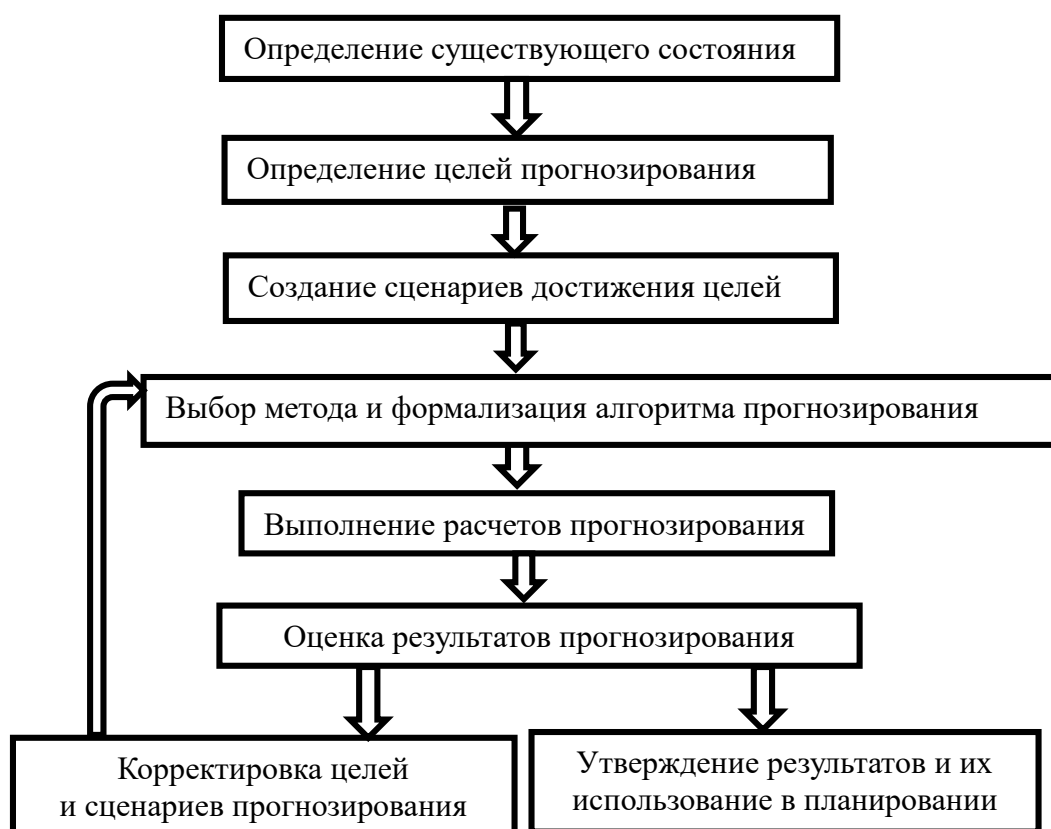


Рисунок 6.1 – Процесс прогнозирования

Поэтому при составлении и реализации прогнозов следует стремиться к определенной совместимости целей и, кроме того, требуется обстоятельный учет всех видов ресурсов, необходимых для достижения той или иной цели. Для современных экономических систем важно соблюдение следующих основных условий:

1. Обязательное достижение соответствия между спросом и предложением на все виды ресурсов, товаров и услуг и создание условий и предпосылок для достижения подобного равновесия. Внешним проявлением достижения подобной сбалансированности в рыночной экономике выступает отсутствие излишков и в меньшей степени дефицитов.

2. Составление системы целей прогнозов и затем их конкретная реализация на основе достижений научно-технического прогресса и научно-технической революции.

3. Соблюдение соответствия между факторами производства на всех уровнях, от специфики каждого зависит и оптимальная их комбинация.

Результаты прогнозирования должны удовлетворять определенным требованиям (рис. 6.2).

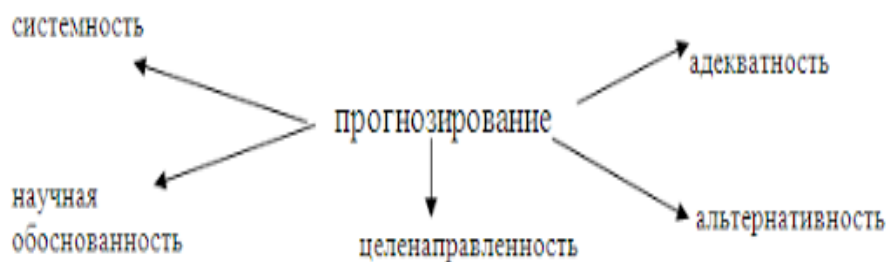


Рисунок 6.2 – Требования к результатам прогнозирования

Методы прогнозирования можно рассматривать с различных точек зрения (рис. 6.3). В задачу прогнозирования входят выявление и исследование потребностей всех экономических объектов и субъектов с последующим адекватным их отражением в системах прогнозов различного уровня и детализацией. При этом следует учитывать, что уровень и структура частных, коллективных и общественных потребностей находятся в непрерывном развитии и должны соответствовать реальным условиям воспроизводственных процессов и требованиям экономических законов рыночных отношений.

Научно разработанные и обоснованные прогнозы, объективно необходимые в экономике, позволяют оптимально достичь цели общественного развития в каждый конкретный промежуток времени, служат основой для создания эффективной сбалансированности экономических систем. На качестве прогнозов сказывается ряд обстоятельств (рис. 6.4).



Рисунок 6.3 – Методы прогнозирования

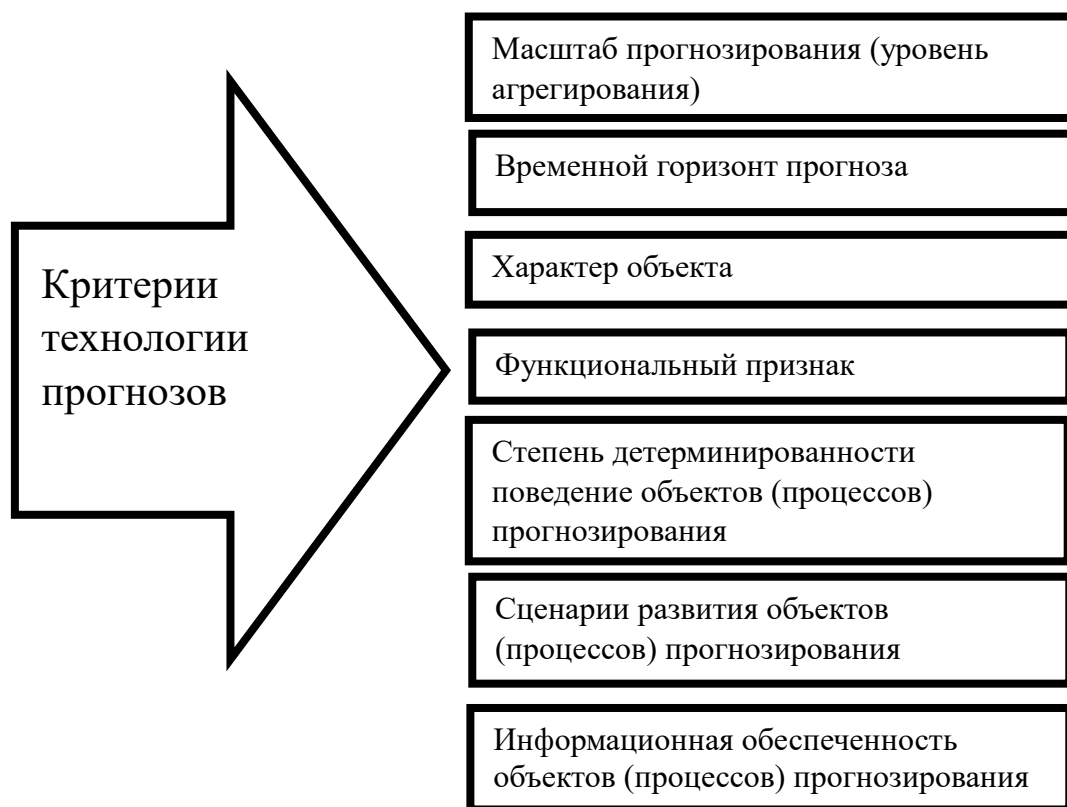


Рисунок 6.4 – Критерии и признаки технологии прогнозов

Прогнозирование в экономике – это всегда попытка предсказать исход события, однако это жизненно важный процесс, поскольку он дает возможность планировать деятельность предприятия. Тем не менее существуют некоторые проблемы, которые необходимо учитывать:

1. Устаревшие данные. Все сведения, применяемые для расчета, будут считаться историческими, и нет никаких гарантий, что аналогичные тенденции сохраняться в будущем.

2. Невозможно учесть абсолютно все факторы, которые смогут повлиять на будущие прогнозные результаты. Так, нельзя предсказать результаты биржевых торгов, рост уровня инфляции, ключевых ставок, процентов по кредитам и вкладам и т. д.

3. Существует риск не принять во внимание фактор, который в настоящее время является незначительным или отсутствует вовсе, но уже завтра его влияние окажется значимым и повлияет на итоговый прогнозируемый результат.

Выбор методов прогнозирования зависит от целей, наличия и качества исходной информации, опыта экспертов и технических возможностей. Краткосрочные прогнозы не требуют применения большого количества методов или привлечения большого количества специалистов. Для них важна оперативность. В долгосрочной перспективе составление прогноза требует большого числа методов с применением компьютерного моделирования, математических и статистических расчетов.

На практике, при выборе метода следует учитывать несколько моментов:

1. Точность и достоверность составления прогноза.
2. Сроки составления и период прогноза.
3. Затраты на процесс прогнозирования.

Проблема выбора заключается в том, что методов очень много, в них сложно ориентироваться, и трудно подобрать те, которые полностью удовлетворяют целям составления прогноза. Сложность объектов прогнозирования повышается, увеличивается число факторов, влияющие на итоговый ожидаемый результат, необходимо оценивать как внешнюю, так и внутреннюю среду.

Успех составления прогноза зависит от качества, полноты и достоверности исходных данных. При отсутствии информации прогнозирование может стать невозможным.

Результаты прогнозирования должны быть простыми и удобными для анализа и использования, быть своевременными и максимально точными, надежными (дающими наибольшую вероятность достижения).

### **6.3. Классификация методов прогнозирования**

Единого универсального метода прогнозирования не существует. Огромное разнообразие прогнозируемых ситуаций приводит к необходимости столь же большого разнообразия методов прогнозирования. В настоящее время насчитывается свыше 150 методов прогнозирования, но наиболее широко используются около 20 методов. Для всех без исключения методов прогнозирования характерно использование статистических данных. Только одни методы ориентированы на непосредственную обработку рядов динамики, а другие – на косвенное использование статистических данных.

В самом общем виде методы прогнозирования можно разделить на три класса:

- 1) фактографические методы;
- 2) экспертные методы;
- 3) комбинированные методы.

Фактографические методы базируются на статистической информации, представляющей ряды динамики. Наиболее эффективно применение этих методов при эволюционном развитии прогнозируемых процессов.

Экспертные (интуитивные) методы основаны на использовании знаний специалистов-экспертов об объекте прогнозирования и обобщении их мнений о поведении этого объекта в будущем. Статистическая информация используется опосредованно и проявляется в системе знаний эксперта. Экспертные методы более успешно применяются для прогнозирования скачкообразных процессов.

Комбинированные методы используют как фактографическую, так и экспертную информацию.

Кроме рассмотренной выше существуют и другие классификации методов прогнозирования. Например, существующее многообразие методов

прогнозирования условно представлено на рис. 6.5.

Наконец, существуют эвристические методы прогнозирования, основанные на приемах вычисления и процедурах, вытекающих из опыта и интуиции специалистов, осуществляющих прогноз. Эвристические методы используются в тех случаях, когда применение строгих математических моделей не обеспечивает достоверных результатов прогноза из-за того, что лежащие в их основе предпосылки не соответствуют реальным свойствам поведения прогнозируемого процесса или объекта. При применении эвристических методов предполагают, что подходы, используемые для формирования прогноза, не изложены в явной форме и неотделимы от лица, делающего прогноз. При формировании прогноза доминируют интуиция и опыт. К данной категории относятся методы социологических исследований и экспертные методы. При этом специалист использует субъективные оценки как свои собственные, так и полученные в результате опросов, а также определенные причинно-следственные связи, результаты обработки и анализа статистических данных.

| Методы прогнозирования             |                             |                                |                        |
|------------------------------------|-----------------------------|--------------------------------|------------------------|
| Экспертные методы                  |                             | Метод<br>экстраполяции         | Метод<br>моделирования |
| индивидуальные                     | коллективные                |                                |                        |
| интервью                           | комиссий                    | аналитического<br>выравнивания | имитационные           |
| аналитическая<br>докладная записка | «мозговой атаки»            | скользящих<br>средних          | логические             |
|                                    | Дельфи                      | огibaющей<br>кривой            | статистические         |
|                                    | морфологического<br>анализа | регрессионный<br>анализ        | информационные         |

Рисунок 6.5 – Условная классификация методов прогнозирования

При прогнозировании на основе моделирования широко применяется метод исторической и математической аналогии.

Метод исторической аналогии основан на установлении и использовании аналогии объекта прогнозирования с другими, но одинаковыми по закономерности поведения объектами, которые опередили в развитии рассматриваемый объект. В качестве примера можно использовать прогнозирование развития экономики стран, находящихся на более раннем этапе развития по сравнению со странами, имеющими развитую рыночную экономику.

Метод математической аналогии основан на установлении общих закономерностей развития рассматриваемого процесса и процесса-прототипа, для которого уже существует математическое описание. Для прогнозирования используется это уже известное математическое описание.

В условиях динамично изменяющейся среды эффективность использования методов прогнозирования, основанных на предыстории развития процесса, не всегда оправдана. Тогда при прогнозировании приходится опираться на опыт и интуицию экспертов, т. е. специалистов, имеющих значительный опыт в интересующей области. Методы прогнозирования, основанные на обработке суждений и предположений, высказанных экспертами, называются методами экспертных оценок или просто экспертными. Одним из центральных аспектов экспертных методов является системный анализ прогнозируемого состояния объекта или процесса. Этот анализ заключается в представлении процесса как результата взаимодействия нескольких более простых явлений, поведение которых более предсказуемо. Оценка по отдельным элементам значительно облегчает процесс прогнозирования и делает его более объяснимым. Задача эксперта заключается в выявлении наиболее существенных компонентов процесса и их оценки. Такой подход является общим для всего прогнозирования, а не прерогативой только экспертных методов. Различие заключается в том, что высказывания экспертов базируются на их личном опыте и не могут быть формально обоснованы.

Следующим центральным моментом экспертного прогнозирования является проведение опроса экспертов. В зависимости от целей и задач экспертизы, существа и сложности проблемы, отведенного времени, выделенных средств, а также квалификации участников выбирается метод опроса, который может различаться по следующим позициям:

- опрос может быть индивидуальным или групповым;
- опрос может быть экспедиционным (интервью) или корреспондентским (рассылка анкет или опросных писем);
- в случае анкеты вопросы могут быть открытые или закрытые.

Индивидуальный опрос позволяет максимально использовать способности и знания каждого специалиста. В отличие от индивидуального при групповом опросе происходит обмен мнениями между экспертами, уточняются оценки и корректируются выводы. Коллективный опрос позволяет стимулировать выдвижение новых точек зрения и выявлять новые аспекты прогнозируемого процесса. В то же время недостатком такого опроса является сильное влияние лидеров на большинство участников экспертизы.

Интервью предполагает проведение беседы с экспертом организатора процесса прогнозирования. Этот метод требует соответствующей квалификации не только от эксперта, но и от интервьюера, который должен быстро, фактически экспромтом реагировать на вопросы и ответы.

Метод аналитических докладных записок заключается в предоставлении экспертом в письменной форме экспертных оценок и их обоснований. Эксперт должен оценить состояние и тенденции развития прогнозируемого процесса. Для оценки важности проблем и решений широко используются ранговые шкалы, позволяющие выявить предпочтения эксперта, а для обработки высказанных предпочтений используются методы многомерного шкалирования (методы рангов).

Метод «мозговой атаки» («мозгового штурма») используется для

коллективной генерации идей и творческого решения проблем и представляет собой неструктурированный процесс генерирования и оценки идей по поставленной проблеме. Формы применения этого метода могут варьироваться.

При одном варианте эксперты высказываются поочередно. Затем с высказываниями знакомятся все участники, и на втором этапе производится оценка предложенных идей.

В другом варианте заседание экспертов проводится по «круговой» системе. Группа специалистов подразделяется на подгруппы из 3–4 человек, каждый из которых должен зафиксировать на выданной карточке по две-три идеи. Затем внутри подгруппы происходит обмен карточками, которые дополняются и детализируются остальными членами подгруппы. Затем на рассмотрение всей группы экспертов представляется отчет о проделанной работе и производится оценка и выбор представленных вариантов прогноза.

Метод «мозговой атаки наоборот» отличается от предыдущего метода тем, что экспертам предлагается выявить недостатки предлагаемых идей и предложить меры для их устранения.

Метод «мысленного группового анализа реальной ситуации» применяется для большой группы экспертов (около 20 человек) при количественной оценке ситуации или процесса (например, фигурное катание и другие аналогичные виды спорта). Оценка ситуации обсуждается и уточняется в ходе последующего заседания группы.

Метод «дерева целей» широко применяется для прогнозирования возможных направлений развития науки, техники, технологии. «Дерево целей» позволяет увязывать перспективные цели и конкретные задачи на каждом уровне иерархии. При этом цель высшего порядка соответствует вершине «дерева», а «ниже», в несколько ярусов, располагаются локальные цели и задачи, с помощью которых обеспечивается достижение целей верхнего уровня. Таким образом устанавливается причинно-следственная связь между целями и условиями их достижения. Следствием такого построения является логическая последовательность условий, при которых может быть достигнуто то или иное состояние системы. Далее остается только оценить вероятность описанных событий и другие необходимые в конкретной ситуации параметры.

В основу морфологического анализа положено построение матрицы характеристик процесса, например рынка, и их возможных значений. Далее на основе перебора характеристик процесса получают различные варианты прогноза.

Матричный метод широко используется в маркетинге для прогнозирования оценки позиции фирмы на рынке. Это позволяет в дальнейшем планировать деятельность фирмы и принимать решение о выборе одной из возможных стратегий развития фирмы, включая инвестиционную и инновационную деятельность. Многочисленные примеры таких матриц приведены в учебниках маркетинга (например, матрица BCG – Boston Consulting Group, матрица Маккинзи – стратегическая маркетинговая матрица).

Метод Дельфи является наиболее формальным из всех методов экспертного прогнозирования. Это групповой метод, при котором проводится индивидуальный опрос группы экспертов, как правило, относительно их

предположений о будущих событиях. Опрос проводится анонимно, т. е. личные контакты и коллективные обсуждения исключаются. Полученные ответы сопоставляются, и обобщенные результаты снова направляются той же группе экспертов. На основе этой информации каждый эксперт, сохраняя анонимность, делает дальнейшие предположения о будущем, причем этот процесс может повторяться несколько раз до совпадения мнений. Полученные результаты используются в качестве прогноза.

Метод составления сценариев является наиболее популярным за последнее десятилетие. Термин «сценарий» впервые предложен в 1960 году футурологом Х. Каном при прогнозе состояния общества будущего для решения стратегических вопросов в военной области. Сценарий – это описание будущего процесса или состояния системы с учетом правдоподобных предположений (гипотез). В большинстве случаев составляется три сценария: оптимистический, пессимистический и наиболее вероятный. Для прогнозирования характерно наличие нескольких вариантов развития событий. Для каждого варианта развития необходимо составить свой сценарий. Сценарий включает составление списка существенных факторов, определение их значений, генерирование гипотезы об изменении показателя и его расчет, определение значения прогнозируемого показателя.

По всем сценариям производится выделение наиболее вероятного значения прогнозируемого показателя на основе анализа полученных результатов. Метод составления сценариев может использоваться для качественной оценки состояния процесса или в качестве обязательной компоненты при использовании других способов прогнозирования.

Методы прогнозирования можно классифицировать по двум измерениям: степень свободы прогнозирования от субъективности и большей или меньшей степени формализации этого процесса. Совмещение этих двух измерений позволяет идентифицировать четыре типа прогнозных методов (рис. 6.6).

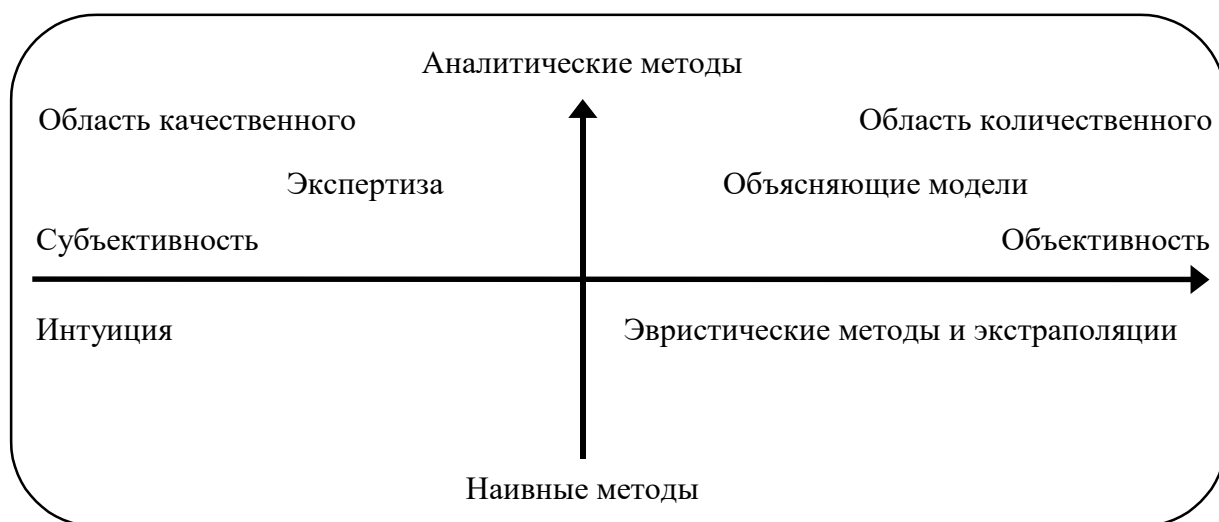


Рисунок 6.6 – Типология методов прогнозирования

При использовании *субъективных* методов предполагается, что процессы,



используемые для получения прогноза, не формализованы и неотделимы от лица, делающего прогноз. Для *объективных* методов характерно наличие четкой формулировки и алгоритма, а также возможность воспроизведения такого же результата другими лицами, действующими согласно заданному алгоритму. Эта классификация (по оси абсцисс) как бы противопоставляет количественные методы качественным. Для *наивных* методов характерно формирование прогноза на основе наблюдения за прошлыми эволюциями переменной, без учета в явной форме существенных факторов. *Причинно-следственные* методы построены на идентификации существенных факторов, будущие значения которых уже спрогнозированы. На этой основе выводится вероятное значение прогнозируемого показателя при условии реализации принятого сценария. Второе измерение противопоставляет методы экстраполяции методам, позволяющим объяснить взаимосвязи, независимо от их качественного или количественного характера. Но резкого разграничения между применяемыми методами прогнозирования провести нельзя, так как при любом формировании прогноза явно или неявно составляется сценарий, выдвигаются гипотезы, используется опыт и интуиция специалиста в данной области. Единственным исключением из этого правила является формальное применение того или иного метода без проведения обоснования использования способа и всеобъемлющего анализа полученного результата.

Таким образом, общим для всех методов прогнозирования является использование в той или иной степени статистических данных, описывающих предысторию состояния процесса. Рассмотрим более подробно статистические методы прогнозирования.

#### **6.4. Регрессия в прогнозировании и планировании**

Парная и множественная регрессия широко используется в решении проблем спроса, доходности акций, при изучении функции издержек производства, в макроэкономических расчетах и в ряде других вопросов экономики. В настоящее время регрессия – это один из наиболее распространенных методов в эконометрических расчетах. Использование регрессионной модели для прогнозирования состоит в подстановке в уравнение регрессии ожидаемых значений факторных признаков для расчета точечного прогноза результативного признака или (и) его доверительного интервала с заданной вероятностью.

Прогнозирование – это предсказание будущего на основании накопленного опыта и текущих предположений относительно него. Прогнозирование представляет сложный процесс, по ходу которого необходимо решать большое количество различных вопросов.

Существующие в экономике методы прогнозирования на основе регрессии классифицируются:

- По признаку оценки. Выделяют качественные и количественные методы. Качественные – это, как правило, экспертные оценки, основанные на суждениях и умозаключениях, когда отсутствуют ретроспективные данные для

расчетов. Количественные подразумевают возможность проведения расчетов на основе числовых показателей прошлых и текущего периодов. Базируются на математических методах.

– По информационной составляющей. Статические (или формализованные) основаны на использовании сведений о динамике развития явления, наиболее достоверны, имеют короткие сроки выполнения, точность расчетов, минимизируют проблемы при сборе и оценке сведений. Интуитивные методы основаны на проведении аналогий, оценке имеющихся закономерностей. Необходимость экспертной оценки в соответствующей области. Оцениваются схожие процессы и явления.

– По степени охвата задач. Все методы прогнозирования можно условно разделить на общие и специализированные методы. Общие методы применимы для широко спектра проблем и могут применены в различных сферах деятельности. Специализированные методы ориентированы на определенную сферу деятельности и не могут быть применены в иных целях, поскольку не дадут достоверных и точных результатов, либо же необходимы специфические вводные данные.

Для прогнозирования в сочетании со сценарным подходом применяются различные методы, которых на сегодняшний день существует огромное множество, но на практике используются всего 15–20. Остановимся на наиболее популярных из них (рис. 6.7):

1. Метод экспертных оценок. Суть данного метода заключается в том, что в основе прогноза лежит мнение одного специалиста или группы специалистов, которое основано на профессиональном, практическом и научном опыте. Различают коллективные и индивидуальные экспертные оценки, часто используется при оценке персонала.

2. Метод экстраполяции. Основная идея экстраполяции – изучение сложившихся как в прошлом, так и настоящем стойких тенденций развития предприятия и перенос их на будущее. Различают прогнозную и формальную экстраполяцию. Формальная – основывается на предположении о том, что в будущем сохранятся прошлые и настоящие тенденции развития предприятия; при прогнозной – настоящее развитие увязывают с гипотезами о динамике предприятия с учетом того, что в будущем изменится влияние на него различных факторов. Следует знать, что методы экстраполяции лучше применять на начальной стадии прогнозирования, чтобы выявить тенденции изменения показателей.

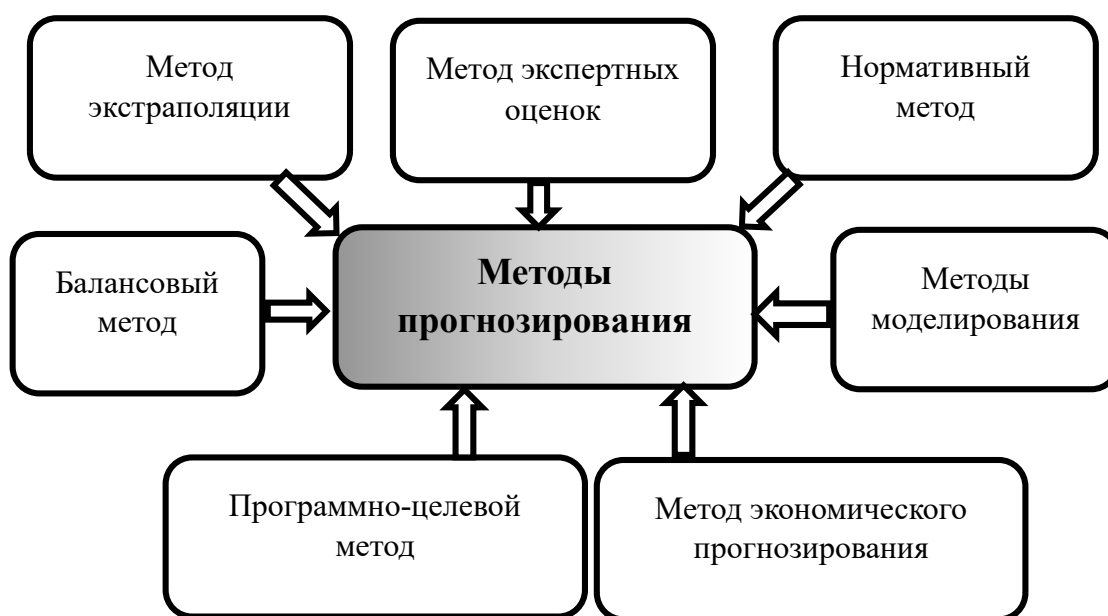


Рисунок 6.7 – Наиболее популярные методы прогнозирования

3. Методы моделирования. Моделирование – это конструирование модели на основании предварительного изучения объекта и процессов, выделение его существенных признаков и характеристик. Прогнозирование с использованием моделей включает ее разработку, экспериментальный анализ, сопоставление результатов предварительных прогнозных расчетов с фактическими данными состояния процесса или объекта, уточнение и корректировку модели.

4. Метод экономического прогнозирования (экономический анализ) заключается в том, что какой-либо экономический процесс или явление, имеющие место на предприятии, расчленяются на части, после чего выявляется влияние и взаимосвязь этих частей на ход и развитие процесса, а также друг на друга. При помощи анализа можно раскрыть сущность такого процесса, а также определить закономерности его изменения в будущем, всесторонне оценить пути достижения поставленных целей. Поскольку экономический анализ – это необъемлемая часть и один из элементов логики прогнозирования, он должен осуществляться на макро-, мезо- и микроуровнях. Используется при планировании производства на предприятии. Процесс экономического анализа можно подразделить на несколько стадий:

- а) постановка проблемы, определение критериев оценки и целей;
- б) подготовка необходимой для анализа информации;
- в) аналитическая обработка информации после ее изучения;
- г) разработка рекомендаций о возможных путях достижения целей;
- д) оформление результатов.

5. Балансовый метод. Данный метод основан на разработке балансов, которые представляют систему показателей, где первая часть, характеризующая ресурсы по источникам их поступления, равна второй, отражающей распределение их по всем направлениям расхода. При помощи балансового

метода воплощается в жизнь принцип пропорциональности и сбалансированности, который применяется при разработке прогнозов. Его суть заключается в увязке потребностей предприятия в различных видах сырьевых, материальных, финансовых и трудовых ресурсах с возможностями производства продукта и источниками ресурсов. Таким образом, система балансов, которую используют в прогнозировании, включает: финансовые, материальные и трудовые балансы. В каждую из данных групп входит еще ряд балансов.

6. Нормативный метод – один из основных методов прогнозирования. В настоящее время ему стало придаваться большое значение. Его сущность заключается в технико-экономических обоснованиях прогнозов с использованием нормативов и норм. Последние применяются при расчете потребности в ресурсах, а также показателей их использования.

7. Программно-целевой метод (ПЦМ). В сравнении с другими методами данный метод является сравнительно новым и недостаточно разработанным. Он начал широко применяться только в последние годы. ПЦМ тесно связан с уже рассмотренными методами и предполагает разработку прогноза начиная с оценки итоговых потребностей на основании целей развития предприятия при дальнейшем определении и поиске эффективных средств и путей их достижения, а также ресурсного обеспечения. Суть ПМЦ заключается определении основных целей развития предприятия, разработки взаимосвязанных мероприятий по их достижению в заранее определенные сроки при сбалансированном обеспечении ресурсами, а также с учетом эффективного их использования.

Кроме прогнозирования, ПМЦ применяется при создании комплексных целевых программ, которые представляют документ, где отражены цель и комплекс производственных, организационно-хозяйственных, социальных и других мероприятий и заданий, увязанных по исполнителям, срокам осуществления и ресурсам.

Основные этапы построения эконометрических моделей и использования их для прогнозирования:

1. Выбрать адекватную теорию, которая объясняет поведение экономической системы и нужны переменные. Переменные, значения которых определяются в модели являются эндогенными, а значения которых определяются вне модели – экзогенными.

2. Отразить теорию в виде системы уравнений, которое связывает выбранные переменные.

3. Найти данные о значениях переменных, следуя насколько это возможно, теоретических концепций.

4. Используя соответствующие эконометрические методы оценить числовые значения неизвестных параметров, входящих в уравнение.

5. Спрогнозировать значения экзогенных переменных. На основе уравнений с оцененными параметрами и прогноза экзогенных переменных сделать предсказания значений эндогенных переменных.

Существующие методы прогнозирования имеет как преимущества, так и недостатки. Абсолютно одинаковые прогнозы получить с помощью разных методов практически очень сложно, поэтому если прогнозные значения,

полученные двумя методами, не совпадают, можно говорить о возможности и/или необходимости проведения комбинированного прогноза. Для получения комбинированной оценки прогноза необходимо решить две задачи:

1. Установить область, внутри которой прогнозные результаты, полученные разными методами, могут считаться согласованными.

2. Установить такое соотношение между результатами прогнозов, которое наиболее адекватно отражало бы их связь с наиболее вероятным результатом прогнозирования.

Процесс прогнозирования проходит в несколько этапов:

1. Определение цели составления прогноза и постановка конкретных задач. Помимо этого, определяется состав рабочей команды, формулируются задания для них, определяется информационная составляющая (источники данных, какие сведения необходимы).

2. Проведение анализа: определяются критерии оценки, качественные и количественные показатели, анализируются внутренние и внешние условия.

3. Выбор метода прогнозирования. Начиная от момента сбора необходимых данных, до проведения анализа, расчетов, экспертных оценок.

4. Аналитическая оценка полученных результатов.

5. Принятие управленческих решений, разработка рекомендаций на основе полученного прогноза.

Прогнозирование по корреляционной модели предполагает оценку ожидаемых значений зависимой переменной при заданных значениях независимых переменных, входящих в уравнение. Различают точечный и интервальный прогнозы.

Точечный прогноз – это:

- прогноз, которым указывается единственное значение прогнозируемого показателя.

- расчетное значение зависимой переменной, полученное подстановкой в уравнение множественной линейной регрессии прогнозных (заданных исследователем) значений независимых переменных.

Если заданы значения  $x_1^{пр}, x_2^{пр}, \dots, x_p^{пр}$ , то прогнозное значение зависимой переменной (точечный прогноз) будет равно:

$$\hat{y}_{пр} = a_0 + a_1 x_1^{пр} + a_2 x_2^{пр} + \dots + a_p x_p^{пр}. \quad (6.1)$$

Интервальный прогноз – это минимальное и максимальное значения зависимой переменной, в промежуток между которыми она попадает с заданной долей вероятности и при заданных значениях независимых переменных.

Интервальный прогноз для линейной функции вычисляется по формуле:

$$\hat{y}_{пр} = \pm t_T s_y, \quad (6.2)$$

где  $t_T$  – теоретическое значение критерия Стьюдента при  $df = n - (m-1)$  степенях свободы;

$s_y$  – стандартная ошибка прогноза.

Найдем прогнозные значения поступления налогов, при условии, что связь между показателями описывается уравнением:

$$y = 3226,44 + 2,45x_1 + 0,06x_2 + 0,31x_3.$$

Зададим прогнозные значения независимых переменных:

- количество занятых  $x_1$ : 500 тыс. человек;
- объем отгрузки в обрабатывающих производствах  $x_2$ : 65 000 млн руб.;
- производство энергии  $x_3$ : 15 000 млн руб.

Найдем точечный и интервальный прогноз поступления налогов. При заданных значениях независимых переменных поступление налогов в среднем составит:

$$y = 3226,44 + 2,45 \times 500 + 0,06 \times 65000 + 0,31 \times 15000 = 18013 \text{ млн руб.}$$

Вектор прогнозных значений независимых переменных будет иметь вид:

$$x_{\text{пр}} = \begin{pmatrix} 1 \\ 500 \\ 65000 \\ 15000 \end{pmatrix}.$$

Ошибка прогноза составила 5556,7. Табличное значение t-критерия при числе степеней свободы  $df = 44$  и уровне значимости  $\alpha = 0,05$  равно 2,0154. Следовательно, прогнозные значения поступления налогов будут с вероятностью 0,95 находиться в границах:

- от  $18\,013,69 - 2,0154 \cdot 5556,7 = 6814,1$  млн руб.;
- до  $18\,013,69 + 2,0154 \cdot 5556,7 = 29\,212$  млн руб.

Прогнозирование по нелинейным моделям множественной регрессии также можно осуществлять по формулам (6.1)–(6.3), предварительно линеаризовав указанные модели.

## 6.5. Формы многофакторной корреляционной зависимости

При определении параметров линейной многофакторной корреляционной зависимости  $y = a_0 + a_1x + a_2z$  для определения параметров  $a_0, a_1, a_2$  используется система нормальных уравнений, имеющая определенную структуру:

$$\begin{cases} na_0 + a_1 \sum x + a_2 \sum z = \sum y; \\ a_0 \sum x + a_1 \sum x^2 + a_2 \sum zx = \sum yx; \\ a_0 \sum z + a_1 \sum xz + a_2 \sum z^2 = \sum yz. \end{cases}$$

Поэтому для получения параметров нелинейных корреляционных уравнений иногда достаточно осуществить замену переменной, а затем, получив систему линейных уравнений, определить неизвестные параметры.

Для полулогарифмической многофакторной корреляционной зависимости:

$$y = a_0 + a_1 \ln x + a_2 \ln z,$$

чтобы определить значения параметров искомого корреляционного уравнения, достаточно заменить переменные  $x' = \ln x, z' = \ln z$ .

Аналогичная ситуация возникает и во многих других случаях, например:

1. Для гиперболической многофакторной корреляционной зависимости (между тремя факторами):

$$y = a_0 + \frac{a_1}{x} + \frac{a_2}{z}$$

осуществляется замена переменной  $x' = \frac{1}{x}, z' = \frac{1}{z}$ .

2. Для показательной многофакторной корреляционной зависимости:

$$y = a_0 a_1^x a_2^z$$

сначала производится процедура логарифмирования, т. е. для определения параметров корреляционное уравнение преобразуется к виду:

$$\log y = \log a_0 + x \log a_1 + z \log a_2.$$

3. Аналогично степенная многофакторная корреляционная зависимость:

$$y = a_0 x^{a_1} z^{a_2}$$

для определения параметров приводится к виду:

$$\log y = \log a_0 + a_1 \log x + a_2 \log z.$$

4. Для параболической многофакторной корреляционной зависимости (между тремя факторами):

$$y = a_0 + a_1 x + a_2 z + a_3 x^2 + a_4 z^2$$

система нормальных уравнений для определения параметров  $a_0, a_1, a_2, a_3, a_4$  имеет вид:

$$\begin{cases} na_0 + a_1 \sum x + a_2 \sum z + a_3 \sum x^2 + a_4 \sum z^2 = \sum y; \\ a_0 \sum x + a_1 \sum x^2 + a_2 \sum xz + a_3 \sum x^3 + a_4 \sum xz^2 = \sum yx; \\ a_0 \sum z + a_1 \sum xz + a_2 \sum z^2 + a_3 \sum x^2 z + a_4 \sum z^3 = \sum yz; \\ a_0 \sum x^2 + a_1 \sum x^3 + a_2 \sum xz^2 + a_3 \sum x^4 + a_4 \sum x^2 z^2 = \sum x^2 y; \\ a_0 \sum z^2 + a_1 \sum xz^2 + a_2 \sum z^3 + a_3 \sum x^2 z^2 + a_4 \sum z^4 = \sum z^2 y. \end{cases}$$

## 6.6. Мультиколлинеарность данных

Мультиколлинеарность – это тесная корреляционная взаимосвязь между отбираемыми для анализа факторами, совместно воздействующими на общий результат, которая затрудняет оценивание регрессионных параметров.

Выделяют два вида мультиколлинеарности:

- строгая мультиколлинеарность (ее также называют полной или точной);
- нестрогая мультиколлинеарность (ее также называют частичной).

Под строгой мультиколлинеарностью понимается ситуация, когда между регрессорами в модели есть точная линейная связь, т. е. когда одна объясняющая переменная точным образом линейно выражается через другие.

Представим, например, что при анализе макроэкономической модели в качестве переменных в нее включили экспорт, импорт и чистый экспорт. Чистый экспорт равен разности между экспортом и импортом и, следовательно, при включении в модель этих трех переменных окажется, что регрессоры модели линейно выражаются друг через друга. *(Примечание: Регрессор – это имя, данное любой переменной в регрессионной модели, которая используется для прогнозирования переменной отклика. Когда есть только один регрессор, модель называется простой моделью линейной регрессии, а когда есть несколько регрессоров, модель называется моделью множественной линейной регрессии, чтобы указать, что существует несколько регрессоров. В модели  $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + z$ ,  $x_i$  – регрессоры;  $\beta_i$  – регрессоров;  $z$  – ошибка;  $y$  – переменная ответа).*

В терминах матричной записи точная мультиколлинеарность предполагает линейную зависимость столбцов матрицы регрессоров, откуда следует неполный ранг матрицы регрессоров. Это означает, что при полной мультиколлинеарности невозможно вычислить МНК-оценки коэффициентов, потому что матрица  $X'X$  является вырожденной, и матрица  $(X'X)^{-1}$  не определена.

Из определения и из приведенного выше примера легко догадаться, как можно решить проблему строгой мультиколлинеарности. Для этого следует исключить лишнюю переменную. Например, если в модели уже учтены экспорт и импорт, то понятно, что включение еще и чистого экспорта не принесет никакой дополнительной информации, и этой третьей переменной можно безболезненно пожертвовать.

Современные эконометрические пакеты при возникновении чистой мультиколлинеарности сами избавляются от одной из линейно зависимых переменных, чтобы вычисление МНК-оценок стало технически возможным.

Частичная мультиколлинеарность – это ситуация, когда между объясняющими переменными нет точной линейной связи, но эти переменные сильно коррелируют между собой. Иными словами, они не линейно зависимы, но «почти» линейно зависимы. При частичной мультиколлинеарности вычислить МНК-оценки можно, однако стандартные ошибки оценок коэффициентов оказываются высокими, а точность оценок коэффициентов – низкой. Так происходит потому, что при сильной корреляции двух регрессоров в выборке они, как правило, меняются одновременно, и оказывается трудно отличить влияние одного регрессора на зависимую переменную от влияния



другого. Таким образом, основным негативным последствием мультиколлинеарности является снижение точности оценки отдельных коэффициентов.

Частичная мультиколлинеарность не нарушает ни одну из предпосылок классической линейной модели множественной регрессии и поэтому не приводит к смещению оценок коэффициентов модели.

Это особенно хорошая новость потому, что на практике почти в любой множественной регрессии объясняющие переменные в той или иной степени коррелированы. Поэтому частичная мультиколлинеарность в данных наблюдается очень часто. Представьте, например, что вы моделируете выпуск фирмы некоторой отрасли в зависимости от количества используемых фирмой труда и физического капитала (то есть, как сказали бы экономисты, моделируете производственную функцию). Скорее всего, в вашей выборке будут большие фирмы и маленькие, причем большие фирмы в среднем будут использовать относительно много каждого из факторов производства, а маленькие, напротив, относительно мало. В результате переменные, характеризующие количество труда и количество капитала, будут положительно коррелированы друг с другом.

Среди последствий частичной мультиколлинеарности можно выделить следующие:

- увеличение дисперсий оценок параметров;
- уменьшение значений  $t$ -статистик для параметров, что приводит к неправильному выводу об их статистической значимости (*Примечание:  $t$ -статистика – это отношение разности в оценочном значении числа от его предполагаемого значения к его стандартной ошибке. Она используется при проверке гипотез с помощью  $t$ -критерия Стьюдента.  $t$ -статистика используется в  $t$ -критерии для определения того, следует ли поддерживать или отвергать нулевую гипотезу*);
- получение неустойчивых оценок параметров модели и их дисперсий;
- возможность получения неверного с точки зрения теории знака у оценки параметра.

Точные количественные критерии для обнаружения частичной мультиколлинеарности отсутствуют. В качестве признаков ее наличия чаще всего используют следующие:

- превышение некоего порога модулем парного коэффициента корреляции между регрессорами  $x_i$  и  $x_j$ ;
- близость к нулю определителя матрицы  $X'X$ ;
- большое количество статистически незначимых параметров в модели.

Есть несколько способов выявить мультиколлинеарность на этапе предварительного анализа данных (то есть еще до оценки параметров уравнения регрессии). О наличии существенной частичной мультиколлинеарности в модели говорят:

- 1) большие по абсолютной величине (больше 0,9) парные коэффициенты корреляции между регрессорами;
- 2) близость к нулю определителя матрицы  $X'X$ ;

- 3) большие (больше 10) значения коэффициентов VIF. (Примечание: Коэффициент инфляции дисперсии или VIF – это мера, которая помогает обнаружить степень мультиколлинеарности в множественной регрессионной модели. Он показывает, насколько увеличивается дисперсия коэффициента оценки из-за линейной зависимости с другими предикторами).

Коэффициенты VIF (variance inflation factor) показывают, насколько сильно связаны друг с другом регрессоры модели. Чтобы определить коэффициент VIF, соответствующий регрессору  $x_i$ , нужно оценить вспомогательную регрессию, в которой слева стоит  $x_i$ , а справа – все остальные объясняющие переменные исходной модели. После этого нужно вычислить коэффициент VIF по формуле:

$$VIF = \frac{1}{1-R^2},$$

где  $R^2$  – это коэффициент детерминации из оцененной вспомогательной регрессии.

Если коэффициенты VIF для всех регрессоров оказались меньше 10, это значит, что существенной мультиколлинеарности в модели не наблюдается. В противном случае стоит сделать вывод о том, что в модели есть мультиколлинеарность.

*Пример. Мультиколлинеарность.* Исследуется модель:

$$y_i = \beta_1 + \beta_2 x_i + \beta_3 w_i + \beta_4 z_i + \varepsilon_i.$$

На этапе предварительного анализа данных оцениваются следующие вспомогательные уравнения:

$$\hat{x}_i = 10,1 + 1,9w_i + 2,3z_i, R^2 = 0,95$$

$$\hat{w}_i = 18,7 + 0,8x_i + 4,8z_i, R^2 = 0,99$$

$$\hat{z}_i = -5,0 + 0,1w_i + 0,3x_i, R^2 = 0,20.$$

Что можно сказать о наличии мультиколлинеарности в исходной модели?

*Решение:*

Коэффициенты VIF для переменных  $x$ ,  $w$ ,  $z$  равны, соответственно:

$$1/(1 - 0,95) = 20 \quad 1/(1 - 0,99) = 100 \quad \text{и} \quad 1/(1 - 0,20) = 1,25.$$

Так как некоторые из коэффициентов больше 10, можно заключить, что в модели присутствует существенная мультиколлинеарность.

Некоторые признаки мультиколлинеарности можно увидеть уже после оценки параметров модели. Перечислим их:

- Неустойчивость результатов. Небольшое изменение исходных данных приводит к существенному изменению оценок коэффициентов. Например, если после оценки уравнения по 200 наблюдениям вы исключили из выборки несколько точек, оценили модель заново и обнаружили сильное изменение результатов.

- Незначимость большинства переменных. Каждая переменная в отдельности является незначимой, а уравнение в целом является значимым и характеризуется близким к единице коэффициентом  $R^2$ .

- Неправдоподобность результатов. Оценки коэффициентов имеют неправильные с точки зрения экономической теории знаки или неоправданно большие значения. Стоит отметить, что причиной возникновения такой ситуации может быть не только потеря точности оценивания из-за мультиколлинеарности,

но и гораздо более серьезные проблемы, например, смещение из-за пропуска существенной переменной (не говоря уж о том, что не все экономические теории прошлого проходят испытание современными данными).

Что можно предпринять, если при столкновении с негативными последствиями мультиколлинеарности в рассматриваемой модели? Существует несколько путей решения этой проблемы.

Если есть возможность увеличить количество наблюдений, то это отличный вариант, так как больший размер выборки увеличит точность результатов, компенсировав ее потерю из-за мультиколлинеарности.

Мультиколлинеарность будет устранена, если исключить из уравнения тот регрессор, который сильно коррелирован с остальными объясняющими переменными модели. Однако следует помнить, что применение этого способа не всегда целесообразно, так как может привести к гораздо более серьезным последствиям: смещению оценок в результате пропуска существенной переменной. В данном примере про производственную функцию ни труд, ни капитал из уравнения исключать не хотелось бы, так как ясно, что выпуск фирмы зависит от каждого из этих факторов производства.

Решением проблемы может быть использование вместо отдельных переменных их линейных комбинаций. Заметим, что переменные, включенные в модель по отдельности, могут быть причиной мультиколлинеарности, так как обычно коррелированы друг с другом. Однако, заменив их на линейную комбинацию, можно избежать этой проблемы. Другой пример: представим, что вы оцениваете зависимость успеваемости студента физического факультета от баллов за ЕГЭ по математике и по физике, которые этот студент получил, будучи школьником. Так как два этих регрессора наверняка коррелированы, то вместо включения в модель каждого из них по отдельности можно бы оставить в уравнении одну переменную – средний балл ЕГЭ по двум этим предметам.

Использование альтернативных (нелинейных) форм зависимостей в некоторых случаях также может снизить остроту проблемы мультиколлинеарности.

В заключение еще раз подчеркнем, что мультиколлинеарность сама по себе не вызывает смещения оценок коэффициентов. Поэтому бороться с ней нужно только в том случае, если она приводит к существенным проблемам (например, к огромным стандартным ошибкам оценок коэффициентов или заведомой неадекватности полученных результатов). Во всех остальных случаях данную проблему можно игнорировать.

## 6.7. Выбор наилучшей функции регрессии

Если функциональная форма связи между показателями неизвестна, необходимо построить несколько функций регрессии и выбрать из них наилучшую. Перечень функций, которые будут рассмотрены, определяется знаниями и предположениями, экспертными оценками относительно характера связи между показателями.

При прочих равных условиях критериями выбора наилучшей функции являются следующие.

1. Статистическая значимость всех параметров при независимых переменных.
2. Значимость всей функции в целом.
3. Выполнение требований Гаусса–Маркова, предъявляемых к случайным остаткам модели, в первую очередь, постоянство дисперсии и независимость друг от друга.
4. Минимальная доля остаточной дисперсии в общей дисперсии, т. е. максимальная величина коэффициента детерминации  $R^2$ . Если модели регрессии содержат разное количество параметров, вместо  $R^2$  следует сравнивать скорректированные коэффициенты детерминации  $R^2_{\text{скорр}}$ .

Сопоставление функций по коэффициентам детерминации можно проводить только в том случае, если зависимая переменная представлена в моделях в одной и той же форме. Например, в линейной, полулогарифмической, полиномиальной функциях зависимая переменная  $y$  представлена в исходной форме:

$y = a_0 + a_1x_1 + \dots + a_px_p + \varepsilon$  – линейная функция;

$y = a_0 + a_1\ln x_1 + \dots + a_p\ln x_p + \varepsilon$  – полулогарифмическая функция;

$y = a_0 + a_{1,1}x_1 + a_{1,2}x_1^2 + \dots + a_{p,1}x_p + a_{p,2}x_p^2 + \varepsilon$  – полином второй степени.

В нелинейных функциях, линеаризация которых связана с преобразованием зависимой переменной, коэффициент детерминации рассчитывается для значений преобразованной переменной. Например, при логарифмировании зависимой переменной:

$$R^2 = 1 - \frac{\sum(\ln y - \ln \hat{y})^2}{\sum(\ln y - \overline{\ln y})^2},$$

при преобразовании зависимой переменной в обратной функции:

$$R^2 = 1 - \frac{\sum(1/y - 1/\hat{y})^2}{\sum(1/y - \overline{1/y})^2}.$$

Очевидно, что показатели детерминации, рассчитанные по исходным и преобразованным значениям зависимой переменной, сравнивать нельзя. Не является корректным решением и расчет коэффициента детерминации по исходным (нелинейным) функциям. Это связано с тем, что при мультипликативной связи переменных (степенная, показательная функции) параметры могут быть найдены с помощью МНК, только если случайные остатки умножаются на выровненные значения зависимой переменной, а не прибавляются к ней:

$$y = a_0 x_1^{a_1} \dots x_p^{a_p};$$

$$y = a_0 x_1^{a_1} \dots x_p^{a_p} \varepsilon.$$

Если рассчитать выровненное значение зависимой переменной по исходной функции:

$$y = a_0 x_1^{a_1} \dots x_p^{a_p};$$

$$\hat{y} = a_0 a_1^{x_1} \dots a_p^{x_p},$$

то случайные остатки  $\varepsilon$  не могут быть вычислены как разность фактического и выровненного значений зависимой переменной, что предполагается при расчете коэффициента детерминации:

$$\varepsilon \neq y - a_0 x_1^{a_1} \dots x_p^{a_p};$$

$$\varepsilon \neq y - a_0 a_1^{x_1} \dots a_p^{x_p}.$$

Аналогичный вывод можно сделать для регрессии, построенной на основе обратной функции. Случайные остатки для этой функции могут быть вычислены как:

$$\varepsilon = \frac{1}{y} - a_0 - \sum_{i=1}^p a_i x_i = \frac{1}{y} - \frac{1}{\hat{y}} \neq y - \hat{y}.$$

Иногда коэффициент детерминации по исходным данным можно вычислить в качестве меры приближения выровненных значений зависимой переменной к исходным. В этом случае используют формулу:

$$R^2 = 1 - \frac{\sum (y - \hat{y})^2}{\sum (y - \bar{y})^2}.$$

Полученный показатель называют *квази- $R^2$* . Для функций, линеаризуемых путем логарифмирования правой и левой части, значения коэффициента детерминации, рассчитанные по линеаризованной регрессии и по исходной, часто имеют близкие значения.

## 6.8. Временные ряды в прогнозировании

Применение временных рядов является необходимым условием для разработки и использования методов прогнозирования, обеспечивающим высокие требования к обоснованности, действительности и временности прогнозов. Математические модели временных рядов, используя апостериорную информацию, позволяют воспроизвести запаздывающее отражение, переводя его в опережающее отражение. На анализе временных рядов основаны многие методы прогнозирования. На первом этапе анализа обобщаются данные временного ряда и строится прогностическая модель, на основании которой на втором этапе находят прогнозные значения показателя. В качестве прогностических наиболее часто используют:

- трендовые модели на основе аналитического выравнивания;
- численные модели, полученные при помощи методов скользящих и экспоненциальных средних;
- модели на основе авторегрессии;
- эконометрические модели.

Прогноз возможен только в случае осмысления и переработки информации, полученной с помощью моделей, активным субъектом – лицом, принимающим решение (ЛПР).

Для повышения роли прогноза в процессе управления необходимо формировать прогностические оценки с ориентацией на их непосредственное использование при составлении долгосрочных планов. При этом длительность периода, ассоциированного с прогнозом, должна быть достаточной для выработки соответствующего управляющего решения и претворения этого решения в жизнь. Практически совершенно бесполезно разрабатывать прогнозы на слишком короткие отрезки времени, в течение которых принятие сколь угодно эффективных управленческих мер оказывается невозможным.

Прогнозирование представляет многоступенчатый процесс, включающий составление предварительного и конечного (формального) прогнозов.

Приступая к прогнозированию, составляют, как отмечалось выше, один или несколько сценариев будущих событий. Это заключается, прежде всего, в выявлении факторов, оказывающих существенное влияние на интересующий показатель. Первым шагом к составлению прогноза является определение внешних и внутренних факторов, которые могут оказать влияние на поведение фирмы. Например, изменения общего уровня спроса вытекают из изменений самого рынка и формируются под воздействием трех групп факторов.

1. Поведения потребителей, включая:
  - изменения числа потенциальных покупателей;
  - изменения потребностей потенциальных покупателей;
  - изменения их покупательной способности.
2. Внешних факторов, не поддающихся контролю со стороны фирмы:
  - общая экономическая ситуация (темпы инфляции, уровень процентных ставок, уровень безработицы);
  - возможные государственные мероприятия (снижение налогов, введение новых законодательных актов);
  - оценка вероятного поведения основных конкурентов (совершенствование производственных мощностей, увеличение объема реализации);
  - влияние социальных и культурных факторов (изменение темпов рождаемости, сокращение числа замужних женщин, занятых трудовой деятельностью, сдвиги в системе ценностей и образе жизни отдельных социальных слоев и т. п.).
3. Внутренних факторов, присущих фирме:
  - совершенствование производственных мощностей фирмы;
  - планирование закупок;
  - выпуск дополнительной продукции.

После определения существенных факторов составляется первоначальный прогноз путем оценки степени влияния тех или иных внутренних и внешних факторов на рассматриваемый показатель.

Как было отмечено выше, техника разработки прогноза разнообразна: от составления сложной многофакторной регрессионной модели до относительно

простой процедуры опроса отдельных агентов по сбыту за последние 12 месяцев. При выполнении прогнозирования выбор адекватной модели прогнозирования является важнейшим условием, обеспечивающим достаточную степень достоверности прогноза. При выборе модели прогнозирования учитываются:

- цель прогнозирования (получение информации о характере доминирующих тенденций (рост, снижение) и подробная картина предстоящих изменений в рыночной конъюнктуре);
- временные рамки (краткосрочный прогноз – до 3 месяцев, среднесрочный – от 3 месяцев до 2 лет, долгосрочный – свыше 2 лет) (чем меньше временной отрезок, тем выше достоверность прогноза);
- необходимый уровень конкретизации:
  - а) прогнозирование одного, группы товаров;
  - б) в рамках страны, региона, района;
  - в) последствия для всей фирмы, отдельного подразделения;
- допустимый уровень погрешности;
- характер исходных данных;
- допустимый объем затрат на прогнозирование:
  - а) на разработку самой модели;
  - б) на сбор и хранение данных, используемых в модели;
  - в) на непосредственную обработку данных;
- объем исходной информации – для определения типов моделей требуется соответствующий объем данных (например, для многофакторной регрессионной модели с семью независимыми переменными требуется, по меньшей мере, 28–35 наблюдений);
- простота в применении (большинство фирм используют именно те модели, которые им наиболее доступны, хотя, может быть, и менее эффективны);
- стадия «жизненного цикла» товара.

Первоначальный прогноз является наиболее важной ступенью процесса прогнозирования, поскольку непосредственно влияет на выбор и формирование стратегии фирмы.

На следующем этапе переходят к разработке стратегии. Первоначальный прогноз служит ориентиром для формирования стратегии, по крайней мере, в той части, которая касается цели прогнозирования.

Например, первоначальный прогноз объема реализации фирмы *A* в следующем году – 15 000 единиц. Если руководство фирмы эти цифры не устраивают, оно может, например, предпринять два шага:

- а) увеличить расходы на продвижение товара на рынок на 18 %, причем главный акцент сделать на усиление заинтересованности торговцев розничной сети;
- б) снизить цену товара на 5 % с тем, чтобы уравнивать ее с розничной ценой главного конкурента.

По расчетам руководства, данные меры позволят поднять объем продаж до 19000, и эта цифра войдет в план действий фирмы.

Первоначальный прогноз подвергается анализу руководства и

специалистов фирмы и корректируется по их замечаниям. Его корректировка приводит к получению формального прогноза. После этого формальный прогноз передается на рассмотрение другим ведущим специалистам фирмы. На основе анализа обратного потока информации, получаемой от потребителей, они могут еще раз скорректировать данные формального прогноза. В конечном счете, выводится окончательная цифра, которая ложится в основу планирования производственной и сбытовой деятельности фирмы.

При реализации прогноза производится его проверка на адекватность. Это заключается в том, что с началом фактической реализации продукции данные формального прогноза подвергаются постоянной сверке на предмет выявления их адекватности реальному процессу. В случае значительных отклонений расчетных данных от фактических формальный прогноз подвергается корректировке.

Например, аспекты, которые следует принимать во внимание при прогнозировании уровня сбыта:

- колебания спроса на рынке в целом;
- колебания спроса на рынке, обслуживаемом компанией.

Принципы прогнозирования на сформировавшемся рынке оказываются предельно простыми, например:

- 1) общий спрос в следующем году останется таким же, как и в прошедшем, если на него не повлияют какие-либо факторы;
- 2) доля компании на рынке останется той же, если она или ее конкуренты не предпримут каких-либо шагов, направленных на ее изменение.

Задача прогнозирования можно разделить на три блока:

1. Предварительное исследование:
  - а) исследовать объект прогнозирования;
  - б) собрать данные.
2. Тестирование моделей:
  - а) выбрать модели, способные дать прогноз;
  - б) оценить модели, дать по ним прогноз;
  - в) оценить прогнозную точность этих моделей.
3. Непосредственное прогнозирование:
  - а) выбрать подходящую модель;
  - б) дать прогноз по выбранной модели;
  - в) откорректировать прогноз в соответствии с экспертным суждением.

Точность прогнозов во многом зависит от выполнения условий, гипотез и предположений, положенных в основу прогностической модели, например, наличия в социально-экономических процессах определенной степени инерционности.

Кроме статистических способов, используются непараметрические методы статистики, которые применяются для исследования генеральной совокупности с неизвестной функциональной формой распределения. Например, при анализе нечисловой информации широко используются методы, основанные на рангах, порядковых статистиках, выборочных долях, оцифровке данных.

На основе выявленных особенностей и закономерностей изменения явлений в прошлом исследователи стремились предугадать поведение ряда в



будущем, т. е. пытались строить различные прогнозы путем экстраполяции (продления) рядов. Экстраполяция ряда динамики – это распространение тенденций, установленных в прошлом, на будущий период. Экстраполяция может использоваться в предположении, что

- тенденции, сложившиеся в прошлом, не изменятся и в будущем;
- аналитическое представление ряда динамики достаточно точно и надежно отражает эти тенденции.

Математические методы экстраполирования сводятся к определению значений, которые будет принимать показатель, если известен ряд значений в прошлые моменты времени.

В ряде случаев используется только несколько последних точек ряда динамики. Если их две – экстраполяция сводится к проведению через них прямой и называется линейной.

Во многих случаях перед экстраполяцией производится выравнивание рядов динамики в целях устранения нехарактерных для них отклонений. Среди наиболее распространенных методов решения этой задачи – *метод наименьших квадратов*.

Модели экстраполяции успешно используются в рамках среднесрочного горизонта прогнозирования, но не обеспечивают надежных результатов на более отдаленную перспективу. Под горизонтом прогнозирования понимается период, для которого прогноз действителен с заданной точностью. Иногда этот термин трактуется как промежуток времени, на который рассчитывается прогноз. Для последней трактовки лучше использовать термин «глубина прогнозирования».

Таким образом, экстраполяцию ряда динамики можно осуществлять различными способами. Но независимо от способа любая экстраполяция обязательно основывается на предположении о том, что закономерность (тенденция) изменения изучаемого явления, выявленная в прошлом для определенного периода времени, сохранится на ограниченном отрезке времени в будущем. Поэтому любому прогнозированию в виде экстраполяции ряда должен предшествовать анализ «длительных» рядов динамики для определения тенденций. А так как в действительности тенденция развития, в свою очередь, может изменяться, то данные, полученные путем экстраполяции ряда, надо рассматривать как вероятностные оценки.

Результаты экстраполяции дополняются интервальными оценками. Для определения границ интервалов используется формула:

$$\bar{y}_t \pm t_{\alpha} \sigma_{\varepsilon},$$

где  $t_{\alpha}$  – коэффициент доверия по распределению Стьюдента;

$\sigma_{\varepsilon}$  – остаточное среднее квадратическое отклонение тренда  $\bar{y}_t$ , скорректированное по числу степеней свободы,

$$\sigma_{\varepsilon} = \sqrt{\frac{\sum_t (\bar{y}_t - y_t)^2}{n-m}},$$

где  $n$  – число уровней ряда динамики;

$m$  – число параметров уравнения тренда;

$y_t$  – значения уровня ряда динамики.

Рассмотрим некоторые простейшие приемы экстраполяции рядов динамики, помогающие прогнозировать показатели на определенный отрезок времени в будущем.

#### *Способ № 1*

Если при анализе ряда динамики обнаруживается, что абсолютные приросты уровней примерно постоянны, то в этом случае можно рассчитать средний абсолютный прирост (как среднюю арифметическую величину) и последовательно прибавлять его к последнему уровню ряда столько раз, на сколько периодов экстраполируется ряд.

*Пример.* По данным о затратах населения на отдых за 1998–2003 годы (табл. 6.1) составить прогноз до 2005 года.

Для составления прогноза рассчитаем абсолютные приросты данного ряда динамики (табл. 6.2).

Таблица 6.1 – Затраты населения на услуги за 1998–2003 годы

| Год      | 1998  | 1999  | 2000  | 2001  | 2002  | 2003  |
|----------|-------|-------|-------|-------|-------|-------|
| Млн руб. | 106,5 | 110,3 | 115,9 | 120,7 | 125,6 | 131,5 |

Таблица 6.2 – Абсолютные приросты

| Год      | 1999 | 2000 | 2001 | 2002 | 2003 |
|----------|------|------|------|------|------|
| Млн руб. | 3,8  | 5,6  | 4,8  | 4,9  | 5,9  |

А затем рассчитаем средний абсолютный прирост:

$$\frac{3,8+5,6+4,8+4,9+5,9}{5} = \frac{25}{5} = 5 \text{ млн руб.}$$

Тогда прогнозная оценка затрат населения на отдых в 1995 году равна:

$$131,5 + 2 \times 5 = 141,5 \text{ млн руб.}$$

#### *Способ № 2*

Если за исследуемый ряд лет (или другие периоды) годовые коэффициенты роста более или менее постоянны, то в этом случае можно рассчитать средний коэффициент роста и последний уровень ряда умножить на средний коэффициент роста в степени, соответствующей периоду экстраполяции.

*Пример.* Определить численность населения района (табл. 6.3) в 2004 и 2005 годах.

Таблица 6.3 – Изменение численности населения района в 1998–2003 гг.

| Год                                               | 1998 | 1999 | 2000 | 2001  | 2002 | 2003  |
|---------------------------------------------------|------|------|------|-------|------|-------|
| Численность населения на 01.01, тыс. чел.         | 52   | 53   | 54,5 | 55,8  | 57,5 | 59,1  |
| Коэффициент роста по отношению к предыдущему году | –    | 1,02 | 1,03 | 1,025 | 1,03 | 1,028 |

Так как коэффициенты роста по годам более или менее стабильны, средний годовой коэффициент роста:

$$\bar{k} = \sqrt[6-1]{\frac{59,1}{52}} = 1,026.$$

Если исходить из предположения о том, что данный темп развития сохранится на определенный отрезок времени в будущем, то можно рассчитать перспективную численность населения, например, на 01.01.2004, 2005 и т. д.

Так, на 01.01.2004 численность населения составит<sup>Λ</sup>

$$59,1 \times 1,026 = 60,6,$$

а на 01.01.2005

$$60,6 \times 1,026 = 62,2.$$

#### *Способ № 3*

Учитывая, что между изменениями нескольких показателей существует зависимость, можно экстраполировать один ряд динамики на основе имеющихся сведений об изменении второго ряда, связанного с ним. Другими словами, на основе корреляции между рядами динамики и найденного соотношения между изменением показателей одного ряда ( $x$ ) и изменениями показателей второго ряда ( $y$ ) для нескольких рядов можно показатели одного ряда рассчитать на основе показателей другого ряда. Например, определив зависимость между изменением объема инвестиций и объемом выпускаемой продукции, можно экстраполировать данные о производстве продукции на основе намечаемых инвестиций.

#### *Способ № 4*

Можно экстраполировать ряды при выравнивании их по какой-либо аналитической формуле. Зная уравнение для исчисления теоретических уровней и подставляя в него значения  $t$  за пределами исследованного ряда, можно рассчитать для данных  $t$  вероятностные оценки  $\bar{y}_t(t)$ . Тенденцию на основе уравнения прямой, параметры которой найдены при выравнивании данных о численности клиентов фирмы за 1999–2000 годы,  $\bar{y}_t = 120,8 + 5,2t$  можно продлить на несколько лет. В частности:

для 2001 года ( $t = 3$ )  $\bar{y}_t = 120,8 + 5,2 \times 3 = 136,4$  (фактически – 136).

для 2002 года ( $t = 4$ )  $\bar{y}_t = 120,8 + 5,2 \times 4 = 141,64$  (фактически – 141).

Так как при выравнивании рядов динамики по аналитическим формулам главным образом определяется тренд, то при прогнозировании целесообразно поступать следующим образом: выровняв ряд по той или иной формуле и определив тренд, найти отклонения фактических от выравненных уровней. Затем можно попытаться определить закономерность (тренд) изменения во времени этих отклонений, т. е. найти формулу для описания их изменения, после этого экстраполировать ряд, накладывая их. Пользуясь этим методом, следует помнить, что экстраполяция динамического ряда на основе уравнения, полученного при выравнивании, только тогда может дать оценки, близкие к действительности, если в эмпирическом ряду случайные колебания, выражающиеся в разности  $(y - \bar{y}_t)$  и измеряемые средним квадратическим отклонением, будут небольшими и если между случайными отклонениями отсутствует автокорреляция.

#### *Способ № 5*

Порой при прогнозировании можно основываться на экстраполяции авторегрессионной функции уровней ряда. При этом методе изучаемый ряд

динамики анализируется с точки зрения автокорреляции. Чем больше автокорреляция между уровнями ряда, тем больше оснований для расчета будущих показателей на основе имеющихся. При этом автокорреляция должна быть исчислена для разных разрывов между уровнями. Установив наличие автокорреляции между уровнями ряда (с определенным лагом), можно найти уравнение, выражающее эту автокорреляционную зависимость, и, пользуясь им, экстраполировать ряд.

Уравнение тренда является моделью аналитического выравнивания уровней ряда динамики. К недостаткам этого подхода следует отнести:

- а) ряды динамики должны быть достаточно длинными;
- б) уровни ряда динамики должны меняться достаточно медленно и плавно;
- в) величина и структура факторов не должны изменяться;
- г) информация о каждом уровне ряда одинаково значима независимо от времени ее получения.

#### *Способ № 6*

Но ряды динамики экономических показателей часто имеют небольшую длину и подвержены значительным колебаниям. Несколько снижают отмеченные недостатки методы экспоненциального сглаживания. При этом временной ряд выравнивается с помощью взвешенной скользящей средней, в которой веса распределяются по экспоненциальному закону. С помощью такой средней уровням, находящимся в конце интервала сглаживания, придается большая значимость. Экспоненциальная средняя величина может быть получена по следующей формуле:

$$\bar{y}_t^{(t)} = \alpha \sum_{i=0}^n (1 - \alpha)^i y_{t-i},$$

где  $\alpha$  – коэффициент сглаживания;

$\bar{y}_n^{(1)}$  – экспоненциальная средняя величина первой степени.

Экспоненциальная средняя величина степени  $k$  определяется следующим образом:

$$\bar{y}_t^{(k)} = \alpha \bar{y}_t^{(k-1)} + (1 - \alpha) \bar{y}_{t-1}^{(k)} \quad \bar{y}_t^{(k)} = \alpha \bar{y}_t^{(k-1)} + (1 - \alpha) \bar{y}_{t-1}^{(k)}.$$

Коэффициенты полиномов, используемых для выравнивания, могут быть получены через сглаженные значения ряда. Для линейного уравнения тренда параметры  $a_0, a_1$  определяются с помощью системы линейных уравнений:

$$\begin{cases} a_0 = 2y_t^1 - \bar{y}_t^2 \\ a_1 = \frac{a}{1-a} (\bar{y}_t^1 - \bar{y}_t^2) \end{cases}$$

Начальные величины  $\bar{y}_{t-1}^1, \bar{y}_{t-1}^2$  могут быть получены при выравнивании динамического ряда по уравнению тренда. Коэффициент сглаживания  $\alpha$  выбирается при проведении анализа исходя из преследуемой цели и относительной ценности данных предшествующих периодов. Если необходимо придать больший вес последним данным, то значение  $\alpha$  выбирается близким к единице. А если необходимо учесть большую часть имеющихся данных, то значения  $\alpha$  относительно невелики. В качестве приема, позволяющего оценить значение  $\alpha$ , используется метод обучающей последовательности. Причем в

качестве обучающей последовательности используется часть заданного ряда динамики, по которой определяется значение  $\alpha$ . Проверка выбранного значения параметра  $\alpha$  осуществляется или для оставшейся части ряда динамики (контрольная последовательность), или для всего ряда динамики. Оптимальное значение  $\alpha$  достигается при минимуме среднеквадратической ошибки, т. е. отклонения фактических значений ряда от расчетных.

Существенным преимуществом подхода, основанного на экспоненциальном сглаживании, является возможность учета ценности информации, относящейся к определенной совокупности периодов. Эта совокупность периодов должна быть аналогичной по предполагаемому сценарию развития прогнозируемого процесса.

Перечисленные методы прогнозирования являются не исчерпывающими, а лишь простейшими. Прогнозирование, основанное только на обработке данных в прошлом, слишком рискованно, если оно не учитывает множества взаимосвязанных факторов и моментов, которые способны изменить тенденцию развития в будущем.

## 7. СИСТЕМЫ ОДНОВРЕМЕННЫХ УРАВНЕНИЙ

### 7.1. Определение системы одновременных уравнений

Одной из причин коррелированности регрессоров со случайными членами могут служить факторы, действующие одновременно и на сами регрессоры, и на объясняемые переменные при фиксированных значениях регрессоров. Иными словами, в экономической ситуации значения объясняемых переменных и регрессоров формируются одновременно под воздействием некоторых внешних факторов. Измерение тесноты связи между переменными, построение изолированных уравнений регрессии недостаточно для объяснения функционирования сложных экономических систем. Изменение одной переменной не может происходить при абсолютной неизменности других. Ее изменение повлечет за собой изменения во всей системе взаимосвязанных признаков. Таким образом отдельно взятое уравнение регрессии не может характеризовать истинное влияние отдельных признаков на вариацию результирующей переменной. Это означает, что рассматриваемая модель не полна: ее следует дополнить уравнениями, в которых объясняемыми переменными выступали бы сами регрессоры. Таким образом, приходим к необходимости рассматривать системы одновременных или регрессионных уравнений.

Поэтому в экономических исследованиях важное место заняла проблема описания структуры связей между системой переменных.

Система одновременных уравнений – это система взаимосвязанных регрессионных уравнений и тождеств, в которых одни и те же переменные в различных регрессионных уравнениях могут одновременно выступать и в роли результирующих показателей, и в роли объясняющих переменных. Тем самым подчеркивается, что в системе одни и те же переменные у одновременно рассматриваются как зависимые в одних уравнениях и как независимые в других. В эконометрике эта система уравнений называется также *структурной формой модели*. В отличие от предыдущих систем каждое уравнение системы одновременных уравнений не может рассматриваться самостоятельно, и для нахождения его параметров традиционный МНК неприменим. С этой целью используются специальные приемы оценивания.

Примером системы одновременных уравнений может служить следующая гипотетическая модель:

$$\begin{cases} Y_1 = a_0 + a_1 Y_2 + a_2 X_1 + \varepsilon_1 \\ Y_2 = b_0 + b_1 Y_1 + b_2 X_2 + \varepsilon_2 \end{cases}'$$

где  $\varepsilon_1, \varepsilon_2$  – случайные ошибки.

Если значения переменной формируются вне модели, то такие переменные называются экзогенными.

Эндогенные переменные – это взаимозависимые переменные, которые определяются внутри модели (системы). Как правило, каждое уравнение модели определяет одну эндогенную переменную, стоящую в левой части уравнения.

Таким образом, число уравнений в системе равно числу эндогенных переменных.

Экзогенные переменные – это независимые переменные, которые определяются вне системы.

В приведенной выше системе одновременных уравнений  $Y_1$  и  $Y_2$  являются эндогенными, а  $X_1$  и  $X_2$  – экзогенными переменными.

Предопределенные переменные – это экзогенные и лаговые (за предшествующие промежутки или моменты времени) эндогенные переменные системы.

С математической точки зрения, главное отличие между экзогенными и эндогенными переменными заключается в том, что экзогенные переменные не коррелируют с ошибками регрессии, между тем как эндогенные могут коррелировать (и, как правило, коррелируют). Естественно предположить, что схожие случайные факторы действуют как на цену равновесия, так и на спрос на товар. Причинная зависимость между переменными и приводит, очевидно, к коррелированности их со случайными членами.

Структурная форма модели – это система уравнений, отражающая взаимосвязь между переменными в соответствии с положениями экономической теории и характеризующая структуру экономики или ее сектора.

Параметры структурной формы модели называют структурными параметрами.

Если модель содержит тождества, то их можно назвать уравнениями, в которых структурные параметры при переменных равны 1.

Таким образом структурная форма модели или система одновременных уравнений – это система уравнений, в каждом из которых помимо объясняющих (независимых переменных) могут содержаться объясняемые (зависимые переменные) из других уравнений. Уравнения, составляющие исходную модель, называются структурными уравнениями модели.

Простейшая структурная форма модели имеет вид:

$$\begin{aligned}y_1 &= \alpha_1 + \beta_{1,2}y_2 + \alpha_{1,1}x_1 + \alpha_{1,2}x_2 + \varepsilon_1; \\y_2 &= \alpha_2 + \beta_{2,1}y_1 + \alpha_{2,1}x_1 + \alpha_{2,2}x_2 + \varepsilon_2,\end{aligned}$$

где  $y_1, y_2$  и  $x_1, x_2$  – зависимые и независимые переменные;

$\varepsilon_1, \varepsilon_2$  – случайные члены;

$\alpha_1, \beta_{12}, \alpha_{11}, \alpha_{12}, \alpha_2, \beta_{21}, \alpha_{21}, \alpha_{22}$  – параметры модели.

Параметры структурной формы модели называются структурными коэффициентами.

Структурная форма модели включает в систему не только уравнения, отражающие взаимосвязи между отдельными переменными, но и уравнения, отражающие тенденцию развития явления – функции времени, а также разного рода уравнения-тождества. Тождества не содержат каких-либо подлежащих оценке параметров, а также не включают случайных членов.

Структурные коэффициенты можно однозначно выражать через приведенные коэффициенты, или они могут иметь несколько разных оценок, но совсем не выражаться через них.

Структурный коэффициент называется идентифицируемым, если его можно вычислить на основе приведенных коэффициентов, причем точно идентифицируемым, если он имеет единственную оценку, и сверхидентифицируемым, если он имеет несколько разных оценок. В противном случае он называется неидентифицируемым.

Какое-либо структурное уравнение является идентифицируемым, если идентифицируемы все его коэффициенты. Если хотя бы один структурный коэффициент неидентифицируем, то и все уравнение является неидентифицируемым.

Модель считается идентифицируемой, если каждое ее уравнение идентифицируемо. Если хотя бы одно из уравнений системы неидентифицируемо, то и вся модель неидентифицируема.

Классическим примером является одновременное формирование спроса  $Q^d$  и предложения  $Q^s$  товара в зависимости от его цены  $P$ :

$$\begin{aligned} Q^d &= \beta_1 + \beta_2 P + \beta_3 I + \varepsilon_1; \\ Q^s &= \beta_4 + \beta_5 P + \varepsilon_2. \end{aligned} \quad (7.1)$$

Здесь  $I$  – доход.

Если предположить, что рынок находится в состоянии равновесия, то в равенствах (7.1) следует положить  $Q^d = Q^s = Q$ . В этом случае наблюдаемое значение  $P$  – это цена равновесия, которая формируется одновременно со спросом и предложением. Таким образом, можно считать  $P$  и  $Q$  объясняемыми переменными, а величину дохода  $I$  – объясняющей переменной.

Разделение ролей между переменными в системе одновременных уравнений может быть проинтерпретировано следующим образом: переменные  $Q$  и  $P$  формируют свои значения, подчиняясь уравнениям (7.1), т. е. внутри модели. Такие переменные называются эндогенными. Между тем переменная  $I$  считается в уравнениях (7.1) заданной, ее значения формируются вне модели. Такие переменные называются экзогенными.

Приведем общий вид системы одновременных уравнений. Пусть  $Y_j$  – эндогенные переменные,  $X_i$  – экзогенные переменные. Введем блочные матрицы  $B$  и  $\Gamma$  вида:

$$B = \begin{pmatrix} \beta_{1,1} & \dots & \beta_{1,m} \\ \dots & \dots & \dots \\ \beta_{m,1} & \dots & \beta_{m,m} \end{pmatrix}; \quad \Gamma = \begin{pmatrix} \gamma_{1,1} & \dots & \gamma_{1,l} \\ \dots & \dots & \dots \\ \gamma_{m,1} & \dots & \gamma_{m,l} \end{pmatrix}.$$

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

$$BY + \Gamma X = \varepsilon, \quad (7.2)$$

$$Y = \begin{pmatrix} Y_1 \\ \dots \\ Y_m \end{pmatrix}; \quad X = \begin{pmatrix} X_1 \\ \dots \\ X_l \end{pmatrix}; \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \dots \\ \varepsilon_m \end{pmatrix}.$$

Кроме регрессионных уравнений (они называются также *поведенческими* уравнениями) модель может содержать *тождества*, которые представляют алгебраические соотношения между эндогенными переменными.



Например, для модели формирования спроса и предложения и цены равновесия имеем два поведенческих уравнения (7.1) и одно тождество  $Q^s = Q^d$ .

Тождества, вообще говоря, позволяют исключить некоторые эндогенные переменные и рассматривать систему регрессионных уравнений меньшей размерности. Так, в модели спроса и предложения можно положить:

$$Q^s = Q^d = Q$$

и рассматривать структурную форму (7.2), где

$$Y = \begin{pmatrix} Q \\ P \end{pmatrix}; X = \begin{pmatrix} 1 \\ I \end{pmatrix}; \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix};$$

$$B = \begin{pmatrix} 1 & -\beta_2 \\ 1 & -\beta_5 \end{pmatrix}; \Gamma = \begin{pmatrix} -\beta_1 & -\beta_3 \\ -\beta_4 & 0 \end{pmatrix}.$$

Оценивание систем одновременных уравнений требует специальных методов.

Рассмотрим широко известную модель спроса и предложения:

$$\begin{cases} Q_D(t) = \beta_1 + \beta_2 P(t) + \beta_3 Y(t) + u_t, \\ Q_S(t) = \alpha_1 + \alpha_2 P(t) + \alpha_3 P(t-1) + \varepsilon_t \\ Q_S(t) = Q_D(t), \end{cases}$$

где  $Q_D(t)$  – спрос на товар в момент времени  $t$ ;

$Q_S(t)$  – предложение товара в момент времени  $t$ ;

$P(t)$  – цена товара в момент времени  $t$ ;

$Y(t)$  – доход в момент времени  $t$ .

Первое из приведенных уравнений системы называют уравнением спроса, второе – уравнением предложения, третье – уравнением равновесия. Цена товара  $P(t)$  и спрос на товар

$$Q(t) = Q_D(t) = Q_S(t)$$

определяются из уравнений системы (т. е. являются эндогенными переменными). Предопределенными переменными являются доход  $Y(t)$  и значение цены товара в предыдущий момент времени  $P(t-1)$ .

К моделям рассматриваемой (второй) группы следует отнести и весьма широко применяемые модели систем массового обслуживания. Популярность этих моделей объясняется, прежде всего, наличием развитого математического аппарата теории массового обслуживания, содержащего многие формальные схемы описания различных реальных систем. Подробнее о некоторых практических применениях таких моделей самого разного уровня (моделей фирмы; бензоколонки; торговой точки; звена управления и др.). Весьма интересные модели важных и сложных процессов (прогнозирование в промышленности; управления трудовыми ресурсами и запасами; повышения надежности и улучшения ремонта оборудования и др.).

В качестве примера использования математической модели для анализа одного из сложнейших вопросов коммерции – ценообразования можно привести следующий: пусть требуется обосновать возможный подход к определению рыночной цены на некоторый программный продукт (заметим, что данная проблема остро стоит на нашем рынке). Для любого элемента программного обеспечения можно указать на две особенности:

- оригинальный программный продукт является авторским произведением и должен рассматриваться как интеллектуальная собственность;
- созданный программный продукт может быть легко размножен, причем расходы на размножение ничтожно малы по сравнению с расходами на разработку оригинала.

Таким образом, существуют два вида цен: цена на оригинал программы и цена на ее копию (легко провести аналогию с аудио- и видеопродукцией, картинами и т. п.). Все программы можно подразделить на уникальные (создаваемые по заказу и не предусматривающие тиражирование); специализированные (разрабатываемые для ограниченного контингента пользователей) и универсальные или рыночные (предназначенные для широкого потребления). В дальнейшем будем рассматривать именно универсальные программы и цены на них. Под рыночной ценой продукта будем понимать цену, устанавливаемую продавцом (автором программы или его представителем на рынке) на копию программного продукта, готовую к реализации. При этих допущениях первоначальные затраты на разработку программ являются постоянными и их вообще можно не рассматривать при определении рыночной цены. Действительно, сколько бы ни стоила разработка продукта, его цена определяется потребностью в программах данного типа; количеством потенциальных покупателей и их финансовыми возможностями; наличием конкурентов; качеством предлагаемого товара; известностью автора (фирмы-производителя) продукта; эффективностью рекламной кампании и т. п. Упростим рассматриваемую ситуацию: пусть у нашего продукта нет аналогов (а, следовательно, и конкурентов); нам известны все потенциальные покупатели, а им – информация о нашем программном продукте; каждый покупатель готов купить одну копию нашей программы по любой цене, не превышающей его собственной предельной цены на данную программу. Итак, необходимо установить такую цену на программный продукт, чтобы сумма выручки от его продажи была максимальной. Подобная ситуация может быть описана оптимизационной экономической математической моделью вида:

$$Y = P \times K_p \rightarrow \max,$$

где  $Y$  – доход продавца;

$P$  – искомая цена продукта;

$K_p$  – количество копий, которые были проданы по цене  $P$ .

Очевидно, что количество проданных копий зависит от цены на продукт, и сложность заключается именно в определении вида этой зависимости (функции  $K_p(P)$ ). Обозначим:  $P_0$  – нижний предел рыночной цены,  $P_1$  – верхний предел рыночной цены, назначаемые потенциальными покупателями (их, как правило, оценить возможно). Тогда наша модель дополняется ограничением  $P_0 < P < P_1$ .

На отрезке  $P_0, P_1$  функция  $K_p(P)$  является монотонно невозрастающей, и, более того, она кусочно-постоянна и меняется тогда, когда цена достигает такого значения, при котором очередной потенциальный покупатель уже не может купить программу.

Пример. Пусть есть три группы покупателей, каждая из которых готова приобрести 100 копий программного продукта, причем их собственные предельные цены составляют 600, 1000 и 1500 руб., соответственно. Тогда возможны колебания цены в пределах от 600 до 1500 руб. Очевидно, при цене 600 руб. будет продано 300 копий (всем покупателям); при цене от 601 до 1000 руб. – 200 копий (покупателям второй и третьей групп); при цене от 1001 до 1500 руб. – 100 копий (покупателям третьей группы). Соответственно, доход для этих цен составит:

$$Y_0 = K_{p0} \times P_0 = 300 \times 600 = 180000 \text{ руб.}$$

$$Y = K_p \times P = 200 \times 1000 = 200000 \text{ руб.}$$

$$Y_1 = K_{p1} \times P_1 = 100 \times 1500 = 150000 \text{ руб.}$$

(вообще говоря, доход для второй цены рассчитать сложно, поэтому взята предельная цена для второй группы покупателей).

Если отказаться от допущений о неизменности цен во времени и равенстве их значений для различных групп покупателей, можно наметить два пути увеличения общей прибыли от продаж:

- установление значительной скидки для покупателей, не имеющих возможностей приобрести программу по рыночной цене (школ; вузов; государственных научно-исследовательских учреждений и т. п.) – в нашем примере при выделении первой группы покупателей в особый класс, для которого вводится скидка 50 % от найденной нами цены (1000 руб.), прибыль увеличится на 50 000 руб. и составит 250 000 руб.;

- постепенное снижение цены на программный продукт – если считать, что нет инфляции и за время торговли реклама не устаревает, следует принять такой сценарий продаж: а) сначала установить предельную цену на программу  $P_1$  и удерживать ее до тех пор, пока все покупатели третьей группы не приобретут интересующие их копии; б) затем снизить цену до следующего предельного значения, дав возможность купить программу всем покупателям второй группы; в) снизить цену до минимального порога  $P$ , обеспечив копиями программ всех оставшихся покупателей (такой сценарий принесет продавцу максимальный для условий примера доход в 310 000 руб.).

Понятно, что приведенная модель имеет учебную (иллюстративную) направленность и не может быть непосредственно применена на практике. Вместе с тем учет в ней таких факторов, как неопределенность при нахождении потенциальных покупателей; переход к интервальным оценкам их собственных предельных цен; учет инфляции, нелегального копирования и устаревания рекламы; других неопределенных факторов (после чего данная модель станет полноценным представителем моделей второй группы) позволит использовать результаты моделирования, например, при определении стратегии продаж на некоторый интервал времени.

## 7.2. Классификация систем одновременных уравнений

Статические эконометрические модели систем получают тремя способами: экспериментально-статистическими, детерминированными и смешанными.

Экспериментально-статистические способы требуют проведения активных или пассивных экспериментов на действующем объекте. Стохастические модели используют для решения различных задач, связанных с исследованием и управлением процессами. В большинстве случаев эти модели получают в виде линейных уравнений регрессии.

Исходя из свойств реальных процессов, можно утверждать, что уравнения взаимосвязи переменных процесса должны иметь иную, возможно, более сложную структуру. Чем более «далека» структура уравнений регрессии от «истинной», тем меньше будет точность прогноза с увеличением диапазона изменения переменных процесса. Это ухудшает качество управления и, следовательно, уменьшает качество функционирования объекта в оптимальном режиме.

Детерминированные модели составляют «на основании экономических закономерностей и представлений о процессах». Следовательно, их можно получать еще на стадии анализа или формирования процесса. В настоящее время на базе детерминированного подхода разработано несколько методов построения эконометрических моделей непрерывных процессов. Так, например, при моделировании ряда процессов в промышленной технологии нашел применение метод многомерного фазового пространства. Сущность метода заключается в том, что протекание моделируемого технологического процесса рассматривается как движение некоторых «изображающих точек» в многомерном фазовом пространстве. Это пространство определено как пространство декартовой системы координат, по осям которого отложены пространственные координаты аппарата и внутренние координаты реагирующих твердых частиц. Каждая точка в многомерном фазовом пространстве описывает определенное состояние моделируемого процесса. Число этих точек равно числу частиц в аппарате. Протекание технологического процесса характеризуется изменением потока изображающих точек.

В экономическом анализе проведение системного подхода предполагает сложную структуру взаимосвязей между признаками, когда эффективность деятельности экономического объекта характеризуется несколькими показателями. В этом случае одного регрессионного уравнения может быть недостаточно, и для описания явления или процесса может потребоваться система уравнений и тождеств.

Различают следующие виды систем эконометрических уравнений:

- системы независимых уравнений;
- системы рекурсивных уравнений;
- системы взаимозависимых уравнений.

В системе независимых уравнений, когда каждая зависимая переменная у рассматривается как функция одного и того же набора факторов  $x$ :

$$\begin{cases} y_1 = a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,n}x_n + \varepsilon_1 \\ y_2 = a_{2,1}x_1 + a_{2,2}x_2 + \cdots + a_{2,n}x_n + \varepsilon_2 \\ \quad \quad \quad \dots \dots \dots \\ y_m = a_{m,1}x_1 + a_{m,2}x_2 + \cdots + a_{m,n}x_n + \varepsilon_m. \end{cases} \quad (7.1)$$

По существу, каждое уравнение этой системы является уравнением регрессии. Каждое уравнение такой системы может рассматриваться самостоятельно. Для нахождения его параметров используется метод наименьших квадратов. Так как фактические значения зависимой переменной отличаются от теоретических на величину случайной ошибки, то в каждом уравнении присутствует величина случайной ошибки  $\varepsilon_1$ .

Если зависимая переменная  $y$  одного уравнения выступает в виде фактора  $x$  в другом уравнении, то исследователь может строить модель в виде системы рекурсивных уравнений. Таким образом, в системе рекурсивных уравнений зависимая переменная  $y$  включает в каждое последующее уравнение в качестве факторов все зависимые переменные предшествующих уравнений и набор факторов  $x$ :

$$\begin{cases} y_1 = a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,n}x_n + \varepsilon_1 \\ y_2 = a_{2,1}x_1 + a_{2,2}x_2 + \cdots + a_{2,n}x_n + \varepsilon_2 \\ \quad \quad \quad \dots \dots \dots \\ y_m = a_{m,1}x_1 + a_{m,2}x_2 + \cdots + a_{m,n}x_n + \varepsilon_m. \end{cases} \quad (7.2)$$

В данной системе зависимая переменная увключает в каждое последующее уравнение в качестве факторов все зависимые переменные предшествующих уравнений наряду с набором собственно факторов  $x$ . В таких моделях уравнения оцениваются последовательно (от первого уравнения к последнему) с использованием МНК. А каждое уравнение этой системы рассматривается самостоятельно, и его параметры определяются методом наименьших квадратов (МНК).

Наибольшее распространение в эконометрических исследованиях получила система взаимозависимых уравнений, в которой одни и те же зависимые переменные в одних уравнениях входят в левую часть, а в других – в правую часть системы:

[illegible]

Система взаимосвязанных уравнений получила название *системы совместных, одновременных уравнений*. В эконометрике эта система уравнений называется также *структурной формой модели*. В отличие от предыдущих систем каждое уравнение системы одновременных уравнений не может рассматриваться самостоятельно, и для нахождения его параметров традиционный МНК неприменим. С этой целью используются специальные приемы оценивания.

В рассмотренных классах систем эконометрических уравнений структура матрицы коэффициентов при зависимых переменных различна.

Представим систему эконометрических уравнений в матричном виде:

$$BY + \Gamma X = E,$$

где  $B$  – матрица коэффициентов при зависимых переменных;

$Y$  – вектор зависимых переменных;

$\Gamma$  – матрица параметров при объясняющих переменных;

$X$  – вектор объясняющих переменных;

$E$  – вектор ошибок.

Если матрица  $B$  диагональная, то рассматриваемая модель является системой независимых уравнений. Так, при трех зависимых и трех объясняющих переменных модель имеет вид:

$$\begin{cases} y_1 = a_{0,1} + a_{1,1}x_1 + a_{1,2}x_2 + a_{1,3}x_3 + E_1 \\ y_2 = a_{0,2} + a_{2,1}x_1 + a_{2,2}x_2 + a_{2,3}x_3 + E_2 \\ y_3 = a_{0,3} + a_{3,1}x_1 + a_{3,2}x_2 + a_{3,3}x_3 + E_3. \end{cases}$$

Матрица параметров при зависимых переменных является диагональной:

$$B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Если матрица  $B$  треугольная (или может быть приведена к такому виду), то модель представляет систему рекурсивных уравнений. Так, если модель имеет вид:

$$\begin{cases} y_1 = a_{0,1} + a_{1,1}x_1 + a_{1,2}x_2 + E_1 \\ y_2 = a_{0,2} + a_{2,1}x_1 + a_{2,2}x_2 + E_2 \\ y_3 = a_{0,3} + a_{3,1}x_1 + a_{3,2}x_2 + E_3, \end{cases}$$

т. е. зависимая переменная  $y_1$  первого уравнения участвует как объясняющая переменная во втором уравнении системы, а зависимая переменная  $y_2$  второго уравнения рассматривается как объясняющая переменная в третьем уравнении. Тогда матрица коэффициентов при зависимых переменных модели составит:

$$B = \begin{bmatrix} 1 & 0 & 0 \\ -b_{2,1} & 1 & 0 \\ 0 & -b_{3,2} & 1 \end{bmatrix},$$

т. е. представляет треугольную матрицу.

Если матрица  $B$  не является ни диагональной, ни треугольной, то модель представляет систему одновременных уравнений. Так, для модели вида:

$$\begin{cases} y_1 = a_{0,1} + b_{1,2}y_2 + a_{1,1}x_1 + a_{1,2}x_2 + E_1 \\ y_2 = a_{0,2} + b_{2,1}y_1 + b_{2,3}y_3 + a_{2,3}x_3 + E_2 \\ y_3 = a_{0,3} + b_{3,1}y_1 + a_{3,2}x_2 + a_{3,3}x_3 + E_3 \end{cases}$$

получим матрицу коэффициентов при зависимых переменных:

$$B = \begin{bmatrix} 1 & -b_{1,2} & 0 \\ -b_{2,1} & 1 & -b_{2,3} \\ -b_{3,1} & 0 & 1 \end{bmatrix},$$

которая не является ни диагональной, ни треугольной. Соответственно это отражается на выборе метода оценки параметров эконометрических систем.

### 7.3. Проблема идентификации систем одновременных уравнений

Идентификация систем рассматривается как совокупность методов для построения математических моделей динамической системы по данным наблюдений. Математическая модель в данном контексте означает математическое описание поведения какой-либо системы или процесса в частотной или временной области, к примеру, экономического процесса (реакция биржевых котировок на внешние возмущения) и т. п.

Для построения математических моделей наиболее широко применяется метод многомерного фазового пространства.

Фазовое пространство – это пространство, точка в котором описывает состояние всей системы целиком в фиксированный момент времени. Изменение системы с течением времени соответствует движению точки в фазовом пространстве. Сложные системы описываются большим количеством параметров и требуют рассмотрения многомерных фазовых пространств. Координаты в фазовом пространстве называются фазовыми переменными.

Однако этот метод имеет и недостатки, которые ограничивают область его применения:

- а) многомерное координатное (фазовое) пространство хорошо определяет состояние процесса, но никак не определяет его движения, т. е. возможных изменений переменных процесса, поэтому метод хорошо применим для процессов, начальное и конечное состояние которых заранее известно;
- б) допущение о том, что координаты можно считать независимыми часто является слишком грубым;
- в) применение этого метода вызывает определенные трудности, когда необходимо учитывать различное время взаимодействия подсистем.

Таким образом, из-за перечисленных особенностей, использовать метод многомерного фазового пространства для построения эконометрических моделей затруднительно.

Как правило, в результате теоретического анализа процесса удастся получить математическую модель, параметры которой необходимо уточнять в процессе управления экономическим объектом.

При построении математической модели необходимо:

- а) Существование априорной информации, которая имеется еще до получения измерительной информации и характеризует структуру идентифицируемого объекта. С этой точки зрения по структуре объекты подразделяются на:
  - динамические (поведение выхода зависит от предыдущих значений входа) и статические (поведение выхода зависит только от значений входа в текущий момент времени);
  - стохастические (поведение выхода зависит от случайных факторов) и детерминированные (поведение выхода не зависит от случайных факторов);

- нелинейные (его реакция на два различных возмущений входа не эквивалентна сумме реакций на каждое из этих возмущений в отдельности) и линейные (реакция на два различных возмущений входа эквивалентна сумме реакций на каждое из этих возмущений в отдельности);
  - дискретные (состояние его входов и выходов изменяется лишь в дискретные моменты времени) и непрерывные (состояние его входов и выходов изменяется непрерывно).
- б) Осуществление структурной идентификации, которая включает решение задач:
- Выделение объекта из окружающей его и взаимодействующей с ним среды.
  - Ранжирование входов и выходов объекта по степени их влияния на поведение объекта.
  - Определение оптимального числа входов и выходов объекта, учитываемых в модели.
  - Определение характера связей между входами и выходами модели объекта.
- в) Определение набора данных, полученных в результате нормальной работы изучаемого объекта либо при эксперименте. Для этого входная и выходная информация должна записываться в течение заранее спланированного идентификационного эксперимента, во время которого можно выбрать какую и когда информацию фиксировать. Дисциплина «Планирование экспериментов» может подсказать как сделать экспериментальную информацию наиболее информативной с учетом тех ограничений, которые могут быть наложены на эксперимент. Но, к сожалению, не редка ситуация, когда отсутствует возможность проводить эксперимент, а приходится работать с той информацией, которая предоставляется.
- г) Существования множества моделей-кандидатов для использования. Множество моделей-кандидатов получается на основании решения о том классе моделей, в котором будет проводиться поиск. Без сомнения, этот выбор является самым важным и самым сложным в процедуре идентификации. Именно на этом этапе опираясь на априорную информацию, интуиция должна подсказать принятие решения базируясь на формальных свойствах рассматриваемых вероятных моделей. А множество предполагаемых моделей может быть выявлено или на основе известных экономических законов, или на использовании уже известных моделей, ориентируясь на существенные моменты рассматриваемой экономической деятельности. Такие модели, имеющие параметры, изменяя которые можно добиться приближения к изучаемому объекту, называются моделями черного ящика. Модели же имеющие настраиваемые параметры и опирающиеся на известные законы называются серыми ящиками. Вообще говоря, модельная структура – это параметризованное отображение из множества входов и выходов до



момента времени  $t-1$  включительно на множество выходов текущего момента времени  $t$ :

$$\hat{y}(t|\theta) = g(\theta, Z^{t-1}).$$

- д) Определение правила, по которому каждая модель-кандидат может быть принята или отвергнута. Критерием для выбора модели является ее способность повторять данные, полученные из эксперимента, т. е. соответствовать поведению изучаемого объекта. Но необходимо помнить, что модель никогда не сможет быть принята как «настоящее» или «истинное» описание объекта из-за своей врожденной приблизительности.

Процедура идентификации имеет естественный логический порядок: сначала собираются данные, потом формируется множество моделей и затем выбирается наилучшая модель. Обычным первая выбранная модель не проходит тест на соответствие экспериментальным данным. Тогда следует вернуться назад и выбрать другую модель или изменить критерии поиска. Модель может быть неудовлетворительной по следующим причинам:

- Численный метод не может найти, подходящую к выбранному критерию, модель.
- Неправильно выбранный критерий.
- Неправильно сформированное множество моделей, в нем вообще может не быть качественной модели.
- Собранные данные не информативны.

При идентификации предполагается анализ и сопоставление входных и выходных процессов, и задача идентификации состоит в выборе соответствующей математической модели. Модель должна быть такой, что ее реакция и реакция объекта на один и тот же входной сигнал должны быть, в известном смысле, близкими. Результаты решения задачи идентификации являются исходными данными для проектирования систем управления, оптимизации, анализа параметров систем и т. д.

В настоящее время для определения динамических свойств регулируемых объектов применяют следующие методы:

1. Методы, построенные на моделировании искусственного воздействия на систему непериодическим сигналом, мощность которого велика по сравнению с уровнем помех в системе. В качестве воздействия выбирается обычно скачкообразное изменение регулирующего воздействия, и в результате определяются временные характеристики.
2. Методы, построенные на моделировании искусственного воздействия на систему периодическими сигналами разной частоты, амплитуда которых велика по сравнению с уровнем помех в системе. В результате определяются частотные характеристики.
3. Методы, построенные на моделировании искусственного воздействия на систему синусоидальными сигналами, соизмеримыми с помехами в системе. В результате также определяются частотные характеристики.
4. Методы, не требующие искусственных воздействий, использующие возмущения, которые имеются в процессе нормальной эксплуатации.

При рассмотрении известных подходов к идентификации можно выделить две группы.

**К первой группе** относятся работы, в которых существенным образом используется априорная информация об идентифицируемых параметрах. В основе первого подхода данной группы лежит гипотеза о том, что идентифицируемые параметры являются решениями известных однородных систем разностных уравнений или же представляются в виде случайного процесса, порожденного марковской моделью, т. е. являются решениями известных систем дифференциальных или разностных уравнений с возмущениями типа белого шума, характеризующегося гауссовским распределением, известными средними значениями и интенсивностью. Такой подход правомерен при наличии большого объема априорной информации об искомых параметрах и при несоответствии реальных параметров принятой модели приводит к потере сходимости алгоритма.

Второй подход, относящийся к первой группе, базируется на параметризации нестационарных параметров и использует гипотезу о возможности точного представления нестационарных идентифицируемых параметров на всем интервале или отдельных локальных интервалах идентификации в форме конечной, как правило, линейной комбинации известных функций времени с неизвестными постоянными весовыми коэффициентами, в частности в виде конечной суммы членов ряда Тейлора, гармонического ряда Фурье, обобщенного ряда Фурье по системам ортогональных функций Лагерра, Уолша.

Простейшим случаем параметризации является представление нестационарных параметров постоянными величинами на последовательности отдельных локальных интервалов, покрывающих интервал идентификации.

При текущей идентификации рекомендуют переходить к скользящему интервалу времени  $[t-T, t]$  длительности  $T$  и считать на этом интервале искомые параметры постоянными или же точно представимыми в виде интерполяционного полинома конечной степени, или указанной конечной линейной комбинации. К данному подходу могут быть отнесены работы, базирующиеся на использовании итерационного метода наименьших квадратов. В этих работах из-за использования экспоненциального (с отрицательным показателем) весового множителя в минимизируемом квадратичном функционале, определенном на текущем интервале времени  $[0, t]$ , с течением времени происходит «стирание» старой информации о координатах объекта. Такое положение, по существу, соответствует идее постоянства идентифицируемых параметров на некотором скользящем интервале времени при учете информации о состоянии объекта на этом интервале с экспоненциальным весом.

Данный подход позволяет непосредственно распространить методы идентификации стационарных параметров на случай идентификации нестационарных параметров. Однако на практике основополагающая гипотеза этого подхода не выполняется и можно говорить лишь о приближенном представлении (аппроксимации) искомых параметров конечной линейной

комбинацией известных функций времени с неизвестными постоянными весовыми коэффициентами. Такое положение приводит к возникновению методической ошибки идентификации, которая принципиально изменяет существо обсуждаемого подхода, так как при этом длительность  $T$  интервала аппроксимации и число членов линейной комбинации становятся параметрами регуляризации. Данная методическая ошибка, как правило, не принимается во внимание. В частности, в предположении о прямолинейном законе изменения искомых параметров на больших, значительно превосходящих  $T$ , локальных интервалах времени и ряде ограничений на статистические характеристики координат регрессионной модели объекта и действующей помехи был предложен адаптивный алгоритм коррекции параметра  $T$ . Пренебрежение методической ошибкой приводит к тому, что рассматриваемый подход оказывается, по существу, совершенно не исследованным, область его работоспособности в условиях нестационарности идентифицируемых параметров не определена и можно говорить о применимости этого подхода только в частном случае точного выполнения указанной гипотезы.

**Ко второй группе** относятся методы, использующие значительно меньший объем информации об искомых параметрах, причем данная информация используется лишь на этапе выбора параметров алгоритма идентификации.

Первый подход, относящийся к этой группе, базируется на применении градиентных самонастраивающихся моделей. Такой подход обсуждался в работах по параметрической идентификации линейных и нелинейных динамических объектов. Основное достоинство этого подхода состоит в том, что он приводит к замкнутой системе идентификации и, тем самым, обладает определенными преимуществами в плане помехоустойчивости по сравнению с разомкнутыми методами идентификации. Недостатки указанного подхода связаны с необходимостью измерения компонент градиента критерия настройки, представляющих собой функциональные производные требованием достаточно точной априорной информации о начальных значениях идентифицируемых параметров (для выбора начальных значений параметров модели, гарантирующих устойчивость работы системы идентификации) и отсутствием полного теоретического анализа динамики работы системы идентификации данного типа. Последнее объясняется сложностью системы интегро-дифференциальных уравнений, описывающих процессы в контуре самонастройки, вследствие чего теоретический анализ проводится лишь в предположении медленного изменения параметров объекта и модели. В связи с этим не удастся достаточно полно оценить область устойчивости, быстродействие и точность работы градиентных самонастраивающихся моделей, а тем самым четко определить область применимости систем указанного типа при текущей идентификации нестационарных параметров. Следует, однако, отметить, что с увеличением степени нестационарности искомых параметров значительно возрастают методические погрешности определения компонент градиента критерия настройки, в результате чего

увеличивается ошибка идентификации за пределы зоны глобального экстремума минимизируемого критерия.

Особенно усиливается такой эффект с повышением числа идентифицируемых параметров из-за взаимосвязи каналов идентификации. Поэтому применение градиентных самонастраивающихся моделей принципиально ограничено случаем медленного изменения искомых параметров.

Второй подход базируется на использовании алгоритма Качмажа. Известно, что основной алгоритм этого типа обладает слабой помехоустойчивостью и низким быстродействием. Такое положение побудило к созданию различных модификаций данного алгоритма, характеризующихся повышенным быстродействием. Тем не менее, быстродействие указанных модификаций по-прежнему является невысоким, что априори ограничивает область применимости второго подхода к случаю идентификации медленно изменяющихся параметров.

Ко второй группе также могут быть отнесены методы, предназначенные для идентификации только линейных динамических объектов и характеризуются дополнительными ограничениями (необходимость использования тестовых входных сигналов в виде совокупности гармоник или псевдослучайного периодического бинарного сигнала конечность интервала идентификации, наличие полной информации о входных и выходных сигналах объекта на всем интервале идентификации и возможность идентификации коэффициентов только левой части дифференциального уравнения). Из-за этого возможны значительные ошибки идентификации на отдельных конечных подинтервалах времени, при этом также необходимо решать сложную краевую задачу.

Перечисленные сложности идентификации реальных функционирующих систем определяют наиболее ориентированный на широкое использование подход к моделированию нелинейных объектов, заключающийся в выборе вида математической модели в виде эволюционного уравнения и последующей идентификации параметров, либо непараметрической идентификации модели. Модель считается адекватной, если оценка заданного критерия адекватности, вычисленная как зависимость невязки модели от экспериментальных данных находится в допустимых пределах.

В эконометрике идентификация – это единственность соответствия между приведенной и структурной формами модели.

С позиции идентифицируемости структурные модели можно подразделить на три вида:

- 1) идентифицируемые;
- 2) неидентифицируемые;
- 3) сверхидентифицируемые.

Модель *идентифицируема*, если все структурные ее коэффициенты определяются однозначно, единственным образом по коэффициентам приведенной формы модели, т. е. если число параметров структурной модели равно числу параметров приведенной формы модели. В этом случае структурные

коэффициенты модели оцениваются через параметры приведенной формы модели и модель идентифицируема.

Модель *неидентифицируема*, если число приведенных коэффициентов меньше числа структурных коэффициентов, и в результате структурные коэффициенты не могут быть оценены через коэффициенты приведенной формы модели.

Модель *сверхидентифицируема*, если число приведенных коэффициентов больше числа структурных коэффициентов. В этом случае на основе коэффициентов приведенной формы можно получить два или более значений одного структурного коэффициента. В этой модели число структурных коэффициентов меньше числа коэффициентов приведенной формы. Сверхидентифицируемая модель в отличие от неидентифицируемой модели практически решаема, но требует для этого специальных методов исчисления параметров.

Структурная модель всегда представляет систему совместных уравнений, каждое из которых требуется проверять на идентификацию. Модель считается идентифицируемой, если каждое уравнение системы идентифицируемо. Если хотя бы одно из уравнений системы неидентифицируемо, то и вся модель считается неидентифицируемой. Сверхидентифицируемая модель содержит хотя бы одно сверхидентифицируемое уравнение.

Выполнение условия идентифицируемости модели проверяется для каждого уравнения системы. Чтобы уравнение было идентифицируемо, необходимо, чтобы число предопределенных переменных, отсутствующих в данном уравнении, но присутствующих в системе, было равно числу эндогенных переменных в данном уравнении без одного.

Если обозначить число эндогенных переменных в уравнении  $i$  системы через  $H$ , а число экзогенных (предопределенных) переменных, которые содержатся в системе, но не входят в данное уравнение, – через  $D$ , то условие идентифицируемости модели может быть записано в виде следующего счетного правила:

Таблица 7.1 – Счетное правило

|             |                                |
|-------------|--------------------------------|
| $D + 1 = H$ | уравнение идентифицируемо      |
| $D + 1 < H$ | уравнение неидентифицируемо    |
| $D + 1 > H$ | уравнение сверхидентифицируемо |

Для оценки параметров структурной модели система должна быть идентифицируема или сверхидентифицируема.

Если необходимое условие выполнено, то далее проверяется достаточное условие идентификации.

*Достаточное условие идентификации.* Уравнение идентифицируемо, если определитель матрицы, составленной из коэффициентов при переменных, отсутствующих в исследуемом уравнении, не равен нулю, и ранг этой матрицы не менее числа эндогенных переменных системы без единицы.

#### 7.4. Метод максимального правдоподобия

Принцип максимального правдоподобия является спорным принципом статистического вывода, который предполагает, что вся информация о статистической выборке содержится в функции правдоподобия.

Функция правдоподобия основывается на условной вероятности взятием ее как функции от второго аргумента при фиксировании первого. Например рассмотрим модель в которой плотностью вероятности случайной величины  $X$  зависит от параметра  $\theta$ . Тогда для некоторого конкретного значения  $x$  случайной величины  $X$  функция

$$L(\theta|x)=P(X=x|\theta)$$

и есть функция правдоподобия  $\theta$ , определяя насколько правдоподобно каждое конкретное значение параметра  $\theta$  при условии, что известно значение  $x$  величины  $X$ . Две функции правдоподобия являются равными, если одна есть произведение второй на некоторую скалярную величину.

С принципом максимального правдоподобия связана концепция правдоподобия, формулируемая как **закон максимального правдоподобия**, согласно которому значения истинного и рассчитанного параметра равняется отношению их функций правдоподобия. Тогда отношение:

$$\lambda = \frac{L(a|X=x)}{L(b|X=x)} = \frac{P(X=x|a)}{P(X=x|b)}$$

является мерой того, насколько величина  $x$  принимает параметр  $a$  в отношении к  $b$ . Таким образом, если отношение равняется 1, то разницы нет, а если больше 1, то  $a$  предпочтительней  $b$  и наоборот.

Из принципа максимального правдоподобия и закона максимального правдоподобия следует, что параметр, который максимизирует функцию правдоподобия является лучшим. Это и является основой широко известного метода максимального правдоподобия.

Принцип максимального правдоподобия принимается не всеми. Некоторые широко используемые методы традиционной статистики, как например проверка статистических гипотез противоречат принципу максимального правдоподобия.

Для фиксированного набора данных и базовой вероятной модели, используя метод максимального правдоподобия, можно получить значения параметров модели, которые делают данные «более близкими» к реальным. Оценка максимального правдоподобия дает уникальный и простой способ определить решения в случае нормального распределения. Метод оценки максимального правдоподобия применяется для широкого круга статистических моделей, в том числе:

- линейные модели и обобщенные линейные модели;
- факторный анализ;
- моделирования структурных уравнений;
- многие ситуации, в рамках проверки гипотезы и доверительного интервала формирования;

- дискретные модели выбора.

Метод применяется в широких областях науки, в том числе:

- системы связи;
- психометрия;
- эконометрика;
- время задержки в акустических и электромагнитных системах;
- моделирование в ядерной физике и физике элементарных частиц;
- вычислительная филогенетика;
- моделирование каналов в транспортных сетях.

#### Пример 1

Пусть  $X_1, \dots, X_n \sim U[0, \theta]$  независимая выборка из непрерывного равномерного распределения на отрезке  $[0, \theta]$ , где  $\theta > 0$  – неизвестный параметр. Тогда функция правдоподобия имеет вид:

$$f(x|\theta) = \begin{cases} \frac{1}{\theta^n}, & x \in [0, \theta]^n \\ 0, & \neg x \in [0, \theta]^n \end{cases}.$$

Последнее равенство может быть переписано в виде:

$$f(x|\theta) = \begin{cases} \frac{1}{\theta^n}, & \theta \geq \max(x_1, \dots, x_n) \\ 0, & \theta < \max(x_1, \dots, x_n) \end{cases},$$

где  $x = (x_1, \dots, x_n)^T$ , откуда видно, что своего максимума функция правдоподобия достигает в точке  $\theta = \max(x_1, \dots, x_n)$ . Таким образом:

$$\hat{\theta}_{МП} = \max(X_1 \dots X_n).$$

#### Пример 2

Пусть  $x_1, \dots, x_n \sim N(\mu, \sigma^2)$  – независимая выборка из нормального распределения с неизвестными средним и дисперсией. Построим оценку максимального правдоподобия для неизвестного вектора параметров  $(\mu, \sigma^2)$ . Логарифмическая функция правдоподобия принимает вид:

$$L(x|\mu, \sigma^2) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \mu)^2.$$

Чтобы найти ее максимум, приравняем к нулю частные производные:

$$\begin{cases} \frac{\partial}{\partial \mu} L(x|\mu, \sigma^2) = 0 \\ \frac{\partial}{\partial \sigma^2} L(x|\mu, \sigma^2) = 0 \end{cases} \rightarrow \begin{cases} \frac{\sum_{i=1}^n X_i - n\mu}{\sigma^2} = 0 \\ -\frac{n}{2\sigma^2} + \frac{\sum_{i=1}^n (X_i - \mu)^2}{2(\sigma^2)^2} = 0 \end{cases},$$

откуда

$\mu$  – выборочное среднее,

$\sigma^2$  – выборочная дисперсия.

Следует добавить, что название «метод максимального правдоподобия» является калькой с английского «maximum likelihood method».

## **8. РЯДЫ ДИНАМИКИ, ОПРЕДЕЛЕНИЯ И ПОКАЗАТЕЛИ**

### **8.1. Временные ряды (ряды динамики), определения**

Экономическая деятельность фирм развивается во времени. Поэтому время можно рассматривать как признак-фактор, определяющий развитие социально-экономических процессов. Следовательно, для выявления статистических закономерностей необходимо знать характеристики изменения и тенденции их развития. Достигается это при построении, анализе и расчете аналитических показателей соответствующих временных рядов (рядов динамики).

В каждой сфере экономики встречаются явления, которые интересно и важно изучать в их развитии, т.к. они эволюционируют и флуктуируют во времени. С течением времени изменяются цены, экономические условия, режим протекания того или иного производственного процесса. Совокупность измерений подобного рода показателей в течение некоторого периода времени и представляет временной ряд.

Цели изучения временных рядов могут быть различными. Можно, например, стремиться предсказать будущее на основании знаний прошлого, управлять процессом, порождающим ряд, пытаться выяснить механизм, лежащий в основе процесса, очистить временной ряд от компонент, которые затемняют его динамику, или просто сжато описать характерные особенности ряда. Временной ряд – это упорядоченная последовательность значений какого-либо показателя за несколько периодов времени. Основная характеристика, которая отличает временной ряд от простой выборки данных, – указанное время измерения или номер изменения по порядку.

Таким образом, временной ряд (динамический ряд, ряд динамики) – статистическое описание изменения явлений во времени. Более точно: динамическим рядом называют ряд числовых характеристик изучаемого явления за последовательно взятые отрезки или моменты времени. Основной чертой, выделяющей анализ временных рядов среди других видов статистического анализа, является существенность порядка, в котором производятся наблюдения.

Временные ряды используются для аналитики и прогнозирования, когда важно определить, что будет происходить с показателями в ближайший час/день/месяц/год: например, сколько пользователей скачают за день мобильное приложение. Показатели для составления временных рядов могут быть не только техническими, но и экономическими, социальными и даже природными.

Выбор соответствующих приемов представления и способов анализа таких рядов определяется характером исходных данных и зависит от задач исследования. Поэтому отнесение ряда динамики к тому или иному виду имеет важное значение. Каждый ряд динамики содержит:

- 1) показатель времени;
- 2) уровень развития явления.

В качестве показателей времени используют либо определенные даты



(моменты) времени, либо отдельные периоды (годы, кварталы, сутки и т. д.), которые характеризуют последовательность изменения значений показателя, фиксирующего состояние явления или процесса. Статистические показатели, характеризующие изучаемый объект во временном аспекте, называются уровнями ряда, т. е. уровни ряда отображают значение показателя явления, соответствующее определенному времени.

Различают два вида временных рядов. Измерение некоторых величин (температуры, напряжения и т. д.) производится непрерывно, по крайней мере, теоретически. При этом наблюдения можно фиксировать в виде графика. Но даже в том случае, когда изучаемые величины регистрируются (или могут регистрироваться) непрерывно, практически при их обработке используются только те значения, которые соответствуют дискретному множеству моментов времени. Следовательно, если время измеряется непрерывно, временной ряд называется непрерывным, если же время фиксируется дискретно (т. е. через фиксированный интервал времени), то временной ряд дискретен. В дальнейшем будем иметь дело только с дискретными временными рядами. Дискретные временные ряды получаются двумя способами:

- выборкой из непрерывных временных рядов через регулярные промежутки времени (например, численность населения, величина собственного капитала фирмы, объем денежной массы, курс акции) – такие временные ряды называются моментными;
- накоплением переменной в течение некоторого периода времени (примеры: объем производства какого-либо вида продукции, количество осадков, объем импорта), — в этом случае временные ряды называются интервальными.

В зависимости от способа выражения уровней (в виде абсолютных, относительных или средних величин) ряды динамики подразделяются на ряды абсолютных, относительных или средних величин.

Ряды динамики могут быть с равноотстоящими (по времени) уровнями и с неравноотстоящими (по времени) уровнями.

Статистическое отображение развития процесса во времени может быть представлено в виде ряда динамики с нарастающими итогами. Такое представление позволяет отобразить результаты развития не только за текущий, но и за предшествующие периоды времени (табл. 8.1).

Таблица 8.1 – Данные об объемах продаж товаров в регионе

| Год       | 1998  | 1999  | 2000  | 2001  | 2002  |
|-----------|-------|-------|-------|-------|-------|
| Тыс. руб. | 223,7 | 195,7 | 237,4 | 179,2 | 189,1 |

Это интервальный ряд динамики абсолютных величин с равноотстоящими по времени уровнями. Его уровни характеризуют суммарный итог за определенный отрезок времени – за каждый год. Уровни интервального ряда могут суммироваться, так как не содержат элементов повторного счета, например, доходы фирмы за ряд лет.

Примером моментного ряда абсолютных величин с неравноотстоящими

(по времени) уровнями может служить ряд динамики, показывающий число фирм в стране (на конец года) (табл. 8.2).

Уровни этого ряда – обобщенные итоги учета числа фирм по состоянию на определенную дату (конец каждого года). Отдельные уровни моментного ряда содержат элементы повторного счета, так как отдельные фирмы, учитываемые в 1990 году, существовали и в 2002 году, являясь единицами совокупности. Это делает бессмысленным суммирование уровней моментных рядов динамики.

Таблица 8.2 – Число фирм в регионе (на 31.12 каждого года)

| Год              | 1990 | 1995 | 2000 | 2001 | 2002 | 2003 |
|------------------|------|------|------|------|------|------|
| Число фирм, тыс. | 33,0 | 28,5 | 25,9 | 25,9 | 25,9 | 26,6 |

Примером интервального ряда динамики средних величин с неравноотстоящими уровнями (по времени) может служить ряд динамики средней месячной денежной заработной платы лиц наемного труда страны (табл. 8.3).

Таблица 8.3 – Средняя месячная заработная плата лиц наемного труда

| Год            | 1969  | 1970  | 1975  | 1980  | 1981  | 1982  | 1984  |
|----------------|-------|-------|-------|-------|-------|-------|-------|
| Зарплата, руб. | 116,9 | 122,0 | 145,8 | 168,9 | 172,5 | 177,3 | 184,8 |

Его уровни относятся к годовым интервалам времени, но суммирование их самостоятельного значения не имеет.

Примером моментного ряда динамики относительных величин с равноотстоящими уровнями по времени может служить ряд динамики, характеризующий удельный вес городского населения в общей численности населения на начало года (табл. 8.4).

Таблица 8.4 – Удельный вес городского населения на начало года

| Период | Население региона, % |           |          |
|--------|----------------------|-----------|----------|
|        | Общее количество     | Городское | Сельское |
| 1      | 100                  | 63,0      | 37,0     |
| 2      | 100                  | 64,0      | 36,0     |
| 3      | 100                  | 64,4      | 35,6     |
| 4      | 100                  | 64,8      | 35,2     |
| 5      | 100                  | 65,2      | 34,7     |

Суммирование уровней данного ряда не имеет смысла.

С помощью рядов динамики решаются вопросы:

- характеристики уровней развития процессов во времени;
- измерения динамики изучаемых процессов с помощью системы показателей;
- выявления и количественной оценки основной тенденции развития;
- изучения периодических колебаний;

- прогнозирования с помощью экстраполяции (продолжения) рядов динамики.

В эконометрии принято моделировать временной ряд как случайный процесс, называемый также стохастическим процессом, под которым понимается статистическое явление, развивающееся во времени согласно законам теории вероятностей. Случайный процесс – это случайная последовательность. Обычно предполагают, что эта последовательность идет от минус до плюс бесконечности:  $\{X_t\}_{t=-\infty, \dots, +\infty}$ . Временной ряд – это лишь одна частная реализация такого теоретического стохастического процесса:

$$x = \{x_t\}_{t=1, \dots, T} = (x_1, \dots, x_T)'.$$

Здесь  $T$  – это длина временного ряда.

Временной ряд  $x = (x_1, \dots, x_T)'$  также часто неформально называют выборкой. Обычно стоит задача по данному ряду сделать какие-то заключения о свойствах лежащего в его основе случайного процесса, оценить параметры, сделать прогнозы и т. п. В литературе по временным рядам существует некоторая неоднозначность, и иногда временным рядом называют сам случайный процесс, либо его отрезок  $t = 1, \dots, T$ , а иногда статистическую модель, которая порождает данный случайный процесс. В дальнейшем не будем в явном виде, посредством особых обозначений (как это только что сделали, используя для записи теоретического процесса большие буквы), различать случайный процесс и его реализацию. Из контекста каждый раз будет ясно, о чем идет речь.

Возможные значения временного ряда в данный момент времени  $t$  описываются с помощью случайной величины  $x_t$  и связанного с ней распределения вероятностей  $p(x_t)$ . Тогда наблюдаемое значение  $x_t$  временного ряда в момент  $t$  рассматривается как одно из множества значений, которые могла бы принять случайная величина  $x_t$  в этот момент времени. Следует отметить, однако, что, как правило, наблюдения временного ряда взаимосвязаны, и для корректного его описания следует рассматривать совместную вероятность  $p(x_1, \dots, x_T)$ .

Для удобства можно провести классификацию случайных процессов и соответствующих им временных рядов на детерминированные и случайные процессы (временные ряды). Детерминированным называют процесс, который принимает заданное значение с вероятностью единица. Например, его значения могут точно определяться какой-либо математической функцией от момента времени  $t$ , как в следующем примере:

$$x_t = A \cos(2\pi f t - \theta).$$

Когда же будем говорить о случайном процессе и случайном временном ряде, то, как правило, будем подразумевать, что он существенно случаен, то есть не является детерминированным.

Стохастические процессы подразделяются на стационарные и нестационарные. Стохастический процесс является стационарным, если он находится в определенном смысле в статистическом равновесии, т. е. его свойства с вероятностной точки зрения не зависят от времени. Процесс будет не стационарным, если эти условия нарушаются.

Как правило, термин «временной ряд» и сам по себе подразумевает, что этот ряд является одномерным (скалярным). Часто бывает важно рассмотреть совместную динамику набора временных рядов  $x_t = (x_{1t}, \dots, x_{kt})$ . Такой набор называют многомерным временным рядом или векторным временным рядом. Соответственно, говорят также о многомерных (векторных) случайных процессах.

При анализе экономических временных рядов традиционно различают разные виды эволюции (динамики). Эти виды динамики могут, вообще говоря, комбинироваться. Тем самым задается разложение временного ряда на составляющие (компоненты), которые с экономической точки зрения несут разную содержательную нагрузку. Перечислим наиболее важные:

- Тенденция соответствует медленному изменению, проходящему в некотором определенном направлении, которое сохраняется в течение значительного промежутка времени. Тенденцию называют также трендом или долговременным движением.
- Циклические колебания – это более быстрая, чем тенденция, квазипериодическая динамика, в которой есть фаза возрастания и фаза убывания. Наиболее часто цикл связан с флуктуациями экономической активности.
- Сезонные колебания соответствуют изменениям, которые происходят регулярно в течение года, недели или суток. Они связаны с сезонами и ритмами человеческой активности.
- Календарные эффекты – это отклонения, связанные с определенными предсказуемыми календарными событиями, такими как праздничные дни, количество рабочих дней за месяц, високосность года и т.п.
- Случайные флуктуации – беспорядочные движения относительно большой частоты. Они порождаются влиянием разнородных событий на изучаемую величину (несистематический или случайный эффект).
- Выбросы – это аномальные движения временного ряда, связанные с редко происходящими событиями, которые резко, но лишь очень кратковременно отклоняют ряд от общего закона, по которому он движется.
- Структурные сдвиги – это аномальные движения временного ряда, связанные с редко происходящими событиями, имеющие скачкообразный характер и меняющие тенденцию.

Некоторые экономические ряды можно считать представляющими те или иные виды таких движений почти в чистом виде. Но большая часть их имеет очень сложный вид. В них могут проявляться, например, как общая тенденция возрастания, так и некоторые сезонные изменения, на которые могут накладываться случайные флуктуации. Часто для анализа временных рядов оказывается полезным изолированное рассмотрение отдельных компонент. Для того, чтобы можно было разложить конкретный ряд на эти составляющие, требуется сделать какие-то допущения о том, какими свойствами должны обладать эти составляющие. Желательно также построить сначала формальную

статистическую модель, которая бы включала в себя в каком-то виде эти составляющие, затем оценить ее, а после этого на основании полученных оценок вычленил составляющие. Однако построение формальной модели является сложной задачей. В частности, из содержательного описания не всегда ясно, как моделировать те или иные компоненты. Например, тренд может быть детерминированным или стохастическим. Аналогично, сезонные колебания можно комбинировать с помощью детерминированных переменных или с помощью стохастического процесса определенного вида. Компоненты временного ряда могут входить в него аддитивно или мультипликативно. Более того, далеко не все временные ряды имеют достаточно простую структуру, чтобы можно было разложить их на указанные составляющие.

Существует две основные категории методов разложения временных рядов на компоненты. Первая категория использует множественные регрессии с факторами, являющимися функциями времени, вторая основана на применении линейных фильтров.

## **8.2. Сопоставимость и смыкание рядов динамики**

Важнейшим условием правильного построения рядов динамики является сопоставимость всех входящих в него уровней. Данное условие решается либо в процессе сбора и обработки данных, либо путем их пересчета.

В процессе развития во времени, прежде всего, происходят количественные изменения явлений, а затем на определенных ступенях совершаются качественные скачки, приводящие к изменению закономерностей явления. В связи с этим научный подход к изучению рядов динамики заключается в том, чтобы ряды, охватывающие большие периоды времени, разделять на такие, которые бы объединяли лишь однородные периоды развития совокупности, характеризующейся одной закономерностью развития.

Процесс выделения однородных этапов развития рядов динамики носит название периодизации динамики. Вопрос о том, какие этапы развития прошлого или иное явление нужно исследовать за определенный исторический отрезок времени, решается теорией той науки, к области которой относится изучаемая совокупность явлений.

Важно также, чтобы в ряду динамики интервалы или моменты, по которым определены уровни, имели одинаковый экономический смысл. Скажем, при изучении роста поголовья скота бессмысленно сравнивать цифры поголовья по состоянию на 1 октября с данными 1 января, так как первая цифра включает в себя не только скот, оставшийся на зимовку, но и предназначенный к убою, а вторая цифра отражает только скот, оставленный на зимовку.

Уровни ряда динамики могут оказаться несопоставимыми по кругу охватываемых объектов вследствие перехода ряда объектов из одного подчинения в другое.

Несопоставимость уровней ряда может возникнуть вследствие изменений территориальных границ областей, районов и т. д.

Следовательно, прежде чем анализировать временной ряд, надо, исходя из цели исследования, убедиться в сопоставимости уровней ряда и, если последняя отсутствует, добиться ее дополнительными расчетами. Для того чтобы привести уровни ряда динамики к сопоставимому виду, иногда приходится прибегать к приему, который носит название **смыкания временных рядов**. Под смыканием понимают объединение в один ряд (более длинный) двух или нескольких рядов динамики, уровни которых являются несопоставимыми. Для осуществления смыкания необходимо, чтобы для одного из периодов (переходного) имелись данные, исчисленные по разной методологии (или в разных границах).

Таким образом, при составлении рядов динамики должны выполняться следующие требования.

1. Периодизации развития, т. е. расчленения явления во времени на однородные этапы, в пределах которых показатель подчиняется одному закону развития. Периодизация может осуществляться несколькими методами:

- Исторический метод. Периодизация осуществляется на основе «узаконенной» структуры динамики. При этом обращают внимание на значимые даты и события, а именно на: время принятия управленческих решений по данному показателю; изменение хозяйственного механизма, смену руководства и т. д. Недостатком этого метода является то, что точные временные границы периодов путем теоретического анализа удастся получить крайне редко.
- Метод параллельной периодизации. Идея этого метода заключается в том, что показатели рядов изменяются однокачественно.
- Методы многомерного статистического анализа. Часто требуется выделить однокачественные периоды в развитии явлений или процессов, получить адекватное отображение которых с помощью одного лишь показателя трудно. Поэтому для анализа системы показателей и периодизации используются методы многомерной средней и факторного анализа.

2. Соответствие величины временных интервалов интенсивности изучаемых процессов. Чем больше вариация уровней во времени, тем чаще следует регистрировать значения уровней ряда. Соответственно, для стабильных процессов интервалы можно увеличивать.

3. Упорядоченность во времени числовых уровней ряда динамики. Не допускается анализ рядов с пропусками отдельных уровней. Если же такие пропуски неизбежны, то их восполняют условными расчетными значениями.

4. Сопоставимость статистических данных по территории, кругу охватываемых объектов, единицами измерения, времени регистрации, ценам, методологии расчета.

Правильные выводы при анализе рядов динамики можно получить, если обеспечивается сопоставимость его элементов. Соблюдение требования сопоставимости уровней ряда означает, что правомерно будет сравнение только тех показателей, которые однородны по экономическому содержанию. Чтобы обеспечить требование сопоставимости, необходимо обеспечить единство:

- явления, когда уровни ряда за отдельные периоды должны характеризовать

полноту охвата одинаковых частей явления, один и тот же его состав. Например, при характеристике динамики физического объема внешней торговли целесообразно рассматривать только экспорт и импорт товаров, для производства которых используется одинаковая структура факторов производства;

- территории (пространства), когда уровни ряда должны характеризовать распространенность явления в одних и тех же границах. Территориальные изменения административных единиц и стран, слияния и разделения фирм приводят к несопоставимости уровней ряда, так как определяют разные и несопоставимые явления. Например, если до поглощения у фирмы был один объем продаж, а за счет поглощения другой фирмы он резко увеличился, то полученный ряд динамики будет содержать несопоставимые уровни ряда. В этом случае необходимо рассматривать по периодам объем продаж двух фирм до и после слияния. Использование инструментария динамических рядов подразумевает эволюционный путь развития явления, т. е., как правило, отсутствие резкого изменения (скачкообразного) уровней;

- методологии получения показателей, составляющих ряд динамики, т. е. если в разные периоды показатели рассчитывались по отличным друг от друга методикам, то полученные уровни ряда также являются несопоставимыми. Особенно часто эта проблема возникает при международных сопоставлениях;

- единиц измерения в натуральной, трудовой и стоимостной форме. Например, сопоставление объемов выпуска текстильной промышленности при разных стандартах ее ширины не позволяет делать достаточно обоснованных выводов. То же самое относится и к сопоставлению производительности труда при разной длительности рабочей недели. В стоимостной форме необходимо учитывать уровень инфляции и различную покупательную способность основных рынков сбыта, на которых реализуется продукция;

- времени, т. е. равенство периодов, определяющих уровни ряда. Например, ритмичность выпуска продукции должна характеризоваться периодами, в которых одинаковое количество рабочих дней.

Динамические ряды формируются в результате систематизации результатов наблюдений в хронологическом порядке. При этом в отдельные периоды времени могут происходить изменения, приводящие к несопоставимости уровней ряда. Но, с другой стороны, чем больше периодов захватывает ряд динамики, тем с большей вероятностью может быть выявлена присущая ему закономерность, которая необходима при прогнозировании. Поэтому при анализе рядов динамики возникает необходимость смыкания рядов, т. е. объединения двух и более рядов, характеризующих изменение процесса, в один ряд с одновременным приведением уровней ряда к сопоставимому виду.

Несопоставимость вызывается разными причинами. Это могут быть разновеликость показаний времени, неоднородность состава изучаемых совокупностей по времени, изменения методики проведения и обработки результатов наблюдения, различия применяемых в отдельные периоды единиц измерения изучаемых параметров. Необходимость смыкания возникает в случае территориальных, ведомственных, организационных изменений, например,

слияния фирм, изменения стандартов и т. п.

При изучении динамики деятельности фирм возникает несопоставимость по внутригодовым периодам:

- 1) при неодинаковой продолжительности показаний времени (месяцев, кварталов, полугодий);
- 2) при сравнении деятельности фирм, действующих в разных странах из-за разного количества праздничных дней.

Для анализа интенсивности деятельности фирм объемные показатели за разновеликие периоды пересчитываются в среднесуточные показатели с учетом продолжительности рабочего дня. Это позволяет устранить несопоставимость уровней ряда.

Несопоставимость в рядах динамики может произойти в связи с административно-территориальными изменениями из-за слияния фирм, расширения деятельности на другие регионы и страны, закрытия или открытия филиалов фирмы. Существует несколько способов приведения рядов динамики к сопоставимому виду.

*Пример*

Таблица 8.5 – Объем продаж до и после слияния фирм (в год)

| Год                     | 1987 | 1988 | 1989 | 1990 | 1991 | 1992 | 1993 | 1994 |
|-------------------------|------|------|------|------|------|------|------|------|
| До слияния, млн руб.    | 19,1 | 19,7 | 20,0 | 21,2 | –    | –    | –    | –    |
| После слияния, млн руб. | –    | –    | –    | 22,8 | 23,6 | 24,5 | 26,2 | 28,1 |

При приведении ряда динамики к сопоставимому виду определим для 1990 года коэффициент соотношения уровней двух рядов:

$$22,8/21,2 = 1,1.$$

Умножая на этот коэффициент уровни первого ряда, получаем их сопоставимость с уровнями второго ряда:

$$1987 - 19,1 \cdot 1,1 = 21,0 \text{ млн руб.},$$

$$1988 - 19,7 \cdot 1,1 = 21,7 \text{ млн руб.},$$

$$1989 - 20,0 \cdot 1,1 = 22,0 \text{ млн руб.}$$

Таким образом получен сопоставимый ряд динамики после слияния фирм (табл. 8.6).

Таблица 8.6 – Ряд динамики абсолютных величин с сопоставимыми уровнями

| Год                    | 1987 | 1988 | 1989 | 1990 | 1991 | 1992 | 1993 | 1994 |
|------------------------|------|------|------|------|------|------|------|------|
| Объем продаж, млн руб. | 21,0 | 21,7 | 22,0 | 22,8 | 23,6 | 24,5 | 26,2 | 28,1 |

Другой способ смыкания рядов динамики заключается в том, что уровни года, в котором произошли изменения, принимаются за 100 %, а остальные пересчитываются в процентах по отношению к этим уровням (до изменения – по отношению к 21,2, а после изменений – по отношению к 22,8). В результате получаем сомкнутый ряд (табл. 8.7).



Таблица 8.7 – Ряд динамики относительных величин с сопоставимыми уровнями

| Год             | 1987 | 1988 | 1989 | 1990  | 1991  | 1992  | 1993  | 1994  |
|-----------------|------|------|------|-------|-------|-------|-------|-------|
| Объем продаж, % | 90,1 | 92,9 | 94,3 | 100,0 | 103,5 | 107,5 | 114,9 | 123,2 |

Проблема сопоставимости в рядах динамики возникает в связи с применением различных по экономическому значению денежных измерителей из-за инфляции, деноминации, перехода к новым валютам. Так как уровни цен изменяются во времени, то для стоимостной оценки объема продаж используются цены соответствующих периодов. Но для изучения динамики физического объема продаж для устранения влияния изменения цен объем реализации пересчитывается в неизменные (базисные) цены.

### 8.3. Показатели временного ряда

Одним из важнейших направлений анализа рядов динамики является определение и прогнозирование особенностей развития явления в отдельные периоды времени. Основные задачи, связанные с определением закономерностей развития явления, связаны с выявлением:

- величины и тенденций изменения уровней ряда динамики;
- интенсивности изменения уровней временного ряда;
- средних значений динамического ряда;
- основных факторов, обуславливающих изменение значений уровней ряда динамики;
- прогноза развития явления.

Для выявления специфики развития изучаемых явлений за отдельные периоды времени и их количественной оценки определяют *абсолютные* и *относительные показатели* изменения ряда динамики:

- прирост;
- значение 1 % прироста;
- темп роста;
- темп прироста;
- темп наращивания.

В основе расчета показателей рядов динамики лежит сравнение его уровней. В зависимости от применяемого способа сопоставления показатели динамики могут определяться на постоянной и переменной базах сравнения. Выбор базы сравнения зависит от цели исследования. При сравнении каждого уровня ряда с предыдущим получают *цепные показатели*; при сравнении каждого уровня с одним и тем же уровнем (базой) получают *базисные показатели*.

*Пример.* Проведем анализ динамики затрат населения на отдых за период с 1985 по 1990 годы. Данные и показатели приведены в табл. 8.8.

Таблица 8.8 – Анализ динамики затрат потребителей за период с 1985 по 1990 годы

| Год  | Объем затрат, млн руб. | Абсолютные приросты (снижения) |        | Темпы роста, % |        | Темпы прироста, % |        | Абсолютное значение 1 % прироста |
|------|------------------------|--------------------------------|--------|----------------|--------|-------------------|--------|----------------------------------|
|      |                        | к предыдущему                  | к 1985 | к предыдущему  | к 1985 | к предыдущему     | к 1985 |                                  |
| 1985 | 103,0                  | –                              | –      | 100,00         | –      | –                 | –      | –                                |
| 1986 | 105,4                  | 2,4                            | 2,4    | 102,30         | 102,3  | 2,30              | 2,3    | 1,03                             |
| 1987 | 107,4                  | 2,0                            | 4,4    | 101,90         | 104,3  | 1,90              | 4,3    | 1,05                             |
| 1988 | 110,7                  | 3,3                            | 7,7    | 103,10         | 107,5  | 3,10              | 7,5    | 1,07                             |
| 1989 | 109,0                  | –1,7                           | 6,0    | 98,50          | 105,8  | –1,50             | 5,8    | 1,11                             |
| 1990 | 107,3                  | –1,7                           | 4,3    | 98,44          | 104,2  | –1,56             | 4,2    | 1,09                             |

Важнейшим статистическим показателем динамики является абсолютный прирост ( $\Delta$ ), который используется для вычисления абсолютной скорости роста (снижения) уровней ряда динамики. Поэтому этот показатель еще определяют как скорость роста. Величина его определяется как разность двух сравниваемых уровней ряда динамики:

базисный абсолютный прирост исчисляется как разность между сравниваемым уровнем и уровнем, принятым за базу сравнения,

$$\Delta = y_i - y_0;$$

цепной абсолютный прирост – это разность между сравниваемым уровнем и уровнем, который ему предшествует,

$$\Delta = y_i - y_{i-1},$$

где  $y_i$  – уровень периода  $i$ ;

$y_0$  – уровень базисного периода.

Так, например, абсолютный прирост за 1987 год по сравнению с 1986 годом составил:

$$107,4 - 105,4 = 2,0 \text{ млн руб.},$$

а по сравнению с базисным периодом (1985 годом) – 4,3 млн руб.

Абсолютный прирост может быть и отрицательным. Тогда он показывает, на сколько уровень изучаемого периода ниже базисного или предшествующего.

Между базисными и цепными абсолютными приростами существует связь: сумма цепных абсолютных приростов равна базисному абсолютному приросту последнего периода ряда динамики.

Распространенным статистическим показателем динамики является темп роста ( $T_p$ ), который характеризует отношение двух уровней ряда и может выражаться в процентах или коэффициентах. Этот показатель используется для оценки интенсивности изменения уровней ряда динамики.

Базисные темпы роста оцениваются отношением текущего уровня к базисному:

$$T_p = \frac{y_i}{y_0} \times 100.$$

Цепные темпы роста определяются отношением текущего уровня к предыдущему:

$$T_p = \frac{y_i}{y_{i-1}} \times 100.$$

Этот показатель всегда представляет положительное число. Так, для 1990 года темп роста по сравнению с 1989 годом составил

$$\frac{107,3}{109,0} \times 100 = 98,44 \, \%.$$

Темп роста может быть выражен и в виде коэффициента. В этом случае он показывает, во сколько раз данный уровень ряда больше уровня базисного года или какую часть его он составляет. Если темп роста больше единицы (или 100 %), то это указывает на увеличение изучаемого уровня по сравнению с базисным. Если показатель равен единице (или 100 %), то изучаемый уровень не изменился по сравнению с уровнем базисного периода. При величине показателя меньше единицы (100 %) изучаемый уровень меньше уровня базисного периода.

Между показателями базисного и цепного темпа роста существует взаимосвязь:

- произведение последовательных цепных темпов роста равно базисному темпу роста;
- частное от деления последующего базисного темпа роста на предыдущий равно соответствующему цепному темпу роста.

Для выражения изменения величины абсолютного прироста уровней ряда динамики в относительных величинах определяется темп прироста ( $T_{пр}$ ). Исчисленный в процентах темп прироста показывает, на сколько изменился сравниваемый уровень с уровнем, принятым за базу. Этот показатель определяется как отношение абсолютного прироста к предыдущему или базисному уровню.

Базисный темп прироста – это отношение сравниваемого базисного абсолютного прироста к уровню, принятому за базу сравнения,

$$T_{пр} = \frac{\Delta}{y_0} \times 100.$$

Цепной темп прироста – это отношение сравниваемого цепного абсолютного прироста к предыдущему уровню

$$T_{пр} = \frac{\Delta}{y_{i-1}} \times 100.$$

Между показателями темпа роста и темпа прироста существует взаимосвязь, т. е. темп прироста может быть получен вычитанием из темпов роста 100 % (или единицы в зависимости от формы представления как коэффициента или в процентах):

$$T_{пр} = T_p - 100.$$

В нашем примере темп прироста показывает, на сколько (например, в процентах) затраты в 1990 году возросли по сравнению с 1985 годом:

$$\frac{4,3}{103} 100 = 4,2 \, \% \text{ или } 104,2 - 100 = 4,2.$$

Если уровни ряда уменьшаются, то соответствующие показатели темпа прироста будут отрицательными, так как характеризуют относительное уменьшение прироста уровня ряда динамики.

Важным показателем динамики процессов является темп наращивания, который представляет собой отношение цепного абсолютного прироста к уровню, принятому за постоянную базу сравнения,

$$T_n = \frac{\Delta}{y_0}.$$

Этот показатель во многом может противоречить значениям показателя цепного темпа прироста из-за особенностей его метода расчета. Темп наращивания можно непосредственно определять по базисным темпам роста:

$$T_n^i = \frac{y_i - y_{i-1}}{y_0} = T_p^i - T_p^{i-1}.$$

Анализ относительных показателей динамики (темпов роста и прироста) необходимо рассматривать вместе с абсолютными показателями (уровнями ряда и абсолютными приростами). Сравнение абсолютного прироста и темпа роста за одни и те же периоды времени показывает, что замедление темпов прироста не всегда сопровождается уменьшением абсолютных приростов. Поэтому для адекватной оценки значения темпа прироста его сравнивают с показателем абсолютного прироста. Полученный показатель, называемый абсолютным значением одного процента прироста (%), определяется как результат деления абсолютного прироста на соответствующий темп прироста, выраженный в процентах, т. е.

$$e[\%] = \frac{\Delta}{T_{пр}} \text{ или } 0,01y_{i-1}.$$

Расчет этого показателя имеет экономический смысл только на цепной основе. Для 1990 года абсолютное значение 1 % прироста равно:

$$0,01 \times 109 = 1,09 \text{ млн руб.}$$

В тех случаях, когда сравнение производится с достаточно отдаленным периодом времени, принятым за базу сравнения, рассчитывают пункты роста (снижения) как разность базисных темпов роста (прироста) для двух смежных периодов в процентах. Например, для 1990 года пункт роста (снижения) составит  $4,2 - 5,8 = -1,6$  (табл. 8.9).

Результатом суммирования пунктов роста является темп прироста текущего по сравнению с базисным периодом. По данным табл. 8.9 сумма пунктов роста (снижения) равна:

Таблица 8.9 – Сравнение темпов прироста и пунктов роста (снижения)

| Периоды | Темп прироста по сравнению с 1985 г., % | Сумма пунктов роста (снижения), % |
|---------|-----------------------------------------|-----------------------------------|
| 1985    | –                                       | –                                 |
| 1986    | 2,3                                     | 2,3                               |
| 1987    | 4,3                                     | 2,0                               |
| 1988    | 7,5                                     | 3,2                               |
| 1989    | 5,8                                     | –1,7                              |
| 1990    | 4,2                                     | –1,6                              |
| Итого   | $x$                                     | 4,2                               |

Для сопоставления двух рядов динамики используются коэффициенты опережения (отставания), которые представляют собой отношение темпов роста

или прироста за одинаковые периоды времени:

$$K_{\text{о.п.}} = \frac{T_p^1}{T_p^2} \text{ или } K_{\text{о.п.}} = \frac{T_{\text{пр}}^1}{T_{\text{пр}}^2},$$

где  $T_p^1, T_p^2$  – темпы роста двух сравниваемых рядов динамики;

$T_{\text{пр}}^1, T_{\text{пр}}^2$  – темпы прироста двух сравниваемых рядов динамики.

Эти показатели используются при сопоставлении двух изменяющихся по времени признаков одной совокупности или при изменении одного признака двух совокупностей. С помощью этих коэффициентов могут осуществляться сравнения разных территорий, фирм, отраслей, стран, регионов, а также структурные изменения внутри совокупности. Ниже в табл. 8.10 приводится пример такого анализа.

Таблица 8.10 – Изменение платежеспособности и потребления

| Период | Доходы, тыс. руб. | Абсолютный прирост доходов | Расходы, тыс. руб. | Абсолютный прирост расходов | Темп роста доходов по отношению к 1-му периоду |          | Темп прироста расходов по отношению к 1-му периоду |          | Коэффициенты опережения по |                 |
|--------|-------------------|----------------------------|--------------------|-----------------------------|------------------------------------------------|----------|----------------------------------------------------|----------|----------------------------|-----------------|
|        |                   |                            |                    |                             | доходов                                        | расходов | доходов                                            | расходов | темпам роста               | темпам прироста |
| 1      | 400               | –                          | 380                | –                           | –                                              | –        | –                                                  | –        | –                          | –               |
| 2      | 420               | 20                         | 390                | 10                          | 1,0500                                         | 1,026    | 5                                                  | 2,631    | 0,977                      | 0,526           |
| 3      | 450               | 30                         | 410                | 20                          | 1,0714                                         | 1,051    | 7,1429                                             | 5,128    | 0,981                      | 0,717           |
| 4      | 390               | –60                        | 430                | 20                          | 0,8667                                         | 1,048    | –13,333                                            | 4,878    | 1,210                      | –0,365          |
| 5      | 430               | 40                         | 435                | 5                           | 1,1026                                         | 1,011    | 10,256                                             | 1,162    | 0,917                      | 0,113           |
| 6      | 460               | 30                         | 440                | 5                           | 1,0698                                         | 1,011    | 6,9767                                             | 1,144    | 0,945                      | 0,164           |

Коэффициенты опережения, рассчитанные разными способами, отличаются по абсолютной величине и отображению тенденции (рис. 8.1).

#### 8.4. Средние показатели рядов динамики

Средние показатели рядов динамики являются обобщающей характеристикой его абсолютных уровней, абсолютной скорости и интенсивности изменения уровней ряда динамики. Различают следующие средние показатели: средний уровень ряда динамики, средний абсолютный прирост, средний темп роста и прироста.

Средний уровень ряда динамики характеризует типическую величину абсолютных уровней и обобщает итоги развития явления за интервал или момент. Методы расчета среднего уровня ряда динамики зависят от его вида и способов получения статистических данных.

В интервальном ряду динамики с равноотстоящими уровнями во времени расчет среднего уровня ряда ( $\bar{y}$ ) производится по формуле средней арифметической простой

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n}.$$

В нашем случае среднее значение затрат за 5 лет составило (табл. 8.8):

$$\bar{y} = 539, 8/5 = 107,96 \text{ млн руб.}$$

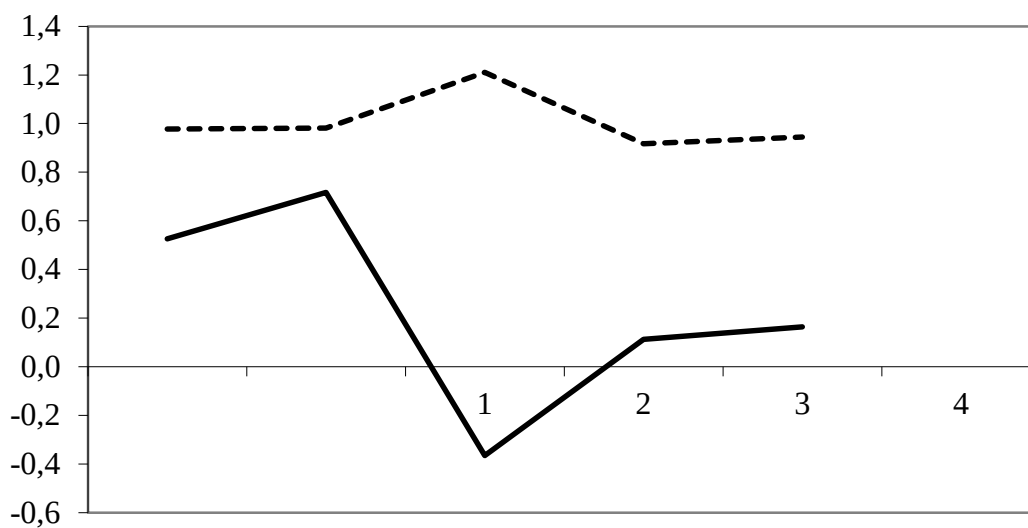


Рисунок 8.1 – Коэффициенты опережения по темпам роста и прироста:

----- по темпу роста  
———— по темпу прироста

Если интервальный ряд динамики представлен неравноотстоящими по времени уровнями, то средний уровень ряда вычисляется по формуле:

$$\bar{y} = \frac{\sum_{i=1}^n y_i t_i}{\sum_{i=1}^n t_i},$$

где  $t_i$  – число периодов времени, в течение которых уровень не изменяется.

Для моментного ряда с равноотстоящими по времени уровнями рассчитывается средняя хронологическая

$$\bar{y} = \frac{0,5y_1 + y_2 + y_3 + \dots + 0,5y_n}{n-1},$$

где  $n$  – число уровней ряда.

Средняя хронологическая для неравноотстоящих по времени уровней моментного ряда динамики рассчитывается по формуле

$$\bar{y} = \frac{(y_1 + y_2)t_1 + (y_2 + y_3)t_2 + (y_3 + y_4)t_3 + \dots + (y_{n-1} + y_n)t_{n-1}}{2\sum t_{n-1}}.$$

Средний абсолютный прирост (или средняя скорость роста) представляет обобщенную характеристику индивидуальных абсолютных приростов ряда динамики. Показатель определяется по абсолютным уровням ряда динамики и по цепным абсолютным приростам:

$$\bar{\Delta} = \frac{y_n - y_0}{n-1} \text{ или } \bar{\Delta} = \frac{\sum_{i=1}^n \Delta_i}{n}.$$

Среднегодовой абсолютный прирост затрат равен (см. табл. 8.8):

$$\bar{\Delta} = \frac{4,3}{5} = 0,86 \text{ млн. руб. или } \bar{\Delta} = \frac{107,3 - 103,0}{5} = 0,86 \text{ млн руб.}$$

Средний темп роста – обобщающая характеристика индивидуальных темпов роста ряда динамики и вычисляется по формуле средней геометрической

$$\bar{T}_p = \sqrt[m]{T_1 T_2 T_3 \dots T_m} \text{ или } \bar{T}_p = \sqrt[n-1]{\frac{y_n}{y_1}},$$

где  $m$  – число коэффициентов роста;

$n$  – число уровней ряда динамики

Среднегодовой темп роста можно рассчитать двумя способами:

*1-й способ*

$$\bar{T}_p = \sqrt[5]{1,023 \cdot 1,019 \cdot 1,031 \cdot 0,985 \cdot 0,984} = 1,008, \text{ или } 100,8 \%;$$

*2-й способ*

$$\bar{T}_p = \sqrt[5]{\frac{107,3}{103,0}} = 1,008.$$

Среднегодовой темп прироста определяется по формулам:

$$\bar{T}_{\text{пр}} = \bar{T}_p - 100 = 100,8\% - 100\% = 0,8 \% \text{ или } T_{\text{пр}} = T_p - 1.$$

Средний темп роста, определенный по второму способу, т. е. по начальному и конечному уровню ряда, используется только в случае монотонного поведения уровней. В случае сильной вариации уровней используется первый способ. При изменении тенденции поведения уровней ряда, например, до определенного периода уровни ряда увеличиваются, а после достижения некоторого максимума наблюдается тенденция к снижению уровней ряда, целесообразно эти интервалы периодов рассматривать с точки зрения средних характеристик отдельно.

## 8.5. Выявление основной тенденции ряда динамики

За колебаниями уровней в отдельные годы может не просматриваться закономерность изменения величины. Она может быть выявлена статистическими методами. Уровни ряда динамики формируются под влиянием множества длительно и кратковременно действующих факторов. Выявление основной закономерности (тенденции) изменения уровней ряда предполагает ее количественное измерение, в известной мере свободное от случайных воздействий. Тренд – это основная тенденция уровней ряда динамики, а ее выявление называется выравниванием временного ряда, методы выявления тренда называются методами выравнивания. Методы выравнивания рядов динамики заключаются в замене фактических уровней ряда динамики расчетными значениями, отражающими основную тенденцию развития процесса во времени. При изучении в рядах динамики основной тенденции развития (тренда) решаются две взаимосвязанные задачи:

- 1) выявление в изучаемом процессе наличия тренда с описанием его качественных особенностей;
- 2) измерение выявленного тренда, т. е. получение обобщающей количественной оценки основной тенденции развития.

Наиболее распространенными методами статистического изучения тренда являются:

- укрупнение интервалов;
- сглаживание скользящей средней;
- аналитическое выравнивание.

Существуют и другие методы, которые можно рассматривать как разновидность перечисленных выше, например:

- метод экспоненциального сглаживания, который аналогичен методу сглаживания скользящей средней, но вместо простых (арифметических) средних величин, используются экспоненциальные средние;
- экспоненциальное сглаживание временных рядов – модификация метода аналитического выравнивания, в котором более поздним наблюдениям придается больший вес, иными словами, веса точек ряда убывают (экспоненциально) по мере удаления в прошлое.

Одним из наиболее простых методов выявления основной тенденции является *метод укрупнения интервалов*. Этот способ основан на укрупнении периодов времени, к которым относятся уровни ряда. Метод применяется для выявления тренда в рядах динамики, в которых колебания уровней не позволяют определить основную тенденцию развития. Суть метода заключается в преобразовании первоначального ряда динамики в ряд с более продолжительными периодами. Например, ежемесячный объем продаж заменяется рядом объема продаж за квартал (рис. 8.2).



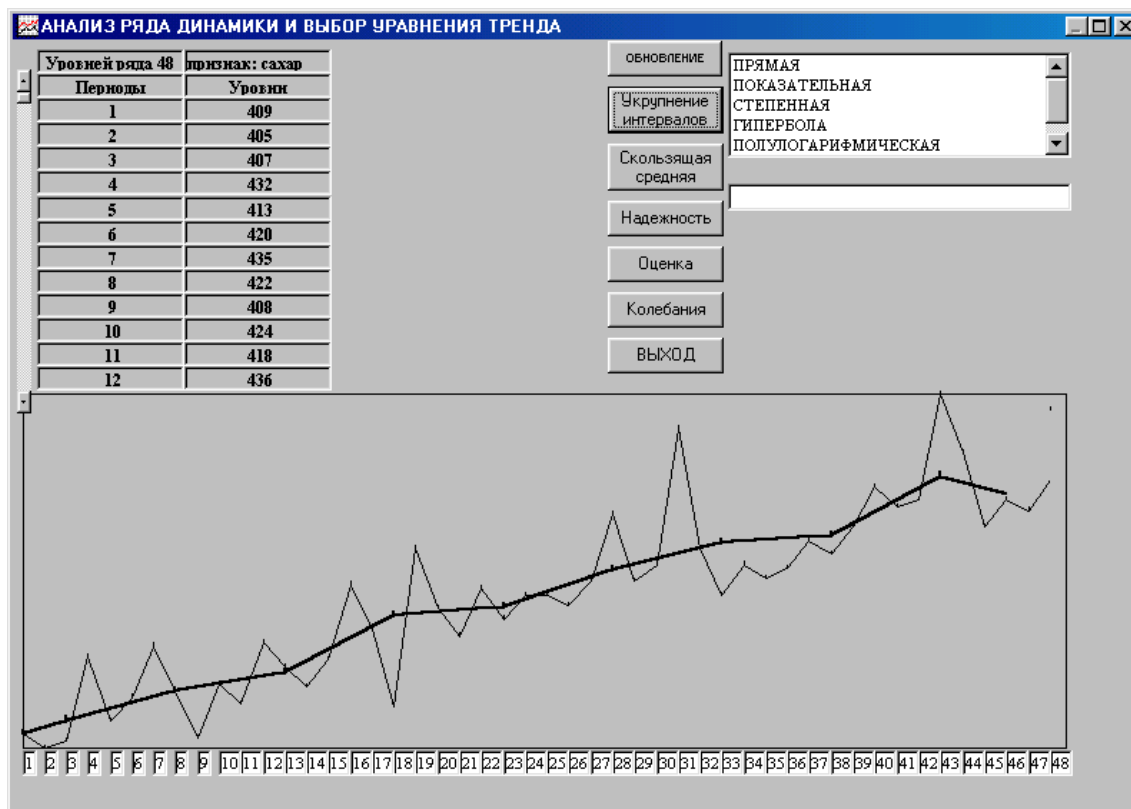


Рисунок 8.2 – Определение тенденции динамического ряда

Полностью уровни ряда динамики приведены на рис. 8.3.

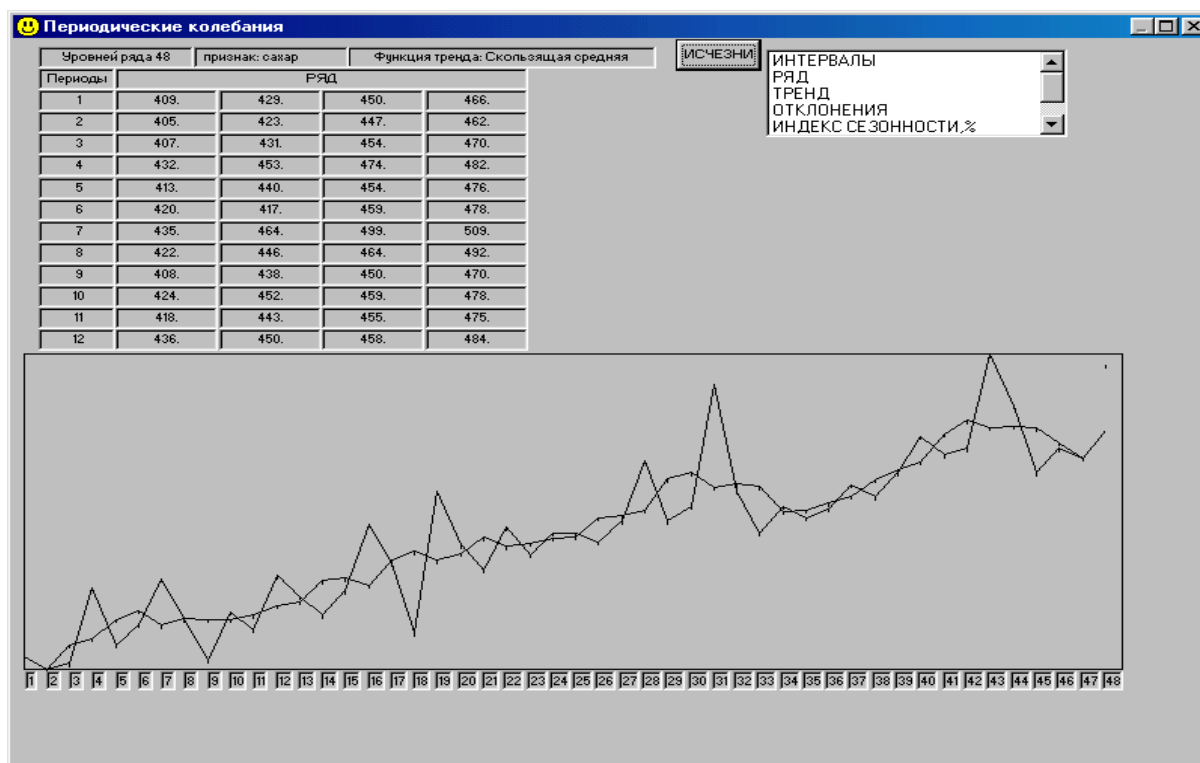


Рисунок 8.3 – Применение метода скользящей средней

Другой метод – *метод скользящей средней*, суть которого состоит в замене исходных уровней их средними арифметическими величинами за определенные периоды. Расчет средних величин ведется способом скольжения, т. е. постепенным исключением из принятого периода скольжения первого уровня и включением следующего. Допустим, что  $y_1, y_2, \dots, y_n$  исходный ряд динамики. Тогда для каждой  $m$  соседних уровней может быть рассчитано среднее арифметическое значение:

$$\bar{y} = \frac{1}{m} \sum_{i=k}^{i=k+m-1} y_i; k = 1, 2, \dots, n - m + 1,$$

где  $m$  – величина интервала сглаживания.

Например, если дан ряд ежегодных уровней:  $x_1, x_2, \dots, x_9$ , то трехлетняя скользящая средняя выглядит следующим образом:

$$\text{для первого интервала } \bar{y}_1 = \frac{y_1 + y_2 + y_3}{3};$$

$$\text{для второго интервала } \bar{y}_2 = \frac{y_2 + y_3 + y_4}{3};$$

$$\text{для третьего интервала } \bar{y}_3 = \frac{y_3 + y_4 + y_5}{3} \text{ и т. д.}$$

В результате сглаживания получается ряд динамики, количество уровней которого на два меньше, чем у исходного (теряются два крайних значения).

В случае если  $m$  нечетное число, то рассчитанным на основе приведенной формы значением заменяют уровень ряда, находящийся в середине интервала сглаживания. В основу метода положено определение по исходным уровням ряда теоретических уровней, в которых случайные колебания погашаются, а основная тенденция выражается в виде некоторой более плавной линии. Для выявления основной тенденции методом скользящей средней, прежде всего, устанавливается число уровней исходного ряда, которые будут заменяться их средней величиной. Обычно рекомендуется использовать нечетное число уровней исходного ряда динамики, заменяемых на среднюю величину.

*Пример.* На основе данных об объеме продаж изделий за 15 дней месяца (табл. 8.11) произвести сглаживание ряда методом пятичленной скользящей средней.

Таблица 8.11 – Сглаживание ряда методом пятичленной скользящей средней

| День<br>месяца | Объем<br>продаж,<br>шт. | Пяти-<br>дневные<br>скользящие<br>суммы | Пяти-<br>дневные<br>скользящие<br>средние $\bar{y}_i$ | День<br>месяца | Объем<br>продаж,<br>шт. | Пяти-<br>дневные<br>скользящие<br>суммы | Пятидневные<br>скользящие<br>средние $\bar{y}_i$ |
|----------------|-------------------------|-----------------------------------------|-------------------------------------------------------|----------------|-------------------------|-----------------------------------------|--------------------------------------------------|
| 1              | 30                      | –                                       | –                                                     | 9              | 33                      | 156                                     | 31,6                                             |
| 2              | 31                      | –                                       | –                                                     | 10             | 31                      | 157                                     | 31,8                                             |
| 3              | 31                      | –                                       | 30,8                                                  | 11             | 31                      | 158                                     | 32,0                                             |
| 4              | 32                      | –                                       | 30,8                                                  | 12             | 33                      | 159                                     | 32,0                                             |
| 5              | 30                      | 154                                     | 31,0                                                  | 13             | 32                      | 160                                     | 32,6                                             |
| 6              | 30                      | 154                                     | 31,0                                                  | 14             | 33                      | 160                                     | –                                                |
| 7              | 32                      | 155                                     | 31,2                                                  | 15             | 34                      | 163                                     | –                                                |
| 8              | 31                      | 155                                     | 31,4                                                  |                |                         |                                         |                                                  |

Взяв данные за первые 5 дней, исчислим пятидневные суммы, а затем среднюю величину.

$$\bar{y}_1 = \frac{30+31+31+32+30}{5} = \frac{154}{5} = 30,8;$$

$$\bar{y}_2 = \frac{31+31+32+30+30}{5} = \frac{154}{5} = 30,8 \quad \text{и т. д.}$$

Ниже приведен другой пример применения метода скользящей средней, в котором используются скользящие средние, также заменяющие пять членов ряда динамики (рис. 8.3). Интервал скользящего можно брать четный, но в этом случае определение скользящей средней по четному числу членов осложняется, так как средняя величина может быть отнесена только к середине между двумя периодами. Чтобы ликвидировать этот сдвиг, применяется центрирование, т. е. нахождение средней из средних величин для отнесения полученного уровня к определенному периоду. При четном интервале скользящего необходимо также находить скользящие суммы, скользящие средние по этим суммам и средние из средних. Если при сглаживании рядов динамики скользящая средняя составляется из нечетного числа уровней, то необходимость в центрировании отпадает. Интервал сглаживания должен быть достаточно большим для погашения случайных колебаний уровней ряда, чаще всего это 5–10 периодов. Недостатком метода является то, что число выравненных уровней меньше исходного их числа на величину  $(m - 1)$ . В случае использования взвешенных скользящих средних каждому из уровней при усреднении придается определенный вес. Способ задания весов зависит от целей анализа. Если степень важности информации возрастает в хронологическом порядке, то более поздним наблюдениям придается больший вес. При использовании метода экспоненциального сглаживания средние величины рассчитываются по следующей формуле:

$$\bar{y}_t = \sum_{i=k}^{k=n} \alpha_i y_{t-i} \quad \text{для любого } t,$$

где  $\sum_{i=1}^n \alpha_i = 1$ ,  $n$  – период скользящего, принимается нечетный.

Если период скользящего  $n = 3$ , то выравненный ряд начинается со второго уровня, как это показано на примере, представленном в табл. 8.12.

Таблица 8.12 – Выравнивание по методу экспоненциального сглаживания

| Период $t$ | Уровень ряда $y_t$ | Выравненный уровень ряда $\bar{y}_t = 0,4y_{t+1} + 0,3y_t + 0,3y_{t-1}$ |
|------------|--------------------|-------------------------------------------------------------------------|
| 1          | 30                 |                                                                         |
| 2          | 31                 | 30,7                                                                    |
| 3          | 31                 | 31,4                                                                    |
| 4          | 32                 | 30,9                                                                    |
| 5          | 30                 | 30,6                                                                    |
| 6          | 30                 | 30,8                                                                    |
| 7          | 32                 | 31,0                                                                    |
| 8          | 31                 | 32,1                                                                    |
| 9          | 33                 | 31,6                                                                    |

|    |    |      |
|----|----|------|
| 10 | 31 | 31,6 |
| 11 | 31 | 31,8 |
| 12 | 33 | 32,0 |
| 13 | 32 | 32,7 |
| 14 | 33 | 33,1 |
| 15 | 34 |      |

Применение методов укрупнения интервалов и скользящей средней позволяет выявить тренд. Но получить обобщенную статистическую оценку тренда при использовании этих методов невозможно. Для решения задачи измерения тренда используется *метод аналитического выравнивания*. При использовании этого метода уровни ряда динамики выражаются в виде функции времени  $\bar{y}_t = f(t)$ .

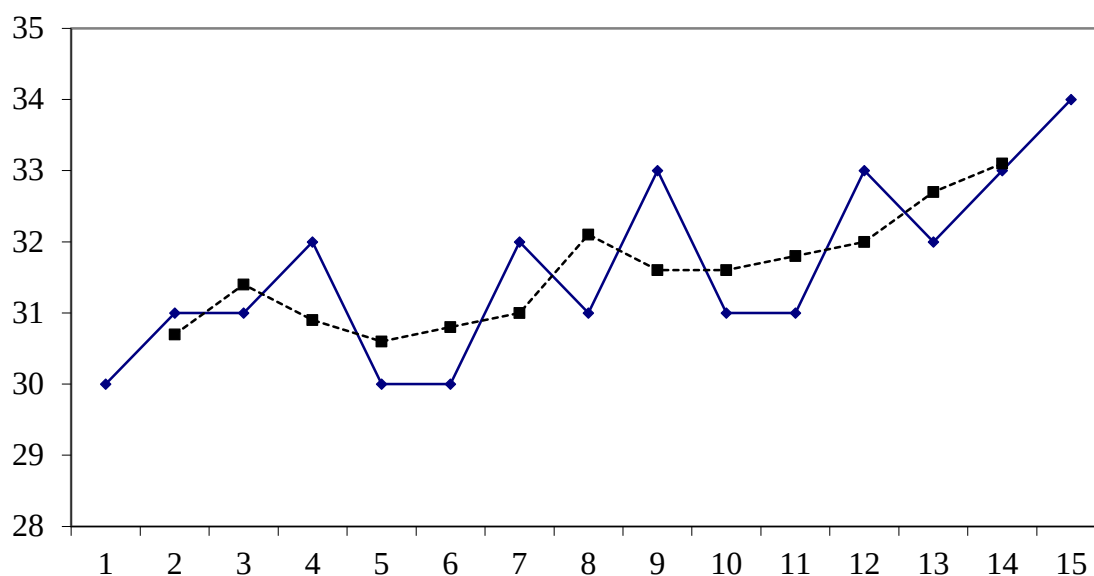


Рисунок 8.4 – Применение метода экспоненциального сглаживания:

—◆— уровень ряда;  
 ---■--- выровненный уровень ряда

Определение теоретических (расчетных) уровней  $\bar{y}_t = f(t)$  производится на основе так называемой адекватной математической функции, которая наилучшим образом отображает основную тенденцию ряда динамики. Выбор функции, характеризующей форму основной тенденции, может осуществляться на основе анализа по методу укрупнения интервалов.

Расчет параметров выбранной «адекватной» функции осуществляется методом наименьших квадратов, для которого формальным критерием выбора параметров является минимум остаточной дисперсии отклонения выровненных уровней ряда  $\bar{y}_t$  от полученных в результате наблюдений  $y_t$ :

$$\sum_{i=1}^n (y_i - \bar{y}_{t,i})^2 = \min.$$

Аналитическое выравнивание может быть осуществлено по любому

рациональному многочлену. Выбор функции производится на основе анализа характера закономерностей динамики данного ряда. В качестве уравнения тренда могут быть использованы разные функции. Часто рассматриваются полиномы разных степеней. Для примера рассмотрим полином первой и второй степени.

Наиболее часто рассматривается полином первой степени, т. е. уравнение прямой  $y_t = a_0 + a_1 t$ , где  $a_0, a_1$  – неизвестные параметры. Для определения параметров уравнения тренда ряда динамики с равноотстоящими по времени уровнями может быть использован метод конечных разностей. Для линейного уравнения схема использования метода конечных разностей базируется на том, что:

– первые разности являются постоянными, т. е.

$$\Delta'_t = y_t - y_{t-1} = \text{const для } \forall t;$$

– вторые разности равны нулю, т. е.  $\Delta''_t = \Delta'_t - \Delta'_{t-1} = 0$ .

Постоянство первых разностей и равенство нулю вторых разностей при вычислении по фактическим данным должно выполняться для всего ряда динамики. Для уравнения прямой схема конечных разностей представлена в табл. 8.13.

Таблица 8.13 – Схема конечных разностей для линейного уравнения

| $t$     | Уровни ряда        | Первые разности               | Вторые разности                          |
|---------|--------------------|-------------------------------|------------------------------------------|
| $t = 0$ | $y_0 = a_0$        |                               |                                          |
| $t = 1$ | $y_1 = a_0 + a_1$  | $\Delta'_1 = y_1 - y_0 = a_1$ |                                          |
| $t = 2$ | $y_2 = a_0 + 2a_1$ | $\Delta'_2 = y_2 - y_1 = a_1$ | $\Delta''_1 = \Delta'_2 - \Delta'_1 = 0$ |
| $t = 3$ | $y_3 = a_0 + 3a_1$ | $\Delta'_3 = y_3 - y_2 = a_1$ | $\Delta''_2 = \Delta'_3 - \Delta'_2 = 0$ |

Для полинома второй степени, т. е. параболы  $y_t = a_0 + a_1 t + a_2 t^2$ , постоянными являются вторые разности, а равными нулю – третьи (табл. 8.14).

Кроме определения параметров метод конечных разностей может быть использован для выбора формы уравнения тренда и оценки погрешности выбранной формы. Порядок равных конечных разностей определяет выбор степени полинома, который наилучшим образом может быть использован в качестве уравнения тренда.

Таблица 8.14 – Схема конечных разностей для уравнения параболы

| $t$     | Уровни ряда                | Первые разности          | Вторые разности                             | Третьи разности   |
|---------|----------------------------|--------------------------|---------------------------------------------|-------------------|
| $t = 0$ |                            |                          |                                             |                   |
| $t = 1$ | $y_1 = a_0 + a_1 + a_2$    | $\Delta'_1 = a_1 + a_2$  | $\Delta''_1 = \Delta'_2 - \Delta'_1 = 2a_2$ | $\Delta'''_1 = 0$ |
| $t = 2$ | $y_2 = a_0 + 2a_1 + 4a_2$  | $\Delta'_2 = a_1 + 3a_2$ | $\Delta''_2 = \Delta'_3 - \Delta'_2 = 2a_2$ | $\Delta'''_2 = 0$ |
| $t = 3$ | $y_3 = a_0 + 3a_1 + 9a_2$  | $\Delta'_3 = a_1 + 5a_2$ | $\Delta''_3 = \Delta'_4 - \Delta'_3 = 2a_2$ |                   |
| $t = 4$ | $y_4 = a_0 + 4a_1 + 16a_2$ | $\Delta'_4 = a_1 + 7a_2$ |                                             |                   |

Для выбора формы уравнения тренда используются и другие критерии:

- если темпы роста колеблются мало, то для уравнения тренда подходит показательная функция;
- если первые конечные разности имеют тенденцию уменьшаться с постоянным темпом, то целесообразно использовать модифицированную экспоненту;
- если первые конечные разности обратных значений средних уровней изменяются на один и тот же процент, то можно использовать логистическую функцию.

Таким образом, сделать вывод о наличии устойчивой тенденции развития и обосновать выбор формы уравнения тренда можно на основе анализа скорости роста, темпа роста, пунктов роста и других показателей ряда динамики. Если анализ показателей ряда динамики выявил, что поведение уровней ряда обусловлено изменением условий протекания рассматриваемого процесса, то этот процесс разбивается на ряд характерных в поведенческом отношении этапов. Каждый этап может иметь свое аналитическое представление. Анализ рядов динамики обычно преследует две цели:

- 1) выявление основных причин и степени влияния основных факторов, оказывающих существенное влияние на поведение экономического объекта;
- 2) прогнозирование поведения объекта.

Если выявленная тенденция поведения объекта может быть представлена аналитическим представлением за весь имеющийся в распоряжении исследователя период, то возможно воспользоваться уравнением тренда для прогнозирования. В том случае, если приходится процесс представить совокупностью этапов, характерных в поведении объекта, то для прогнозирования используется только одно уравнение тренда (по одному из этапов), представляющее, по предположению, то же повторяющееся поведение объекта в будущем.

## 8.6. Выбор формы уравнения тренда

При выборе формы уравнения тренда следует учитывать объем имеющийся информации. Чем больше параметров в уравнении тренда, тем больше требуется уровней ряда динамики для одной и той же степени надежности. Выбор формы уравнения тренда может осуществляться и на основе формального критерия. В качестве такого критерия часто используется сумма квадратов отклонений, полученных при расчете уравнения тренда, от фактических уровней ряда. В целях удобства выбора различают следующие типичные случаи социально-экономических процессов:

- 1) равномерное развитие;
- 2) равноускоренное развитие;
- 3) развитие с переменным ускорением;
- 4) развитие со стабильными темпами роста;
- 5) развитие с замедлением роста в конце периода.

Равномерное развитие характеризуется практически постоянными абсолютными приростами:

$$\Delta = y_i - y_{i-1} \cong \text{const для } \forall i.$$

Основная тенденция развития в рядах динамики с постоянными абсолютными приростами отображается линейным уравнением

$$y_t = a_0 + a_1 t,$$

где  $a_0, a_1$  – неизвестные параметры, для определения которых составляется следующая система нормальных уравнений:

$$\begin{cases} na_0 + a_1 \sum_i t_i = \sum_i y_i; \\ a_0 \sum_i t_i + a_1 \sum_i t_i^2 = \sum_i t_i y_i \end{cases}$$

*Пример.* Определить уравнение тренда по данным, представленным в табл. 8.15.

Таблица 8.15 – Динамика затрат на производство с 1999 по 2003 г.

| Год   | Объем затрат, тыс. руб. | $t$ | $t^2$ | $ty$  | $\bar{y}_t = a_0 + a_1 T$ | $y - \bar{y}_t$ | $(y - \bar{y}_t)^2$ |
|-------|-------------------------|-----|-------|-------|---------------------------|-----------------|---------------------|
| 1999  | 15,5                    | -2  | 4     | -31   | 15,18                     | 0,3             | 0,09                |
| 2000  | 15,1                    | -1  | 1     | -15,1 | 15,31                     | -0,21           | 0,044               |
| 2001  | 15,2                    | 0   | 0     | 0     | 15,44                     | -0,2            | 0,04                |
| 2002  | 15,4                    | 1   | 1     | 15,4  | 15,57                     | -0,2            | 0,04                |
| 2003  | 16,0                    | 2   | 4     | 32    | 15,7                      | -0,3            | 0,09                |
| Итого | 77,2                    | –   | 10    | 1,3   | –                         | –               | 0,3                 |

Для выравнивания ряда динамики используется линейное уравнение:

$$y_t = a_0 + a_1 t.$$

Для расчета параметров  $a_0$  и  $a_1$  по способу наименьших квадратов необходимо решить систему нормальных уравнений:

$$\begin{cases} na_0 + a_1 \sum_{t=1}^n t = \sum_{t=1}^n y_t; \\ a_0 \sum_{t=1}^n t + a_1 \sum_{t=1}^n t^2 = \sum_{t=1}^n ty_t, \end{cases}$$

где  $y_t$  – исходные уровни ряда динамики;

$n$  – число членов ряда;

$t$  – показатель времени, который обозначается порядковыми номерами, как это представлено в табл. 8.16.

Таблица 8.16 – Представление показателя времени

| Год | 1989 | 1990 | 1991 | 1992 | 1993 |
|-----|------|------|------|------|------|
| $t$ | 1    | 2    | 3    | 4    | 5    |

Решение системы дает следующее выражение для  $a_0$  и  $a_1$ :

$$\begin{cases} a_0 = \frac{\sum_t t^2 \sum_t y_t - \sum_t t \sum_t ty_t}{n \sum_t t^2 - \sum_t t \sum_t t} \\ a_1 = \frac{n \sum_t ty_t - \sum_t t \sum_t y_t}{n \sum_t t^2 - \sum_t t \sum_t t} \end{cases}.$$

В рядах динамики техника расчета параметров уравнения может быть упрощена. Для этой цели показателям времени придают такие значения, чтобы их сумма была равна нулю, т. е.  $\sum t = 0$ . При этом уравнения примут вид

$$\begin{aligned} a_0 n &= \sum_t y_t; \\ a_1 \sum_t t^2 &= \sum_t ty_t; \end{aligned}$$

откуда  $a_0 = \frac{\sum_t y_t}{n}$  представляет средний уровень динамики ряда ( $\bar{y}$ ); параметр  $a_1 = \frac{\sum_t ty_t}{\sum_t t^2}$  является коэффициентом регрессии, определяющим направление и темпы развития. Если  $a_1 > 0$ , то уровни выравненного ряда динамики равномерно возрастают, а при  $a_1 < 0$  происходит равномерное снижение.

По итоговым данным нашего примера (табл. 8.15) получим:

$$a_0 = 77,2:5 = 15,44;$$

$$a_1 = 1,3:10 = 0,13;$$

уравнение основной тенденции за 1989–1993 годы.

$$\bar{y}_t = 15,44 + 0,13t.$$

Подставляя в уравнение принятые значения  $t$ , получим:

$$1989 \text{ год} - \bar{y}_t = 15,44 + 0,13(-2) = 15,18;$$

$$1990 \text{ год} - \bar{y}_t = 15,44 + 0,13(-1) = 15,31 \text{ и т. д.}$$

Целесообразно полученные результаты изобразить графически (рис. 8.5).

Если число уровней в ряду динамики – четное, то условное обозначение показателя времени может принимать вид, представленный в табл. 8.17.



Таблица 8.17 – Представление показателя времени

| Год | 1978 | 1979 | 1980 | 1981 | 1982 | 1983 |
|-----|------|------|------|------|------|------|
| $t$ | -5   | -3   | -1   | 1    | 3    | 5    |

Этот прием позволяет вести расчет по полугодиям.

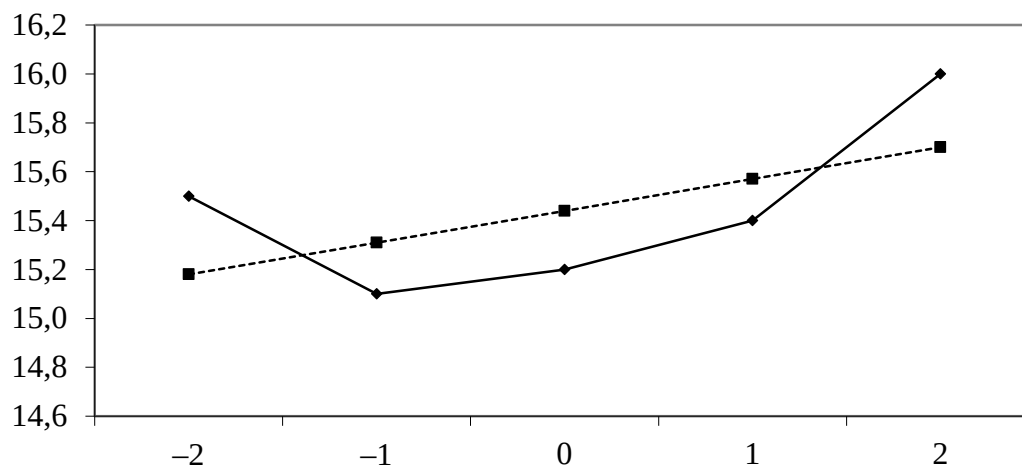


Рисунок 8.5 – Исходный ряд и уравнение тренда:

—◆— исходный ряд;  
 ---■--- уравнение тренда

Величину  $\sum t^2$  определяют по формулам:

при нечетном числе уровней

$$\sum t^2 = \frac{(n-1)n(n+1)}{12};$$

при четном числе уровней

$$\sum t^2 = \frac{(n-1)n(n+1)}{3}.$$

*Равноускоренному (равнозамедленному)* типу процессов свойственно постоянное во времени увеличение (замедление) развития. Уровни таких рядов изменяются с постоянными темпами прироста:

$$T_{\text{пр}} = \frac{\Delta}{y_{i-1}} 100.$$

Основная тенденция развития в рядах динамики со стабильными темпами прироста отображается функцией параболы второй степени:

$$\bar{y}_t = a_0 + a_1 t + a_2 t^2.$$

В этом уравнении значения параметров  $a_0$  и  $a_1$  идентичны параметрам, используемым в формуле линейного тренда. Параметр  $a_2$  характеризует постоянное изменение интенсивности развития в единицу времени. При  $a_2 > 0$  происходит ускорение развития, а при  $a_2 < 0$ , соответственно, замедление. Параметр  $a_1$  может быть с любым знаком. Для определения параметров составляется следующая система нормальных уравнений:

$$\begin{cases} na_0 + a_1 \sum_t t + a_2 \sum_t t^2 = \sum_t y_t; \\ a_0 \sum_t t + a_1 \sum_t t^2 + a_2 \sum_t t^3 = \sum_t ty_t; \\ a_0 \sum_t t^2 + a_1 \sum_t t^3 + a_2 \sum_t t^4 = \sum_t y_t t^2 \end{cases}$$

Развитие *с переменным ускорением (замедлением)* выражается уравнением параболы третьей степени:

$$\bar{y}_t = a_0 + a_1 t + a_2 t^2 + a_3 t^3.$$

В этом уравнении параметр  $a_3$  показывает изменение ускорения. При  $a_3 > 0$  ускорение увеличивается, а при  $a_3 < 0$  ускорение замедляется. Для определения параметров составляется следующая система нормальных уравнений:

$$\begin{cases} na_0 + a_1 \sum_t t + a_2 \sum_t t^2 + a_3 \sum_t t^3 = \sum_t y_t; \\ a_0 \sum_t t + a_1 \sum_t t^2 + a_2 \sum_t t^3 + a_3 \sum_t t^4 = \sum_t ty_t; \\ a_0 \sum_t t^2 + a_1 \sum_t t^3 + a_2 \sum_t t^4 + a_3 \sum_t t^5 = \sum_t y_t t^2; \\ a_0 \sum_t t^3 + a_1 \sum_t t^4 + a_2 \sum_t t^5 + a_3 \sum_t t^6 = \sum_t y_t t^3. \end{cases}$$

$$\begin{cases} na_0 + a_1 \sum_t t + a_2 \sum_t t^2 + a_3 \sum_t t^3 = \sum_t y_t; \\ a_0 \sum_t t + a_1 \sum_t t^2 + a_2 \sum_t t^3 + a_3 \sum_t t^4 = \sum_t ty_t; \\ a_0 \sum_t t^2 + a_1 \sum_t t^3 + a_2 \sum_t t^4 + a_3 \sum_t t^5 = \sum_t y_t t^2; \\ a_0 \sum_t t^3 + a_1 \sum_t t^4 + a_2 \sum_t t^5 + a_3 \sum_t t^6 = \sum_t y_t t^3 \end{cases}$$

Развитие *со стабильными темпами роста* (развитие по экспоненте) соответствует практически постоянному показателю темпов роста:

$$T_p = \frac{y_i}{y_{i-1}} 100.$$

В этом случае основная тенденция развития отображается показательной функцией  $\bar{y}_t = a_0 a_1^t$ , в которой для определения параметров  $a_0, a_1$  сначала производится логарифмирование  $\lg y_t = \lg a_0 + t \lg a_1$ , а затем составляется следующая система нормальных уравнений:

$$\begin{cases} n \lg a_0 + \lg a_1 \sum_t t = \sum_t \lg y_t; \\ \lg a_0 \sum_t t + \lg a_1 \sum_t t^2 = \sum_t t \lg y_t. \end{cases}$$

В этих уравнениях параметр  $a_1$  – темп роста (снижения) изучаемого процесса в единицу времени, т. е. интенсивность развития. Встречаются попытки использовать модифицированную экспоненту:

$$\bar{y}_t = a_0 + a_1 a_2^t.$$

Развитие с замедлением роста в конце периода характеризуется показателем цепного абсолютного прироста, величина которого в конце периода имеет тенденцию к уменьшению:

$$\Delta = y_i - y_{i-1} \xrightarrow{i \rightarrow n} 0.$$

Основная тенденция развития в таких рядах динамики отображается полулогарифмической функцией:

$$\bar{y}_t = a_0 + a_1 \lg t.$$

Для получения неизвестных параметров  $a_0, a_1$  составляется следующая система нормальных уравнений:

$$\begin{cases} na_0 + a_1 \sum_t \lg t = \sum_t y_t; \\ a_0 \sum_t \lg t + a_1 \sum_t (\lg t)^2 = \sum_t y_t \lg t. \end{cases}$$

В действительности порой невозможно однозначно решить, к какому типу развития относится анализируемый процесс, а использованные для классификации значения показателей ряда динамики имеют такой диапазон изменения, который встречается реже, чем этого бы хотелось. Совокупное действие постоянных, периодических и случайных факторов обуславливает такое изменение уровней ряда динамики, которое часто не согласуется с приведенными выше типовыми случаями. Поэтому выбору уравнения тренда должен предшествовать качественный анализ процесса, который должен быть представлен графически и на этой основе может быть выдвинута гипотеза. Но даже в этой ситуации невозможно однозначно выбрать тип аналитического представления основной тенденции. Поэтому проблема определения наилучшего уравнения тренда может решаться прямым перебором из заданного числа типов уравнений, которые отличаются значительным разнообразием. Кроме типичных случаев для получения уравнения тренда используются и другие функции:

степенная функция:

$$\bar{y}_t = a_0 t^{a_1},$$

для определения параметров которой сначала производится логарифмирование  $\lg y_t = \lg a_0 + a_1 \lg t$ , а затем составляется следующая система нормальных уравнений:

$$\begin{cases} n \lg a_0 + a_1 \sum_t \lg t = \sum_t \lg y_t; \\ \lg a_0 \sum_t \lg t + a_1 \sum_t (\lg t)^2 = \sum_t \lg t \lg y_t. \end{cases}$$

При  $a_1 = 1$  уравнение тренда является линейным, при  $a_1 = 2$  –

параболическим. Степенная форма гибкая, пригодная для отображения изменений с малой мерой пропорциональности. «Жестким» условием является обязательное прохождение через начало координат: при  $t = 0 \rightarrow \bar{y}_t = 0$ . При усложнении уравнения  $\bar{y}_t = a_0 + a_1 t^{a_2}$  могут возникнуть некоторые сложности;

– гипербола:

$$y_t = a_0 + a_1 \frac{1}{t},$$

для определения параметров которой составляется следующая система нормальных уравнений:

$$\begin{cases} na_0 + a_1 \sum_t \left( \frac{1}{t} \right) = \sum_t y_t; \\ a_0 \sum_t \left( \frac{1}{t} \right) + a_1 \sum_t \left( \frac{1}{t} \right)^2 = \sum_t \left( \frac{y_t}{t} \right). \end{cases}$$

Если  $a_1 > 0$ , гиперболический тренд выражает тенденцию замедляющегося снижения уровня, стремящегося к пределу  $a_0$ . Эта форма тренда подходит для отображения процессов, ограниченных предельным значением уровня. Существует несколько разновидностей гиперболы, обладающих разными свойствами:

- равносторонняя гипербола, асимптотическая к линии, параллельной оси  $t$ :

$$y = \frac{1}{a_0 + a_1 t};$$

- равносторонняя гипербола, асимптотическая к линии, параллельной оси  $y$ :

$$y = a_0 + a_1 \left( \frac{1}{t} \right);$$

- равносторонняя гипербола, асимптотическая к линиям, параллельным обоим осям:

$$\frac{1}{y} = a_0 + a_1 \left( \frac{1}{t} \right);$$

– логистическая функция:

$$\bar{y} = \frac{y_{\max} - y_{\min}}{1 + e^{a_0 + a_1 t}} + y_{\min},$$

где  $y_{\max}, y_{\min}$  – максимальное и минимальное значения уровней ряда динамики;  $a_0, a_1$  – неизвестные параметры уравнения тренда.

Данная функция имеет две точки перегиба: от ускоряющегося роста к равномерному (вогнутость) и от равномерного роста к замедляющемуся (рис. 8.6, табл. 8.18).

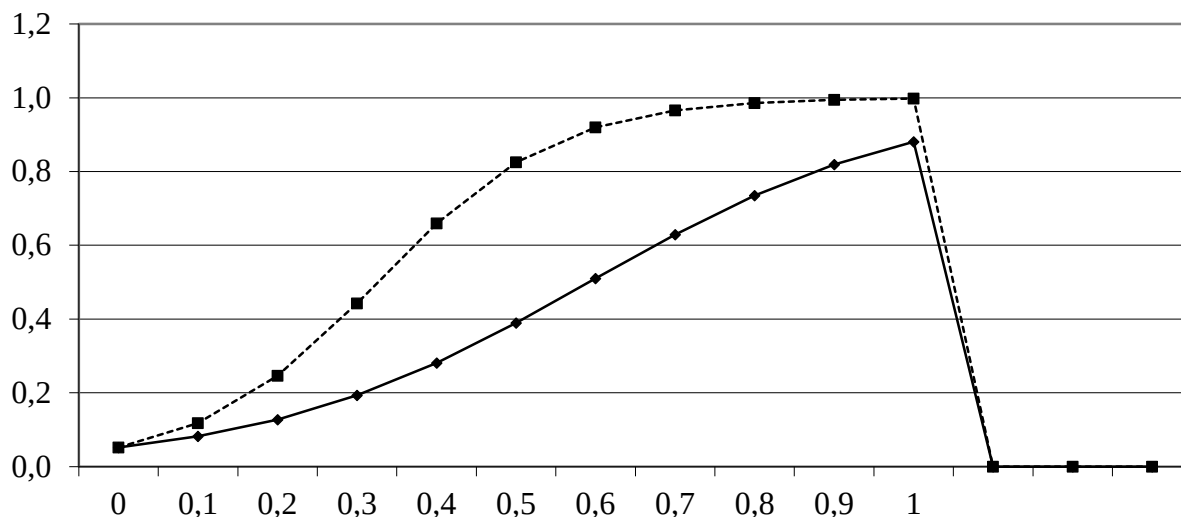


Рисунок 8.6 – Графики логистических кривых:

—◆— 1-я логиста  
 ----■---- 2-я логиста

Данная функция может отображать процесс развития в течение длительного периода, проходя все фазы насыщения рынка товарами: сначала медленный, но все ускоряющийся рост доли потребителей, затем равномерный (от 30–40 до 70–80 %), а в заключение замедляющийся по мере приближения к 100 %.

Таблица 8.18 – Данные по графикам рис. 8.6

| $t$ | 1-я логиста                                  | 2-я логиста                                  |
|-----|----------------------------------------------|----------------------------------------------|
|     | $\bar{y}_t = \frac{1}{1 + \exp(2,9 - 4,9t)}$ | $\bar{y}_t = \frac{1}{1 + \exp(2,9 - 8,9t)}$ |
| 0,0 | 0,052154                                     | 0,052154                                     |
| 0,1 | 0,082413                                     | 0,118157                                     |
| 0,2 | 0,127862                                     | 0,246011                                     |
| 0,3 | 0,193099                                     | 0,442752                                     |
| 0,4 | 0,280900                                     | 0,659260                                     |
| 0,5 | 0,389361                                     | 0,824914                                     |
| 0,6 | 0,509999                                     | 0,919827                                     |
| 0,7 | 0,629483                                     | 0,965444                                     |
| 0,8 | 0,734973                                     | 0,985514                                     |
| 0,9 | 0,819061                                     | 0,994000                                     |
| 1,0 | 0,880797                                     | 0,997527                                     |

Если  $y_{\min} = 0$ ,  $y_{\max} = 100\%$ , уравнение упрощается до формы:

$$\bar{y}_t = \frac{1}{1 + \exp(a_0 + a_1 t)};$$

- кривая Гомперца:

$$\bar{y}_t = a_0 a_1^{a_3^t};$$

- логарифмическая парабола:

$$\bar{y}_t = a_0 a_1^t a_2 t.$$

При выборе формы уравнения тренда существенное значение имеет объем информации. Чем больше параметров содержит уравнение тренда, тем больше должно быть наблюдений при одной и той же степени надежности оценивания.

## 8.7. Анализ периодических колебаний

Под воздействием систематических и случайных факторов уровни ряда динамики подвергаются изменениям, под воздействием которых формируются основная тенденция и периодические (сезонные) колебания, которые, в свою очередь, могут сопровождаться случайными колебаниями. Первоочередной задачей при изучении рядов динамики является выделение основной тенденции (тренда). Следующим этапом при анализе колебаний рядов динамики является задача изучения сезонных (периодических) колебаний для выявления их влияния на динамику изменения уровней ряда и оценки случайной составляющей. В широком понимании к сезонным, вернее, периодическим колебаниям относят все явления, которые обнаруживают в своем развитии отчетливо выраженную закономерность устойчиво повторяющихся изменений уровней ряда. Для их выявления анализируют месячные, квартальные, годовые уровни ряда динамики за год или несколько лет. Основными задачами исследования периодичности изменения уровней ряда являются:

- определение наличия и численного проявления сезонных колебаний;
- выявление и оценка факторов, вызывающих сезонные колебания;
- определение цикла сезонных колебаний;
- оценка последствий сезонных колебаний;
- моделирование сезонных колебаний при прогнозировании.

Для измерения сезонных колебаний используются методы абсолютных разностей, относительных разностей, индексов сезонности. В таблице 8.19 приводится классификация наиболее распространенных методов измерения сезонных колебаний.

Выбор того или иного метода зависит от характера основной тенденции ряда динамики. При использовании метода абсолютных разностей оперируют непосредственно размерами этих разностей, а при использовании метода относительных разностей используют отношение абсолютных размеров разностей к выравненному уровню. При выявлении основной тенденции используют либо метод скользящей средней (рис. 8.7), либо аналитическое выравнивание (рис. 8.8).

Как показывает практика, выбор между методами аналитического выравнивания и скользящей средней производится после анализа характера колебаний уровней ряда. Причем в одном случае целесообразно применять

метод скользящей средней, а в другом – аналитического выравнивания. И в том, и в другом случае необходимо учитывать величину тренда.

Таблица 8.19 – Классификация методов измерения сезонных колебаний

| Методы измерения сезонных колебаний, основанные на применении | Наименование методов определения сезонных колебаний                                                                                                           |
|---------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1. Средней арифметической                                     | Метод абсолютных разностей<br>Метод отношений средних «помесичных» к средней за весь период<br>Метод отношений «помесичных» уровней к средней данного периода |
| 2. Относительных величин                                      | Метод относительных величин<br>Метод относительных величин на основе медианы<br>Метод У. Персона (цепной метод)                                               |
| 3. Механического выравнивания                                 | Метод скользящих средних<br>Метод скользящих сумм и скользящих средних                                                                                        |
| 4. Аналитического выравнивания                                | Выравнивания по прямой<br>Выравнивание по параболе и экспоненте<br>Выравнивание по ряду Фурье                                                                 |

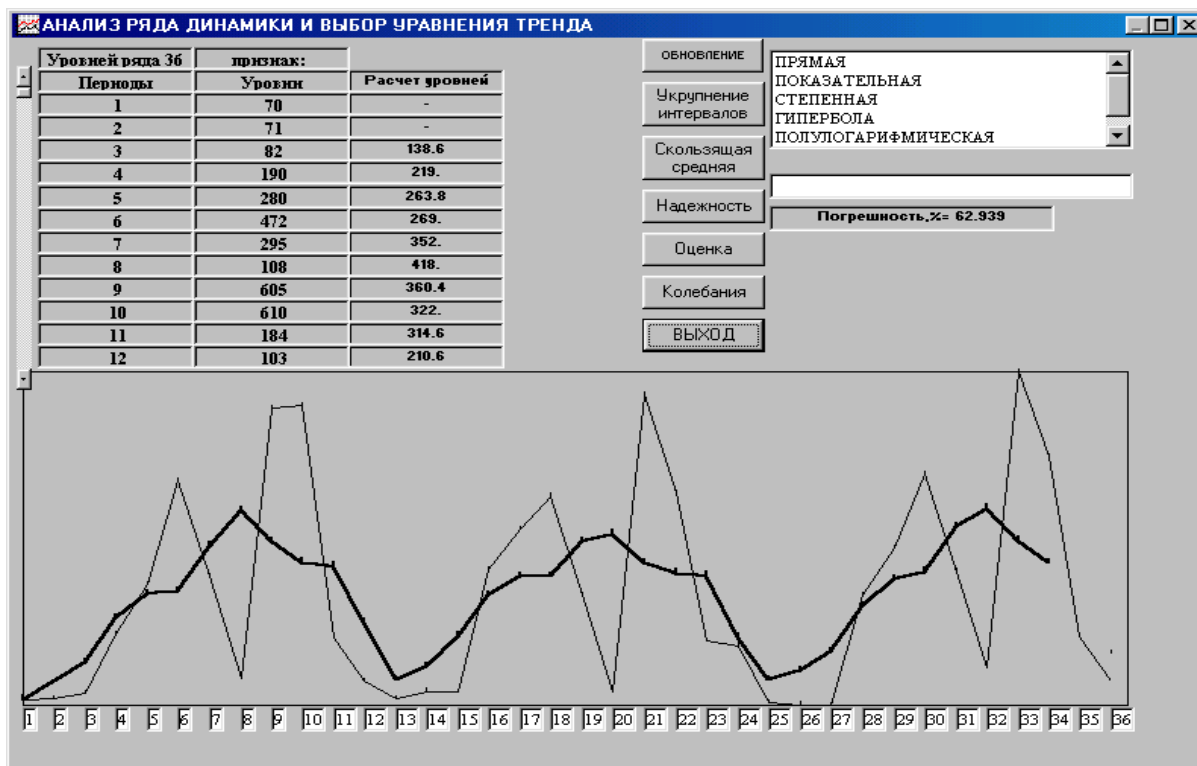


Рисунок 8.7 – Основная тенденция потребления картофеля, полученная методом скользящей средней

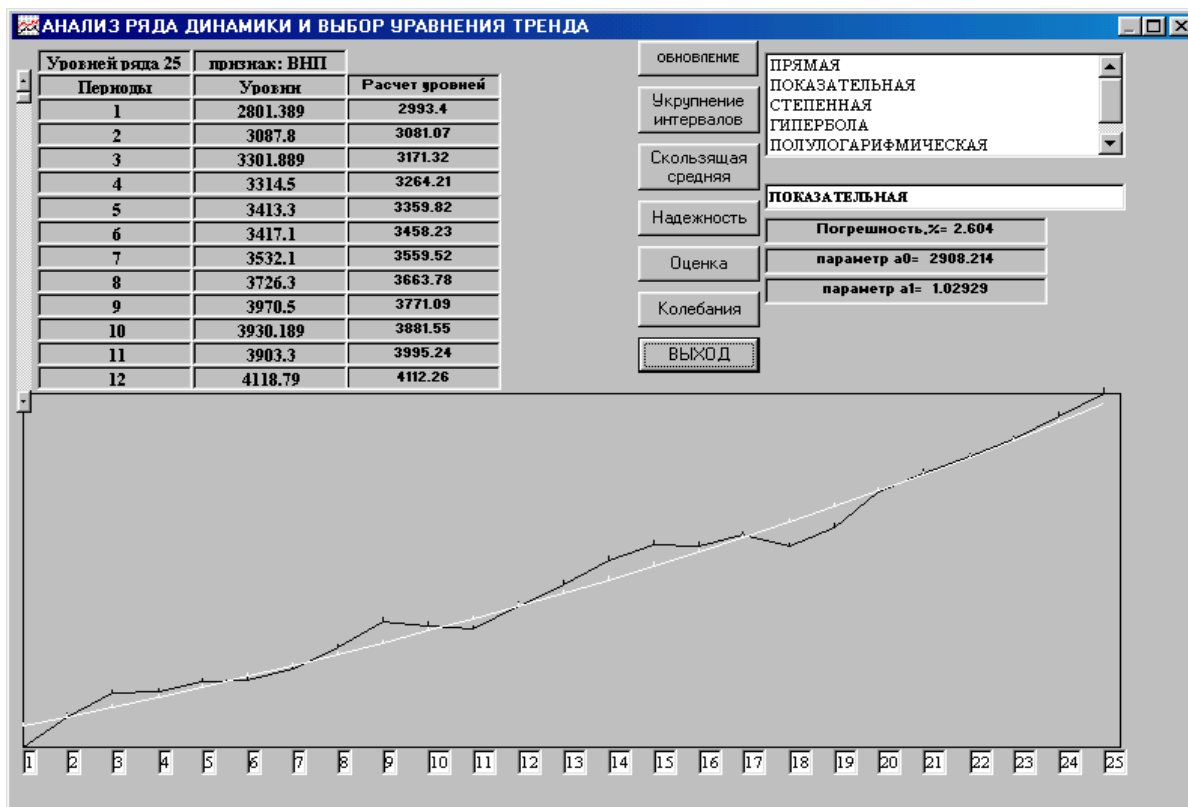


Рисунок 8.8 – Основная тенденция валового внутреннего продукта США, полученная методом аналитического выравнивания (показательная функция)

*Пример.* По данным табл. 8.20 провести выравнивание ряда динамики методами абсолютных и относительных разностей.

Таблица 8.20 – Среднесписочная численность работников по месяцам

| Месяц | Численность работников | Месяц | Численность работников |
|-------|------------------------|-------|------------------------|
| 1     | 620                    | 7     | 990                    |
| 2     | 640                    | 8     | 980                    |
| 3     | 710                    | 9     | 970                    |
| 4     | 730                    | 10    | 870                    |
| 5     | 880                    | 11    | 740                    |
| 6     | 920                    | 12    | 630                    |

В приведенном примере средний уровень ряда:

$$\bar{y} = \frac{9680}{12} \approx 807 \text{ чел.}$$

При использовании метода абсолютных разностей определяется абсолютное отклонение уровней ряда по периодам (например, по месяцам) от среднего уровня ряда. Для января эта величина составит

$$\bar{y}_t - \bar{y} = 620 - 807 = -187 \text{ чел.}$$

При использовании метода относительных разностей рассчитывают



отношения абсолютных отклонений и среднего уровня ряда:

$$\frac{\bar{y}_t - \bar{y}}{\bar{y}} = \frac{-187}{807} \cdot 100 = -23,17\%.$$

Вместо относительных разностей часто используют индекс сезонности, например, индекс сезонности для января составит:

$$I_s = \frac{\bar{y}_t}{\bar{y}} \cdot 100\% = \frac{620}{807} \cdot 100 = 76,8 \%.$$

Выполненный таким образом анализ дает представление о периодических (сезонных) колебаниях уровней ряда (табл. 8.21, рис. 8.9, 8.10).

Таблица 8.21 – Расчет периодических колебаний разными методами

| Период<br>(месяц) | Уровень<br>ряда, чел. | Метод абсолютных<br>разностей, чел. | Метод<br>относительных<br>разностей, % | Индекс<br>сезонности, % |
|-------------------|-----------------------|-------------------------------------|----------------------------------------|-------------------------|
| 1                 | 620                   | –187                                | –23,172200                             | 76,82776                |
| 2                 | 640                   | –167                                | –20,693900                             | 79,30607                |
| 3                 | 710                   | –97                                 | –12,019800                             | 87,98017                |
| 4                 | 730                   | –77                                 | –9,541510                              | 90,45849                |
| 5                 | 880                   | 73                                  | 9,045849                               | 109,04580               |
| 6                 | 920                   | 113                                 | 14,002480                              | 114,00250               |
| 7                 | 990                   | 183                                 | 22,676580                              | 122,67660               |
| 8                 | 980                   | 173                                 | 21,437420                              | 121,43740               |
| 9                 | 970                   | 163                                 | 20,198270                              | 120,19830               |
| 10                | 870                   | 63                                  | 7,806691                               | 107,80670               |
| 11                | 740                   | –67                                 | –8,302350                              | 91,69765                |
| 12                | 603                   | –177                                | –21,933100                             | 78,06691                |

Для ряда внутригодовой динамики, где основная тенденция роста незначительна, изучение сезонности основано на *методе постоянной средней*, полученной из всех фактических уровней.

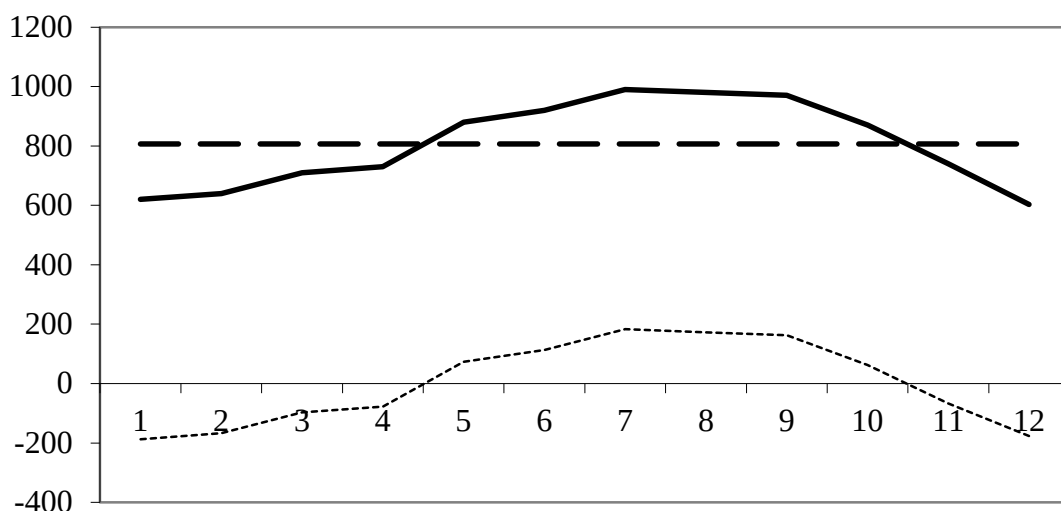


Рисунок 8.9 – Колебания уровней ряда по методу абсолютных разностей:  
 — уровни ряда; — — — средний уровень; - - - - абсолютные разности

Самый простой способ заключается в следующем: для каждого года рассчитывается средний уровень, а затем с ним сопоставляется уровень каждого месяца в процентах. Это отношение именуется индексом сезонности

$$I_s = \frac{y_t}{\bar{y}} \cdot 100\%.$$

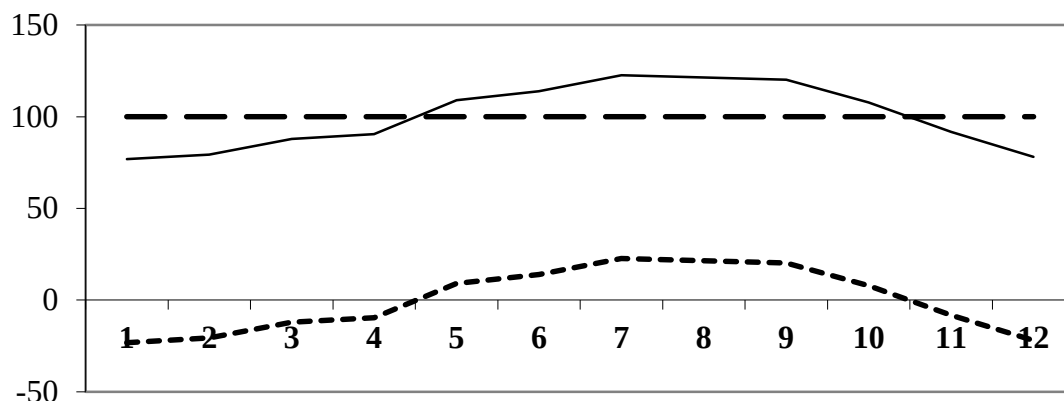


Рисунок 8.10 – Периодические колебания по методу относительных разностей и сезонных колебаний (%):  
 - - - - относительная разность;  
 — индекс сезонности, %;  
 — — — средний уровень

Однако данные одного года ненадежны. Поэтому для выявления сезонности пользуются данными за ряд лет (в основном за три года). Тогда для каждого месяца рассчитывается средняя величина уровня за три года, затем среднемесячный уровень для всего ряда, а в заключение определяется процентное отношение средних для каждого ряда к общему среднемесячному уровню ряда, т. е.:

$$I_s = \frac{y_t}{\bar{y}} \cdot 100\%,$$

где  $\bar{y}$  – средняя для каждого месяца за три года.

При наличии ярко выраженной тенденции к увеличению или уменьшению уровней из года в год применяются другие способы измерения сезонных колебаний. В частности, индексы сезонности определяются на основе методов, позволяющих исключить влияние тенденции роста (падения). При этом для определения основной тенденции может быть использован как метод аналитического выравнивания, так и метод скользящей средней.

При использовании способа *аналитического выравнивания* выполняются следующие этапы:

- определение основной тенденции (уравнение тренда);
- по соответствующему представлению тренда определяются для каждого периода (месяца, квартала) выравненные уровни на момент времени  $t$ ;
- берутся отношения фактических данных по периоду (месяцу, кварталу) ( $y_i$ ) к соответствующим выравненным данным ( $\bar{y}_t$ ) в процентах:

$$\frac{y_i}{\bar{y}_t} \cdot 100 = U_t;$$

- определяется средняя величина из этих отношений для одноименных периодов (месяцев, кварталов):

$$\bar{U}_i = \frac{U_1 + U_2 + \dots + U_n}{n},$$

где  $n$  – число одноименных месяцев;

- из полученных 12 относительных величин по периоду (по месяцу)  $\bar{U}_i$  вычисляется общий среднемесячный уровень  $\bar{U}_t$ ;
- определяются индексы сезонности:

$$I_s = \frac{\bar{U}_i}{\bar{U}_t} \cdot 100 \text{ или } I_s = \frac{\sum \frac{y_i}{\bar{y}_{ti}}}{n},$$

где  $y_i$  – исходные уровни ряда,  
 $\bar{y}_{ti}$  – выравненные уровни ряда;  
 $n$  – число годовых периодов.

Та же методика расчета индексов сезонности и при использовании метода скользящей средней. Индекс сезонности по способу скользящей средней:

$$I_s = \frac{\sum \frac{y_i}{\bar{y}_{ci}}}{n},$$

где  $y_i$  – исходные уровни ряда;  
 $\bar{y}_{ci}$  – выравненные уровни ряда;  
 $n$  – число годовых периодов.

Ниже на рисунках представлены результаты статистического исследования периодичности колебаний уровней ряда с использованием индекса сезонности и разных методов определения основной тенденции (рис. 8.11, 8.12).

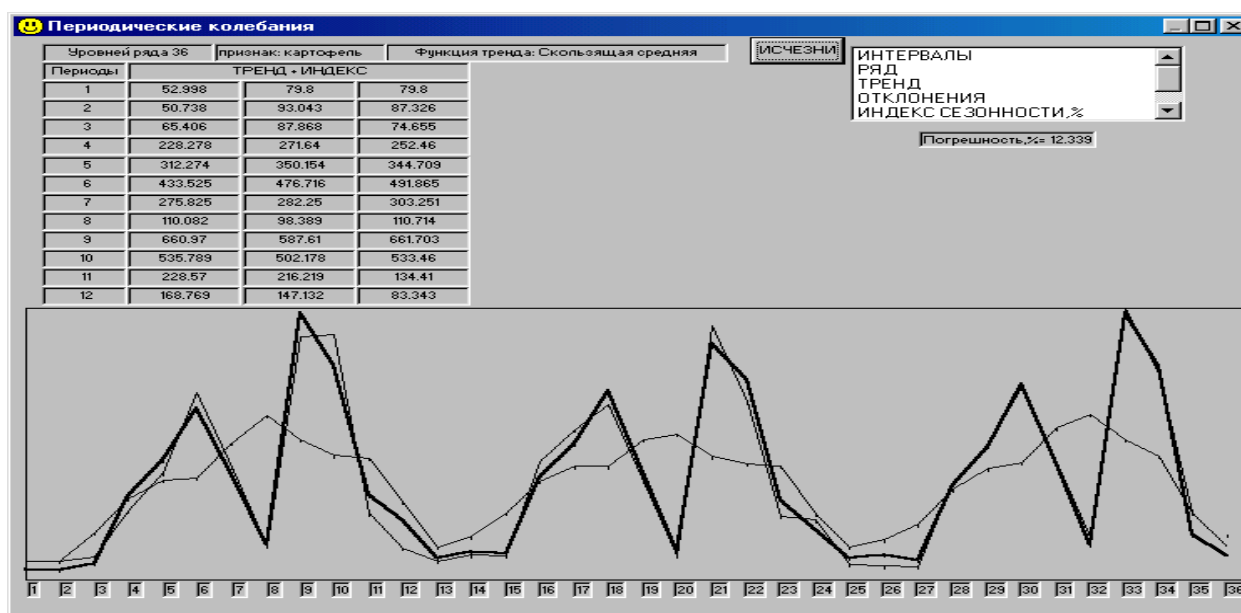


Рисунок 8.11 – Определение тенденции с использованием индекса сезонности и метода скользящей средней

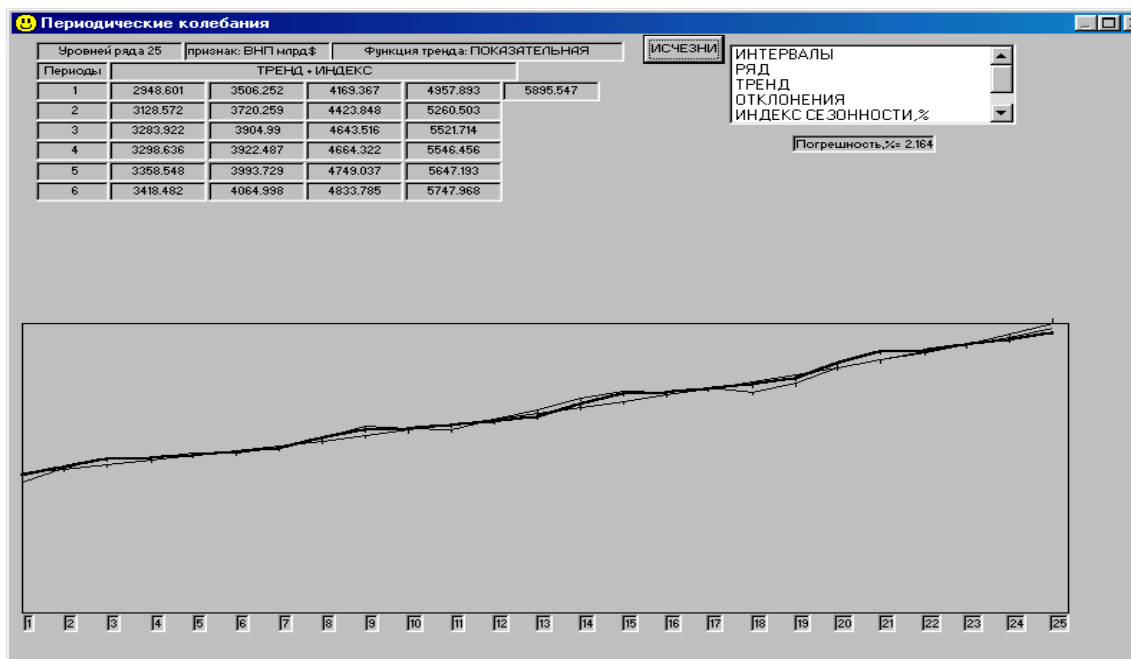


Рисунок 8.12 – Определение тенденции с использованием индекса сезонности и метода аналитического выравнивания

Сравнивая погрешность при использовании только основной тенденции по методу скользящей средней (рис. 8.7) и основной тенденции вместе с индексом

сезонности (рис. 8.11), можно отметить уменьшение погрешности с 63 до 12 %.

Эти же выводы можно сделать при использовании только уравнения тренда (рис. 8.8) и уравнения тренда и индекса сезонности (рис. 8.12). В этом случае погрешность изменяется с 2,6 до 2,16 %.

Обобщающим показателем силы колеблемости динамического ряда из-за периодичности служит среднее отклонение индексов сезонности в процентах от 100 %, т. е.:

$$\sigma_{\text{период}} = \sqrt{\frac{\sum_{s=1}^n (I_s - 100)^2}{n}},$$

где  $n$  – число периодов в цикле периодичности.

Сравнение средних квадратических отклонений для разных периодов показывает сдвиги в периодичности (сезонности). Уменьшение  $\sigma_{\text{период}}$  свидетельствует о сокращении влияния периодичности на динамику ряда.

Выделение сезонной волны можно выполнить на основе построения аналитической модели периодических колебаний. В качестве аналитической формы сезонной волны применяется уравнение следующего вида:

$$\bar{y}_t = a_0 + \sum_{k=1}^m (a_k \cos kt + b_k \sin kt),$$

где  $k$  – степень точности гармоник тригонометрического многочлена;  
 $t$  – время.

Это уравнение представляет собой ряд Фурье, где время  $t$  выражается в радианной мере или градусах. Обычно при выравнивании по ряду Фурье рассчитывают не более 4 гармоник и затем уже определяют, каким уровнем гармоник наилучшим образом отражается периодичность изменения уровней ряда.

Например, при  $k = 1$

$$\bar{y}_t = a_0 + a_1 \cos t + b_1 \sin t;$$

При  $k = 2$

$$\bar{y}_t = a_0 + a_1 \cos t + b_1 \sin t + a_2 \cos 2t + b_2 \sin 2t.$$

Параметры уравнения выравненных уровней, определяемых рядом Фурье, находят по способу наименьших квадратов:

$$a_0 = \frac{\sum y}{n}; a_k = \frac{2 \sum y \cos kt}{n}; b_k = \frac{2 \sum y \sin kt}{n}.$$

Уравнение волны можно использовать как вместе с основной тенденцией, полученной по методу скользящей средней (рис. 8.13), так и для уравнения тренда, полученного при аналитическом выравнивании (рис. 8.14). Хотя аналитическое представление обладает рядом преимуществ, тем не менее, использование уравнения волны в ряде случаев оказывается менее эффективным, чем при представлении периодических колебаний с помощью индекса сезонности.

Использование ряда Фурье в первом случае позволяет уменьшить погрешность до 10,5 %, а во втором случае – до 0,33 %.

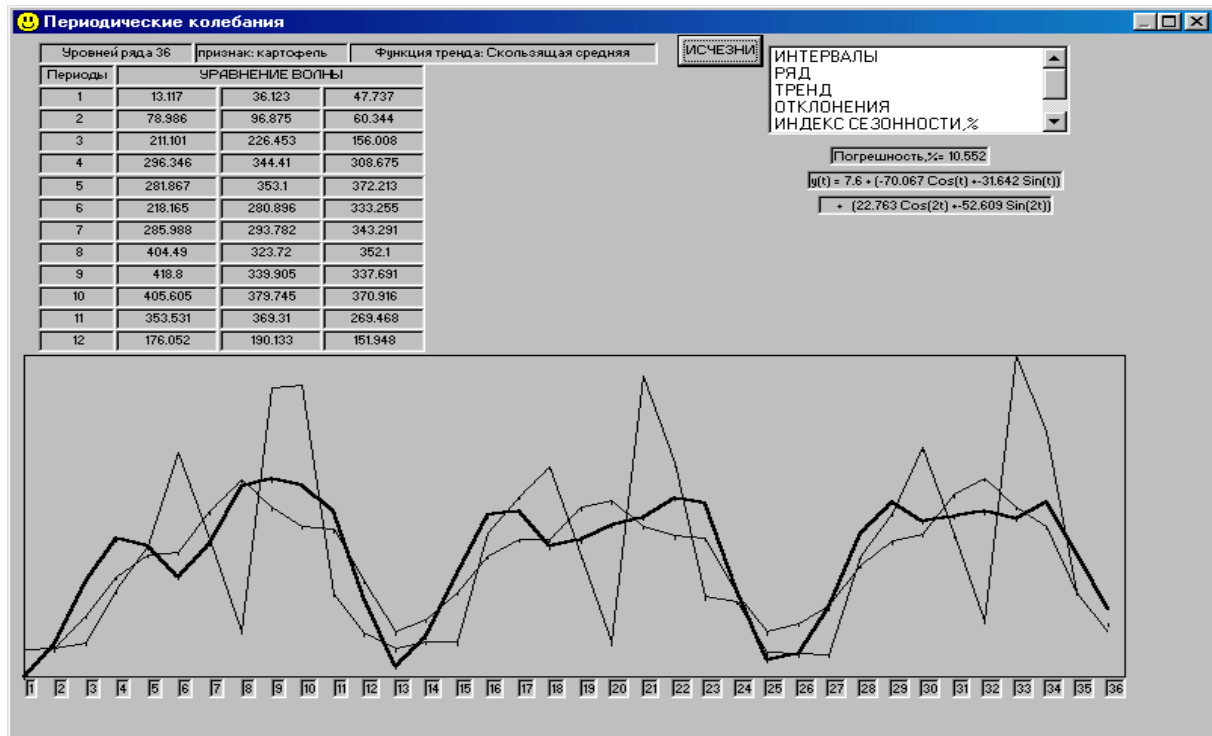


Рисунок 8.13 – Определение тенденции с использованием ряда Фурье и метода скользящей средней

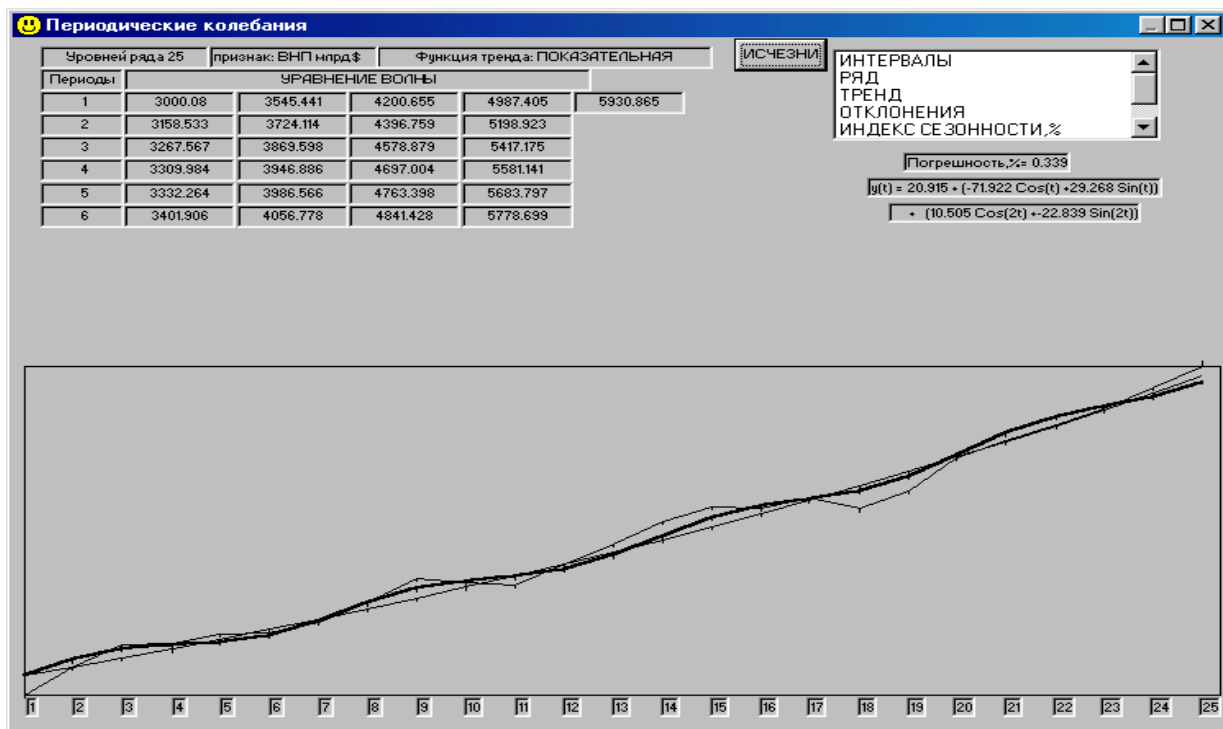


Рисунок 8.14 – Определение тенденции с использованием ряда Фурье и метода аналитического выравнивания

## БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Айвазян, С. А. Методы эконометрики: учебник / С. А. Айвазян. – М.: Магистр: ИНФРА-М, 2010. – 512 с.
2. Айвазян, С. А. Классификация многомерных наблюдений / С. А. Айвазян, З. И. Бежаева, О. В. Староверов. – М., 1974. – 240 с.
3. Айвазян, С. А. Прикладная статистика: основы моделирования и первичная обработка данных: справочное издание / С. А. Айвазян, И. С. Енюков, Л. Д. Мешалкин. – М.: Финансы и статистика, 1983. – 472 с.
4. Айвазян, С. А. Прикладная статистика: классификация и снижение размерности / С. А. Айвазян, В. М. Бухштабер, И. С. Енюков, Л. Д. Мешалкин – М.: Финансы и статистика, 1989. – 608 с.
5. Айвазян, С. А. Прикладная статистика: исследование зависимостей / С. А. Айвазян, И.С. Енюков, Л. Д. Мешалкин. – М.: Финансы и статистика, 1985. – 487 с.
6. Альберт, А., Регрессия, псевдоинверсия и рекуррентное оценивание / А. Альберто. – М.: Наука, 1977. – 224 с.
7. Андерсон, Т. Введение в многомерный статистический анализ / Т. Андерсон; пер. с англ. – М., 1963.
8. Андерсон, Т. Статистический анализ временных рядов / Т. Андерсон; пер. с англ., – М.: Мир, 1976. – 760 с.
9. Благуш, П., Факторный анализ с обобщениями / П. Благуш. – М.: Финансы и статистика, 1989. – 248 с.
10. Болч, Б. Многомерные статистические методы для экономики / Б. Болч, К. Дж. Хуань; пер. с англ. – М.: Статистика, 1979. – 320 с.
11. Бриллинджер, Д. Временные ряды. Обработка данных и теория / Д. Бриллинджер; пер. с англ. – М.: Мир, 1980. – 536 с.
12. Вазан, М. Стохастическая аппроксимация / М. Вазан – М.: Мир, 1972. – 296 с.
13. Вальд, А. Последовательный анализ / А. Вальд. – М.: Физматлит, 1960 – 328 с.
14. Вальд, А. Статистические решающие функции / А. Вальд // Позиционные игры. – М.: Наука 1967. – С. 300–522.
15. Вентцель, А. Д. Курс теории случайных процессов / А. Д. Вентцель. – 2-е изд., доп. – М.: Наука, Физматлит, 1996. – 400 с.
16. Гихман, И. И. Введение в теорию случайных процессов / И. И. Гихман, А. В. Скороход. – М.: Наука, 1977. – 570 с.
17. Доугерти, К. Введение в эконометрику / К. Доугерти. – М.: ИНФРА-М, 1997. – 402 с.
18. Дрейпер, Н. Прикладной регрессионный анализ. Книга 2 / Н. Дрейпер, Г. Смит. – 2-е изд. – М.: Финансы и статистика, 1986. – 351 с.

19. Елисеева, И. И. Эконометрика: учебник для магистров / И. И. Елисеева и др.; под ред. И. И. Елисеевой. – М.: Издательство Юрайт, 2014. – 453 с.
20. Ермаков, С. М. Статистическое моделирование / С. М. Ермаков, Г. А. Михайлов. – 2-е изд. – М.: ФИЗМАТЛИТ, 1982. – 296 с.
21. Ибрагимов, И. А. Асимптотическая теория оценивания / И. А. Ибрагимов, Р. З. Хасьминский. – М.: Наука, 1979. – 527 с.
22. Иберла, К., Факторный анализ / К. Иберла. – М.: «Статистика», 1980. – 53 с.
23. Кендалл, М. Дж. Многомерный статистический анализ и временные ряды / М. Дж. Кендалл, А. Стюарт; пер. с англ., – М.: Наука, 1976 – 375 с.
24. Крамер, Г. Математические методы статистики / Г. Крамер; пер. с англ. – 2-е изд. – М.: Мир, 1975. – 648 с.
25. Леман, Э. Проверка статистических гипотез / Э. Леман. – М.: Наука, 1979. – 408 с.
26. Линник, Ю. В. Метод наименьших квадратов и основы математико-статистической теории обработки наблюдений / Ю. В. Линник. – 2-е изд., доп. и испр. – М.: Физматгиз, 1962. – 349 с.
27. Липцер, Р. Ш. Статистика случайных процессов / Р. Ш. Липцер, А. Н. Ширяев. – М.: Наука, 1974. – 696 с.
28. Ли, Ц. Оценивание параметров марковских моделей по агрегированным временным рядам / Ц. Ли, Д. Джадж, А. Зельнер. – М., Статистика, 1977. – 224 с.
29. Магнус, Я. Р. Эконометрика. Начальный курс: учеб. / Я. Р. Магнус, П. К. Катышев, А. А. Пересецкий. – 6-е изд., перераб. и доп. – М.: Дело, 2004. – 576 с.
30. Модельные эксперименты с механизмами экономического управления. Front Cover. / под ред. В. Л. Макаров, В. А. Житков. – Наука, 1989. – 221 с.
31. Невельсон, М. Б. Стохастическая аппроксимация и рекуррентное оценивание / М. Б. Невельсон, Р. З. Хасьминский. – М.: Наука, 1972. – 304 с.
32. Нэйлор, Т., Машинные имитационные эксперименты с моделями экономических систем / Т. Нэйлор. – М.: Мир, 1975. – 500 с.
33. Рао, С. Р. Линейные статистические методы и их применение / С. Р. Рао; пер. с англ. – М.: Наука, 1968. – 547 с.
34. Розанов, Ю. А. Стационарные случайные процессы / Ю. А. Розанов. – М.: ФИЗМАТЛИТ, 1990. – 272 с.
35. Себер, Дж., Линейный регрессионный анализ / Дж. Себер; пер. с англ. – М.: Мир, 1980. – 456 с.
36. Соболев, И. М., Численные методы Монте-Карло / И. М. Соболев. – М.: Наука, 1973. – 312 с.
37. Файоль, А. Общее и промышленное управление / А. Файоль. – М.: Дело, 1991. – 388 с.
38. Шимко, П. Д. Статистика / П. Д. Шимко, М. П. Власов. – Ростов н/Д: Феникс, 2003. – 448 с. – (Высшее образование. Учебники, учебные пособия).



- 39. Bol'shev, L. N., Cluster analysis. Bull. Int. Stat. Inst., 1969, № 43, pp. 411–425.
- 40. Berger, J. O, Statistical Decision theory, N. Y. - Berlin, 1984, 617 p.
- 41. Billingsley, P., Statistical Methods in Markov chains, Ann. Math. Stat., v. 32, № 1, 1961.
- 42. Fisher, R. A., Ann. of Eugenics, 1936, v. 7, pp.179–88.
- 43. Kruskal, J. B., «Psychometrika», 1964, v. 29, pp. 1–27.
- 44. Hotelling, H. «Ann. Vath. Stat.», 1931, v. 2, pp. 360–78.
- 45. Wishart, J., «Biometrika», 1928, v. 20A, pp. 32–52.