БАКАЛАВР. АКАДЕМИЧЕСКИЙ КУРС

И. С. Светуньков, С. Г. Светуньков

МЕТОДЫ СОЦИАЛЬНО-ЭКОНОМИЧЕСКОГО ПРОГНОЗИРОВАНИЯ

Том 1. Теория и методология

УЧЕБНИК и ПРАКТИКУМ

ВЫБОР ВУЗОВ РОССИИ

УМО ВО РЕКОМЕНДУЕТ

Юрайт издательство biblio-online.ru

И. С. Светуньков, С. Г. Светуньков

МЕТОДЫ СОЦИАЛЬНО-ЭКОНОМИЧЕСКОГО ПРОГНОЗИРОВАНИЯ

Том 1. Теория и методология

УЧЕБНИК И ПРАКТИКУМ
ДЛЯ АКАДЕМИЧЕСКОГО БАКАЛАВРИАТА

Рекомендовано Учебно-методическим отделом высшего образования в качестве учебника и практикума для студентов высших учебных заведений, обучающихся по экономическим направлениям

Книга доступна в электронной библиотеке biblio-online.ru, а также в мобильном приложении «Юрайт.Библиотека»

ВЫБОР ВУЗОВ РОССИИ 2015

Москва • Юрайт • 2019

УДК 33(075.8)
ББК 65я73
С24

Авторы:

Светуньков Иван Сергеевич — кандидат экономических наук, PhD in Management Science, лектор департамента Management Science, Lancaster University, UK;

Светуньков Сергей Геннадьевич — профессор, доктор экономических наук, профессор Высшей школы управления и бизнеса Санкт-Петербургского политехнического университета Петра Великого.

Рецензенты:

Торопцев Е. Л. — доктор экономических наук, профессор Северо-Кавказского федерального университета;

Иванов Е. Е. — кандидат технических наук, профессор Санкт-Петербургского государственного экономического университета.

Светуньков, И. С.

С24 Методы социально-экономического прогнозирования. В 2 т. Том 1. Теория и методология : учебник и практикум для академического бакалавриата / И. С. Светуньков, С. Г. Светуньков. — М. : Издательство Юрайт, 2019. — 351 с. — Серия : Бакалавр. Академический курс.

ISBN 978-5-534-02801-0 (т. 1)
ISBN 978-5-534-02802-7

В учебнике рассмотрены современные методы социально-экономического прогнозирования, чаще всего используемые на практике. Существенная часть экономических решений нацелена на получение результатов в будущем, поэтому для принятия верного управленческого решения необходимо грамотное социально-экономическое прогнозирование, невозможное без знания соответствующих методов и моделей. В связи с этим обязательным условием подготовки высококвалифицированного экономиста и менеджера является изучение им дисциплины «Методы социально-экономического прогнозирования».

Первый том учебника содержит основные положения современной теории прогностики и ее методологические основы. Здесь же определяются отличительные особенности социально-экономических процессов от процессов другой природы и выделяются те свойства экономической динамики, которые следует учитывать при прогнозировании. Рассматриваются вопросы сбора и обработки данных об объекте прогнозирования; объясняются методы прогнозирования стационарных однородных и неоднородных социально-экономических процессов.

Соответствует актуальным требованиям Федерального государственного образовательного стандарта высшего образования.

Для студентов академического бакалавриата, но может быть полезен магистрантам, аспирантам и докторантам, а также практикующим специалистам, занимающимся прогнозированием социально-экономических процессов.

УДК 33(075.8)
ББК 65я73

ISBN 978-5-534-02801-0 (т. 1)
ISBN 978-5-534-02802-7

Оглавление

Введение **5**
Глава 1. Общетеоретические основы социально-экономического прогнозирования **10**
1.1. Предсказание и прогнозирование 10
1.2. Теория социально-экономического прогнозирования 16
1.3. Понятийный аппарат теории прогнозирования 29
1.4. Прогнозирование как функция управления 48
1.5. Временные ряды социально-экономической динамики 53
1.6. Обратимые и необратимые процессы 66
Вопросы и задания для самоконтроля 77
Глава 2. Анализ и первичная обработка данных **78**
2.1. Шкалы измерения социально-экономической информации 78
2.2. Измерение социально-экономических отношений для их прогнозирования 94
2.3. Предварительный анализ и обработка данных 110
2.4. Математическая обработка и прогнозирование информации, измеренной в неметрических шкалах 130
2.5. Оценка адекватности прогнозных моделей 147
2.5.1. Графический анализ модели 148
2.5.2. Основные коэффициенты оценки качества модели 153
2.5.3. Процедура ретропрогноза 171
Практикум 172
Глава 3. Теоретические основы прогнозирования стационарных социально-экономических процессов **176**
3.1. Генеральная совокупность, выборка и выборочный метод 177
3.2. Средние величины в прогнозировании однородных стационарных процессов 186
3.3. Общие принципы определения доверительных границ для выборочных значений из генеральной совокупности 195
3.4. Статистическая проверка гипотез 203

3.4.1. Общие принципы проверки статистических гипотез....204
3.4.2. Проверка гипотез с помощью нормального распределения....209
3.4.3. Проверка гипотез с помощью распределения Стьюдента....211
3.4.4. Проверка гипотез с помощью распределения хи-квадрат....214
3.4.5. Проверка гипотез с помощью распределения Фишера....216
3.4.6. Другие способы проверки гипотез....218
3.5. Основы регрессионного анализа в прогнозировании....221
3.6. Корреляционный анализ при построении однофакторных прогнозных моделей....239
3.7. Коэффициент согласия в динамике....254
Практикум....268
Глава 4. Методы и методики прогнозирования стационарных социально-экономических процессов....271
4.1. Типовые прогнозные модели....272
4.2. Метод наименьших квадратов и уравнения в отрезках....283
4.3. Прогнозирование стационарных неоднородных процессов: многофакторные модели....292
4.3.1. Пропущенные переменные....305
4.3.2. Проблема гетероскедастичности....315
4.3.3. Проблема автокорреляции остатков....317
4.3.4. Проблема мультиколлинеарности....319
4.4. Один метод получения устойчивых оценок в частном случае мультиколлинеарности....328
4.5. Синтез однофакторных моделей в многофакторную....335
4.6. Учет качественных характеристик при построении регрессий....340
Практикум....348

Введение

Методы прогнозирования многообразны, как многообразны и объекты, прогнозированием которых занимается человек. Ему всегда было интересно узнать будущее — свое, своих близких, государства, экономики, предугадать погоду, определить как изменятся в ближайшее время курсы валют и т.п.

Но прогнозировать ситуацию важно не только из-за того, что это интересно, но и потому, что от предвидения будущего зависят действия человека. Если, например, на следующий день ожидается дождь, то можно подготовить соответствующую одежду и взять с собой зонт; если прогнозируется скачок цен, то можно заранее закупить тот товар, который подорожает в наибольшей степени и т.п. Фактически, можно сказать, что жизнь современного человека невозможно себе представить без прогнозирования.

Менеджмент выделяет прогнозирование как одну из функций управления, но поскольку на результатах прогнозирования базируются все остальные функции, то понятно, что ошибки в прогнозах могут существенно снизить эффективность управления любым объектом в целом. Так, если в результате выполненного прогноза ожидается высокий спрос на какой-либо товар, то на основе прогнозных значений цен и объемов спроса фирмой разрабатывается план по выпуску этого товара соответствующего объема, вычисляется размер необходимых инвестиций и определяется их источник, создаются организационные структуры для реализации разработанного плана; запускается производство и контролируется его точное соответствие плану. Если же прогноз окажется неточным и цена спроса на товар — меньше прогнозировавшейся, то товар не будет продаваться, а работа фирмы в данном направлении станет убыточной.

Поэтому важность точного прогнозирования социально-экономических объектов для экономической практики сложно переоценить.

Умения правильно прогнозировать социально-экономические процессы формируются у студентов вузов в результате изучения дисциплин по экономическому прогнозированию. В учебных планах вузов они называются по-разному, но во всех случаях направлены на получение знаний в инструментальной базе современного прогнозирования с учетом особенностей социально-экономической динамики.

Задачами изучения студентом *дисциплины* являются:

- освоение навыков правильной диагностики динамических процессов социально-экономического развития;
- получение знаний о сути методов социально-экономического прогнозирования и особенностях их практического применения;
- умение правильно подобрать модель прогнозирования для каждого временно́го ряда и рассчитать ее основные характеристики;
- формирование представления о путях дальнейшего развития теории и методики социально-экономического прогнозирования.

В данном учебнике рассматривается прогнозирование социально-экономических систем. Основной упор при этом делается на методы и модели прогнозирования, а другие вопросы прогнозирования, например, организации этого процесса или взаимосвязи прогнозирования и планирования мы оставили вне рассмотрения. Объясняется это тем, что учебник написан для того, чтобы каждый студент, закончив обучение, знал не только общие понятия о том, как прогнозировать, а владел навыками практического применения конкретных методов прогнозирования, необходимых для его дальнейшей профессиональной деятельности как менеджера и экономиста.

Предполагается, что студенты, изучающие данный курс, уже знают основные разделы высшей математики, теории вероятностей и математической статистики, хорошо знакомы с современной эконометрикой. Впрочем, в некоторых разделах учебника часть уже пройденных ранее в указанных дисциплинах тем пересказывается, но, что называется, «с авторской интонацией». Вызвано это тем, что дополнительный неформальный взгляд на некоторые используемые при прогнозировании инструменты позволяет не допустить ошибку при их применении, ведь существенная часть из них в математической статистике разрабатывалась для задач аппроксимации и интерполяции, а не прогнозирования.

Основная цель учебника — показать задачу социально-экономического прогнозирования как задачу творческую, методическое и методологическое обеспечение которой непрерывно развиваются во времени. Так, например, если к концу 1960-х гг. в отечественной экономике особым научным достижением в этой области считалось использование метода наименьших квадратов для построения той или иной прогнозной модели и вычисления ее статистических характеристик (например, тренда в форме логарифмической функции и доверительных границ этого тренда), то сегодня такую задачу легко решает первокурсник экономического вуза.

В учебнике приводятся устоявшиеся, общепринятые и наиболее распространенные методы социально-экономического прогнозирования, дается их теоретическое обоснование и приводится материал по их практическому применению. В то же время в нем рассматриваются и новые подходы и методы прогнозирования, знания о которых появились в последние годы, широко используется и зарубежный опыт в социально-экономической прогностике, который нечасто описывается в отечественной учебной литературе.

Для более четкого понимания теоретической части материала в учебнике приводится решение соответствующих задач, взятых из реальной экономической жизни и базирующихся на конкретных примерах. В ряде случаев мы обращались к базе данных М3, которая представляет собой общедоступный сборник из 3003 временны́х рядов, по которым прогнозисты обычно сравнивают точность прогнозов математических моделей. База собиралась под эгидой Международного института прогнозистов и может быть скачана по ссылке: URL: http://forecasters.org/resources/time-series-data/m3-competition/.

Сегодня экономисту, занимающемуся прогнозированием в своей хозяйственной деятельности, предлагается довольно большое количество различного рода прикладных программ, в которые встроены различные методы прогнозирования. Учебник не содержит указаний на использование этих пакетов, поскольку не каждый из них является общедоступным. К тому же, как правило, математическое содержание, которое облечено в форму такого программного продукта, чаще всего скрывается разработчиками и обычно является элементарным.

Учебник является двухтомным. Первый его том условно может быть разделен на две взаимосвязанные части.

Первая, теоретико-методологическая часть (первые две главы учебника) включает в себя общую терминологию прогностики, принципы и концепцию современного прогнозирования, необходимые классификации и определения, которые позволят студенту уверенно пользоваться общепринятой терминологией. В ней показывается многообразие информации, используемой для социально-экономического прогнозирования, приводятся наиболее часто встречающиеся шкалы ее измерения и методы ее предварительного анализа.

Во второй части первого тома учебника (третья и четвертая главы) приводятся методы прогнозирования обратимых стационарных процессов, которые довольно часто встречаются в социально-экономическом прогнозировании. Но некоторые люди, работающие в реальной экономике и имеющие поверхностные знания об экономике вообще, считают, что в ней превалируют обратимые стационарные процессы, и используют их в каждом случае, когда сталкиваются с необходимостью прогнозирования социально-экономических процессов. На самом деле сложность экономики как объекта прогнозирования определяется наличием в ней множества необратимых процессов, к моделированию и предсказанию поведения которых предъявляются особые требования. Пожалуй, можно сказать, что необратимые процессы встречаются в экономической практике значительно чаще, чем обратимые. Тем не менее, обратимые стационарные и нестационарные процессы в практике социально-экономического прогнозирования встречаются повсеместно, поэтому методам их прогнозирования также уделено место в учебнике (в третьей и четвертой главах). Мы также разбираем наиболее типичные ошибки, которые совершают экономисты и менеджеры в практике прогнозирования.

В первом томе учебника изложены методы прогнозирования, разработанные для относительно простых объектов, поведение которых можно описать с помощью регрессионных моделей и их модификаций. Более сложные социально-экономические объекты, характеризующиеся необратимыми процессами, как объект для прогнозирования рассматриваются во втором томе.

Изучив материалы первого тома учебника, читатель будет:

знать

- теорию прогностики и ее понятийный аппарат;
- суть методов прогнозирования;

- основные прогнозные модели и методы, используемые в настоящее время экономистами для прогнозирования обратимых процессов;

уметь

- правильно понимать те процессы и объекты, которые он собирается прогнозировать;
- использовать для прогнозирования аппарат современной эконометрики;
- подбирать и применять наиболее эффективные методы прогнозирования;

владеть

- навыками прогнозирования обратимых социально-экономических процессов;
- аппаратом построения прогнозных моделей и статистической оценки их коэффициентов.

Глава 1
ОБЩЕТЕОРЕТИЧЕСКИЕ ОСНОВЫ СОЦИАЛЬНО-ЭКОНОМИЧЕСКОГО ПРОГНОЗИРОВАНИЯ

В результате освоения данной главы студент должен:

знать

- объект, предмет, теоретические и практические задачи прогнозирования;
- основы теории прогнозирования;
- основные результаты новейших исследований по проблемам менеджмента в части прогнозирования как функции управления;

уметь

- правильно классифицировать социально-экономические процессы, прогнозирование которых ему предстоит;
- выделять характеристики и свойства объектов прогнозирования;
- применять системный подход для представления задач прогнозирования;
- выявлять перспективные направления научных исследований в теории социально-экономического прогнозирования;

владеть

- понятийным аппаратом теории прогностики;
- навыками диагностики сути прогнозируемых социально-экономических процессов;
- основами методологии проведения научных исследований.

1.1. Предсказание и прогнозирование

Каждому участнику сложной системы рыночного взаимодействия хочется знать, как в ближайшей и дальней перспективе будут складываться цены на производимые товары, как поведут себя поставщики сырья и оборудования, какие цены они будут запрашивать. Чрезвычайно важно предусмотреть поведение конкурентов с тем, чтобы заранее быть готовым к тем стратегиям, которые они начнут реализовывать; необходимо заранее оценивать состояние экономической конъюнктуры рынка и правильно прогнозировать поведение потребителей и т.п.

Как известно, каждое государство живет на основе бюджета, который разрабатывается, обсуждается и утверждается парламентом заранее, на основе того, какими представляются его разработчикам предстоящий год или более далекие перспективы (каковы будут его доходы, куда можно будет потратить средства бюджета для решения наиболее значимых государственных задач и т.п.). Применительно к российской экономике доходы государственного бюджета на следующий год формируются на основе прогноза цен на минерально-сырьевые ресурсы, в первую очередь, — на основе прогноза того, сколько будет стоить баррель нефти, поскольку существенную часть доходов российского бюджета составляют налоги и сборы от реализации нефти на мировом рынке.

Таким образом, и на уровне предприятий и организаций, и на уровне регионального и государственного управления мы встречаемся с объективной потребностью заглянуть в будущее, предугадать складывающуюся ситуацию и сделать это как можно точнее, потому что такое знание позволяет принимать наиболее эффективные решения из всех возможных альтернатив.

Осознание необходимости прогнозирования возникло на заре появления цивилизаций, поскольку как только человек начал хозяйствовать, он столкнулся с необходимостью думать о будущем: вырастет ли зерно, если бросить его в землю, каким будет урожай, удачной ли будет охота или сбор плодов растений и т.п. И уже на этом этапе человек столкнулся с двумя *способами предугадывания будущего.*

Первый способ появился на основе познания окружающего мира, выявления закономерностей и причинно-следственных связей, присущих изучаемому явлению. После того, как человек понял, что в жарком климате растения начинают лучше плодоносить, если за ними ухаживать (поливать, окучивать и т.п.), он начал использовать систему поливного земледелия. Наблюдения за природой позволили определить наличие различных природных циклов (например, сезонов года) и понять, когда лучше начинать посевную, когда осуществлять прополку растений и собирать урожай. С возникновением обмена продавцы научились принимать правильные решения о том, когда лучше всего отправляться в путь и какой транспорт для этого выбирать. Понимание системы причинно-следственных связей позволило человеку прогнозировать будущее, в том числе и в области хозяйствования. Этот способ предугадывания, базирующийся на научном познании свойств явлений, складывающихся тенденций, наблюдаемых статистических закономерностей, лег в основу современной науки о прогнозировании.

Второй способ предугадывания будущего был связан с незнанием человеком свойств многих окружающих его предметов и явлений. Он наблюдал, что время от времени наступает засуха или идут проливные дожди; начинается массовый падеж скота от различных заболеваний; реки внезапно разливаются, а торговые пути вдруг становятся непроходимыми. Человек не понимал причину происходящих событий и считал, что существуют некоторые силы, управляющие всеми этими процессами, недоступными для человеческого понимания. Эти силы в дальнейшем персонифицировались в соответствующих богов, управляющих миром. Их волей объяснялось все происходящее, к ним обращались с просьбами повлиять на ситуацию наилучшим образом. Возникла особая группа людей, чьей основной задачей было осуществлять посредничество между людьми и богами. Эти люди обращались к богам за советом и передавали остальным их волю. Так появились религии и люди, осуществляющие религиозные культы.

Поскольку на заре цивилизации накопленных знаний было мало и научному объяснению поддавались очень немногие происходящие вокруг процессы, религиозное объяснение мира превалировало. Пытливый человеческий ум стремился найти некоторые приметы или знаки, которые посылают боги, чтобы по ним понять будущее. Эти приметы искали во внутренностях жертвенных животных, в полете птиц, в форме костра и т.п.

История любой древней цивилизации указывает на процветание в ней культа гадания, технологии которого определялись культурными и религиозными особенностями, у всех народов предсказанием будущего жрецы занимались. Поскольку за свои предсказания они получали определенное вознаграждение, их можно считать первыми профессиональными прогнозистами.

У древних греков и римлян гадание было доведено до высшего уровня организации и получило название *мантика*. Система гаданий состояла из таких направлений, как[1]:

1) метеорологические гадания (гром, молния, комета, облака, радуга и т.п.), широко распространенные у греков Гомеровских времен;

2) по предметам (*эмпиромантия* или *пиромантия* — по огню, *омоплатоскопия* или *скапулимантия* — по кости лопатки животного, *клеромания* — посредством жребия, *библиомантия* — по писанному слову, в основном поэтическим произведениям Гомера и Гесиода);

3) по одушевленным предметам (*орнитомантия* — по полету, крику, приземлению и действиям птиц, *гиероскопия* — по внутренностям жертвенных животных, *морфоскопия* — по строению человеческого тела, *хиромантия* — по линиям на ладони и т.п.);

4) по числам (начертание имени числом, астрология и т.п.);

5) через сновидения (*онейромантика*), вызывание души покойников (*некромантия*), религиозные откровения (*хресмология*) и др.

Их жрецы объединялись в коллегии, специализирующиеся на отдельных видах гаданий. Среди этих коллегий самыми многочисленными были *гаруспики*, представители которых гадали по внутренностям жертвенных животных, и *авгуры*, гадавшие по полету птиц, облакам и другим метеорологическим явлениям.

Примеры таких гаданий и их толкований повсеместно встречаются у древнегреческих и римских писателей, философов и историков.

Например, сразу после воцарения в Македонии Александра Великого, оказалось, что «статуя Орфея фракийца, сына Эагра, находящаяся в Пиериде, все время покрывается потом. Одни прорицатели предсказывали одно, другие другое; Александр же, телмесец, тоже прорицатель, сказал Александру: “Дерзай; знамение это значит, что поэтам эпическим и лирическим, а также исполнителям их произведений предстоит великий труд: создать произведения, в которых будут воспевать Александра и его дела”»[2].

---

[1] Гаданіе // Энциклопедическій словарь. Издатели: Ф. А. Брокгауз, И. А. Ефрон. Томъ VII А. СПб., 1892. С. 773–777.

[2] *Арриан.* Поход Александра. СПб. : Алетейя, 1993. С. 26.

Гай Светоний Транквилл в жизнеописаниях римских цезарей также часто приводит примеры предсказаний оракула или гаданий по внутренностям животных. Так, перед одной из битв император Август «совершал жертвоприношения, но не мог добиться добрых знамений и уже велел привести новых жертвенных животных, как вдруг неприятели сделали внезапную вылазку и захватили все принадлежности жертвоприношения. Тогда гадатели единодушно решили, что все беды и опасности, возвещенные жертвователю, должны пасть на того, кто завладел жертвенными внутренностями; и так оно и случилось»[1].

Точность таких предугадываний будущего была крайне невелика. Более того, прекрасно понимая ответственность за результаты предсказаний, жрецы чаще всего облекали предсказания в такие завуалированные формы, что их можно было толковать самыми разными способами. Наиболее яркий пример этого — оракул в храме Аполлона в древних Дельфах: Пифия высказывала свои пророчества стихами, которые сами нуждались в толковании. Именно поэтому «относительно оракулов известно, что вера в них стала утрачиваться еще задолго до пришествия Иисуса Христа»[2].

Переход от предсказаний к прогнозированию постепенно осуществлялся с развитием науки, когда ученые открывали присущие природе законы и закономерности. Пожалуй, наибольшее влияние на формирование прогностики как некоторой научной дисциплины оказал факт появления в науке метода индукции. До его открытия для доказательства ученые использовали лишь аристотелевский метод дедукции — трудоемкий, требующий долгого и тщательного сбора данных, охватывающих все множество состояний изучаемого объекта (известно, что дедуктивный вывод является истинным, если он применяется и используется правильно — формируются истинные посылки, и умозаключение выводится из них логически верно).

Индуктивный метод Бэкона—Милля как мето́да познания заключается в том, что на основе учета всех частных посылок, констатирующих наличие некоторого признака отдельных предметов определенного класса, делается обобщающий вывод о наличии данного признака у всех предметов данного

[1] *Светоний Г. Т.* Жизнь двенадцати цезарей. М. : Эксмо, 2006. С. 102.

[2] *Монтень М.* Опыты. Избранные произведения : в 3 т. Т. 1. М. : Голос, 1992. С. 41.

класса. Индуктивный метод проявляется в двух основных формах: индукции посредством перечисления и индукции посредством исключения. С помощью индукции в научном исследовании выявляются не только сходные, но и отличающиеся факты, подтверждающие гипотезу и отражающие существенные, закономерные связи между исследуемыми свойствами предметов и явлений. Индуктивный метод позволять предполагать, т.е. прогнозировать свойства всей совокупности, если эти свойства выявлены у некоторой части ее элементов.

В отличие от дедуктивного, индуктивный вывод не требует изучения свойств всех элементов совокупности, а ограничивается лишь некоторым представительным множеством выборочных значений из этой совокупности, поэтому он значительно проще в применении и существенно менее трудоемок.

При этом следует иметь в виду следующее: «Об истинности индуктивного умозаключения никогда нельзя говорить с достоверностью... Даже если посылки предполагаются истинными и вывод является правильным индуктивным умозаключением, результат может оказаться ложным. Самое большее, что мы можем сказать, это то, что по отношению к данным посылкам заключение имеет некоторую степень вероятности... Мы знаем, что единичное утверждение факта, полученное путем наблюдения, никогда не является абсолютно достоверным, потому что мы можем сделать ошибки в наших наблюдениях. Но по отношению к законам существует еще бóльшая неопределенность. Любой закон, относящийся к миру, устанавливает, что в любом частном случае, в любом месте и в любое время, если одна вещь истинна, то другая вещь также истинна. Ясно, что здесь речь идет о бесконечном числе возможных случаев... Даже наилучшим образом обоснованные законы физики опираются на конечное число наблюдений. Всегда возможно, что завтра может быть обнаружен противоречащий случай»[1].

Из философских оснований индуктивного вывода возник и математический аппарат получения этого вывода — теория вероятностей и математическая статистика. Если внимательно изучить разделы математической статистики, то можно убедиться в том, что выборочные статистические характеристики и параметры — это не что иное, как результат индуктивного вывода.

---

[1] *Карнап Р.* Философские основания физики. М., 1971. С. 60–61.

Индуктивный метод и соответствующая ему теория вероятностей и математическая статистика позволили ученым прогнозировать с определенной вероятностью наличие вычисляемых, но пока еще не наблюдаемых свойств у объекта исследования. На самой заре появления индуктивного метода и благодаря ему ученые смогли предсказать наличие планет в солнечной системе, даже тех, которые не видны в телескоп; поведение жидкостей при различных давлениях и температурах, не достижимых в реальных опытах и т.п. Наука стала обогащаться новыми знаниями, выявленными законами и закономерностями.

В современной науке прогноз — это результат именно индуктивного вывода, когда по характеру ограниченного множества значений показателей или взаимосвязи факторов делается вывод о том, что и остальные, еще не наблюдаемые значения этих показателей или взаимосвязи будут обладать аналогичными свойствами.

Сегодня человека окружает много явлений, которые еще невозможно объяснить с помощью научных знаний, а тем более — спрогнозировать. Однако среди социально-экономических явлений таких «белых пятен» осталось очень немного, и наступление некоторых непрогнозируемых последствий связано не с действием некоторых высших и неведомых сил, а с действием сил, еще не выявленных и не изученных наукой. Поэтому уже давно никто из практикующих экономистов не прогнозирует состояние конъюнктуры рынка посредством гаданий. В основе всех функций управления в экономике и прогнозирования в том числе лежат научные знания, получаемые как на основе опытного пути (эмпирика), так и на основе научных теорий и генерируемых на их основе гипотез.

1.2. Теория социально-экономического прогнозирования

Как известно, теория представляет собой некоторое объяснение реально существующего явления. Теории могут быть научными или не научными. Научная теория опирается на достоверные факты, в своих выводах и основных положениях использует общенаучные методы и принципы научных исследований. Социально-экономическое прогнозирование, являясь разделом экономической науки, также имеет свою теорию, которая непрерывно развивается во времени за счет получения новых знаний об объекте исследования и разработки новых подходов и методов прогнозирования. Поэтому методы социально-экономического прогнозирования, которые еще 20 лет назад считались научно обоснованными и передовыми, сегодня зачастую представляют лишь исторический интерес и не используются на практике, поскольку имеются значительно более точные методы.

Теория социально-экономического прогнозирования, как и любая другая научная теория, имеет в своем арсенале многообразные, но взаимосвязанные в концептуальном единстве принципы, подходы, способы и методы решения конкретных задач, имеет собственную методологическую основу, с помощью которой формируются новые теоретические положения отдельных разделов прогнозирования.

Эмпирические знания, т.е. знания, полученные человеком в ходе его практической деятельности, могут либо совпадать с теорией и развивать теоретические знания, либо входить с ними в противоречие.

В первом случае человек объясняет ситуацию и выявляет совокупность причинно-следственных связей на базе теоретического знания. При этом, объясняя ситуацию, он может прогнозировать ее развитие, предвидеть последствия каждого варианта своих действий, а значит, принимать правильные и эффективные решения.

Во втором случае, когда практика входит в противоречие с теорией, требуется более глубокое изучение причин этого расхождения. *Причинами противоречия могут быть:*

- некоторое исключение из общего правила, в результате чего теоретическое знание остается без изменений, но дополняется знанием данного исключения;
- наличие таких новых условий и причин, которые еще не были учтены теорией, поэтому теория развивается и уточняется за счет этого нового знания;
- ошибочность теории. В этом случае возникает необходимость создания новой теории или модификации уже существующей.

Эта схема встречается и в ходе практической деятельности прогнозиста — он всегда проверяет изучаемую ситуацию на соответствие ее той или иной известной ему теории, объясняя суть поведения прогнозируемого процесса. Формальный подход, ориентированный на унифицированное применение методов прогнозирования к любому объекту без учета его отличительных свойств, чаще всего ведет к ошибкам в прогнозировании. Если прогнозист не обладает теоретическими знаниями об объекте, то он не в состоянии объяснить ситуацию, понять ее динамику и причинно-следственные связи. В любом случае он будет стараться как-либо объяснить ситуацию, т.е. строить свою теорию. Если эта теория не будет научно обоснованной, то действия прогнозиста, опирающегося на нее, будут неэффективными. Чаще всего теория создается после того, как явно или неявно сформирована некоторая *концепция* — совокупность идей, принципов, положений об объекте исследования, своеобразная форма теоретического знания, пока не получившая всесторонней проверки и систематизации.

С учетом того, что научное знание непрерывно развивается и на смену одним истинам приходят другие, некоторые теоретические положения, которые когда-то считались очевидными, со временем перестают быть таковыми. Более того, становится ясно, что они являются ошибочными. Но если в одном из разделов науки когда-то была разработана теория и ее аксиоматическое ядро не было четко сформулировано, а опиралась она на некоторые постулаты, размытые в научных текстах, при устаревании этого аксиоматического ядра заметить это становится сложно. Это приводит к определенному «замораживанию» соответствующего раздела науки. Прогнозы, выполняемые с помощью подобных частных теорий, все чаще не выполняются, возникают различные парадоксы и назревает объективная потребность разработки новой теории. Поэтому любую теорию, а особенно в обширной области экономики, следует рассматривать в качестве инструмента, время от времени требующего пересмотра и уж точно — постоянного совершенствования.

При создании теории широко используют *гипотезы* — предположения (допущения), выдвигаемые для решения исследовательской задачи. Научная гипотеза формулируется, исходя из имеющегося в распоряжении ученого багажа научных знаний. Гипотезы служат для систематизации данных, их дедуктивного упорядочивания, обобщения и предположительного расширения наличных эмпирических данных. Они также применяются для придания исследованию направленного характера. Научная гипотеза должна быть синтаксически грамотной, семантически осмысленной, логически корректной, теоретически обоснованной и эмпирически проверяемой. При соблюдении данных требований гипотеза будет являться научной вне зависимости от того, окажется ли она впоследствии истинной или ложной. Научная гипотеза, оказавшаяся ложной, все равно имеет большое значение для построения теории, поскольку в ходе исследования высказывается несколько гипотез и опровержение одной из них сужает их круг и позволяет быстрее найти гипотезу, оказавшуюся истинной.

По своему происхождению гипотезы могут быть теоретическими или эмпирическими. Теоретические гипотезы вытекают из имеющейся теории, а эмпирические формулируются на основе опыта. Так, например, исходя из теории Маслоу можно ожидать от потребителя, что он, удовлетворив свои потребности нижнего уровня, перейдет к решению задач удовлетворения потребностей более высокого уровня. Это предположение будет гипотезой, сформулированной на базе теории. Такая теоретическая гипотеза, как легко заметить, может лечь в основу соответствующей прогнозной модели поведения потребителя.

Если же в ходе маркетинговых исследований было выявлено, что потребители, обладающие каким-либо отличительным признаком, одинаково реагируют на товар и его маркетинговое сопровождение, то высказывается предположение (гипотеза) о том, что все другие потребители, которые имеют этот же отличительный признак, будут также реагировать на товар и его маркетинговое сопровождение. Эта гипотеза базируется на эмпирическом знании и может лечь в основу другой прогнозной модели поведения потребителя.

Когда ученый приступает к построению определенной теории, он вынужден начинать работу с некоторого мысленного представления объекта и его свойств. Понятно, что этот образ весьма упрощен. Метод исследования реальных объектов, при котором происходит мысленное отвлечение от всех реальных свойств объектов и акцентируется внимание на наиболее важных из них, называется *абстрагированием*. Без абстракции не существует научного мышления. Любое представление реальности в виде мысленного, словесного, графического или математического описания (модели) является абстрактным.

Любая модель в экономике, как бы хорошо она ни описывала прошлое, никогда не сможет учесть все реально складывающиеся условия развития экономического объекта и включить в себя все реально действующие факторы. Поэтому экономическая модель, как и любая другая, всегда абстрактна. Например, при описании производства часто используют модель производственной функции:

$$Q_t = f(K_t, L_t), \tag{1.1}$$

где Q_t — результат производства; K_t — величина капитальных ресурсов; L_t — величина привлекаемых трудовых ресурсов.

Эта модель часто позволяет довольно точно прогнозировать результаты производства, хотя в ней и не учитывается множество других важных факторов производства (уровень технологии, квалификация исполнителей, количество и качество сырья и материалов и т.п.). На определенном уровне абстрагирования эти подробности для целей исследователя становятся несущественными.

Выделяют *три вида абстрагирования*. Первый из них предполагает, что в предмете выделяются определенные признаки, а все другие остаются за пределами внимания (происходит отвлечение от всех других признаков). Результат применения такого приема есть абстрактно мыслимый, характеризуемый лишь некоторой совокупностью выделенных признаков предмет. Этот прием неразрывно связан с обобщением предметов некоторого класса и поэтому может быть назван *обобщающе-различающим абстрагированием*. К этому виду абстрагирования относится и модель 1.1.

Второй вид — *отождествляющее абстрагирование*. Этот прием состоит в том, что, выделяя некоторые признаки предмета, исследователь игнорирует все остальные как несущественные с той или иной точки зрения, и отождествляет все предметы, обладающие выделенными признаками. Выделяя, например, те или иные слова по их структуре, мы игнорируем все различия, связанные с их написанием или произношением, и рассматриваем все случаи употребления слова одной и той же структуры как различные экземпляры одного и того же слова. Применительно к модели производственной функции отождествляющее абстрагирование будет осуществлено, если прогнозист посчитает приемлемым для моделирования любых производственных процессов только одну разновидность производственных функций — функцию Кобба — Дугласа.

И, наконец, существует так называемое *изолирующее абстрагирование*, при котором отдельные признаки и характеристики предметов отделяются от самих предметов и становятся самостоятельными предметами мысли. Если прогнозист моделирует динамику экономики с помощью совокупности моделей, рассматривая отдельные показатели как самостоятельные объекты прогнозирования, то этот прием приводит к появлению изолирующего абстрагирования.

Метод идеализации предполагает еще бóльшую степень отвлечения от реальных свойств объекта, когда сложные для познания свойства объекта заменяются свойствами, облегчающими его понимание, но в реальности не существующими. Метод идеализации применяется в том случае, когда иначе изучить объект невозможно, и представляет собой только первый шаг в научном исследовании. Идеализация состоит в том, что, имея в виду некоторые предельные случаи (предел уменьшения трения, увеличения упругости и т.д.), мы либо мысленно наделяем предметы какими-то свойствами, которых они в действительности не имеют (например, физические тела — способностью восстанавливать при деформации свой объем или форму, в результате чего появляются понятия типа «идеально упругое тело» или «идеальная жидкость»), либо лишаем их каких-то свойств, которыми они в действительности обладают. Так в нашем сознании возникают «безразмерные» точки, «линии, лишенные ширины», «идеальный газ». Идеализированные объекты познания — результаты определенного типа мысленной «обработки» предметов реальной действительности — идеализации (абсолютно черное тело, абсолютно упругое тело и т.п.).

Многие разделы современной экономической теории рассматривают не абстрактные, а идеализированные объекты. Точно так же и в прогнозировании приходится сталкиваться с идеализированными моделями, которые весьма часто встречаются в таком разделе экономико-математического моделирования, как экономическая динамика, например, предполагается наличие репрезентативного потребителя, *живущего бесконечно долго*, который максимизирует суммарную дисконтированную полезность, получаемую от потребления в течение всей жизни. Курсивом выделено идеализирующее свойство, поскольку понятно, что ни один из потребителей не в состоянии жить «бесконечно долго». Очевидно, что несуществующий потребитель никак не может быть «репрезентативным» и практическая ценность подобных идеализированных моделей очень невелика. Использовать их для прогнозирования реальной экономики нельзя — прогнозист должен тщательно следить за тем, чтобы в процессе абстрагирования не построить идеализированную, не имеющую ничего общего с реальной экономикой прогнозную модель, иначе его прогнозы не будут осуществляться.

Следует отметить, что в результате абстрагирования либо идеализации исследователь работает не с самим реальным объектом, а с его упрощенным образом. Этот образ выступает в научном исследовании как модель. В зависимости от способа представления образа модель может принимать самые разные формы. В экономике чаще всего используют модели математические (основные характеристики объекта и его взаимосвязи описывают математическими уравнениями и неравенствами), графические (основные характеристики объекта и его взаимосвязи объясняют с помощью рисунков и графиков) и словесные (основные характеристики объекта и его взаимосвязи описывают словами). Наибольшую ценность для исследователя представляют *математические модели*, поскольку с их помощью можно провести глубокие и многовариантные исследования объекта. Но, к сожалению, модели далеко не всех объектов экономических исследований можно построить в математической форме. Очень многие показатели и взаимосвязи экономических объектов просто невозможно измерить в количественной шкале, т.е. дать им количественную оценку, поэтому невозможно и их математическое описание.

Менее богатым по возможностям исследования является класс *графических моделей*. Как правило, в экономике эти модели не имеют четких масштабов и пропорций, хотя встречаются модели, строящиеся на фактических материалах реальной статистики. Но чаще всего приходится иметь дело с моделями, которые описывают основные закономерности изменения одного (или нескольких) показателя при изменении другого показателя (или нескольких показателей). Здесь сложно проводить многовариантные расчеты и чаще всего невозможно получить точные количественные оценки прогнозируемых показателей, но с помощью графических моделей удается тщательно изучить основные закономерности и свойства объектов экономических исследований и построить на этой основе теории, которые обладают хорошей прогнозирующей способностью. С помощью графических прогнозных моделей можно определить общие тенденции и направления развития прогнозируемых объектов.

К классу графических моделей следует в том числе отнести разнообразные столбиковые, круговые, точечные диаграммы, схемы и т.п. С их помощью можно получить представление об отношениях между изучаемыми объектами, определить возможные направления изменений этих взаимоотношений, но дать количественную оценку на их основе затруднительно.

Существенным ограничением области применения графических моделей является и то, что с их помощью можно изучить двухфакторные, максимум — трехфакторные взаимосвязи. В первом случае анализ осуществляется с помощью графических моделей, которые изображаются в двухмерной системе координат, во втором — в трехмерной, причем подобные пространственные построения, естественно, реализуются на плоскости. С использованием современной компьютерной техники можно широко и успешно использовать трехмерную графику, но она требует соответствующих аппаратных средств, которые пока еще не стали для экономиста таким же обыденным средством, как бумага и карандаш. Использование графических моделей для изучения четырех- и более факторных моделей очень затруднено, и в экономике подобные модели практически не используются.

В случаях, когда зависимых факторов больше трех, а их измерение затруднительно, используют *словесные модели*. Для их описания и обоснования применяются методы дедукции и индукции. Встречающаяся иногда эклектичность, для которой характерен разрыв связей между элементами теории, нарушающий ее целостность, свидетельствует о том, что теория находится в стадии создания и в дальнейшем, после появления новых знаний, между элементами появится дедуктивная или индуктивная связь и теория станет научно обоснованной.

Нельзя рассматривать процесс построения теории в отрыве от динамики общего научного знания. Сегодня очевидно, что гром и молния — результаты действия статического электричества, возникновения разности потенциалов и, при достижении критической величины этой разности, прохождения электрического тока между точками, имеющими эти различные электрические потенциалы. Это положение может быть высказано сегодня в качестве аксиомы. Но 500 лет назад причину грома и молнии ученые видели в божьем гневе, и это казалось им столь же очевидным, как нам — электростатическая природа молнии.

Любая прогнозная модель, базирующаяся на той или иной теории, лежит в русле некоторой узкопредметной методологии.

Методология — это система определенных принципов, приемов и операций, применяемых в определенной сфере деятельности. Она включает в себя общие основания, выявленные законы и закономерности, принципы и методы организации научного познания, развития и применения результатов научной деятельности в практике.

Следует различать теорию и методологию, являющиеся взаимосвязанными понятиями. На самом раннем этапе развития наука еще не имеет своей методологии. Ее основу составляет формирующаяся частная теория, которая базируется на методологии некой общей науки. Например, теория маркетинга должна базироваться на методологии современной экономики, социологии и психологии. После того, как сформирована первая частная научная теория, на ее основе формируются новые элементы науки — возникают общие принципы исследования в данной науке, разрабатываются способы решения задач, методы и т.п., т.е. формируется методология данной частной науки, которая предопределяет направления дальнейшего развития науки — на этой основе возникают новые частные теории (например, на базе методологии маркетинга возникают теория сегментации, теория коммуникаций, теория бенчмаркинга и т.п.). Эти теории, очевидно, расширяют «тело» методологии. Процесс этот бесконечен и диалектичен.

Ценность любой научной теории заключается, прежде всего, в возможности использования ее положений на практике, решении конкретных практических задач. Последовательность действий по решению таких задач называют способом. Способ включает в себя систему предписаний, рекомендаций, предостережений, последовательность операций, выполнение которых способствует успешному достижению поставленной цели. Если этот способ является универсальным, т.е. применимым к целому ряду аналогичных задач, то его называют **методом**.

Под **методикой** понимается совокупность приемов, связанных с применением конкретного метода, их последовательность и взаимосвязь. Любая научная методология включает в себя теории, способы и методы решения различных задач. Вооруженный таким теоретическим инструментарием практикующий специалист разрабатывает и использует различные методики. Иногда методика может вызывать самостоятельный научный интерес, но чаще всего она — инструмент практической реализации научных знаний, приведших к появлению метода.

Теория прогнозирования, как и любая другая научная теория, использует различные общенаучные методы, среди множества которых для прогнозирования принципиальное значение имеет *метод аналогии*. Он основан на том, что из сходства некоторых признаков двух или нескольких объектов делается индуктивный вывод о сходстве других признаков этих объектов. Любая прогнозная модель — это некий аналог реально существующего явления или процесса.

Аналогия эффективна, если общие признаки сравниваемых предметов разнообразны и существенны и есть основания считать, что они обеспечат сходство и по признаку, интересующему исследователя. В научной аналогии общие признаки у сравниваемых предметов должны быть в точности одинаковыми, связь признаков не должна зависеть от обстоятельств и специфики сравниваемых предметов. Кроме того, для повышения вероятности выводов по аналогии необходимо учитывать сходства внутренних, а не внешних свойств сопоставляемых объектов. Приемы аналогии широко используются в тех случаях, когда выдвинутые теоретические положения не удается доказать с помощью имеющихся знаний об объекте. К аналогии также прибегают в процессе творческого поиска новых идей, в методах математического моделирования, в прикладных исследованиях и др.

В наиболее общем виде *схема аналогии* может быть представлена так:

a имеет признаки P, Q, S, T;

b имеет признаки P, Q, S, ...;

b, очевидно, имеет признак T.

Существует *три вида аналогии*.

1. *Аналогия свойств* заключается в поиске сходства в свойствах двух сравниваемых объектов, когда основные свойства объекта А совпадают с основными свойствами объекта Б. Так, например, свойства потребителей одного сегмента совпадают друг с другом, поэтому, изучив свойства типичного потребителя из этого сегмента, можно перенести эти свойства на аналогичного потребителя того же сегмента.

2. *Аналогия отношений* уместна в условиях сравнения двух или более связанных некоторыми отношениями объектов. При этом если объект А относится к объекту Б так

же, как и объект В — к объекту Г, то особенности отношений А к Б могут быть перенесены на особенности отношений В к Г. Например, одно предприятие имеет филиал в другом городе. Между головным предприятием и филиалом имеются определенные финансовые и производственные отношения. Изучив эти отношения, можно предполагать, что и у другого предприятия, имеющего филиал, будут иметься такие же отношения.

3. *Аналогия структуры* предполагает ситуацию, когда совокупность основных элементов и взаимосвязей между ними объекта А подобна совокупности основных элементов и взаимосвязей между ними объекта Б. Например, структуры экономики Франции и Германии имеют много общего. Значит, при наступлении каких-либо событий в экономике Франции, связанных с некоторым изменением состояния внешних или внутренних факторов, можно ожидать аналогичных изменений в экономике Германии, если в состоянии ее внешних или внутренних факторов происходят схожие изменения.

Аналогии играют важную эвристическую роль, однако умозаключения по аналогии, как и другие эвристические методы познания, ведут лишь к вероятностному, а не к окончательному и достоверному знанию, поскольку они являются результатом индуктивного вывода.

Частным и весьма активно используемым на практике методом исследования является синтез метода аналогии и исторического метода — *метод исторической аналогии*. Здесь в качестве аналога для объекта исследования используют объекты одинаковой физической природы, опережающие во времени развитие прогнозируемого объекта.

Иногда в качестве аналога используются объекты другой физической природы, другой области науки или отрасли техники. При этом необходимым элементом обоснованности таких параллелей и доказательности результатов аналогии является идентичность процессов развития объектов разной природы. Прежде чем использовать подобную аналогию, необходимо доказать, что сравниваемые объекты подобны друг другу. Для этого часто используется инструментарий *теории подобия* — дисциплины, изучающей условия подобия и подобные преобразования различных явлений, а также методы их математического описания.

Фактором, обусловившим исторически значимый рост научного знания в области прогностики, явилось формирование такого раздела экономической науки как *эконометрика*. Этот термин был введен в научный оборот в 1926 г. норвежским экономистом Р. Фришем. Под эконометрикой понимают раздел экономики, занимающийся построением экономических моделей, описывающих экономические объекты на основе фактических данных с помощью методов математической статистики. Впрочем, в последние годы в число эконометрических моделей и, соответственно, методов все чаще включают модели, коэффициенты которых оценены на множестве статистических данных, но не методами математической статистики, а, например, численными методами. Поэтому под эконометрикой сегодня можно понимать раздел экономико-математического моделирования, в котором развитие экономических процессов моделируется с помощью математических методов на основе обработки имеющихся статистических данных.

Таким образом, задача прогнозирования в наши дни базируется на основательном фундаменте общенаучных знаний и знаний в конкретных разделах науки.

Появление в науке кибернетики и кибернетического подхода вновь способствовало существенному развитию прогностики. Более того, представление о кибернетике как о наиболее общей науке об управлении породило желание создать и ее раздел — прогностику как науку о наиболее общем подходе к прогнозированию. В этом направлении было получено множество важных результатов — были сформулированы общие принципы прогностики, сформирован терминологический аппарат, разработаны наиболее общие методы прогнозирования, построены модели процесса прогнозирования и выработаны общие подходы к информационному обеспечению процесса решения задачи прогнозирования.

Однако практика показала, что особенности каждого объекта прогнозирования (экономического, социального, биологического, естественно-природного и т.п.) таковы, что одни и те же методы прогнозирования, применяемые к этим объектам, дают прогнозы неудовлетворительной точности. Постепенно возникло понимание того, что в прогностике, как и в других функциях управления, фактор природной специфичности играет принципиально важную роль — нельзя получить точный прогноз объекта, не изучив природу и свойства прогнозируемого объекта. А раз так, то помимо общих положений прогностики должны быть созданы и разделы прогнозирования в каждой конкретной науке. Именно поэтому методы социально-экономического прогнозирования, с одной стороны, имеют много общего с методами прогнозирования в других науках, а с другой, — весьма специфичны и часто применимы исключительно для задач прогнозирования в социально-экономической сфере.

В чем же специфика прогнозирования в экономике вообще и прогнозирования социально-экономических процессов, в частности?

Первая отличительная черта заключается в том, что однозначных и количественно выражаемых законов в экономике нет. Кто не помнит, например, всемирный закон тяготения, открытий Исааком Ньютоном? Он имеет количественное выражение, в его математической формулировке есть неизменная константа и с помощью математической модели этого закона можно достаточно точно спрогнозировать движение любого тела в пространстве. А вот в экономике нет ни одного подобного закона, математически описанного и неизменного. Существующие в ней законы отражают лишь общее направление, количественно не выраженное. Всем известен закон спроса (с ростом цены на товар объемы спроса на него уменьшаются), но любые математические формулировки этого закона носят преходящий характер — они пригодны только для частных случаев. Это тем более верно, если вспомнить о парадоксе Гиффена, который противоречит этому закону.

Вторая отличительная черта социально-экономического прогнозирования заключается в том, что социально-экономические системы развиваются как результат активной человеческой деятельности, которая носит творческий, зачастую новаторский характер. Поэтому любые складывающиеся закономерности меняются во времени. Именно поэтому прогнозная модель, построенная, например, пять лет назад и хорошо прогнозировавшая социально-экономическое развитие в тот период, сегодня будет давать значительно менее точные прогнозы — за пять лет человек успевает внести изменения как в технологии производства, так и в организационно-экономические отношения.

Третья черта, являющаяся логическим следствием первых двух, связана с тем, что социально-экономические процессы в значительной степени носят эволюционный характер. В социально-экономических системах меняются и структура элементов, и взаимосвязи между ними, и внешнее окружение этих систем, причем эти изменения носят и количественный, и качественный характер. И если объект прогнозирования, например, экономика России, и пять, и десять, и пятнадцать лет назад был одним и тем же (экономикой России), то свойства этого объекта сегодня и его свойства пять, десять и пятнадцать лет назад таковы, что в них больше различий, чем сходства. И более разумно говорить о том, что это не одна и та же, а различные социально-экономические системы. Эта особенность существенно отличает задачу социально-экономического прогнозирования от прогнозирования объектов другой природы.

1.3. Понятийный аппарат теории прогнозирования

Любая теория оперирует своими специфическими понятиями. Не является исключением и теория социально-экономического прогнозирования. Она имеет свои устоявшиеся и общепринятые термины и понятия, которые прогнозист должен знать и использовать в своей работе.

Прежде всего, следует определить само базовое понятие «*прогноз*».

Любопытно, что в словаре Брокгауза и Эфрона прогноз определяется так: «прогноз (Prognosis) – предсказание развития и исхода болезни, основанное на верности распознавания»[1]. Это свидетельствует о том, что прогнозирование как особая научная деятельность в конце XIX – начале XX в. не рассматривалось. Лишь к началу XX в. ученые стали относиться к прогнозированию в экономике как самостоятельному разделу экономической науки (как самостоятельный раздел этой науки социально-экономическая прогностика сформировалась лишь к 1970-м гг.). Несмотря на очевидную смысловую и логическую простоту данного слова, его четкое научное определение не является однозначным. Иногда встречаются разновидности следующих определений: прогноз – это *вероятностное* суждение о состоянии какого-либо объекта (процесса или явления) в определенный момент времени в будущем и (или) об альтернативных путях достижения каких-либо результатов; экономический прогноз – это определенная гипотеза, некоторая *вероятностная* оценка протекания экономического

---

[1] Прогнозъ // Энциклопедическій словарь. Издатели: Ф. А. Брокгауз, И. А. Ефрон. Томъ XXV. СПб., 1898. С. 773–777.

процесса в будущем. Эрих Янч в 1974 г. определял прогноз как «*вероятностное* утверждение о будущем с относительно высокой степенью достоверности»[1].

О вероятностном характере прогнозов долгое время говорили практически все отечественные и иностранные прогнозисты. И только в конце XX в. пришло понимание того, что использование при определении прогноза указания на его вероятностный характер ограничивает совокупность применяемых при этом методов прогнозирования, так как в данных определениях понимание термина «вероятности» дается скорее в обыденном смысле, а не в строго научном. Фраза типа: «одна акция, вероятно, будет стоить 1000 рублей», говорит о том, что «1000 рублей» является тем ориентиром будущего, на который следует равняться — она наиболее правдоподобна. Конечно же, возможны отклонения от этой цифры и она, скорее всего, является серединой интервала неопределенности, но вероятность появления этой величины рассчитать невозможно. С позиций же теории вероятностей употребление слова «вероятность» однозначно свидетельствует о том, что цифра «1000 рублей» есть стохастическая оценка прогнозируемого явления, а значит, может быть рассчитана вероятность ее появления, дисперсия и т.п., т.е. практически места для неопределенности не остается. А ведь каждый студент знает о том, что в экономике «сплошь и рядом» встречаются случаи принятия решений в условиях неопределенности — стоит вспомнить лишь такую дисциплину как «Теория игр»!

На самом деле неопределенность объективно возникает при прогнозировании большинства случаев социально-экономической динамики, так как эта динамика протекает в условиях диалектически изменяющегося мира. Зачастую на практике удается в лучшем случае дать лишь граничные значения прогнозируемого явления без указания каких-либо вероятностных оценок внутри этого интервала и методы математический статистики здесь оказываются неуместными. Поэтому при определении прогноза как понятия следует избегать употребления слова «вероятность». Полагаем, что корректным можно назвать определение, согласно которому *прогноз* — это научно обоснованное суждение о воз-

---

[1] *Янч Э.* Прогнозирование научно-технического прогресса. М. : Прогресс, 1974. С. 10.

можных состояниях объекта в будущем и (или) об альтернативных путях и сроках их осуществления[1].

Очевидно, что в соответствии с данным определением не каждое суждение о будущем будет являться прогнозом, а только то, которое является научно обоснованным. Учитывая, что научная мысль не стоит на месте и непрерывно развивается, критерии научной обоснованности тех или иных методов и подходов усложняются. В 1960-х гг. научно обоснованными инструментами прогнозирования в экономике были тренды, оцененные с помощью метода наименьших квадратов (МНК), в 1970-е гг. — многофакторные регрессионные модели, в 1980-е — имитационные динамические модели, в начале XXI в. — нейронные сети. Это значит, что научная обоснованность прогноза — явление динамическое, подверженное непрерывной количественной и качественной ревизии на предмет соответствия с существующими на настоящий момент последними достижениями научной мысли. Научные теории непрерывно эволюционируют, а значит, непрерывно эволюционирует и наука о прогнозировании.

Прогнозирование — это процесс разработки прогноза, состоящий из ряда взаимосвязанных этапов, на каждом из которых решаются оригинальные задачи с помощью присущей только этому этапу совокупности методов и подходов. В целом прогнозирование может быть представлено в виде некоторой системы подходов и методов, используемых для выполнения наиболее точного прогноза.

Период упреждения прогноза — это тот промежуток времени, на который разрабатывается прогноз. Чаще всего в практике прогнозирования для обозначения периода упреждения употребляют другие словосочетания — время упреждения, период прогнозирования, срок прогнозирования, дальность прогноза и т.п. Период прогнозирования является одним из классификационных признаков, по которому можно осуществить группировку прогнозов.

Любой прогноз основан на изучении некоторого множества проведенных в прошлом наблюдений. Промежуток времени, в течение которого проводились наблюдения, на основании которых строится прогноз, получил название *периода основания прогноза*, или *базы прогноза*.

---

[1] *Светуньков С. Г., Светуньков И. С.* Методы социально-экономического прогнозирования : учебник для вузов. Т. 1. СПб. : Изд-во СПбГУЭФ, 2009. С. 25.

Любой прогноз обладает присущими ему *характеристиками*, к которым относятся:

- *точность прогноза* — оценка доверительного интервала прогноза для определенной доверительной вероятности его осуществления (в том случае, когда прогноз имеет вероятностный характер);
- *достоверность прогноза* — оценка вероятности осуществления прогноза для заданного доверительного интервала (в том случае, когда прогноз имеет вероятностный характер);
- *ошибка прогноза* — фактическая величина отклонения прогноза от действительного состояния объекта прогнозирования.

В случае, когда вероятностные оценки прогноза не могут быть даны, точность прогноза и его достоверность определяются качественными, а не количественными характеристиками или задаются границами без указания вероятности попадания прогнозируемой величины в эти границы.

Для получения прогноза в процессе прогнозирования может быть использовано множество методов, каждый из которых имеет свои, присущие только ему особенности. Для выбора лучшей прогнозной модели или метода прогнозирования используют метод верификации.

Верификация в прогнозировании — процесс установления верности, пригодности, правдоподобности модели или метода прогнозирования для описываемого явления или объекта.

При выборе метода прогнозирования следует исходить из ряда обязательных *принципов прогнозирования*:

- *принцип системности*, предполагающий комплексное изучение объекта с позиций единой системы взаимосвязей явлений и факторов, составляющих его прогнозный фон;
- *принцип природной специфичности*, который требует тщательного изучения особенностей объекта прогнозирования, делающих его отличным от других объектов. Именно выявление этих особенностей позволяет избежать ошибки инструментария, когда используемый прогнозный аппарат оказывается непригодным для данного объекта из-за присущих ему специфичных свойств;
- *принцип оптимальности затрат*, состоящий в естественном желании провести анализ, да и осуществить прогноз с минимальными затратами трудовых и материальных ресурсов.

В настоящее время никто не задается вопросом: а сколько всего существует методов прогнозирования? Хотя еще в середине 1970-х гг. этот вопрос волновал многих исследователей, которые, отвечая на него, называли разные цифры. В начале 1970-х гг. Э. Янч утверждал о наличии около 100 методов прогнозирования, в середине 1980-х гг. называли цифру 150, а то и 200. Поскольку применительно к каждому оригинальному объекту прогнозирования чаще всего приходится адаптировать или модифицировать те или иные методы, а иногда и создавать новые, ответ на этот вопрос нисколько не информативен. Одно можно с уверенностью утверждать — что методов социально-экономического прогнозирования очень много — несколько сотен — и изучать каждый из них в отдельности не имеет смысла. Очевидно, что некоторые методы будут похожи друг на друга и базироваться на одних и тех же предположениях и теоретической базе. Поэтому имеет смысл изучать наиболее типичные методы прогнозирования, которые дают исследователю представление обо всей группе методов. Следовательно, сложную совокупность методов прогнозирования необходимо разбить на устойчивые группы, т.е. осуществить их классификацию по какому-либо признаку.

Но, прежде чем приступить к классификации методов прогнозирования, необходимо еще раз уточнить два основных понятия — «метод» и «способ», поскольку практика показывает, что смешение этих двух понятий приводит к недопониманию и к неправильным действиям.

Когда человеку необходимо достичь какой-либо цели, он последовательно предпринимает совокупность некоторых действий для достижения этой цели. Эта совокупность называется способом достижения цели. Таким образом, *способ* — это система предписаний, рекомендаций, предостережений, последовательность операций, выполнение которых способствует достижению поставленной цели. Поскольку достичь поставленной цели можно разными способами, то возникает задача выявления наилучшего из них. Понятно, что выбор наилучшего способа напрямую связан с критерием этого выбора. Любой человек, а тем более — исследователь, старается действовать рационально, и если перед ним вновь возникнет задача достижения цели, с которой он уже сталкивался раньше, то для экономии собственных усилий и средств он вспомнит о том, что уже сталкивался с проблемой выбора наилучшего способа решения в схожей ситуации. Этим же способом он воспользуется вновь. И если каждый раз при возникновении однотипных задач использовать один и тот же способ их решения, признанный наилучшим из множества других возможных способов, то можно говорить об унифицированном способе, или о методе.

Итак, можно сказать, что «*метод* — это способ познавательной или практической деятельности, представляющий собой унифицированную систему предписаний, рекомендаций, предостережений, последовательность операций, выполнение которых способствует успешному достижению поставленной цели. Научный метод есть инструмент познания, с помощью которого появляются новые сведения об изучаемом объекте. Конкретный метод — это не только инструмент познания или преобразования, но и специфическая форма знания о том, как в определенных условиях действовать для достижения желаемого результата. Правильно выбранный метод дисциплинирует поиск истины, позволяет экономить время и силы, двигаться к искомому кратчайшим путем»[1].

В нашем учебнике будут изучены методы прогнозирования и основные методики их использования, ведь один и тот же метод, применяемый в разных задачах, может использоваться по-разному, поэтому методик прогнозирования значительно больше, чем методов прогнозирования.

Многообразие методов прогнозирования вызвано:

- многообразием условий, в которых функционируют объекты прогнозирования;
- своеобразием каждого из этих объектов и значительным отличием их друг от друга.

Вследствие этого эффективность применения каждого метода прогнозирования зависит от того, насколько прогнозируемый объект похож на тот, для которого метод был предложен и проверен. Иногда в социально-экономическом прогнозировании решаются задачи не экстраполяции (перенесения в будущее той закономерности, которая была выявлена и оценена в прошлом), а интерполяции, когда необходимо оценить значения показателя внутри множества имеющихся наблюдений. Чаще всего эти прогнозы связаны не со временем, а с другими факторами. Например, в распоряжении экономиста может быть информация о произ-

---

[1] *Светуньков С. Г., Хан Т. В.* Логико-гносеологическая терминология в экономике (краткий словарь). СПб. : Изд-во СПбГУЭФ, 2004. С. 66—67.

водительности труда рабочих определенной квалификации и соответствующей ей оплате труда в пределах от 1000 до 20 000 руб. и в этом промежутке нет информации о производительности труда при заработке в промежутке от 13 000 до 15 000 руб. Если экономисту необходимо вычислить прогнозное значение производительности труда при его оплате в размере 14 000 руб., то такая задача будет являться интерполяционной. В этом случае он может использовать либо специальные методы интерполяции, либо методы математической статистики. Выбор определяется конкретными условиями, определяющими задачу прогнозирования.

Множество методов прогнозирования чаще всего группируются по критерию того, что является информационной основой для прогнозирования. По этому критерию выделяют группу методов, которые опираются на фактическую информацию, а есть такие, которые опираются на информацию экспертов. Поэтому выделяют две большие группы методов — фактографические и экспертные. Иногда к ним добавляют третью группу — смешанные или комбинированные методы, по определению представляющие собой смесь первых двух. Но поскольку комбинированные методы сочетают в себе преимущества и первой, и второй групп методов, посредством такого синтеза они приобретают некоторые новые свойства, существенно отличающие их от двух основных групп. Поэтому имеет смысл говорить о *трех основных группах прогнозирования* — фактографических, экспертных и комбинированных.

К *фактографическим методам* относят методы прогнозирования, основанные на обработке объективных данных о прогнозируемом объекте. Они опираются на факты, информация о которых получена, измерена и обработана, потому эти методы и называют «фактографическими».

К *экспертным методам* относят методы, базирующиеся на интуиции и опыте специалистов (экспертов). Их информационной основой является мнение экспертов, а не наблюдения или фактические данные. Легко понять, что это мнение не является фактом, и чтобы использовать его для прогнозирования, вначале необходимо добиться того, чтобы эксперт это мнение высказал. Полученная таким образом информация должна быть измерена и обработана, и тогда на этой основе можно осуществлять прогноз.

Доказано, что применение фактографических методов более эффективно, чем применение экспертных методов. Поэтому последние применяют лишь в том случае, когда фактографические методы использовать невозможно. В основном это касается прогнозов качественного состояния той или иной системы, того или иного явления, или же ситуации, о которой вообще нет информации. Применяются экспертные методы прогнозирования и в ситуации, когда необходимо срочно принять какое-либо решение, а времени на сбор и обработку имеющейся информации для использования фактографических методов у лица, принимающего решения, нет.

Комбинированные методы представляют собой синтез фактографических и экспертных методов. Чаще всего логика их применения такова. Вначале обрабатывается имеющийся фактический материал и на основе полученных результатов выполняется прогноз или серия прогнозов. А поскольку полученные прогнозы есть результат довольно формального подхода, достоверность полученных результатов корректируется с учетом экспертной оценки достоверности этих прогнозов.

Если взять всю совокупность методов прогнозирования, осознанно используемых в экономической практике, за 100%, то на долю фактографических методов придется примерно 80%: эта группа методов самая многочисленная и многообразная.

Выделяют следующие группы фактографических методов[1].

1. *Экстраполяционные методы*, основанные на принципе переноса в будущее тенденций, действовавших в прошлом и настоящем, т.е. базирующиеся на индуктивном выводе.

2. *Системно-структурные*, включающие методы функционально-иерархического моделирования, морфологического анализа, имитационного динамического моделирования, сетевого моделирования и другие методы, которые отличаются широтой охвата и стремлением учета всех основных факторов и возможных вариантов. Объект прогнозирования представляется не как «черный ящик», вход и выход из которого являются информационной базой для прогнозирования, а как система, состоящая из взаимосвязанных подсистем и элементов. При этом делаются попытки подробного изучения объекта прогнозирования

[1] Рабочая книга по прогнозированию / отв. ред. И. В. Бестужев-Лада. М. : Мысль, 1982.

с позиций системного подхода через выявление структуры этого объекта и моделирование взаимосвязей между элементами этой структуры.

3. *Методы опережающей информации* включают в себя методы анализа потоков публикаций, патентной информации, изобретений. Они позволяют определить общие тенденции развития научно-технической мысли и социума и предсказать дальнейшие пути их динамики.

Иногда в отдельную группу выделяют так называемые *ассоциативные методы*, подразумевая под ними методы имитационного моделирования и историко-логического анализа. Однако такое выделение, на наш взгляд, не верно. Дело в том, что методы имитационного моделирования в прогнозировании математически описывают структурные взаимосвязи объекта прогнозирования, предполагая, что они не претерпят в будущем особых изменений или что эти изменения будут развиваться известным образом. По своей логике они являются системно-структурными. Методы историко-логического анализа, которые исходят из продолжения в будущее тенденций, которые уже однажды проявили себя или в прошлом, или в аналогичных процессах также следует отнести к методам экстраполяции.

В социально-экономическом прогнозировании основными являются методы экстраполяции. Их теоретической основой является предположение о том, что выявленные и описанные тенденции и характеристики поведения объекта прогнозирования не претерпят каких-либо принципиальных изменений в будущем, а потому их можно экстраполировать в целях получения прогноза.

Выбор метода прогнозирования одного и того же объекта социально-экономического прогнозирования определяется еще и тем, на какую перспективу выполняется прогноз. Действительно, если необходим прогноз на самую ближайшую перспективу, когда объект в силу присущей ему инерционности не успевает изменить свои характеристики, то прогнозируется не столько состояние объекта, сколько отклонения от этого состояния. А вот если необходимо выполнить прогноз на далекую перспективу, то здесь возникает задача рассчитать долговременную тенденцию движения самого объекта, а отклонения от нее оцениваются как некоторый прогнозный фон.

Все это предопределило *различие в методах прогнозирования в зависимости от периода упреждения*. В зависимости от периода упреждения выделяют несколько видов прогнозов, а значит, и присущих им методов:

- оперативные;
- текущие;
- краткосрочные;
- среднесрочные;
- долгосрочные;
- дальнесрочные.

Однако с позиций методологического основания методов прогнозирования можно укрупнить их и выделить три принципиально различные группы методов прогнозирования:

1) краткосрочные;
2) среднесрочные;
3) долгосрочные.

В *первую группу* методологически оправдано отнести оперативные, текущие и краткосрочные прогнозы. Они основаны на прогнозировании на очень малый промежуток времени, в основном на учет и прогнозирование действия случайных факторов. При этом период упреждения, как правило, равен одному шагу наблюдения (если наблюдения осуществляются ежечасно, то прогноз делается на час, если ведутся ежесуточные наблюдения, то прогноз делается на сутки вперед и т.п.). Деление прогнозов на оперативные, текущие и краткосрочное осуществляется лишь при прогнозировании развития некоторых больших систем, например, электрической нагрузки. Для этого случая прогноз потребления электроэнергии на полчаса и час рассматривается как оперативный — он необходим диспетчерам энергосистем, с тем, чтобы они могли вовремя дать команду на электростанции для набора нагрузки или ее сброса. Текущий прогноз — это прогноз электрической нагрузки примерно на сутки вперед. Краткосрочный — на срок до одного месяца. Разделяя виды этих прогнозов по периоду упреждения, с позиций их методологического содержания мы должны отнести их к одной группе — краткосрочных методов прогнозирования, поскольку в каждом из них используются одни и те же методы.

Ко *второй группе* прогнозов можно отнести методы среднесрочного прогнозирования, при котором осуществляется изучение, анализ и прогнозирование как случайных факторов, так и тенденций развития основных, определяющих факторов на перспективу, когда объект не претерпевает качественных изменений.

К *третьей группе* следует отнести методы долгосрочного и дальнесрочного прогнозирования, когда прогнозируются не столько детерминированные и случайные, сколько неопределенные факторы, основные результаты их проявления в далекой перспективе.

Но эта классификация требует уточнения. Необходимо, прежде всего, ответить на принципиальный вопрос: какой именно срок прогнозирования можно считать кратким — час, сутки, месяц или год?

Раньше экономисты пытались найти однозначный ответ на этот вопрос. Например, в одном из учебников советского периода по прогнозированию об этом говорилось так: «оперативный прогноз имеет период упреждения — от одного месяца до года, среднесрочный — от года до пяти лет, долгосрочный от пяти до 15 лет, дальнесрочный — свыше этого периода»[1]. Эти периоды, очевидно, определялись парадигмой планового централизованного хозяйства, когда за основу планового развития брались пятилетние планы, утверждаемые на съездах КПСС. Согласно данной логике пять лет рассматривались как средний срок прогноза. Все, что меньше его, можно было отнести к краткосрочному прогнозированию, а что свыше — к долгосрочному. Сегодня условия экономической жизни на постсоветском пространстве принципиально изменились, но вышеприведенная градация еще иногда встречается в современных российских учебниках по прогнозированию как некоторая рекомендация[2]. На наш взгляд, это неверно. Для разных объектов прогнозирования сроки должны быть различными, ведь эти объекты имеют разную продолжительность жизни в рыночной экономике, обладают разными способностями к адаптации и выживаемости, отражаемыми в периоде инерционности их развития. Ведь один месяц с позиций, например, брокера на рынке Forex — это средний срок, а с позиций мировой экономики он почти незаметен! Поэтому говорить о соответствии срока прогноза его виду следует, исходя из свойств объекта прогнозирования, и, в первую очередь, из периода его инерционности.

---

[1] Основы экономического и социального прогнозирования : учебник для вузов. Д. М. Крук [и др.]. М. : Высшая школа, 1985. С. 7.

[2] *Лапыгин Ю. Н.* Экономическое прогнозирование : учеб. пособие. М. : Эксмо, 2009. С. 23.

Под *инерционностью* понимается способность объекта сохранять прежнее состояние и его характеристики в течение некоторого промежутка времени при незначительных внешних воздействиях на объект. При этом тенденции развития объекта как системы постепенно меняются под воздействием внутренних и внешних факторов. Если в окружении объекта происходят какие-либо изменения и внешние воздействия на него усиливаются, то он меняет динамику своего развития, но не мгновенно, а постепенно, адаптируясь к этим воздействиям.

Период времени, в течение которого объект продолжает развиваться по инерции, можно назвать «периодом инерционности». И если для объектов естественнонаучных дисциплин вычислить этот период не составляет особого труда, то для объектов социально-экономической природы эта задача является непосильной. Дело в том, что вместе с эволюционирующими социально-экономическими объектами меняются и инерционности их развития. Например, в середине 1980-х гг. эксперты утверждали, что период инерционности энергосистемы России составляет 7–8 лет, сегодня же они называют цифру в 3–5 лет. Тем не менее, несмотря на то, что точно определить период инерционности развития объектов социально-экономического прогнозирования невозможно, каждый высококлассный специалист может дать экспертную оценку этого срока, которая будет незначительно меняться в зависимости от личности эксперта.

Поэтому в зависимости от периода упреждения прогноза по отношению к периоду инерционности виды прогноза можно охарактеризовать следующим образом:

Краткосрочный прогноз — это прогноз на такой промежуток времени, который мал по отношению к периоду инерционности. Поэтому за этот период особых изменений в тенденциях развития объекта прогнозирования произойти не может. Интерес для прогнозирования представляют отклонения от тенденции развития, вызванные действием краткосрочных сил и факторов.

Среднесрочный прогноз выполняется на промежуток времени, соизмеримый с периодом инерционности объекта прогнозирования. Поэтому для таких видов прогноза важно правильно выявить и промоделировать динамику развития объекта — модель такой динамики и будет выступать в качестве прогнозной модели.

Долгосрочный прогноз выполняется на период упреждения, значительно превышающий период инерционности. В этот период на динамику объекта прогнозирования могут оказать существенное влияние факторы, свойства которых в момент выполнения прогноза еще не до конца известны или же неизвестны даже сами факторы. Задача прогнозиста при этом — оценить многовариантные сценарии развития объекта прогнозирования, опираясь, конечно, на сложившиеся тенденции его развития и возможные направления их изменения. Именно для такого вида прогноза характерно наличие высокой неопределенности и именно в этом случае используют комбинированные методы прогнозирования — экспертные методы наряду с фактографическими.

Вышесказанное можно продемонстрировать следующим образом[1]. При анализе любой конкретной величины социально-экономического показателя y_t можно выделить следующие составляющие эту величину слагаемые:

1) *регулярную составляющую* $\tilde{y}_t$, величина которой строго обусловлена влиянием конкретных значений известных факторов x_t, и это влияние полностью детерминировано;

2) *случайную составляющую* ε_t, появление которой вызвано влиянием множества случайных факторов;

3) *неопределенную составляющую* u_t, вызванную влиянием целого ряда факторов, действие которых прогнозисту неизвестно.

Регулярная составляющая $\tilde{y}_t$ представляет собой некоторую легко вычисляемую тенденцию, динамика которой полностью объясняется действием на показатель y_t известных факторов x_t. Это могут быть различного рода закономерности, выявленные во время анализа процесса (например, взаимосвязь между курсом доллара США на ММВБ (Московская межбанковская валютная биржа) и индексами инфляции России, взаимосвязь между производительностью труда и оплатой труда и т.п.).

В этом случае существуют не только однозначное толкование сути изучаемой взаимосвязи, но и, как правило, однозначные численные значения коэффициентов взаимосвязи и других характеристик.

В каждой обобщенной экономической характеристике помимо четко объяснимых значений $\tilde{y}_t$ есть некоторая часть

---

[1] *Светуньков С. Г.* Количественные методы прогнозирования эволюционных составляющих экономической динамики. Ульяновск : Изд-во УлГУ, 1999.

(обычно не очень весомая), появление которой носит случайный характер — к регулярной составляющей прибавляются (или вычитаются) значения ε_t, вызванные проявлением множества случайных факторов (ошибки при получении информации, ошибки округления, результат влияния факторов естественно-природной и военно-политической природы и т.п.). Любое социально-экономическое явление происходит в условиях, когда на него воздействует множество разнообразных факторов самой разной природы. Но, во-первых, этих факторов чрезвычайно много; во-вторых, они воздействуют на это явление отнюдь не регулярно, а чаще всего — однократно; в-третьих, эти воздействия на изучаемое явление носят разнонаправленный характер — одни способствуют увеличению показателей, характеризующих явление, другие — их уменьшению. Таким образом, есть основания утверждать, что эта случайная составляющая нормально распределена, а ее математическое ожидание равно нулю.

Третью часть анализируемого процесса, которую мы назвали неопределенной, исследователь объяснить не в состоянии. Составляющая u_t характеризует изменение социально-экономического показателя так, что исследователь это изменение либо не замечает, либо не может понять его. Кстати, одним из таких факторов являются инновационные процессы, непрерывно протекающие во всех экономических системах. Именно они способствуют изменению во времени всех технико-экономических показателей — производительности труда и оборудования, энергоемкости и материалоемкости продукции, качества изделий и их ассортимента и т.п. Например, замена ламп накаливания на лампы люминесцентные приводит не только к тому, что изменяются нормы расхода электроэнергии на освещение помещений, но меняется и качественная характеристика процесса потребления — лампы накаливания никоим образом не реагируют на колебания частоты, а люминесцентные лампы изменяют свою мощность в зависимости от частоты электроэнергии в сети. Таким образом, при нарушении баланса генерирующей и потребляемой мощностей, изменение частоты будет приводить к несколько иным, чем ранее, последствиям для потребителей, а, следовательно, методы учета и прогнозирования количественных характеристик потребителей качественно меняются. Предугадать на перспективу, как скажутся подобные инновационные процессы на изменении социально-экономических и технико-экономических показателей, можно только в общем, с большой долей неопределенности, тем более, что в настоящее время происходит активное внедрение энергосберегающих ламп освещения, выполненных по разным технологиям.

К этой же неопределенной составляющей можно отнести и необъясненные факторы. Поскольку любая модель абстрактна, то в результате ее построения при абстрагировании отбрасывается из рассмотрения часть несущественных с позиций прогнозиста факторов. Некоторые из них, считающиеся на момент разработки прогноза несущественными, в будущем могут оказать определяющее влияние на характер прогнозируемой величины, и их неучет в прогнозировании приводит к значительному снижению точности прогноза.

Таким образом, наблюдаемая на момент времени t величина показателя y может быть представлена в виде суммы этих трех слагаемых:

$$y_t = \tilde{y}_t + \varepsilon_t + u_t. \qquad (1.2)$$

Бóльшую часть величины y_t представляет регулярная составляющая. Воздействие на нее случайных факторов в среднем таково, что величина ε_t имеет нулевое математическое ожидание, а характер ее распределения приближается к нормальному. Необходимо, впрочем, отметить, что в каждый конкретный момент времени суммарное воздействие случайных факторов может и не быть равным нулю. Тем не менее, колебания их суммарного воздействия могут быть определены и учтены с помощью дисперсии. Влияние случайных факторов на результирующую величину необходимо учитывать в первую очередь при краткосрочном прогнозировании (на час, сутки, месяц), так как вариация результирующего признака в таком разрезе времени вызвана именно случайными факторами. При среднесрочном и долгосрочном прогнозировании влияние случайных факторов на вариацию динамики по сравнению с другими составляющими равенства (1.2) незначительно, так как случайные факторы в общем случае не могут вызвать изменения в тенденциях развития, а вероятность того, что они в течение значительного промежутка времени будут складываться только определенным образом, способным изменить динамику, ничтожно мала.

Однако существенную часть результирующей величины (1.2) невозможно объяснить, так как у исследователя не хватает знаний о происходящих процессах, например, о процессах детерминированного характера — они неизвестны в данный момент, но в дальнейшем становятся известными. Тем самым неопределенность снимается. Это могут быть случайные факторы, сложившиеся в данный период времени определенным образом. Но есть и факторы, чье влияние в силу ряда причин невозможно проследить и объяснить — они отнесены в группу неопределенных факторов.

На момент подготовки прогноза неопределенная составляющая представляется некоторым интервалом, характеристики которого неизвестны. Но, рассматривая развитие неопределенной составляющей во времени, можно отметить, что вызываемые этой составляющей изменения в характере развития экономических систем и их факторов экономической динамики во все возрастающих масштабах становятся такими, что появляется знание о них, и из числа неопределенных факторов они переходят в число факторов определенной природы. Влияние наиболее существенных из них изучается и учитывается с помощью различных способов.

Таким образом, часть неопределенной прежде информации становится понятной и включается исследователем в регулярную составляющую модели процесса (1.2). Однако другая часть неопределенных факторов продолжает действовать таким образом, что ее проявления еще долгое время будут неизвестны исследователю. Кроме того, человеческие изобретательность и творчество находятся в непрерывном развитии и движении. А это значит, что в жизнь будут внедряться все новые и новые идеи, которые, вначале незначительно, а затем все сильнее и сильнее будут сказываться на результатах труда и его характеристиках, на способах взаимодействия с окружающей средой.

Следовательно, неопределенная составляющая, в процессе познания переходящая в определенную, является неизменным спутником экономической динамики, ведь именно инновации приводят к изменению структур экономических систем и взаимоотношений в них.

Влияние неопределенной составляющей на изменение факторов экономической динамики или рядов ее отдельных показателей проявляется лишь в долговременных тенденциях, поэтому оно должно быть учтено при средне- и долгосрочном прогнозировании.

Обобщая вышесказанное, можно сделать следующие выводы.

При оперативном, текущем и краткосрочном прогнозировании, т.е. при прогнозировании на срок, ничтожно малый по сравнению со сроком инерционности социально-экономического явления, динамика регулярной составляющей $\bar{y}_t$ и динамика неопределенной составляющей u_t практически неизменны — в силу инерционности они не успевают претерпеть заметных изменений. Поэтому можно учесть результирующую их суммарной величины, не выделяя их по отдельности:

$$\hat{y}_t = \bar{y}_t + u_t. \tag{1.3}$$

В данном случае основная вариация показателя y_t в краткосрочном аспекте вызвана воздействиями случайной составляющей ε_t. Поэтому величину y_t в случае изучения краткосрочных показателей вполне обоснованно можно представить в виде двух слагаемых:

$$y_t = \hat{y}_t + \varepsilon_t. \tag{1.4}$$

С учетом того, что $\hat{y}_t$ определяется в данном случае достаточно точно, основные направления в прогнозировании результирующего признака y_t осуществляются в области исследования и прогнозирования случайной составляющей ε_t. Здесь с успехом могут быть использованы и методы теории вероятностей, и методы математической статистики, и ряд других методов прогнозирования случайных процессов. В случае, когда инерционность некоторых эволюционных процессов, воздействующих на исследуемый ряд, сравнима со сроком краткосрочного прогнозирования, следует использовать другие методы, например, адаптивные, основанные на принципах вычисления скользящих взвешенных средних. Выбор метода прогнозирования при этом определяется уже характером самого процесса.

При выполнении средне- и долгосрочных прогнозов вариация регулярной и неопределенной составляющих столь значительна по сравнению со случайной составляющей, что акцент делается именно на исследование и выявление тенденций изменения этих двух составляющих. При этом влиянием случайной составляющей можно пренебречь и учесть ее вариацию через дисперсию. Основное же внимание следует уделять анализу долговременных тенденций развития процесса, с тем, чтобы попытаться выделить не только регулярную составляющую, но и те воздействия, которые оказывают на нее инновационные процессы.

Опыт, однако, показывает, что выделить раздельно тенденции двух слагаемых $\bar{y}_t$ и u_t невозможно. Но с учетом того, что математическое ожидание случайной составляющей в среднем равно нулю, можно определить их результирующую (1.3). Нельзя забывать и о том, что эта составляющая есть синтез регулярной и неопределенной составляющих, который приводит к изменению тенденций развития. Это означает, что в данном случае динамический ряд значений показателя отражает изменение не только количественных тенденций, но и качественных изменений в самих системах и состоянии факторов экономической динамики, т.е. является эволюционным.

Статистические данные развития социально-экономической динамики, таким образом, могут не только отражать количественные изменения, но и характеризовать качественную динамику, которая, к сожалению, скрыта в потоке цифр и наблюдений, и о которой в настоящее время можно судить только на основе экспертных заключений. Поэтому и выделить эти составляющие не представляется возможным. Значит, при математической обработке таких рядов следует использовать апостериорную информацию о свойствах изучаемого процесса и под выявленные свойства подбирать соответствующий метод прогнозирования, понимая, что построенная прогнозная модель представляет собой лишь более или менее удачное приближение к оценке сути социально-экономической динамики, но вовсе не закон, описывающий эту динамику. Любая, самая сверхсложная прогнозная модель не выделяет регулярную и неопределенную составляющую, учитывает лишь их совместное влияние. Следовательно, прогноз социально-экономической динамики, имеющий некоторое численное значение, уже содержит в себе неопределенность, и найти на этой основе точное прогнозное значение невозможно. Этот прогноз помимо доверительных границ, оценивающих влияние случайных факторов, должен содержать в себе интервал неопределенности.

Долгосрочные и дальнесрочные прогнозы нацелены на такую перспективу, что влияние неопределенности здесь превалирует. За срок, на который в этом случае прогнозируется социально-экономическая динамика, могут произойти такие качественные и количественные изменения, что попытка измерения количественных показателей на этот срок просто бессмысленна. В этом случае прогнозируются различные варианты социально-экономического развития и по каждому из вариантов определяются возможные последствия. Здесь происходит как раз то, что изучает теория принятия решений в условиях неопределенности — формируются различные «состояния природы» и высчитываются разные варианты возможных действий по каждому из состояний, после чего формируется наилучшая стратегия управления социально-экономическим процессом.

Итак, теперь понятны особенности видов прогноза, выделяемых в зависимости от периода прогнозирования, и те задачи, которые выполняются при прогнозировании в каждом из случаев. Естественно, что для решения этих задач используются и соответствующие методы прогнозирования.

При выполнении прогнозов достигаются самые разнообразные цели, которые используются в различных задачах хозяйствования. Поэтому возможны классификации прогнозов и по другим критериям. Например, в зависимости от того, что является *содержанием прогноза*, прогнозы делятся на поисковые и нормативные.

Поисковый прогноз — это прогноз, содержанием которого является определение возможных состояний объекта прогнозирования в будущем. Примером неудачного поискового прогноза является прогноз, сделанный в конце 1950-х гг. руководителем бывшего СССР Н. С. Хрущевым, о том, что к 1980 г. в СССР будет построен коммунизм, или высказывание Президента РФ В. В. Путина в начале 2000-х гг. о твердой уверенности в удвоении валового внутреннего продукта (ВВП) России к 2010 г. В СССР вместо ранее объявленного коммунизма, как говорили в то время шутники, были проведены Олимпийские игры, а Россия вместо удвоения ВВП столкнулась с мировым финансовым кризисом. Эти примеры вовсе не говорят о том, что все поисковые прогнозы бессмысленны, напротив, чаще всего выполняются именно они.

Нормативный прогноз — это прогноз, содержанием которого является определение путей и сроков достижения возможных состояний объекта прогнозирования в будущем, принимаемых в качестве цели. Так, например, в рамках нормативного прогноза Н. С. Хрущев мог бы говорить о том, какими должны быть показатели народного хозяйства страны и в какие сроки необходимо их достичь для того, чтобы к 1980 г. в СССР был построен коммунизм.

Очевидно, что нормативный прогноз более тесно связан с оптимальным планированием, чем поисковый. Хотя такое противопоставление и не совсем корректно — поисковые прогнозы осуществляются в тех случаях, когда объект прогноза не поддается воздействию со стороны субъекта прогнозирования, а объект нормативного прогноза вполне управляем. Вот как это понимается специалистами: «Различают поисковые и нормативные прогнозы. Первые обычно предшествуют вторым. Поисковый прогноз выявляет не одну единственную или несколько возможностей, а их широкий спектр вне зависимости от целей управления. Это позволяет выбрать затем такое управленческое решение, которое наилучшим образом соответствует как целям и особенностям системы управления, так и законам объекта управления»[1].

Понятно, что деление прогнозов на поисковые и нормативные, скорее, относится к организации прогнозирования, нежели к задачам выбора метода прогнозирования, поэтому в дальнейшем мы не будем использовать эту классификацию.

Иногда выделяют условные и безусловные прогнозы.

Условный прогноз осуществляется, исходя из постановки задачи, структура которой может быть выражена условием: «Если A примет значение A_1, то B примет значение B_1». Такой прогноз позволяет выяснить возможные состояния объекта при тех или иных условиях.

Безусловный прогноз определяет будущее объекта без учета каких-либо условий, например: «B примет значение B_1». Полагаем, что можно согласиться с мнением о том, что безусловный прогноз — это разновидность условного прогноза, и его действительно можно сформулировать так: «Если ничего не будет меняться, то B примет значение B_1».

1.4. Прогнозирование как функция управления

Социально-экономическое прогнозирование как функция управления относительно автономно от других функций управления. В нем имеются оригинальные, на других этапах управления не встречающиеся, модели и методы, подходы и принципы. Но эта автономность условна, поскольку прогнозы выполняются чаще всего не для того, чтобы удовлетворить собственное любопытство, но для того, чтобы, воспользовавшись неким предположительным знанием о будущем, подготовиться к нему.

---

[1] *Алексеев Г. Н.* Прогнозное ориентирование развития энергоустановок. М. : Наука, 1978. С. 9.

В общем случае в управлении выделяют последовательные и взаимосвязанные функции: целеполагание, прогнозирование, планирование, организацию, контроль и регулирование. Прогнозирование, как можно увидеть, следует за целеполаганием, т.е. за формулировкой цели управления. Поэтому и прогнозирование осуществляется не само по себе, а в рамках, определенных целеполаганием: из целеполагания следует круг задач, которые необходимо выполнить при прогнозировании, и объекты, прогнозное состояние которых следует определить.

Поскольку на основании прогнозных значений в последующем разрабатываются планы и осуществляются все функции управления, становится понятно, что точность прогноза напрямую влияет на эффективность планирования действий по достижению поставленных целей и, в конечном итоге — на эффективность управления в целом. Покажем это, используя формальную постановку задачи принятия экономического решения.

Если разобрать ситуацию принятия любого решения, в том числе и в социально-экономической области, то можно выделить его составные части.

Обозначим через b_j параметры решения, которые в терминах формальной теории принятия решений могут быть названы условиями принятия решений или ограничивающими условиями. Здесь j — номер параметра решения ($j = 1, 2, ..., M$). Такими параметрами могут быть, например, количество потребителей товара или емкость целевого рынка; сроки реализации хозяйственного решения или располагаемый бюджет на реализацию проекта; размер располагаемых ресурсов и т.д. Наличие параметров решения является обязательным условием задачи принятия решений. Действительно, если отсутствуют какие-либо ограничения, разве возникает задача выбора?

Второй составляющей частью ситуации принятия решений является наличие альтернативных вариантов решения, количественные характеристики которых можно обозначить как x_i. Здесь $i = 1, 2, 3, ... N$ — число этих альтернативных вариантов. Число N в общем случае не совпадает с числом параметров решения M ($N \neq M$), но это не обязательное условие, поскольку число альтернатив и параметров решения никак друг с другом не связаны. Если бы в момент принятия хозяйственного решения не существовало альтернативных вариантов решения, а имелось бы только одно, то и выбора как такового также не было бы. Потому и сама задача выбора наилучшего решения в таком случае не ставится. Если есть одно единственное решение, то из чего выбирать?

Область допустимых альтернативных вариантов определяется величинами параметров решения. Каждый из вариантов x_i взаимосвязан с параметрами решения некоторым соотношением $f(x_i)$. Например, выпуск x_1 единиц первого товара требует затрат времени, равного $f(x_1)$, а для выпуска x_2 единиц второго товара — $f(x_2)$. Это и есть величины параметров решения.

Бесцельных решений не бывает. Если нет цели, то нет и необходимости принятия решения. Но каждый из альтернативных вариантов решения способствует достижению поставленной цели в разной степени. Поэтому целевая установка, формализуемая в виде некоторого критерия оптимальности, может быть представлена в форме зависимости от значений альтернативных вариантов решений ($Q = F(x_i)$).

Тогда задача принятия решения сводится к нахождению оптимального значения набора альтернативных вариантов решения с учетом параметров решения с целью достижения оптимума целевой установки.

Пусть для определенности рассматривается задача поиска наилучшего управленческого решения, максимизирующего целевую функцию:

$$Q = F(x_i) \to \max \tag{1.5}$$

при наличии ограничений по параметрам решений:

$$f_j(x_i) \le b_j \tag{1.6}$$

и положительности величины каждого из альтернативных вариантов:

$$x_i \ge 0. \tag{1.7}$$

По сути, задача поиска наилучшего варианта решения сведена к задаче оптимизации, с которой знаком любой бакалавр экономики еще с первых курсов обучения. Если задача уже сформирована и представлена, например, в виде задачи линейного программирования, ее можно довольно просто решить, используя симплекс-метод. Но в реальной экономической практике не бывает такого, чтобы кто-то взял и представил для решения задачу в удобной форме (1.5)—(1.7). Само решение задачи по выбору лучшего варианта — заурядное математическое действие в сложном процессе формирования и принятия решения, его логичное завершение. Сложность состоит в формировании этой задачи, в наполнении ее смыслом и реальными данными. Для того чтобы понять сложность этого процесса, рассмотрим последовательность этапов сведения задачи к виду, удобному для выбора наилучшего решения.

Пусть цель принятия хозяйственного решения сформулирована. Формализованная цель может быть представлена в виде критерия, который обычно позволяет осуществить точный расчет того, в какой степени достигнута сформулированная целевая установка. Так, например, цель, сформулированная как «завоевание ведущих позиций на целевом рынке», вполне может быть интерпретирована как цель достижения максимально возможного объема валовой прибыли от продаж товара на этом рынке. Если переменные задачи принятия решения x_i представляют собой неизвестные объемы производства товаров i-го типа, а p_i — прибыль от продажи единицы изделия, то критерий (1.5) в модели принятия решения будет иметь вид максимума элементарной суммы:

$$Q = \sum_{i=1}^{N} p_i x_i \to \max. \qquad (1.8)$$

Параметры решения b_j представляются в форме тех условий, которые составляют ситуацию принятия решения. Их разделяют на две группы — внутренние и внешние параметры.

К внешним параметрам принятия решений относят факторы рынка, на котором работает предприятие, чаще всего — конъюнктуру рынка, действия конкурентов, поведение потребителей и т.п., а к внутренним — производственные, экономические, социальные и другие факторы, определяющие внутреннюю структуру и производственные возможности объекта, на котором возникла задача принятия решения.

Поскольку реализация практически любого решения осуществляется не в момент его принятия, а спустя некоторое время, очевидно, что для решения задачи следует не просто собрать и понять информацию о параметрах решения в настоящий момент, а предсказать их количественное и качественное состояние в будущем, к моменту реализации принятого варианта решения. Иначе говоря, стоит задача прогнозирования состояния внешних и внутренних факторов, составляющих ситуацию принятия решений.

Применительно к задаче (1.5)–(1.7) это означает, что необходимо выполнить прогноз цен на единицу выпускаемой продукции и по этим ценам спрогнозировать прибыль, приходящуюся на единицу продукции; далее нужно спрогнозировать все значения параметров решения b_j, и то, как каждый из альтернативных вариантов решения определен в области параметров решения, т.е. спрогнозировать, какими на прогнозный период будут зависимости $f_j(x_i)$.

Предположим, что по каким-то причинам не удалось выполнить прогнозы параметров решения и характеристик альтернативных решений в области этих параметров, а потому их значения для принятия решения приходится указать в некотором интервале, вероятность нахождения внутри которого не известна. Тогда в формализованной математической постановке ограничения по параметрам решений примут вид:

$$f_j(x_i) \pm \Delta f_j(x_i) \leq b_j \pm \Delta b_j, \tag{1.9}$$

где $\Delta f_j(x_i)$ характеризует разброс возможных значений состояний альтернатив решений x_i в области j-го параметра решения; Δb_j определяет вариацию параметра решения, вызванную действием случайных факторов и неопределенности, которую не удалось уменьшить из-за отсутствия достоверных прогнозов.

Очевидно, что задачи (1.5), (1.7) при ограничениях (1.9) решить невозможно — в зависимости от того, какие границы параметров решений или функций альтернатив брать, получаются самые разные варианты. Но поскольку будущее не определено, совсем не ясно, какому же из вариантов состояния параметров решений и функций отдать предпочтение? В этой ситуации следует оставить надежды на получение оптимального решения, а довольствоваться лишь поиском рационального решения с помощью методов теории игр. А поскольку принимается не наилучшее, а не самое худшее решение, то очевидно, что эффективность и решения, и его реализации, и результатов управления будет не самой высокой.

Следовательно, социально-экономическое прогнозирование можно и нужно рассматривать как метод уменьшения неопределенности в принятии управленческих решений. Предприятия и организации, не занимающиеся прогнозированием или делающие это не на высоком профессиональном уровне, обречены на неэффективность своей работы, и как следствие этого — на проигрыш в конкурентной борьбе.

Поэтому в процессе управления любой организацией выполняются социально-экономические прогнозы. Другое дело, что в зависимости от вида организации эта важнейшая функция управления либо выделяется в отдельную постоянно повторяющуюся задачу, решение которой поручается некоторой организационной структуре, либо выполняется несколькими исполнителями по мере надобности. Второе характерно для малых и средних организационных форм. Крупные предприятия и организации используют первый вариант осуществления прогнозирования (для этого создаются специальные отделы или секторы, приобретается соответствующая техника и программное обеспечение, принимаются на работу специалисты нужной квалификации и т.п.). В государственном управлении на любом уровне (региональном, общегосударственном) прогнозы социально-экономического развития выполняют специально созданные для этого структуры. Так, на федеральном уровне управления социально-экономические прогнозы выполняются в структуре Министерства экономического развития РФ. На основе этих прогнозов правительство РФ принимает соответствующие решения.

1.5. Временные ряды социально-экономической динамики

Экономика работает с тремя *типами данных*:

1) *пространственными* – данными, собранными в какой-то момент времени по нескольким объектам (например, данные, полученные в результате разового опроса населения);

2) *временны́ми* – данными, собранными в разные моменты времени по одному объекту (например, месячные данные по продажам валенок Гатчинского промкомбината);

3) *панельными* – данными, собранными в разные моменты времени по нескольким объектам (например, данные, полученные в результате ежегодных опросов домохозяйств).

Задачи социально-экономического прогнозирования при этом можно разделить на три класса:

1) прогнозирование пространственной экономики;

2) прогнозирование экономической динамики;

3) прогнозирование динамики экономического пространства.

Пространственная экономика представляет собой совокупность хозяйственных систем, подсистем и элементов, взаимосвязь между которыми рассматривается в пространстве (региональные взаимосвязи, межотраслевые взаимодействия, поведение потребителей, результаты их опроса и т.п.). При прогнозировании пространственной экономики не учитывается фактор времени.

При прогнозировании экономической динамики все социально-экономические системы, подсистемы и элементы, а также взаимосвязи между ними рассматриваются во времени. Модели и методы прогнозирования социально-экономической динамики используют временные ряды. Поэтому важно определить сущностные, отличительные характеристики этого понятия — «временной ряд» — и дать ему определение.

Прежде всего, следует указать на то, что в учебной и научной литературе часто вместо понятия «временной ряд» употребляют понятие «динамический ряд». Очевидно, что это синонимические понятия, поэтому мы будем использовать их в одном и том же контексте.

Анализу и прогнозированию временных рядов уделено очень много внимания, как в отечественной, так и в зарубежной литературе. Представленные в ней характеристики временных рядов отличаются друг от друга в зависимости от того, о временных рядах каких показателей ведется речь. Часто ученые, не вдаваясь в суть временно́го ряда, определяют его формально, например: «временные ряды — это барометр; они наглядно отображают наиболее существенные тенденции развития»[1], или «одни ряды представляют собой последовательность чисел, отражающих величину того или иного показателя во времени. Это так называемые временные ряды, или ряды динамики»[2], или «временным рядом, хронологическим рядом, проще хроникой называют последовательность упорядоченных во времени наблюдений, обозначаемых, например, $x_1, x_2, \dots, x_t, \dots, x_T$»[3]; «ряд динамики — это такой способ записи случайной величины (признака, фактора) Y, при котором ее значения (уровни) $y_i = y(t_i)$ приводятся в зависимости от времени t_i»[4].

---

[1] *Хауштейн Г.* Методы прогнозирования в социалистической экономике. М. : Прогресс, 1971. С. 149.

[2] *Громыко Г. Л.* Статистические ряды в экономических и экономико-географических исследованиях. М. : Изд-во МГУ, 1974. С. 4.

[3] *Маленво Э.* Статистические методы эконометрии. Вып. 2. М. : Статистика, 1976. С. 32.

[4] *Лапыгин Ю. Н., Крылов В. Е., Чернявский А. П.* Экономическое прогнозирование : учеб. пособие. М. : Эксмо, 2009. С. 105.

Понятно, что если мы наблюдаем изменение какого-либо показателя во времени и фиксируем его значения по этому признаку, то, в соответствии с вышеприведенными определениями, это будет временно́й ряд. Но тогда временным рядом можно назвать любой наблюдаемый во времени процесс — нагревание жидкости в колбе при проведении химических опытов, плавка металла в печах, движение небесных тел и т.д. Но интуитивно каждый экономист понимает, что к одному и тому же классу временных рядов нельзя отнести, например, фиксируемые через равные промежутки времени показатели изменения длины металлического сердечника при его нагревании и ряд изменения цен на рынке *Forex*, фиксируемых за те же промежутки времени. В первом случае время только характеризует точки отсчета, а во втором — оно означает нечто гораздо большее. Следовательно, для того, чтобы при прогнозировании временных рядов не сделать ошибки методологического плана, следует более тщательно изучить сам феномен временно́го ряда и точно определить, все ли протекающие во времени процессы могут представлять собой временные ряды или только часть из них. Ответ на этот вопрос принципиально важен, ведь если перед прогнозистом имеется ряд, называющийся «временным», и он знает о методах прогнозирования временных рядов, изложенных в многочисленных учебниках, то что ему мешает использовать эти методы для прогнозирования данного ряда? Если же временные ряды представляют собой не простое упорядочивание во времени, а обладают принципиально отличными характеристиками, что и делает их временными, то эти отличия следует четко определить.

Итак, посмотрим, как различные ученые определяют само понятие «временно́й ряд». При этом, конечно, следует обратиться и к работам ученых, которые вводили это понятие в научный оборот.

В одной из первых работ в этом направлении — в монографии Т. Андерсона «Статистический анализ временных рядов», которая получила большую популярность у прогнозистов, об этом говорится так: «временным рядом называют последовательность наблюдений, обычно упорядоченную во времени, хотя возможно упорядочение и по какому-то другому параметру... Если во многих задачах наблюдения статистически независимы, то во временных рядах они, как правило, зависимы, и характер этой зависимости может определяться положением в последовательности»[1]. И далее: «в статистическом анализе последовательность T наблюдений, образующих временной ряд, часто рассматривают как выборку T последовательных наблюдений через равные промежутки времени из существенно более продолжительной (генеральной) последовательности случайных величин»[2].

Это определение подводит базу для использования в прогнозировании временных рядов методов математической статистики, поскольку временной ряд представляется как некоторая выборка из генеральной совокупности случайных величин и для его моделирования приемлем выборочный метод, как раз и лежащий в основе математической статистики.

Необходимо отметить, что эта точка зрения не только очень распространена, но и превалирует в прикладных работах по прогнозированию, в том числе по прогнозированию социально-экономических систем. Так, например, в справочнике по математике для экономистов, выпущенном в 2007 г. в разделе «Временные ряды», говорится следующее:

«Большинство экономических задач связано с оценкой основных экономических показателей во времени и с прогнозом этих показателей на будущие моменты времени. Это значит, что основные экономические характеристики необходимо рассматривать как случайные функции.

Но так как статистика оперирует выборочными значениями показателей, то из случайных функций производится выборка в дискретные моменты времени. В результате получаются так называемые временные ряды»[3].

Разберем более тщательно эти два абзаца, ведь нам следует определиться с тем, что представляет собой временнóй ряд и понять логику того, как авторы учебника обосновывают случайность временных рядов.

Первый абзац состоит из двух предложений, которые, если их внимательно изучить, никак друг с другом не взаимосвязаны. Первое из них говорит о том, что в экономике необходимо оценивать показатели во времени и прогнозировать их. Второе — содержит утверждение о том, что эти харак-

[1] *Андерсон Т.* Статистический анализ временных рядов. М. : Мир, 1976. С. 11.

[2] Там же. С. 406.

[3] Справочник по математике для экономистов : учеб. пособие / под ред. проф. В. И. Ермакова. М. : ИНФРА-М, 2007. С. 383.

теристики следует рассматривать как случайные функции. Очевидно нарушение логики вывода, поскольку из первого не следует второе. В этом выводе, если рассматривать его как дедуктивный, отсутствует большая посылка. Ее смысл легко понимается из сути вывода. Применительно к рассматриваемому случаю она должна быть сформулирована так: «все оцениваемые во времени и прогнозируемые показатели являются случайными функциями». Тогда понятно, что раз «большинство экономических задач связано с оценкой основных экономических показателей во времени и с прогнозом», то это значит, что основные экономические характеристики необходимо рассматривать как случайные функции.

Но вывод по силлогизму верен в том случае, когда верна бóльшая посылка. Но верна ли она? Разве все оцениваемые во времени и прогнозируемые показатели являются случайными функциями? Например, если в 2012 г. оценить, каким днем недели будет 9 мая 2015 г., разве эта оценка будет случайной? То, что это будет субботой, вовсе не случайно, а со всей очевидностью следует из принятого во всем мире календаря. А поскольку бóльшая посылка неверна, неверен и сам вывод о том, что «основные экономические характеристики необходимо рассматривать как случайные функции». Но именно этот недоказанный вывод является большей посылкой для силлогизма второго абзаца. Из нее делается вывод о том, что, если экономические процессы и их характеристики «случайны», то для их анализа надо использовать методы математической статистики. В их основе лежит выборочный метод, а поэтому значения, которые экономист получает о процессах, протекающих в дискретные моменты времени, являются выборочными и, значит, временной ряд представляет собой совокупность выборочных значений!

Но так как посылка о том, что характер экономических процессов является случайным, не обоснована, то и все остальные выводы, базирующиеся на ней, не являются обоснованными. Поэтому и определение временнóго ряда, которое предлагается в упомянутом учебном пособии и следующее из этих ошибочных рассуждений, не является обоснованным.

Отношение к временному ряду под влиянием авторитета математической статистики и ее представителей в экономике до сих пор остается прежним — и в современных учебниках по эконометрике, и в некоторых учебниках по прогнозированию «временным рядом называют последовательность наблюдений, обычно упорядоченных во времени»[1]. Не возражая в принципе по поводу этого определения, следует отметить, что оно неполное — с его помощью невозможно отличить динамические ряды от рядов, упорядоченных во времени.

Социально-экономические процессы можно рассматривать с позиций двух подходов — статического и динамического.

В классической эконометрике наибольшее распространение получил именно статический подход, хотя исследователи, использующие эконометрические методы, убеждены в их соответствии реальным процессам и динамичности. Практически все эконометрические модели отражают экономическую реальность в некоторой застывшей неизменности структуры, взаимосвязей и равновесия элементов. В эконометрии оперируют динамическими данными и в подавляющем большинстве случаев эту динамику представляют как некоторый упорядоченный во времени набор срезов экономических явлений. При этом считается, что структура среза, его количественные и качественные изменения пропорциональны масштабу упорядочения, в качестве которого выступает время. Если при этом используются методы математической статистики, оперирующие такими понятиями колеблемости, как дисперсия, средняя, размах колебаний и т.п., то создается видимость динамического подхода. *Но динамический подход рассматривает экономические* явления в процессе изменения не только самих экономических систем и их элементов, но и соотношений между ними и ищет закономерности в ходе самих изменений. Для статики основной предпосылкой анализа экономического процесса является неизменяемость, тождественность происходящих процессов. Для динамики — непрерывность процессов изменения всех взаимосвязей и показателей.

Н. Д. Кондратьев еще в 1924 г., говоря о статике, динамике и конъюнктуре, указывал: «В тех случаях, когда элементы экономической жизни или их связи подвергаются изменениям, не исчерпывающимся изменением их числа, объема и вообще не сводимых к количественным изменениям, мы говорим о наличии качественных изменений. Сюда относятся, например, изменения в технике производства, в орга-

---

[1] *Лапыгин Ю. Н.* Экономическое прогнозирование : учеб. пособие. М. : Эксмо, 2009. С. 32.

низации хозяйства, в составе и характере общественных потребностей и т.д.»[1]. Именно качественные изменения приводят к необратимости изменения структуры экономических систем. Сами процессы динамики экономических систем представляют собой синтез целого ряда динамических процессов. Выделив и смоделировав эти процессы, можно приблизиться к адекватному отражению экономической реальности.

Н. Д. Кондратьев выделял эволюционные и волнообразные динамические процессы как наиболее существенные в экономическом развитии. К эволюционным он относил процессы, происходящие постепенно, незаметно меняя качественную характеристику, делая невозможным переход от новой структуры к первоначальной, из которой началось эволюционное развитие. Под волнообразными Н. Д. Кондратьевым понимались процессы, которые непрерывно меняют направление своего развития, постепенно возвращаясь к первоначальному моменту.

Естественно, что в процессе развития любой экономической системы на нее воздействует множество случайных факторов, поэтому, рассматривая составляющие динамики, необходимо отметить еще и случайность некоторых процессов.

Таким образом, можно сделать вывод о том, что, по мнению Н. Д. Кондратьева, развитие любой экономической системы складывается из эволюционного, волнообразного и случайного компонентов. Однако на практике этот вывод искажается, превращаясь из динамического в статический. Классифицируя временные ряды и верно отмечая их нестационарность, приверженцы такой статической точки зрения выводят *четыре типа временных рядов*.

1. Ряды с тенденцией роста, но без периодической составляющей. Такие ряды имеют тренд, который в среднем может быть описан с помощью известных непериодических функций (чаще всего, — с помощью полинома N-й степени).

2. Ряды, имеющие помимо тренда ярко выраженные сезонные колебания.

3. Ряды без периодической составляющей и тенденции роста.

4. Ряды со сложной структурой, включающие всевозможные виды колебаний, в частности, сезонные и циклические.

---

[1] *Кондратьев Н. Д.* Проблемы экономической динамики / под ред. Л. И. Абалкина [и др.]. М. : Экономика, 1989. С. 58.

Структура таких временных рядов в общем случае не может быть однозначно описана с помощью известных функций, поскольку для разных участков временного ряда набор этих функций будет различным, т.е. в этой структуре можно говорить о временных рядах с переменной структурой. Важная отличительная черта этих временных рядов, по мнению автора классификации, состоит в том, что «тренд в среднем значении и дисперсии должен рассматриваться не как детерминированная функция времени, а как случайная функция, изменяющаяся по мере развития процесса»[1].

Может сложиться впечатление, что указанная выше классификация не статична, а динамична, однако это не так. Обращает на себя внимание тот факт, что в данной классификации ни разу не был упомянут эволюционный характер развития, присущий динамическому подходу. В качестве показателя такого эволюционного развития мог бы служить тренд, но автор не зря упомянул по отношению к нему слово «в среднем».

«Итак, уже при рассмотрении математического ожидания и дисперсии случайного процесса видно, что он как бы разбивается на некоторую систематическую составляющую (среднюю) и случайные отклонения от нее. При анализе временных рядов это находит свое практическое выражение в представлении ряда y_t в виде суммы $y_t = f(t) + \varepsilon_t$, где $f(t)$ — некоторая неслучайная функция времени; ε_t — случайная величина с нулевой средней и дисперсией.

Функцию $f(t)$, характеризующую детерминированную часть временного ряда y_t, назовем трендом»[2].

Таким образом, тренд, который должен отражать эволюционный характер развития, в соответствии с данной точкой зрения является средней или систематической составляющей (иначе говоря — оценкой математического ожидания), да к тому же, еще и детерминированной части временного ряда. Поэтому тренд здесь следует рассматривать как некоторую детерминированную модель, у которой при данной совокупности входных значений на выходе может быть получен единственный результат (за исключением случайных колебаний). Структура такой модели, естественно, является застывшей, статичной.

---

[1] *Френкель А. А.* Прогнозирование производительности труда: методы и модели. М. : Экономика, 1989. С. 25.

[2] Там же. С. 15.

Такой подход часто используется при статистическом моделировании в экономике. Так, например, в одном из учебников по прогнозированию при описании методологии экстраполяционных методов утверждается, что «рассматриваемый процесс изменения переменной представляет собой сочетание двух составляющих — регулярной и случайной... Считается, что регулярная составляющая $f(A, x)$ представляет собой гладкую функцию от аргумента, описываемую конечномерным вектором параметров A, которые сохраняют свои значения на период упреждения прогноза. Эта составляющая называется также трендом, уровнем, детерминированной основой процесса, тенденцией»[1].

Можно продолжить ряд примеров попыток моделирования динамики с использованием именно статического подхода, называемого при этом «динамическим» — многие работы (и не только отечественные) по прогнозированию и моделированию экономической динамики опираются на методологию статического подхода.

Читатель может возразить, что выше были приведены взгляды сторонников статического подхода, отраженные в литературе прошлых лет, а сегодня авторы публикаций наверняка разделяют динамическую точку зрения (по Кондратьеву). В ответ приведем пример из одного учебника по эконометрике, достаточно популярного в наши дни: «Выше, говоря о расположенной в хронологическом порядке последовательности наблюденных значений ...какого-либо признака, мы, по существу, уже дали определение временного ряда... нас будет интересовать лишь некоторый подкласс подобных последовательностей, а именно тот, который связан с наблюдениями стохастических по своей природе признаков. Другими словами, мы исключаем из рассмотрения детерминированные схемы динамических наблюдений, при которых элементы последовательности могут быть в точности вычислены как значения некоторой неслучайной функции $f(t)$, т.е. $x(t) = f(t)$. Поэтому уточним понятие временного ряда, принятое в данной главе... Ряд наблюдений $x(t_1)$, $x(t_2)$, ..., $x(t_n)$ анализируемой случайной величины $\xi(t)$, произведенных в последовательные моменты времени t_1, t_2, ..., t_n, называется временным рядом»[2].

---

[1] Теория прогнозирования и принятия решений / под ред. С. А. Саркисяна. М. : Высшая школа, 1977. С. 82.

[2] *Айвазян С. А., Мхитарян В. С.* Прикладная статистика и основы эконометрики : учебник для вузов. М. : ЮНИТИ, 1998. С. 780.

Из приведенного отрывка следует, что авторы рассматривают временные ряды, имеющие только «стохастические по своей природе признаки» и «детерминированные схемы». Про эволюционные процессы здесь и не упоминается, следовательно, авторы учебника придерживаются статического подхода.

Читатель может самостоятельно изучить определения понятия «временной ряд», например, в Интернете и убедиться в том, что такой подход к временному ряду в современной науке превалирует.

Следует отметить, что взгляд на временной ряд как на простую упорядоченную во времени последовательность наблюдений подвергался критическому переосмыслению со стороны экономистов. Так, например, еще в 1984 г. отечественный ученый Н. К. Дружинин, не считавший временные ряды простой упорядоченной совокупностью наблюдений, по этому поводу писал: «можно ли сказать об экономическом временном ряде, что он представляет собой совокупность, формирование которой происходит по принципу случайности, допускающему применение корреляционных исчислений? По-видимому, нельзя, так как члены этого ряда упорядочены во времени и между ними обычно обнаруживается определенная связь, имеющая в своей основе экономические причины и статистически выражаемая так называемой автокорреляцией»[1].

Примерно такой же точки зрения придерживаются В. Н. Афанасьев и М. М. Юзбашев, которые считают временным рядом «последовательность упорядоченных во времени числовых характеристик показателей, характеризующих уровень состояния и изменения изучаемого явления»[2]. По их мнению, время не просто индекс упорядочивания показателей, а активная идентифицирующая переменная.

Поясним, что именно, на наш взгляд, дает основание тот или иной подход в экономике отнести к динамическому. Для этого воспользуемся некоторыми понятиями системного подхода — одного из методологических принципов познания. Важнейшим системным свойством является эмерджентность. Именно это свойство представляет собой отражение в эко-

---

[1] *Дружинин Н. К.* О ложной корреляции // Экономика и математические методы. Т. XX. 1984. Вып. 4. С. 723.

[2] *Афанасьев В. Н., Юзбашев М. М.* Анализ временных рядов и прогнозирование : учебник. М. : Финансы и статистика ; ИНФРА-М, 2010. С. 6.

номике одного из главных законов диалектики — перехода количественных изменений в качественные. В соответствии с этим свойством при объединении элементов в систему она приобретает качественно иные свойства, которых нет у элементов, находящихся в изолированном состоянии. С учетом другого свойства экономических систем — динамичности развития — мы должны прийти к выводу о том, что количественное изменение элементов системы во времени приводит и к возникновению новых элементов, которые, присоединяясь к структуре системы, меняют ее качество. Таким образом, к динамическому можно отнести только тот подход, который опирается на положение о непрерывном изменении качественных свойств рассматриваемой системы и систем, с которыми она находится во взаимосвязи. При этом сила и направление взаимодействий, как между системами, так и между отдельными элементами самой системы, непрерывно меняются, приводя к изменению структуры системы и количественных показателей ее развития. Именно в этом проявляется динамичность всех экономических систем, как и диалектика самой природы.

Безусловно, динамика развития больших систем (физических, технических, биологических, экономических) имеет ряд одинаковых свойств. Но различия между системами настолько значительны, что эта динамика для каждой из них приобретает ряд таких отличительных черт, которые делают несерьезными попытки говорить о моделировании и разработке единых методов исследования большой системы вообще и переносить какие-либо свойства и методы моделирования динамики системы одной природы на конкретные системы другой природы. Даже побудительные мотивы развития больших систем существенно различаются, не говоря уже о внутренней структуре, силе и характере взаимодействий между элементами.

Действительно, если для экономических систем характерно развитие с целью достижения некоторого известного (пусть даже не количественно, но качественно) оптимума, как правило, многоцелевого, то для физических систем такого оптимума нет, и их развитие осуществляется в соответствии с действующими физическими законами, а не в результате некоторого целенаправленного управления. При этом надо иметь в виду, что цели экономического управления со временем чаще всего меняются. Так разве можно при моделировании развития настолько различных и несхожих экономических и физических систем использовать одинаковые методы? Конечно же, нет!

Но, как было показано, практически все исследователи определяют динамический ряд как некоторую последовательность значений показателя во времени. Таким образом, статистические данные о развитии экономических, технических и физических систем, упорядоченные во времени, в одинаковой степени можно назвать динамическими. Но из общности названия вовсе не следует, что это ряды, обладающие одинаковыми свойствами, и их обработку необходимо производить одними и теми же статистическими методами.

Здесь за видимой простотой скрывается достаточно сложная проблема. Для ее разрешения вновь придется вернуться к двум принципиальным позициям (статической и динамической) при анализе количественных показателей, отражающих различные качественные процессы.

Технические, физические и другие системы, относящиеся к классу естественнонаучных дисциплин, могут развиваться, изменяясь во времени. Но этот процесс развития практически никогда не является эволюционным. Для динамических рядов, отражающих изменение таких систем, время является лишь индексом упорядочения этих данных. При этом практически неважно, какое численное значение приобретает индекс t. Важно лишь то, что в череде развития того или иного процесса показатель с этим индексом стоит на втором, двадцатом или двухтысячном месте (впрочем, в целом ряде *случаев изучения технических систем и физических явлений* и это неважно). Все остальные показатели также занимают свое место в соответствии с полученным индексом.

Когда же дело касается органической природы, общества или экономики, время становится уже не только индексом, но и показателем эволюционного развития. Никто из экономистов никогда не скажет, что ему все равно, сколько ему лет — 20 или 120! Так почему же показатели роста экономических систем, которые так же эволюционируют, как и организм человека, упорядочивая во времени, рассматривают только как отмасштабированные и систематизированные данные? Очевидно, что рост человека в пять и в 50 лет отражает не только процесс развития организма, но и его качественно различные состояния. Точно так же данные производства промышленности любой страны в 1975 г. и 2005 г. отражают и процесс развития системы промышленности, и качественно различные ее состояния.

Ни один нормальный врач при осмотре 70-летнего пенсионера на медицинском приеме в поликлинике не поинтересуется ростом пациента в семимесячном возрасте, так как качественное состояние пациента в семимесячном возрасте отличается от его качественного состояния в момент приема, если, конечно, не рассматривается случай врожденной патологии. Однако сторонники статичного подхода, исследуя отнюдь не патологические случаи экономического роста, при моделировании экономики стремятся собрать как можно больше статистической информации, считая, что чем больший временной период охватывают статистические данные, тем точнее окажется предсказание будущего!

Если при этом временные данные считаются одинаково важными, то понятен вывод о том, что «последовательным значениям t соответствуют разные наборы значений признаков того же самого "характера". Действительно, ведь это те же самые признаки, только относящиеся к различным моментам времени»[1]. Так как это «те же самые признаки», то они отражают «тот же самый» процесс, который не претерпевает эволюционных изменений. А признание неизменности процесса есть не что иное, как следование методологии статичного подхода.

Если использовать динамический подход к временным рядам социально-экономической динамики, то отношение к ним будет принципиально иным. Необходимо отличать временной ряд y_t, где t — время, и упорядоченный ряд y_i, где i — порядковый номер (упорядочивание может происходить и во времени, тогда $i = t$). Как это сделать?

Сначала зафиксируем отличия временного ряда от любого другого ряда. Нет смысла называть ряд временным, если время t просто выступает как индекс i и служит для упорядочивания значений ряда. Для этого достаточно просто заявить о том, что перед исследователем находится упорядоченный по какому-то признаку ряд наблюдений. Если вводить в название ряда прилагательное, определяющее его, то это должно иметь принципиальный смысл, который приводит к отличию данного ряда от других упорядоченных рядов.

Время в динамическом ряде выполняет не просто роль некоторого индекса и номера наблюдения, а характеризует конкретное место нахождения значения данного показателя

[1] *Плюта В.* Сравнительный многомерный анализ в эконометрическом моделировании. М. : Финансы и статистика, 1989. С. 25.

в имеющемся ряде наблюдений. Что произойдет, если в ряде наблюдений y_i сами наблюдения поменять местами? Изменятся ли характеристики этого ряда? Нет, не изменятся. А что произойдет, если поменять местами значения временного ряда y_t? Ряд приобретет совсем другой вид и другое содержание, поскольку время во временном ряду отражает не только характеристики самого ряда, но и условия, сложившиеся к этому моменту наблюдения. В первом случае простого ряда, если вместо индекса i использовать индекс t, это будет характеризовать просто номер наблюдения в последовательности значений, наблюдаемых друг за другом в дискретном времени. Ни о каких условиях, соответствующих этому t, говорить нельзя — и индекс i, и заменяющий его индекс t играют пассивную роль способа фиксации наблюдения. Во втором случае время играет активную роль в идентификации происходящих процессов, поэтому и переставлять значения временных рядов нельзя. Если этого не будет, то ряд не является временным. Это особенно важно для рядов социально-экономической динамики — можно ли, например, для анализа значений курса доллара на ММВБ взять и перемешать все эти наблюдения, переставив их местами? Нельзя. А можно ли перемешать, например, значения ряда урожайности зерновых на одинаковых полях с внесением в них одинакового количества удобрений? Можно, так как суть ряда не изменится.

Тогда можно определить временной ряд социально-экономических показателей следующим образом: *временным (динамическим) можно назвать упорядоченный во времени ряд наблюдений, в котором время наблюдения характеризует особенность состояния внешних и внутренних факторов поведения объекта наблюдения, в результате чего формирование ряда осуществляется неслучайным образом.*

1.6. Обратимые и необратимые процессы

Среди всех возможных задач социально-экономического прогнозирования существенную часть занимают задачи прогнозирования динамики, задачи, методы решения которых опираются на обработку временных рядов. Временной ряд может формироваться в разных условиях, которые предопределяют его характеристики и свойства. В наиболее общем случае он может отражать два типа процессов — обратимые и необратимые.

Обратимые процессы представляют собой такой тип развития объекта, который характеризуется постоянством качественных характеристик при изменении его количественных показателей как результат реакции на изменение внешних условий. Если внешние условия остаются стабильными, изменение количественных характеристик обратимого процесса определяется только действием случайных факторов ε. Из этого понимания обратимого процесса следует, что если в ходе его развития внешние условия x_t станут такими же, как прежде ($x_t = x_0$), то количественные характеристики процесса y_t вернутся к прежним величинам ($y_t = y_0$), искажаемым разве что влиянием случайных факторов $y_0 + \varepsilon$.

В отличие от обратимого необратимый процесс происходит в условиях, когда претерпевают изменения его количественные и качественные характеристики. Под качественными изменениями в процессе следует понимать изменение его структуры, состава элементов, силы и направления взаимосвязи между ними. Тогда под *необратимым процессом* будем понимать такой тип развития объекта, который характеризуется изменением во времени его количественных характеристик, а также структуры элементов, степени и направления взаимосвязей между ними. Это означает, что если внешние условия протекания процесса остаются стабильными, в нем все равно продолжают происходить изменения, вызванные внутренними факторами и силами, постепенно меняющими структуру процесса. Процесс претерпевает во времени необратимые изменения, поэтому, если в ходе его развития внешние условия x_t станут такими же, как прежде ($x_t = x_0$), то количественные характеристики процесса y_t будут существенно отличаться от прежней величины ($y_t \neq y_0$).

Следует обратить внимание на то, что в определение необратимого процесса введено «время» как ключевое понятие, а в определении обратимого процесса его нет. Это дает дополнительную возможность для толкования понятия «временной ряд», приведенного в предыдущем параграфе.

Любой наблюдаемый процесс протекает во времени. В случае, когда изменения в процессе вызваны его реакцией на действия внешних факторов без изменения его качественных характеристик, этот ряд значений отражает обратимые процессы и называть его «временным» не имеет смысла, поскольку время здесь выступает просто как индекс. Но если протекающий во времени процесс претерпевает в имеющемся ряде наблюдений. Что произойдет, если в ряде наблюдений y_i сами наблюдения поменять местами? Изменятся ли характеристики этого ряда? Нет, не изменятся. А что произойдет, если поменять местами значения временного ряда y_t? Ряд приобретет совсем другой вид и другое содержание, поскольку время во временном ряду отражает не только характеристики самого ряда, но и условия, сложившиеся к этому моменту наблюдения. В первом случае простого ряда, если вместо индекса i использовать индекс t, это будет характеризовать просто номер наблюдения в последовательности значений, наблюдаемых друг за другом в дискретном времени. Ни о каких условиях, соответствующих этому t, говорить нельзя — и индекс i, и заменяющий его индекс t играют пассивную роль способа фиксации наблюдения. Во втором случае время играет активную роль в идентификации происходящих процессов, поэтому и переставлять значения временных рядов нельзя. Если этого не будет, то ряд не является временным. Это особенно важно для рядов социально-экономической динамики — можно ли, например, для анализа значений курса доллара на ММВБ взять и перемешать все эти наблюдения, переставив их местами? Нельзя. А можно ли перемешать, например, значения ряда урожайности зерновых на одинаковых полях с внесением в них одинакового количества удобрений? Можно, так как суть ряда не изменится.

Тогда можно определить временной ряд социально-экономических показателей следующим образом: *временным (динамическим) можно назвать упорядоченный во времени ряд наблюдений, в котором время наблюдения характеризует особенность состояния внешних и внутренних факторов поведения объекта наблюдения, в результате чего формирование ряда осуществляется неслучайным образом.*

1.6. Обратимые и необратимые процессы

Среди всех возможных задач социально-экономического прогнозирования существенную часть занимают задачи прогнозирования динамики, задачи, методы решения которых опираются на обработку временных рядов. Временной ряд может формироваться в разных условиях, которые предопределяют его характеристики и свойства. В наиболее общем случае он может отражать два типа процессов — обратимые и необратимые.

Обратимые процессы представляют собой такой тип развития объекта, который характеризуется постоянством качественных характеристик при изменении его количественных показателей как результат реакции на изменение внешних условий. Если внешние условия остаются стабильными, изменение количественных характеристик обратимого процесса определяется только действием случайных факторов ε. Из этого понимания обратимого процесса следует, что если в ходе его развития внешние условия x_t станут такими же, как прежде ($x_t = x_0$), то количественные характеристики процесса y_t вернутся к прежним величинам ($y_t = y_0$), искажаемым разве что влиянием случайных факторов $y_0 + \varepsilon$.

В отличие от обратимого необратимый процесс происходит в условиях, когда претерпевают изменения его количественные и качественные характеристики. Под качественными изменениями в процессе следует понимать изменение его структуры, состава элементов, силы и направления взаимосвязи между ними. Тогда под *необратимым процессом* будем понимать такой тип развития объекта, который характеризуется изменением во времени его количественных характеристик, а также структуры элементов, степени и направления взаимосвязей между ними. Это означает, что если внешние условия протекания процесса остаются стабильными, в нем все равно продолжают происходить изменения, вызванные внутренними факторами и силами, постепенно меняющими структуру процесса. Процесс претерпевает во времени необратимые изменения, поэтому, если в ходе его развития внешние условия x_t станут такими же, как прежде ($x_t = x_0$), то количественные характеристики процесса y_t будут существенно отличаться от прежней величины ($y_t \neq y_0$).

Следует обратить внимание на то, что в определение необратимого процесса введено «время» как ключевое понятие, а в определении обратимого процесса его нет. Это дает дополнительную возможность для толкования понятия «временной ряд», приведенного в предыдущем параграфе.

Любой наблюдаемый процесс протекает во времени. В случае, когда изменения в процессе вызваны его реакцией на действия внешних факторов без изменения его качественных характеристик, этот ряд значений отражает обратимые процессы и называть его «временным» не имеет смысла, поскольку время здесь выступает просто как индекс. Но если протекающий во времени процесс претерпевает качественные и количественные изменения, то время служит фиксатором этих качественных, а значит, необратимых изменений. Такой ряд мы и будем в дальнейшем рассматривать как «временной».

Обратимые процессы могут иметь самую разную природу и самый разный вид, но и эту сложную совокупность можно разделить на две группы: стационарные и нестационарные обратимые процессы.

В учебной и научной литературе стационарный процесс определяют следующим образом.

«Случайные процессы, протекающие во времени приблизительно однородно и имеющие вид непрерывных случайных колебаний вокруг некоторого среднего значения, причем ни средняя амплитуда, ни характер этих колебаний не обнаруживают сущностных изменений с течением времени, называются стационарными.

Условия стационарности ξ_t заключаются в следующем.

1. $M(\xi_t) = \text{const}$;
2. $r\xi_t(t_i, t_j) = r\xi_t(\tau)$, где $\tau = t_i - t_j$.

Из второй формулы видно, что величина автокорреляционной функции $r\xi_t(t_i, t_j)$ не зависит от начала отсчета, а только от промежутка τ, т.е. числа сдвигов. Одним из важнейших свойств стационарного случайного процесса является эргодичность, состоящая в том, что каждая отдельная реализация случайного процесса является как бы полномочным представителем всей совокупности возможных реализаций»[1].

«Процессы, вероятностная структура которых не изменяется со временем, являются стационарными»[2].

«Случайный процесс с дискретным временем называют стационарным, если распределение величин $Y_{t_1}, Y_{t_2}, ..., Y_{t_n}$ совпадает с распределением $Y_{t_1+t}, Y_{t_2+t}, ..., Y_{t_n+t}$ для любого конечного множества целых чисел $\{t_1, ... t_n\}$ и любого целого t»[3].

«Случайная функция $X(t)$ называется стационарной в широком смысле, если ее математическое ожидание постоянно, а корреляционная функция зависит только от разности аргументов t_1 и t_2: $K_x(t_1, t_2) = K_x(\tau)$, где $\tau = r_2 - t_1$.

---

[1] *Кильдишев Г. С., Френкель А. А.* Анализ временных рядов и прогнозирование. М. : Статистика, 1973. С. 16–17.

[2] *Андерсон Т.* Статистический анализ временных рядов. М. : Мир, 1976. С. 189.

[3] Там же. С. 409.

Случайная функция $X(t)$ называется стационарной в узком смысле, если ее n-мерный закон распределения при любом n зависит только от интервалов $t_2 - t_1$, ... и совсем не зависит от положения этих интервалов в области изменения аргумента t.

В практических задачах обычно применяют понятие стационарной функции в широком смысле»[1].

«Стохастический процесс Y_t называется стационарным в сильном смысле (строго стационарным или стационарным в узком смысле), если совместное распределение вероятностей всех переменных $y_{t_1}, y_{t_2}, ..., y_{t_n}$ точно такое же, что и для переменных $y_{t_{1+\tau}}, y_{t_{2+\tau}}, ..., y_{t_{n+\tau}}$.

Под стационарным процессом в слабом смысле (в широком смысле) понимается стохастический процесс, для которого среднее и дисперсия независимо от рассматриваемого периода времени имеют постоянное значение, а автоковариация зависит только от длины лага между рассматриваемыми переменными»[2].

Итак, для стационарного случайного процесса характерна неизменность во времени его основных вероятностных характеристик, таких, как математическое ожидание и дисперсия. Понятно, что оценки этих величин с увеличением членов ряда будут только улучшаться и приближаться к их истинным значениям. Зная это, дадим следующее определение стационарного ряда.

Под стационарными рядами понимаются случайные процессы, характеристики которых не меняются с течением времени t, т.е. они инвариантны относительно временных сдвигов:

$$t \to t + \tau \quad y(t) \to y(t + \tau) \tag{1.10}$$

при любом фиксированном действительном τ.

Характеристики этих процессов (математическое ожидание, дисперсия и пр.) следует найти с заданной степенью точности, в результате чего задача прогнозирования таких стационарных процессов становится чрезвычайно простой.

В то же время стационарные процессы могут иметь самый различный характер динамики — изменение одной их части

---

[1] Справочник по математике для экономистов : учеб. пособие / под ред. проф. В. И. Ермакова. М. : ИНФРА-М, 2007. С. 361.

[2] *Афанасьев В. Н., Юзбашев М. М.* Анализ временных рядов и прогнозирование. С. 258—259.

не имеет ярко выраженных тенденций, динамика другой — имеет явно выраженную тенденцию изменения, которая может носить и очень сложный нелинейный характер.

В первом случае обратимый стационарный процесс протекает в однородных условиях, когда внешняя среда стабильна и подвержена разве что воздействию случайных факторов слабой силы. Во втором — внешняя среда не однородна и претерпевает некоторое изменение. В ответ на это неоднородное изменение изучаемый процесс также меняет свои количественные характеристики.

Поэтому стационарная группа типов динамики временного ряда может быть в свою очередь разделена на две подгруппы:

- однородный обратимый процесс;
- неоднородный обратимый процесс.

Для группы факторов однородного типа выполняется условие неизменности во времени их математического ожидания и других характеристик случайных процессов.

Если же математическое ожидание и иные характеристики вероятностного процесса претерпевают изменение во времени под воздействием неоднородной внешней среды, то для таких рядов и процессов их математическое ожидание представляет собой некоторую функциональную зависимость от состояния внешних факторов.

Как следует из условия стационарности, для наиболее полного анализа стационарных процессов следует собрать как можно больше статистических данных о них. В этом случае удастся тем более точно определить и спрогнозировать характеристики процесса, чем более полной будет выборка наблюдений за ними.

Выделив стационарные процессы в указанные две группы, мы тем самым легко определяем и математический аппарат их исследования и прогнозирования. Однородные стационарные процессы применительно к социально-экономическим объектам анализируются и прогнозируются с помощью методов математической статистики. Чаще всего применительно к социально-экономическим процессам для рядов такого типа можно утверждать наличие закона нормального распределения и поэтому основные усилия должны быть направлены на доказательство данного положения с помощью соответствующих статистических гипотез и методов их проверки, а после этого — на вычисление характеристик процесса. Если удалось подтвердить гипотезу о нормальном характере распределения изучаемого ряда, то лучшей оценкой его математического ожидания выступает средняя арифметическая, а лучшей оценкой дисперсии — выборочная дисперсия. Причем, здесь уместен основной принцип выборочного метода — чем больше наблюдений, тем лучше оценки модели.

Неоднородные стационарные процессы свидетельствуют о наличии множества факторов, меняющих свои количественные характеристики и воздействующих на объект, показатели которого в ответ на это воздействие меняются во времени. Поэтому задачей прогнозиста является выявление факторов, влияющих на процесс и построение модели, описывающей их количественное влияние на объект прогнозирования. И в этом случае оценки прогнозных моделей с увеличением числа наблюдений улучшаются. Основная сложность при прогнозировании неоднородных стационарных процессов заключается в том, что часть характеристик этих процессов, в том числе и математическое ожидание процесса, изменяется с изменением внешних условий. Это изменение чаще всего носит сложный нелинейный характер, который априорно неизвестен прогнозисту, перед которым стоит очень непростая задача определить закон изменения математического ожидания, понять и математически описать сложную структуру данного процесса. Здесь исследователь вынужден прибегать к некоторым априорным предположениям — допускать наличие того или иного закона распределения вероятностей, свойств процесса и его взаимосвязей, характера динамики и т.п. В данном случае наиболее эффективно может использоваться раздел экономической науки, получивший название эконометрии.

Нестационарные обратимые процессы, в противоположность стационарным, характеризуются такими изменениями их основных характеристик, которые не вытекают из прежних значений. *Под нестационарными в промежуток времени от t до $t + \tau$ обратимыми процессами понимаются такие из них, характеристики которых меняются вариантно относительно временных сдвигов*:

$$t \to t+\tau \quad y(t) \to y(t+\tau)+\Delta y(t+\tau) \qquad (1.11)$$

при фиксированном τ (действительном или целочисленном), где приращение $\Delta y(t + \tau)$ определяется характеристиками процессов в предыдущие моменты времени.

Нестационарные процессы, происходящие в условиях однородности внешней среды, являются результатом некоторого выхода характеристик объекта прогнозирования за пределы номинальных значений. Следует напомнить, что номинальными считаются значения, которые соответствуют нормальным условиям функционирования. Поскольку функционирование происходит в условиях однородной среды, то причиной нестационарности является такое случайное сочетание факторов внешней среды, которое создает условия, выходящие за номинальные. Понятно, что вероятность действия всех случайных факторов в одном направлении почти равна нулю, но не равна ему в точности. Поэтому крайне не часто, но все же встречаются ситуации, когда все условия одновременно складываются неблагоприятным для объекта прогнозирования образом и это их случайное сочетание приводит к тому, что объект ведет себя не так, как прежде, т.е. нестационарно.

Примером такого поведения может служить ледоход, когда ранней весной на реках вскрывается лед и льдины плавно перемещаются по течению реки. Вероятность того, что все льдины вдруг столкнутся и будут «наползать» одна на другую, образуя затор, невелика. Но иногда действие случайных факторов таково, что это происходит (ветер, изменив направление, сдвигает льдины друг к другу; ручей, стекающий в реку, подмывает корни старого дерева, росшего на берегу реки, и оно упало в воду и др.). В результате всех этих случайно сложившихся условий на реке образуется затор. Плавное стационарное течение реки перешло в нестационарное — образовавшийся затор задерживает все новые льдины, которые закупоривают реку и она выходит из берегов, заливая окрестности. Со временем под воздействием температуры затор разрушается, и вся накопившаяся вода потоком устремляется вниз по течению, сея разрушения и наводняя сушу.

Это и будет нестационарным процессом, возникшим в условиях однородности. Процессы такого типа хорошо изучены в естественнонаучных и инженерных дисциплинах, где они называются «переходными процессами». Поскольку эти процессы обратимы, исследователь может неоднократно создавать условия для их возникновения, тщательно изучать их и описывать с помощью математических моделей.

В математической статистике процессы подобного рода рассматривают как переход стационарного процесса от одного уровня к другому. Чаще всего говорят о нестационарности случайного процесса «за счет переменного математического ожидания»[1] — в таком случае следует просто определить форму и характеристики этого математического ожидания и вычесть его от уровней ряда. Довольно часто переход от одного состояния стабильности к другому характеризуется постоянством первой или второй производной. Поэтому для прогнозирования таких нестационарных процессов используют различные процедуры их приведения к стационарному виду — исследуют и моделируют первые или вторые разности вместо исходных уровней ряда, или, обнаружив явление автокорреляции, устраняют ее.

Основной характеристикой нестационарных процессов в условиях однородности, в соответствии с (1.11) является приращение $\Delta y(t+\tau)$, которое представляет собой некоторую функцию от уровня ряда в предыдущие значения времени, т.е.

$$\Delta y(t+\tau) = f\big(y(t, t+1, t+2, ..., t+\tau)\big) \tag{1.12}$$

при фиксированном τ.

Ситуация значительно усложняется, когда нестационарный процесс протекает в условиях неоднородности. Факторы внешней среды при этом претерпевают значительные изменения, что оказывает влияние на объект прогнозирования и его поведение, причем эти изменения не вытекают ни из свойств самого процесса, ни из состояния внешней среды. Сам объект, адаптируясь к этим внешним изменениям, меняет свои характеристики — качественные и количественные. Наступают необратимые изменения в процессе, которые отражены в ряде динамики его показателей.

В зависимости от того, насколько меняются во времени приращения $\Delta y(\tau)$, такие необратимые процессы могут быть выделены в две подгруппы: хаотические и эволюционные процессы.

В случае, когда приращения $\Delta y(\tau)$ не имеют какой-либо достаточно гладкой тенденции во времени и их изменения хаотичны (например, на первом же наблюдении $\Delta y(\tau)$ может быть достаточно велико в сравнении с самим показателем

---

[1] *Вентцель Е. С.* Теория вероятностей : учебник. М. : КноРус, 2010. С. 481.

$y(\tau)$), такие процессы могут быть отнесены к *хаотическим*. Отношение $\Delta y(t + \tau) / y(t + \tau)$ непредсказуемо в любой момент времени — по модулю оно может быть очень большим либо близиться к нулю. Моделированием подобных нестационарных неоднородных процессов занимаются теория хаоса и теория катастроф.

Если приращения $\Delta y(\tau)$ постепенно нарастают с течением времени в результате количественных и качественных изменений, происходящих в объекте, чьей реализацией является нестационарный ряд, то эти процессы могут быть названы *эволюционными*. При этом отношение $\Delta y(\tau) / y(t + \tau)$, характеризующее нарастание неопределенности, имеет увеличивающуюся со временем τ динамику — от нуля до бесконечности:

$$\lim_{\tau \to 0}\left|\frac{\Delta y(\tau)}{y(t+\tau)}\right| = 0,\ \lim_{\tau \to \infty}\left|\frac{\Delta y(\tau)}{y(t+\tau)}\right| = \infty, \qquad (1.13)$$

причем эта динамика может быть как положительной, так и отрицательной

Эволюционные процессы в социально-экономической динамике обусловлены влиянием различных факторов, но основной из них обусловлен особенностями человеческого фактора. Человек накапливает знания на собственном опыте, получает знания от других, рационализирует свой труд и труд других людей, на основе полученных знаний меняет свое отношение к потребляемым товарам, к труду и др. Всю жизнь человеку сопутствуют новшества, которые ему предлагает общество, или которые он сам предлагает обществу. Поэтому не будет ошибкой считать, что главной причиной эволюционного характера развития социально-экономических объектов выступают инновационные процессы, непрерывно протекающие в обществе.

Под влиянием инноваций происходит постепенная смена технологий, уклада хозяйствования, структуры и методов управления. Величина этих «возмущающих» воздействий инноваций на характер динамики социально-экономического объекта ограничивается инерционностью объекта. И именно поэтому отклонения в тенденциях носят постепенный характер, предопределяя эволюционный характер развития процесса.

Хаотический характер динамики возникает в тех случаях, когда или сам процесс не инерционен и легко меняет под воздействием внешних или внутренних факторов динамику развития, или же когда на инерционный процесс воздействуют внешние факторы такой силы, что под их воздействием ломаются и внутренняя структура процесса, и его взаимосвязи, и его динамика.

Иначе говоря, эволюционная динамика характеризует процесс адаптации объекта к внешним и внутренним воздействиям, а хаотическая динамика — отсутствие способности объекта к адаптации.

Сложный характер нестационарной динамики предопределяет и сложность аппарата моделирования и прогнозирования этой динамики.

Прогнозирование эволюционных процессов до последнего времени не попадало в поле зрения специалистов по социально-экономическому прогнозированию как самостоятельная задача, и только в последние годы в учебники по прогнозированию стали включаться соответствующие разделы. На практике эволюционные процессы просто не выделяли в отдельную группу, и для их анализа и прогнозирования использовали приемы классической эконометрии, не задумываясь над корректностью такого применения. Использование аппарата прогнозирования, методологически несовместимого со свойствами объекта прогнозирования, приводит к возникновению серьезных ошибок инструментария и существенной дисперсии прогноза в практике прогнозирования социально-экономической динамики. Для прогнозирования временных рядов социально-экономических показателей эволюционного типа методологически обоснованным является применение адаптивных методов прогнозирования[1].

Теперь, выяснив особенности социально-экономических процессов, их рассмотренную выше классификацию можно представить в виде схемы (рис. 1.1). На ее основе в дальнейшем и будет излагаться материал настоящего учебника. Исходя из общенаучного принципа «от простого — к сложному», вначале будут изложены методы прогнозирования стационарных социально-экономических процессов.

---

[1] *Светуньков С. Г.* Эконометрические методы прогнозирования спроса (на примере промышленной электроэнергетики). М. : Изд-во МГУ, 1993.

Рис. 1.1. **Классификация социально-экономических процессов по типу динамики**

Затем мы перейдем к рассмотрению методов прогнозирования нестационарных социально-экономических процессов, уделяя особое внимание необратимым процессам эволюционного характера. Хаотическим процессам и их прогнозированию в данном учебнике внимание не уделяется — это сложная тема, выходящая за рамки учебника для студентов экономических специальностей и направлений вузов.

При изложении методов прогнозирования по каждому из типов социально-экономической динамики мы будем рассматривать методы краткосрочного, а затем среднесрочного прогнозирования. Дальнесрочные методы прогнозирования, нацеленные на оценку далекой перспективы, получаются с помощью экспертных и комбинированных методов прогнозирования, методами имитационного динамического моделирования. Поэтому они рассматриваются в конце издания отдельно.

Вопросы и задания для самоконтроля

1. В чем отличие между предсказанием и прогнозированием?

2. Почему применение индуктивного метода в науке привело к существенному развитию науки?

3. В чем суть дедуктивного и индуктивного методов?

4. Приведите примеры реальных социально-экономических прогнозов и оцените их успешность.

5. Определите следующие понятия: прогноз, период прогноза, база прогноза, нормативный прогноз, поисковый прогноз, точность прогнозирования.

6. Как можно классифицировать методы прогнозирования?

7. Какие виды прогноза по периоду упреждения необходимо выделить?

8. Как взаимосвязаны друг с другом период инерционности и период прогнозирования?

9. Почему точность прогнозирования определяет эффективность управления любым объектом?

10. Укажите, какова принципиальная особенность правильного выбора метода прогнозирования социально-экономических объектов (в отличие от объектов другой природы).

11. Какие процессы называются стационарными?

12. Какие процессы называются обратимыми, а какие — необратимыми?

13. Расскажите, с какими типами социально-экономических процессов (обратимыми или необратимыми) приходится иметь дело при прогнозировании их динамики.

14. Почему выборочный метод, примененный к прогнозированию необратимых рядов социально-экономической динамики, приведет к получению неточного прогноза?

15. Чем отличаются временные ряды в экономике и временные ряды показателей естественно-научных дисциплин?

Глава 2
АНАЛИЗ
И ПЕРВИЧНАЯ ОБРАБОТКА ДАННЫХ

В результате освоения данной главы студент должен:

знать

- уровень сложности задачи измерения социально-экономической информации;
- суть задачи измерения собранных для прогнозирования данных;
- в каких шкалах следует измерять эти данные;
- основные методы и инструменты количественного и качественного анализа прогнозируемых процессов;

уметь

- правильно подбирать шкалу измерения к данным разной природы;
- устранять ошибки информации или уменьшить их влияние на прогнозируемый результат;
- оценивать адекватность прогнозных моделей;
- обрабатывать эмпирические и экспериментальные данные;

владеть

- навыками самостоятельной научной и исследовательской работы в части сбора, систематизации и первичной обработки эмпирических данных;
- прикладными методами измерения социально-экономических данных;
- методами предварительного анализа данных;
- методами и подходами, с помощью которых можно оценить адекватность построенных моделей прогнозирования.

2.1. Шкалы измерения социально-экономической информации

При прогнозировании социально-экономических явлений приходится сталкиваться не только с тем, что не все из них могут быть промоделированы с помощью наиболее распространенного выборочного метода, но и с тем, что полученная информация вообще не может быть обработана с помощью привычного математического аппарата. Чаще всего у прогнозиста имеется информация, измеренная в метрической шкале. К этой шкале применимо все богатство математического аппарата современной статистики и других разделов математики. Но в экономике так же часто встречаются ситуации, когда приходится прогнозировать некие показатели, определяющиеся мнениями или отношениями и другой неметрической информацией. Например, часто встречаются случаи, когда маркетологов интересует прогноз изменения отношения потребителей к изменяющимся свойствам товара. Это отношение, определенное как, например, лояльное, безразличное или отрицательное, сложно выразить математическим языком, а уж как эту информацию использовать для прогнозирования, без специальных знаний понять невозможно. Следовательно, для решения подобных задач социально-экономического прогнозирования необходимо вначале выяснить, в какой шкале измерена информация, а затем подбирать соответствующий инструмент прогнозирования.

Поэтому прогнозист должен обладать знаниями из теории информации с тем, чтобы уметь различить тип информации, с которой ему предстоит работать. Первоначально прогнозист работает с данными — некоторыми сведениями об объекте. Проанализировав их, дав им некоторую смысловую оценку, он получает информацию об объекте, т.е. осмысленные данные о нем. Информация в той или иной степени уменьшает неопределенность знаний об объекте. Если она полностью снимает неопределенность, то говорят о полной информации. Полная информация дает исследователю знания об объекте. Но для того, чтобы понять имеющуюся информацию, правильно интерпретировать ее значения, необходимо информацию оценить и только после этого появится возможность дать ей соответствующую интерпретацию[1]. Поэтому любая информация представляет собой ценность только при возможности ее обработки. Если такой возможности не представляется, ценность информации ничтожно мала.

Необходимо отметить, что зачастую имеющейся в наличии у экономиста информации сложно дать соответству-

---

[1] *Светуньков С. Г.* Методы маркетинговых исследований. СПб. : Издательство ДНК, 2003.

ющее толкование. Если, например, у исследователя имеется информация о том, что потребитель оценил свое отношение к товару в 76 баллов, то данной информации явно недостаточно для того, чтобы осуществить действия с ней. Ведь может статься, что шкала, предложенная потребителю, имела диапазон от 0 до 80, тогда 76 баллов – очень высокий результат; но если диапазон шкалы был от 50 до 200 баллов, то результат в 76 баллов для производителя данного товара неутешителен. Поэтому для того, чтобы дать интерпретацию полученному результату, необходимо получить знания как минимум еще и о том, в какой шкале было измерено это отношение, какова ее градация, чему равны ее начальная и конечная точки и т.п. Следовательно, совершенно недостаточно получить в ходе исследований информацию, необходимо еще и измерить ее.

О том, что измерение – не такая простая процедура, как это кажется на первый взгляд, ярко демонстрирует пример, приведенный в работе К. Берки[1]. Представим себе, что имеется некое тело, имеющее форму прямоугольного равнобедренного треугольника с катетами $A = B = 1$ м. Требуется измерить длину гипотенузы C. Если при этом применить теорему Пифагора, то в нашем случае окажется, что длина C будет равна величине $\sqrt{2}$ м. Но эту же длину можно измерить с помощью линейки, и будет получена величина, находящаяся в промежутке между 1,41 и 1,42 м. Однако хорошо известно, что это число иррационально и математически выражается бесконечной апериодической десятичной дробью – 1,4112135... м. Следовательно, длина грани измеряемого тела, соответствующая гипотенузе C прямоугольного равнобедренного треугольника (при катетах $A = B = 1$ м), неизмерима с помощью измерительного прибора со сколь угодно малыми долями основной единицы измерения (м, дм, см, мм, мкм, нм и т.п.).

В практике экономических исследований приходится иметь дело с объектами, измерение которых еще более затруднительно, чем в указанном примере. Информация об объекте может в самом общем случае носить качественный или количественный характер, поэтому она может быть отнесена к классифицирующей (качественной), топологической (сравнительной) и метрической (количественной) информации.

---

[1] *Берка К.* Измерения: понятия, теории, проблемы. М. : Прогресс, 1987. С. 12–13.

Классифицирующая информация, определяемая лишь качественно, служит для классификации объектов на основе их общих характеристик. Классификация — это первый шаг в анализе свойств любого объекта и важнейший элемент экономического исследования. Выделение общих свойств объектов в отдельные группы позволяет сделать ряд важнейших выводов, но действия с классифицирующей информацией ограничиваются только идентификацией объектов экономического исследования и отнесением этих объектов к тому или иному классу.

Если общие характеристики классифицирующей информации носят характер свойств, обладающих какими-либо градациями, то упорядочение такой информации в соответствии с данными градациями позволяет получить *топологическую (или сравнительную) информацию*. В отличие от классифицирующей, топологическая информация имеет уже бóльшую познавательную ценность, поскольку позволяет не только устанавливать тождество между объектами, но и сравнивать их друг с другом, получая в результате их расположение в определенном порядке.

Топологическая информация образует переход от классифицирующей информации к метрической. С точки зрения теории топологическая информация имеет гораздо больше общего с информацией классифицирующей, чем с метрической. Это вызвано тем обстоятельством, что упорядочение информации по какому-либо признаку в большей степени относится к классификации, чем к измерению. В дальнейшем будет показано, что топологическая информация включает несколько различных типов, отличающихся друг от друга близостью или удаленностью от крайних значений — классифицирующей или метрической информации.

Метрическая информация не только выражает качественную характеристику объекта, но и содержит его точные и полные количественные характеристики. Очевидно, что метрическая информация исторически является результатом эволюции классифицирующей информации в метрическую. В процессе познания природы и ее измерения человек последовательно осуществляет с имеющейся в его распоряжении информацией ряд иерархически взаимосвязанных действий:

- проводит классификацию и тем самым осуществляет первичный анализ информации;
- сравнивает полученные классы и ранжирует их, получая топологическую информацию об объекте;

• разрабатывает критерий измерения информации и, сделав это, получает метрическую информацию.

Из вышесказанного следует, что объективный переход от классифицирующей информации к метрической нельзя воспринимать как простой перевод качественной информации в количественную — происходит существенное расширение знаний об объекте исследований, ведь метрическая информация содержит в себе в том числе и классифицирующую, и топологическую информацию. Действительно, информация, например, о стоимости товара в рублях говорит исследователю о том, что из всего множества окружающих его вещей он имеет дело с предметом, принадлежащим к системе экономических отношений, причем этот предмет — товар, который может быть куплен или продан. Тем самым осуществляется классификация информации. К тому же, цена товара позволяет говорить о том, что он дороже или дешевле других товаров. Это позволяет осуществить сравнение элементов данного класса предметов (исследователь имеет дело и с информацией топологической). С учетом того, что информацию о цене товара исследователь может складывать и отнимать, умножать ее на объемы или делить цену на величины физических свойств товара (например, при нормировании), то информация, безусловно, относится к классу метрической.

В реальной практике экономических исследований приходится встречаться с информацией всех трех типов. В случае, когда изучаемый объект описывается известными количественными характеристиками, для которых существует общепринятая шкала измерений, особых проблем со сбором и обработкой информации не возникает. Доходы потребителей измеряются в денежных единицах, а объемы потребления товаров — в штуках, килограммах или литрах. Проблемы возникают, когда объект описывается показателями, которые выразить в количественных элементах нельзя и возможны лишь некоторые качественные интерпретации. Но и эти показатели должны быть измерены, поскольку без измерения невозможны никакие действия с информацией. Подобные измерения могут быть осуществлены с помощью различных шкал. Совокупность всех возможных шкальных значений образует одномерный континуум (многомерные измерения не рассматриваются). В процессе измерения информации с помощью той или иной шкалы отношения между объектами измерения отображаются на отношения между числами. И только после этого с информацией можно осуществлять какие-либо действия: сравнение, отношение, складывание и вычитание и т.п.

Как и любой элементарный процесс, измерение, несмотря на очевидность его характеристик, может быть определено различными способами.

Американский ученый Н. Р. Кэмпбелл сформулировал *четыре типа определений понятия «измерение»*[1]:

1) присваивание цифр для представления свойств;

2) процесс присваивания чисел для представления качеств;

3) присваивание цифр для представления свойств в согласии с научными законами;

4) присваивание цифр вещам так, чтобы они представляли факты или конвенции о них.

В данных определениях в качестве синонимов используются, в общем-то, разные понятия — «число» и «цифра», что не совсем верно. Цифра — это способ выражения числа. С помощью одних и тех же цифр можно описать различные числа (например, число 23 и число 32 описаны с помощью цифр 2 и 3).

Неточность присуща и следующему определению: «под измерением понимается определение количественной меры или плотности некоторой характеристики (свойства), представляющей интерес для исследователя»[2]. «Определение количественной меры» априорно отрицает возможность измерения качественных характеристик, что приводит к невозможности измерения классифицирующей информации. А измерение классифицирующей информации — важнейший элемент измерения в экономике, например, при сегментации потребителя по признаку пола.

В «Рабочей книге социолога» под измерением понимается «процедура, с помощью которой объекты измерения, рассматриваемые как носители определенных соотношений, отображаются в некоторую математическую систему с соответствующими отношениями между элементами этой системы»[3]. Это определение уже более точно подходит к понятию про-

---

[1] *Cempbell N. R.* Foundations of Science. The Philosophy of Theory and Experiment. N. Y., 1957. P. 38.

[2] *Голубков Е. П.* Маркетинговые исследования: теория, методология и практика. М. : Финпресс, 1998. С. 172.

[3] Рабочая книга социолога / под ред. Г. В. Осипова. М. : Наука, 1983. С. 142.

цесса измерения, но ограничивает класс способов измерения математическими системами, что также не совсем верно, так как измерение может осуществляться не только с помощью математических, но и, например, с помощью графических символов.

Отечественный социолог В. А. Ядов приводит наиболее корректное, на наш взгляд, определение: «Измерение — это процедура, с помощью которой измеряемый объект сравнивается с некоторым эталоном и получает числовое выражение в определенном масштабе и шкале»[1]. Здесь понятие «число» выступает в широком смысле и тем самым включает в себя возможность измерения всей совокупности типов информации.

Наиболее полно и в то же время кратко понятие «измерение» определяется в математике. Согласно теории алгебраических систем под измерением можно понимать гомоморфное отображение некой эмпирической реляционной системы на некоторую числовую реляционную систему[2].

Обобщая вышесказанное, можно сделать вывод о том, что измерение представляет собой процесс моделирования, в ходе которого оригинал (объект измерения) получает гомоморфное отображение в некоторой модели, которая представляет собой соответствующую шкалу измерений. Напомним, что гомоморфизм — это понятие, означающее такое соотношение между двумя системами, когда каждому элементу и каждому отношению между элементами первой системы соответствует один элемент и одно отношение второй, но не наоборот[3]. Шкала измерения представляет собой гомоморфный образ измеряемого объекта.

Вопросами изучения и решения проблем измерения занимается теория измерения, предлагающая исследователю значительный арсенал методов измерения и соответствующих шкал измерения информации. В формальной теории измерений понятие шкалы интерпретируется в том же значении,

---

[1] *Ядов В. А.* Стратегия социологического исследования. Описание, объяснение, понимание социальной реальности. М. : Добросвет ; Книжный дом "Университет", 1998. С. 81.

[2] *Берка К.* Измерения: понятия, теории, проблемы. М. : Прогресс, 1987. С. 38.

[3] Малая математическая энциклопедия. Будапешт : Изд-во АН Венгрии, 1976. С. 149.

что и понятие измерения. За шкалу здесь принимается упорядоченная тройка

$$\langle E, N, F \rangle, \tag{2.1}$$

где E означает эмпирическую реляционную систему; N — числовую реляционную систему, F — гомоморфное отображение E на N.

В зависимости от указанных трех характеристик можно выделить достаточно большое количество шкал. В экономических исследованиях используется лишь некоторая их часть, которая может быть классифицирована в зависимости от характеристик, составляющих шкалу (2.1).

Так, если в эмпирической реляционной системе (объекте измерения E) выявляют и измеряют числа, служащие для характеристики объектов, внешних по отношению к субъекту измерения, такие шкалы называют оценочными. А если в объекте измерения шкалируют числа, характеризующие внутренние свойства индивидуума, такие шкалы называют шкалами установок.

Числовая реляционная система N (способ отражения свойств объекта) может быть выражена с помощью цифр, вербального или графического представления. Соответствующие шкалы будут называться числовыми, вербальными или графическими. В практике экономических исследований в подавляющем большинстве случаев используются числовые шкалы.

Числовые шкалы, используемые для получения экономической информации, классифицируются по способу преобразования информации в числа. Таких способов достаточно много, поэтому в теории измерений задают некоторую точность данного преобразования, ограничивающую данное множество. Ограниченное множество преобразований называют допустимым. Таким образом, тип шкалы определяется соответствующим этой шкале множеством допустимых преобразований. Из всего множества теоретически возможных шкал в практике социально-экономического прогнозирования чаще всего используются четыре типа шкал:

1) номинальная;
2) порядковая;
3) интервальная;
4) метрическая.

Данное деление было предложено американским психологом С. Стивенсом в 1946 г. В соответствии с его классификацией номинальная и порядковая шкалы относятся к классу качественных, а интервальная и метрическая — количественных шкал. Заметим, что в оригинале в предложенной Стивенсом классификации последняя шкала называлась шкалой «отношений» или «пропорций» (в английском — «ratio»). В данном учебнике мы используем обозначение К. Берка. Существуют и более подробные классификации шкал, однако для целей моделирования социально-экономических процессов вполне достаточно и этой.

Остановимся более подробно на свойствах этих шкал. Каждая из них определяется наличием или отсутствием четырех характеристик:

- описание;
- порядок;
- расстояние;
- начальная точка.

Описание шкалы предполагает использование единого способа записи информации, т.е. характеризует составляющие шкалу элементы, например, степень или способ согласия («да», «нет», «не знаю»). При этом между данными элементами не вводится какая-либо характеристика сравнений — осуществляется только идентификация информации.

Порядок характеризует наличие отношений в способах записи информации, крайних точек зрения («не согласен», «не совсем согласен», «согласен» и т.п.). При этом предусматриваются некоторые сравнительные характеристики, позволяющие упорядочить отношение к предмету исследования.

Расстояние шкалы может быть измерено. Это значит, что оно существует только в тех шкалах, в которых элементы шкалы определены количественно и между ними имеются интервалы, расстояние между которыми имеет смысловое значение.

Начальная точка задает тот или иной уровень соотношений между элементами шкалы. Ее не надо путать с точкой отсчета. Каждая начальная точка является точкой отсчета, но не каждая точка отсчета может быть начальной точкой. Например, начальная точка шкалы измерения массы тела в килограммах говорит о том, что нулевое значение на этой шкале (нуль килограммов) свидетельствует об отсутствии массы тела вообще. Поэтому эта точка, являясь точкой отсчета, одновременно является и начальной точкой шкалы измерения массы. Если же за начальную точку измерения массы тела взять массу тела любого из авторов этой книги, и назвать ее нулевой точкой, то эта точка будет только точкой отсчета новой шкалы измерения массы тела, но не начальной точкой.

Шкала имеет начальную точку, если она имеет единственное начало[1]. Чаще всего эта начальная точка является нулевой и характеризует отсутствие измеряемого свойства (например, если прибыль предприятия равна нулю, это означает, что у предприятия нет прибыли как таковой).

В зависимости от наличия или отсутствия этих четырех характеристик, а также от способов их задания и определяются различные типы шкал.

Самая простая шкала — **номинальная шкала**. Иногда ее называют иначе («шкала наименований», «категориальная шкала», «классификационная шкала»). Сталкиваясь с номинальной шкалой, исследователь имеет дело с самым простым случаем измерения, так как в рамках этой шкалы моделируются самые простые действия с информацией — отношения «равенства — неравенства». Эта шкала обладает только характеристикой шкал — ***описанием*** шкалы, с помощью которого из множества элементов, характеризующих объект измерения, указывается только один элемент (может быть и многофакторным), причем как результат идентификации, а не сравнения. Как пример приведем вопрос: «укажите свой пол: мужской / женский». Данной шкале не присущи порядок, расстояние и начальная точка — их нет.

Существуют различные способы описания элементов номинальной шкалы. В качестве инструментов описания данной шкалы могут выступать различные объекты — слова, словосочетания, набор букв, набор цифр, символы, знаки, рисунки, численные индексы и т.п.

Например, такой элемент, как отрасль знаний может быть описан словами: «экономика», «право», «физика», «электротехника», «история». Но если для целей последующей обработки данных с помощью вычислительной техники оказывается более предпочтительно закодировать эту информацию с помощью некоторого числового кода, то отрасли знаний «экономика» будет присвоен, например, код 1; отрасли знаний «право» — код 2; отрасли знаний «физика» — 3 и т.д. Полученная шкала информации позволяет присвоить каж-

---

[1] *Голубков Е. П.* Маркетинговые исследования: теория, методология и практика. М. : Финпресс, 1998. С. 173.

дому из элементов информации соответствующий код, но не позволяет осуществлять математические действия с этими элементами шкалы. Действительно, нельзя сравнить друг с другом такие элементы, как «экономика» и «физика» на предмет того, какой из элементов больше, а какой меньше, бессмысленным будет и сравнение их кодов 1 и 3. Нелепой будет и попытка осуществлять какие-либо арифметические действия с этими элементами номинальной шкалы, например, суммировать коды 1 и 3. Полученное в результате суммирования число 4 не будет иметь никакого смысла, поскольку в данной шкале оно присвоено в качестве кода понятию «электротехника», которая, очевидно, не представляет собой сумму знаний экономики и физики.

Следует отметить, что, несмотря на кажущуюся простоту измерения экономической информации по этой шкале, ее использование следует осуществлять с большой осторожностью, исходя из целей исследования — нельзя допустить излишней детализации, так как это затрудняет процесс исследования, нельзя ограничиваться и слишком укрупненными элементами. Если экономиста интересуют характеристики потребителя и, например, характер его рабочей специальности, то в отдельных случаях достаточно просто ограничиться группами (служащий, рабочий, крестьянин и т.п.), а можно и детализировать специальности, например, служащих (руководитель, заведующий сектором, старший инструктор, инструктор и т.п.). Степень детализации информации, описываемой с помощью шкалы наименований, определяется в каждом конкретном случае. В подавляющем большинстве случаев шкала наименований используется в целях различного рода группировок экономической информации, ее классификации и типологизации, что является первым и очень ответственным шагом анализа объекта экономического исследования. Здесь уместно напомнить, что «анализ» представляет собой в дословном переводе «расчленение, разделение». Любой экономический анализ начинается именно с того, что сложный объект исследования упрощается за счет мысленного выделения из него относительно самостоятельных элементов. И именно возможность измерения в номинальной шкале позволяет отнести те или иные элементы к различным группам.

Пример

Студенты, обучающиеся в российских вузах, как и все люди, отличаются друг от друга самыми разными характеристиками. Если перед прогнозистом стоит задача выполнить прогноз успеваемости в некотором вузе для того, чтобы заранее оценить необходимый стипендиальный фонд вуза, то общую совокупность студентов следует подразделить на отдельные классы, важные с позиций получения или неполучения студентом стипендии. Неуспевающий студент — это студент, который в ходе сессии заимел некоторую задолженность хотя бы по одному предмету. Он не получает стипендию. Успевающий студент не имеет задолженностей и вовремя получил оценки за все экзамены и зачеты. Он имеет право на простую стипендию. Отличник — это студент, который не только вовремя «сдал» сессию, но и получил только отличные оценки. Он будет получать повышенную стипендию.

Таким образом, формируется номинальная шкала — выделяются три класса студентов: «неуспевающие», «успевающие» и «отличники». Используя перечисленные выше характеристики каждой шкалы (описание шкалы), можно с их помощью оценить информацию о каждом студенте вуза, после чего отнести каждого из них к тому или иному классу. Как вы понимаете, с самими понятиями «неуспевающие», «успевающие» и «отличники» никаких математических действий осуществлять нельзя, но можно собрать количественные характеристики каждого из классов (например, сколько человек в вузе относятся к каждому из трех классов), и уже с этими количественными характеристиками выполнять соответствующие математические действия.

Шкала порядка наряду с *описанием* имеет еще и *порядок*, в результате чего возможно установление приоритетов или сравнений. При этом шкала имеет тем или иным образом сформулированные ранги. Чаще всего с этой шкалой экономист сталкивается в маркетинге, когда изучается поведение потребителя и его отношение к товару или маркетинговому сопровождению этого товара. В маркетинговых исследованиях часто используют опросы потребителей, проводящиеся с помощью анкет, в которых, например, может содержаться просьба — указать предпочтительность одного товара по сравнению с другими, или, иначе говоря, определить степень предпочтения одного объекта другому. Эта шкала в литературе иногда называется «ранговой» или «ординальной шкалой рангов». Способ ее описания определяется целями исследования.

Важно отметить, что шкала порядка не позволяет давать толкование расстояниям между ее элементами — они могут быть только сравнимы друг с другом по рангам. Информация шкалы может позволить утверждать, что, например, элемент А больше элемента Б, а элемент Б в свою очередь больше элемента В. Однако эта шкала не позволяет вычислить разность (А – Б) или (Б – В). Более того, как и в случае шкалы наименований, здесь совершенно бессмысленно выполнять с полученной информацией какие-либо арифметические действия, поскольку в шкалу порядка в качестве ее элемента не вводится расстояние.

Пример

Представим себе ситуацию, когда в целях какого-либо исследования мы попросили эксперта расставить сборные команды стран по хоккею в порядке убывания по мастерству. Это может понадобиться, например, при участии в работе букмекерской фирмы. Эксперт расставил команды так:

1. Сборная России;
2. Сборная Канады;
3. Сборная Чехии;
4. Сборная Финляндии;
5. Сборная Швеции.

Из полученной шкалы рангов команд вовсе не следует, что сборная команда Канады в два раза слабее сборной команды России (2 : 1 = 2), а сборная команда Швеции в пять раз слабее сборной команды России (5 : 1 = 5) и в два с половиной раза слабее сборной Канады (5 : 2 = 2,5). Такие математические действия ни один нормальный человек выполнять не станет, поскольку они бессмысленны.

С помощью порядковой шкалы экономист может не только идентифицировать объект измерения и отнести его к той или иной группе, но и сравнить измеряемый показатель на предмет того, больше он у данного объекта по сравнению с другими или меньше. Например, в терминах потребительского поведения это означает систему предпочтений потребителя.

Для того чтобы ответить на вопрос «Насколько этот показатель данного объекта больше или меньше такого же показателя у другого объекта?», вводится такая дополнительная характеристика, как расстояние между элементами шкалы измерения. Тогда из шкалы порядка будет получена **шкала интервалов**.

В этой шкале есть *описание* (определение свойств объекта, подлежащих измерению), *порядок* (дается описание правил предпочтительности одного элемента другому) и *расстояние* (насколько рядом стоящие деления шкалы отличаются друг от друга). Способов задания этих трех характеристик довольно много, что предопределяет наличие самых разных типов интервальных шкал.

Типичный пример шкалы интервалов — когда отношение высказывается в процентах (например, «на сколько процентов данный товар удовлетворяет эту потребность»). В этой шкале измерение признака на 55% отличается от измерения на 45% на ту же величину, что и 65% от 55%. В шкале интервалов нет только одной характеристики — начальной точки.

Действительно, и 0%, и 100% в данном случае — только крайние, граничные значения информации. Обе точки являются точками отсчета, при желании «процент удовлетворенности» можно поменять на «процент неудовлетворенности» — значения процентов зеркально отобразятся и будут иметь другое численное значение. Поменяется при этом и отношение типа «больше — меньше». Но наличие у измеряемого объекта нулевого значения вовсе не означает отсутствие измеряемого свойства, как это происходит в метрической шкале. Это означает только наличие крайней степени значения измерения информации. Нуль градусов температуры по Цельсию в этой шкале не означает полного отсутствия температуры тела, это только означает, что при этой температуре будет замерзать пресная вода.

Главная трудность при построении интервальных шкал в экономических исследованиях состоит в обосновании равенства или разности дистанций между объектами. Процедуру, позволяющую таким образом преобразовать шкальные значения порядковой шкалы, что равенство расстояний между полученными числами можно будет трактовать как отражение соответствующего равенства «расстояний» между изучаемыми объектами, называют метризацией шкалы или (иногда) «оцифровкой шкальных значений». На практике существует много методов метризации шкал, например, метод парных сравнений, методы шкалирования Терстоуна и т.д.

К числу наиболее часто используемых в экономической практике *интервальных шкал* относят шкалу отношений и шкалу расстояний.

Шкала отношений получается, если учитывается требование о том, чтобы в процессе измерения не только отношения между эмпирическими объектами отображались в соответствующие числовые отношения, но и один и тот же объект отображался в 0. Так, при изучении удовлетворенности респондента товаром в качестве такого объекта выбирается респондент, равнодушный к товару. Фиксацию нулевого объекта можно рассматривать как задание начала отсчета шкальных значений.

Шкалы расстояний получаются из интервальных шкал при фиксации единицы измерения. При этом расстояние между близлежащими элементами шкалы является величиной постоянной вне зависимости от того, на каком участке шкалы осуществляется сравнение. С учетом того, что в шкале интервалов существуют расстояния, величина которых имеет ярко выраженный смысл, на этой шкале возможны арифметические операции с числами. Если в результате таких операций будет получено, например, число 4,75, то это означает, что полученная оценка значительно ближе к 5, чем к 4. При этом расстояние до этих точек имеет смысл силы отношения (если измерялось отношение). В то же время надо помнить, что данная шкала не позволяет выполнять все математические действия с измеренной информацией. Действительно, если, например, шкала имеет интервал от 0 до 7 единиц, то полученное в результате математических действий число 8 не имеет смысла, поскольку информация ограничена интервалом от 0 до 7. Очевидно, что и отрицательные значения, полученные в результате математических вычислений для чисел, измеренных в этой шкале, не имеют смысла, поскольку шкала ограничена нулевым значением, и отрицательные числа не могут быть истолкованы.

Пример

Используя шкалу интервалов, рассмотрим реализацию на рынке соленых огурцов. Попросим потребителя оценить свое отношение к товару в шкале чисел, меняющихся от нуля до, например, 10. Нулевое значение соответствует самому плохому отношению к товару, 10 — оценке товара как совершенного. Расстояние между элементами шкалы задается одинаково. Это означает, что оценка в 5 баллов выше оценки в 4 балла на ту же величину, что и оценка 9 выше оценки 8.

Допустим, что исследование проводилось двумя группами исследователей, каждая из которых по предварительной договоренности использовала исключительно интервальную шкалу измерения полученной информации, но тип шкалы априорно не был определен. Первая собирала информацию в пределах от 0 до 5, а вторая — в шкале от 0 до 10.

Первая группа получила среднюю арифметическую оценку мнения потребителей относительно свойств соленых огурцов, равную 3,75. Вторая — 6,83. О чем это говорит? На первый взгляд, это свидетельствует о том, что первая группа хуже оценила товар, чем вторая (оценка ниже). Но на самом деле это не так — шкалы были заданы с разными интервалами. Поэтому предстоит простая, но трудоемкая задача привести данные обеих групп к единому масштабу. Если бы информация была получена и измерена в метрической шкале, то таких проблем не возникло бы.

В **метрической шкале** есть все четыре характеристики, включая *начальную точку*. Эта шкала является наиболее полной для целей математической обработки информации (например, шкала расстояний между телами, шкала веса тел, шкала стоимости товаров и т.п.). С ее элементами можно выполнять любые математические действия в полном объеме. Иногда эту шкалу называют «количественной», или «абсолютной». Получив информацию в таком виде, ее можно легко обработать для получения каких-либо выводов и рекомендаций с помощью аппарата математической статистики.

Рассмотренные типы информации в зависимости от применяющихся шкал могут при определенных условиях трансформироваться друг в друга. Например, метрическая шкала доходов может быть преобразована в интервальную шкалу. Для этого следует взять максимальный доход за единицу, а доходы всех остальных лиц разделить на эту величину — получится оценка того, какую долю составляет доход каждого лица в сравнении с доходами обладателя максимума. Если при этом умножить каждую величину на 100%, получится еще более удобная для анализа доходов шкала. Но если в распоряжении прогнозиста при наличии подобной шкалы интервалов нет той самой базовой величины, к которой были отнесены все доходы, то получить данные в метрической шкале не получится. Обратный переход от интервальной шкалы в шкалу метрическую оказывается сложным.

Информацию, измеренную в метрической или в интервальной шкале, можно перевести в порядковую шкалу, например, задав четыре группы обладателей тех или иных доходов:

1) с доходом от нуля до размера минимальной заработной платы;

2) от минимальной заработной платы до прожиточного минимума;

3) от прожиточного минимума до средней заработной платы;

4) выше средней заработной платы.

Таким образом, информация о доходах, которая была представлена в метрической шкале, оказалась преобразованной в информацию об уровнях доходов в шкале порядков, в которой в данном случае предусмотрено использование только четырех оценок. При желании полученные данные можно преобразовать и в шкалу более низкого порядка — в номинальную. Так, на основе приведенной выше порядковой шкалы можно получить не более четырех номинальных шкал, в каждой из которых будет измеряться наличие либо отсутствие того или иного признака у респондента (относится к группе получающих доход до минимальной заработной платы или нет, доход составляет меньше прожиточного минимума или больше). Обратное преобразование из шкалы низкого в шкалу более высокого порядка (например, из порядковой в интервальную, а тем более — в метрическую) невозможен без получения дополнительной информации.

Это касается не только рассматриваемого примера с доходами разных лиц. Как правило, чаще всего переход от шкалы низкого к шкале более высокого уровня просто невозможен, хотя потребность в этом на практике очень велика. Действительно, среди задач экономических исследований важнейшими являются задачи обобщения экономической информации, ее агрегирования, а выполнить их в полном объеме можно, только используя всю совокупность математических действий, что возможно только с информацией, измеренной в метрической шкале.

Поэтому прогнозисту при сборе данных о мнении человека или мотивах его поведения желательно придерживаться следующего принципа: задавая вопрос, его нужно формулировать так, чтобы по возможности получать информацию в шкале измерения самого высокого уровня.

2.2. Измерение социально-экономических отношений для их прогнозирования

Несмотря на кажущуюся простоту получения информации от человека, эта задача довольно сложна, а без измерения отношения индивидуума к товару или каким-либо его свойствам невозможно точно спрогнозировать продажи товара, но от того, насколько точно прогнозист сможет выявить это отношение и определить его развитие в будущем, зачастую зависит само существование фирмы.

До сих пор говорилось об информации наблюдаемой, которую необходимо измерить для того, чтобы, проанализировав ее, использовать полученное знание для прогнозирования. Но весьма часто в социально-экономических задачах бывает необходимо спрогнозировать отношения, которые, как известно, не наблюдаются и являются скрытыми в глубинах человеческой психики. Мы будем рассматривать эти отношения (в первую очередь — отношения потребителя к товару) исключительно в контексте социально-экономического взаимодействия.

Само это отношение является результатом взаимодействия сложной структуры социальных установок (аттитюдов). При этом используются все те же основные шкалы измерения информации — номинальная, порядковая, интервальная и метрическая. Однако социальная установка носит сложный характер, поэтому использование одномерной шкалы, какой являются все рассмотренные выше шкалы, в данном случае будет неуместным.

В настоящее время психологи рассматривают трехкомпонентную структуру аттитюда: когнитивный компонент (осознание объекта социальной установки), аффективный компонент (эмоциональная оценка объекта) и конативный компонент (определяющий последовательное поведение по отношению к объекту). Поэтому в ходе элементарного измерения отношения потребителя к изучаемому объекту в целом степень удовлетворенности может быть определена, например, только со стороны когнитивного компонента аттитюда. Измерить социальную установку и отношение к товару в целом с помощью элементарных процедур измерения не удается именно из-за сложности самой структуры социальной установки. Для решения этой проблемы используется процедура глубинного проникновения в проблему, в частности, процедура «логического прямоугольника». Для этого в объекте исследования выделяют его характерные черты, желательно, связанные с вышеназванными тремя компонентами аттитюдов. Например, можно попросить потребителя оценить не степень удовлетворенности товаром вообще, а степень удовлетворенности такими его сторонами, как:

1) соответствие цены потребительским свойствам товара;
2) эстетические свойства товара;
3) потребительские свойства товара.

В первом случае можно получить одну из косвенных оценок когнитивного компонента социальной установки, так как для каждой социальной группы цена и потребительские свойства товара имеют собственные оригинальные отношения. Например, для большинства студентов цена вина сорта «Мадера» 20-летней выдержки (хорошего производителя) не соответствует его потребительским свойствам — она слишком высока и студенты могут высказать неудовлетворение этим отношением. А большинство менеджеров ОАО «Газпром», напротив, выскажет удовлетворенность соотношением между ценой и потребительскими свойствами этого вина.

Во втором случае оценивается аффективный компонент, так как эстетичность товара определяется эмоциями, которые товар вызывает у потребителя. И, вновь возвращаясь к примеру с «Мадерой» 20-летней выдержки, можно заметить, что наличие старой бутылки и выцветшей этикетки у студентов ассоциируется с чем-то несовременным и будет с негодованием отвергнуто, а менеджер газового монополиста, обратив внимание на эти же характеристики, возможно, подумает: «Действительно старое вино, у него, должно быть такой изысканный вкус!»

В третьем случае оценивается конативный компонент аттитюда — определяется отношение к товару с позиций желания и возможности его приобретения для удовлетворения имеющихся потребностей. И с этих позиций отношение студентов и менеджеров к «Мадере» 20-летней выдержки будет различным — букет вина с долгой выдержкой менее разнообразный, чем букет молодого вина, поэтому, попробовав вино, студенты недоуменно пожмут плечами, а менеджер, избалованный самыми разными винами, будет доволен.

Рассмотрим способ комплексной оценки потребительского отношения с помощью процедуры «логического прямоугольника». Пусть степень удовлетворенности будет определяться тремя возможными оценками: 0, 1 и 2, причем 2 означает удовлетворенность, а 0 — неудовлетворенность данными свойствами товара. В таблице, имеющей форму прямоугольника, представлены все возможные варианты сочетания этих оценок (табл. 2.1).

Таблица 2.1

Логический прямоугольник

Вариант сочетания оценок	Удовлетворенность соответствием цены потребительским свойствам товара	Удовлетворенность эстетическими свойствами товара	Удовлетворенность потребительскими свойствами товара	Итоговая оценка
1	2	2	2	6
2	2	2	1	5
3	2	1	2	5
4	1	2	2	5

Окончание табл. 2.1

Вариант сочетания оценок	Удовлетворенность соответствием цены потребительским свойствам товара	Удовлетворенность эстетическими свойствами товара	Удовлетворенность потребительскими свойствами товара	Итоговая оценка
5	2	2	0	4
6	2	0	2	4
7	0	2	2	4
8	2	1	1	4
9	1	2	1	4
...	...	...	...	...
21	0	0	1	1
22	0	1	0	1
23	1	0	0	1
24	0	0	0	0

Как видно из данных «логического прямоугольника», оценка степени удовлетворенности товаром стала значительно более многообразной — вместо трех градаций (от нуля до двух) получено семь градаций степени удовлетворенности товаром (от нуля до шести), что, конечно, само по себе является положительным фактом, поскольку позволяет оценить свойства товара более глубоко. Оценка методом «логического прямоугольника» позволяет выявить уровень задействования всех трех составляющих аттитюдов в единой совокупности. Если итоговая оценка мала (нуль или единица), то ни одна из составляющих аттитюда не задействована полностью в отношении потребителя к товару. Напротив, если итоговая оценка равна шести или пяти, то это говорит о том, что в отношении к товару задействованы все составляющие социальной установки. Очевидно, что можно использовать и более многообразные градации шкалы, например, использовать пять уровней оценки отношения к предмету исследования, и тогда количество оценок и их градация увеличатся в еще большей степени. Легко убедиться в том, что «логический прямоугольник» в этом случае будет содержать оценки от нуля до 15, что позволит углубить степень оценки отношения потребителя к предмету исследования.

Процедура «логического прямоугольника» является наиболее часто используемой процедурой, позволяющей оценить социальные установки и отношение потребителя к товару. Она дает общую схему исследования, которая усовершенствуется и развивается.

Для измерения социальных установок используется и процедура шкалирования суммарных оценок, которую впервые предложил к использованию З. Лайкерт в 1929–1931 гг.[1] Эта процедура весьма успешно используется для измерения социальной установки, например, отношения мужчин к алкоголю, женщин — к услугам системы быстрого питания, юношей — к молодежной моде и т.п. Эти и им подобные социальные установки невозможно измерить посредством «прямых» вопросов, обращенных к респонденту, так как для большей части людей эти установки носят скрытый, латентный характер. Известно, например, что очень часто спивающиеся мужчины категорически не согласны с констатацией этого факта и на прямой вопрос по этому поводу отвечают: «Я в любой момент могу остановиться, значит, я не алкоголик!»

Поэтому при осуществлении опроса потребителя ему предлагается ответить на ряд суждений или утверждений, носящих безличностный характер. Если рассматривается отношение мужчин к алкоголю, то прямой вопрос типа: «как Вы относитесь к алкогольным напиткам» заменяется рядом суждений или утверждений, носящих абстрактный характер, например: «алкоголь — это яд», «пьяный человек неприятен в общении», «употребление спиртных напитков вызывает неприятную реакцию организма» и т.п.

Такой прием позволяет решить важную психологическую проблему. Действительно, когда сложные или неприятные вопросы ставятся напрямую, требуют открытого ответа, сложно добиться искренних ответов. Скорее всего, респондент ответит не так, как он думает на самом деле. Но отвечая на вопросы, которые носят неличностный характер, он будет вести себя более раскованно и свободно. Каждое из утверждений, высказанное относительно объекта исследования, сформулировано не респондентом, а кем-то другим. Задача респондента меняется — надо

---

[1] *Татарова Г. Г.* Методология анализа данных в социологии. М. : NOTA BENE, 1999. С. 67.

согласиться или не согласиться с чужим по самой сути высказыванием. Иначе говоря, ситуация, когда респондент должен сформулировать свое личное мнение об объекте исследования, заменяется ситуацией согласия или несогласия с мнением другого лица. Последнее — значительно более легкая в психологическом плане процедура для любого человека.

Если просуммировать ответы респондента о степени согласия или несогласия с высказанными мнениями, можно определить его мнение по отношению к объекту исследования в целом. При этом необходимо иметь в виду следующее. Важно следить и за тем, чтобы отношение респондента к предлагаемым суждениям высказывалось в одной шкале информации, имеющей одну и ту же степень градаций. Лучше всего при этом использовать шкалу расстояний.

Выбор суждения, с которым респонденту предложено согласиться или не согласиться, представляет собой сложную задачу. Дело в том, что исследователь должен предложить респонденту не вопрос, а готовое суждение, сформулированное заранее. При этом само суждение носит разную степень эмоциональной окраски. Поэтому необходимо подобрать такую совокупность суждений, чтобы они, с одной стороны, были легко понимаемы респондентом, а с другой, — выявляли все возможные стороны отношения респондента к данному объекту. Последнее означает, что суждения должны в одинаковой мере эмоционально определять объект. Например, определения «отвратительный» и «плохой» имеют разную степень эмоциональной окраски, а значит, их применение в разных суждениях и оценивание этих суждений респондентом по одинаковой балльной шкале будет неправильным. Например, суждение «перепой — это отвратительно» и суждение «чрезмерное потребление спиртных напитков — это плохо», имеют разную степень отношения к объекту — потреблению спиртных напитков выше допустимого количества. В первом случае «полное согласие» выскажут практически все опрашиваемые, мнения относительно второго суждения будут менее однозначными. Бóльшая часть опрашиваемых выразит «полное согласие», но часть респондентов отметит «скорее согласен, чем не согласен». И дело здесь вовсе не в разных отношениях к данному явлению, а в том, что категоричность первого суждения приводит к однозначности ответов, структура второго суждения менее категорична, да и оценка его более мягкая — вместо слова «отвратительно» используется слово «плохо».

В том случае, когда приходится формулировать суждения относительно менее эмоциональных предметов, чем это приведено в примере с алкоголем, эта опасность усиливается, так как компоненты аттитюда потребителя ни разу не подвергались ревизии самим потребителем — если об отношении к спиртным напиткам он мог высказываться сам и слышать мнения других, то по отношению к огромному множеству товаров он своего отношения не высказывал и мнения о нем других потребителей не слышал. Его установка еще четко не определена им самим.

При изучении потребительских предпочтений, которые всегда динамичны, специалист иногда задается вопросом о том, как эти предпочтения выражаются конкретными потребителями, преломляется в их сознании, превращается в определенные нормы и образы. Для проникновения во внутренний мир потребителя недостаточно использования заранее заданных схем, однозначно понимаемых категорий. Жестко формализованные, структурированные методы сбора информации в этом случае будут работать плохо. Возникает необходимость привлечения понятий и категорий, которыми пользуются сами люди для упорядочивания своего собственного повседневного опыта. Совокупность этих методов получила название «проективные методы». Эти методы нацелены на проецирование субъективных свойств личности в различных объектах. Респонденту предлагаются различные объекты в виде знаков, текстов, картинок, ситуаций и по его реакциям определяются скрытые неосознаваемые мыслительные процессы, потребности, образы.

Чаще всего в экономической практике используется метод семантического дифференциала Ч. Осгуда, разработанный в середине 1950-х гг. для изучения эмоционального отношения людей к различным понятиям для определения их смысла. Его суть заключается в следующем. Респонденту предлагается выразить отношение к некоторому объекту по совокупности биполярных шкал, в подавляющем большинстве случаев семибалльных (см., например, табл. 2.2).

Таблица 2.2

Пример метода семантического дифференциала Ч. Осгуда

слабый	–3	–2	–1	0	+1	+2	+3	сильный
детский	–3	–2	–1	0	+1	+2	+3	взрослый
пассивный	–3	–2	–1	0	+1	+2	+3	активный
медленный	–3	–2	–1	0	+1	+2	+3	быстрый
простой	–3	–2	–1	0	+1	+2	+3	сложный
ложный	–3	–2	–1	0	+1	+2	+3	правдивый
плохой	–3	–2	–1	0	+1	+2	+3	хороший

Как видно из таблицы, крайние позиции на шкалах описаны *вербальными антонимами*, а *число баллов, изменяющихся* от –3 до 3, составляет 7. Вся совокупность шкал образует исходное пространство шкал. Число градаций на шкале может быть и другим, не равным семи. Помимо вербальных антонимов в качестве крайних значений шкал могут быть использованы и другие знаки, например, пары «черный квадрат — белый квадрат», «улыбка — гримаса огорчения», «стрелка вверх — стрелка вниз» и т.п. Выбор этих знаков, описывающих крайние позиции, определяется конкретными ситуациями. Например, при исследовании отношений детей к детским товарам удобнее пользоваться знаками, так как вербальные антонимы могут быть ими не правильно поняты, а знак улыбки или огорчения легко распознается.

Эксперименты с различными совокупностями биполярных шкал показали, что все они распадаются на три группы факторов, такие как *СИЛА, АКТИВНОСТЬ, ОТНОШЕНИЕ.* Этот феномен и был открыт Ч. Осгудом и назван им синестезией. В приведенном выше примере к первой группе, относящейся к фактору *СИЛА*, принадлежат первая и вторая шкалы; ко второй группе, относящейся к фактору *АКТИВНОСТЬ*, принадлежат третья и четвертая шкалы; последние три шкалы — пятая, шестая и седьмая — принадлежат к фактору *ОТНОШЕНИЕ.*

Ч. Осгуд предложил 20 шкал, которые считаются классическими и предпочтительно используются на практике: слабый — сильный; женский — мужской; пассивный — активный; медленный — быстрый; необычный — обычный;

ложный — правдивый; плохой — хороший; жестокий — добрый; кривой — прямой; разболтанный — пунктуальный; вкусный — безвкусный; неудачный — удачный; твердый — мягкий; глупый — умный; новый — старый; неважный — важный; острый — округлый; хладнокровный — восторженный; бесцветный — красочный; необычный — обычный; красивый — безобразный.

Наличие сложной трехкомпонентной структуры аттитюдов потребителей находит явные параллели с тремя выделенными факторами, которые, являясь по сути ортогональными, составляют собой трехмерное пространство, получившее название семантического. Каждый объект, оцениваемый по трем выделенным факторам, как бы занимает некоторое место в семантическом пространстве. Сравнение положений разных объектов в этом пространстве позволяет получить важную информацию. Это осуществляется следующим образом.

Вначале вычисляется средняя оценка объекта по каждому из трех факторов. Оценка по фактору для объекта равна сумме оценок по всем шкалам, входящим в этот фактор, и по всем респондентам, деленной на величину, равную произведению числа шкал и числа респондентов. После вычисления средней оценки по всем факторам в отдельности осуществляется переход к вычислению близости между объектами. Оценка этой близости представляет собой семантический дифференциал — квадратный корень из суммы квадратов разницы между вычисленными средними для всех трех факторов.

Пример

Пусть получены семантические оценки для товаров А и Б так, как это показано в табл. 2.3 и 2.4. Для каждой из шкал в таблицах приведены средние значения оценок. Пусть первая группа факторов Ч. Осгуда представлена первыми двумя шкалами (в табл. 2.3 и 2.4 выделены подчеркиванием), вторая группа факторов представлена третьей и четвертой шкалами (в табл. 2.3 и 2.4 выделены полужирным шрифтом), третья группа факторов — пятой, шестой и седьмой шкалами (в табл. 2.3 и 2.4 выделены полужирным подчеркиванием).

Тогда легко вычислить средние оценки каждого фактора, данные группой респондентов, и после этого величину семантического дифференциала. Результаты вычислений приведены в табл. 2.5.

Таблица 2.3

Семантические оценки товара А, данные пятью респондентами

Номер шкалы	Оценка 1	Оценка 2	Оценка 3	Оценка 4	Оценка 5	Средняя
1	2	1	3	3	1	2
2	2	2	2	2	2	2
3	**0**	**1**	**–1**	**0**	**0**	**0**
4	**–1**	**0**	**–1**	**–1**	**0**	**–0,6**
5	**–3**	**–2**	**–2**	**–3**	**–2**	**–2,4**
6	**–2**	**–1**	**0**	**0**	**–1**	**–0,8**
7	**–3**	**1**	**1**	**–2**	**–2**	**–1,0**

Таблица 2.4

Семантические оценки товара Б, заданные пятью респондентами

Номер шкалы	Оценка 1	Оценка 2	Оценка 3	Оценка 4	Оценка 5	Средняя
1	1	1	0	1	–1	0,4
2	3	2	0	0	1	1,2
3	**2**	**2**	**2**	**2**	**1**	**1,8**
4	**0**	**1**	**1**	**–1**	**0**	**0,2**
5	**1**	**–1**	**0**	**0**	**–2**	**–0,4**
6	**1**	**1**	**1**	**1**	**–1**	**0,6**
7	**–3**	**–3**	**–3**	**–2**	**–2**	**–2,6**

Таблица 2.5

Результаты вычисления семантического дифференциала

Наименование фактора	Средняя оценка по товару А	Средняя оценка по товару Б	Модуль разности	Квадрат разности
СИЛА	2	0,8	1,2	1,44
АКТИВНОСТЬ	**–0,3**	**1,0**	**1,3**	**1,69**
ОТНОШЕНИЕ	**–1,4**	**–0,8**	**0,6**	**0,36**

Сложив значения последнего столбца табл. 2.5 и найдя корень квадратный из полученного выражения, найдем численное значение семантического дифференциала, которое для данной пары товаров будет равно 1,868. Само по себе это значение не выражает ничего особенного. Единственное, что можно сказать о полученной величине, так это то, что она каким-то образом характеризует различие в оценках товаров со стороны потребителей. Эта величина получает смысловое наполнение только в сравнении с другими подобными величинами, т.е. с другими семантическими дифференциалами. В этом случае можно будет анализировать силу различий между товарами, находить товары, близкие по отношению к ним со стороны потребителей, получать другую информацию.

Следует подчеркнуть, что информация, используемая в шкалах семантического дифференциала Ч. Осгуда, измеряется в шкале отношений. Ее преобразование в информацию, заключающуюся в семантическом дифференциале, относится к шкале расстояний. Именно это обстоятельство делает невозможным сравнение результатов данного исследования с другими исследованиями, поскольку такое сравнение возможно только в универсальной метрической шкале. Это означает, что семантические дифференциалы следует использовать только в рамках одного исследования, когда отношения к различным объектам высказывает одна и та же группа потребителей. В том случае, когда имеются результаты измерения с помощью семантического дифференциала Ч. Осгуда разных групп потребителей, их сравнение не имеет смысла — масштабы, отношения, балльность могут существенно отличаться для разных групп потребителей.

Важнейшей экономической задачей любой фирмы, работающей в условиях конкурентного рынка, является задача прогнозирования потребительского спроса, потребительских предпочтений, а для ее успешного выполнения необходимо выявить устойчивые причинно-следственные связи между факторами. Поэтому собираемая информация должна быть в максимальной степени упорядочена. Рассмотренные ранее шкалы, а также процедуры логического прямоугольника и шкалирования Лайкерта в некоторой степени способствуют достижению этой цели. Однако сама процедура упорядочения любых объектов по возрастанию или убыванию некоторого их свойства представляет собой оригинальную процедуру, получившую название «ранжирование». Поскольку сама эта процедура чрезвычайно важна для получения экономической информации, следует рассмотреть ее более подробно.

Объекты ранжирования многообразны. Поэтому для проведения ранжирования необходимо определить свойство, по которому объекты упорядочиваются, т.е. выявить основание ранжирования. Понятно, что в зависимости от выбранного основания ранжирования упорядочение может быть самым различным. Так, например, если попытаться ранжировать способы проезда студента к университету, то сразу возникнет вопрос: что выбрать в качестве основания для ранжирования? Если им станет минимум времени, затраченного на преодоление расстояния, то упорядочение способов транспортировки будет осуществлено так:

1) вертолет;
2) собственный автомобиль;
3) такси;
4) маршрутное такси;
5) автобус, трамвай или троллейбус;
6) велосипед;
7) пеший ход.

Если же в качестве основания выбрать минимум денежных затрат, то упорядочение способов транспортировки будет осуществлено иначе:

1) пеший ход;
2) велосипед;
3) автобус, трамвай или троллейбус;
4) маршрутное такси;
5) собственный автомобиль;
6) такси;
7) вертолет.

При другом основании ранжирования будет получено другое упорядочение указанных способов. В любом случае в порядке осуществления процедуры упорядочения получается ранжированный ряд объектов ранжирования, в котором каждому объекту приписывается ранг — место в этом ряду. Число мест, как и число рангов, равно числу объектов.

Объекты ранжирования в рассматриваемой совокупности могут различаться с точки зрения выраженности в них заданного свойства, либо некоторые из них могут быть с точки зрения этого свойства неразличимыми. В первом случае все ранги ряда будут различными, а во втором — у нескольких объектов появятся одинаковые ранги. Такие ранги называют связанными.

Пример

В табл. 2.6 приведены результаты ранжирования потребителей по признаку их дохода. В первой строке таблицы буквами А, Б, В, Г и т.д. обозначены потребители. Во второй — приводятся доходы потребителей за определенный промежуток времени (цифры, определяющие доход, приведены в условных единицах). В третьей строке даются ранги каждого из потребителей. Упорядочение объектов ранжирования в таблице приводится по критерию убывания дохода.

Таблица 2.6

Результаты ранжирования по уровням дохода

Потребитель	А	Б	В	Г	Д	Е	Ж	З	И
Доход	100	150	120	100	90	110	80	90	90
Ранг	4,5	1	2	4,5	7	3	9	7	7

Результаты ранжирования, приведенные в таблице, требуют комментариев. На первом месте по рангу стоит гражданин, обозначенный буквой Б — у него максимальный доход, равный 150 единицам. На втором и третьем местах по уровням дохода стоят граждане, обозначенные буквами В и Е. Им присвоены соответствующие ранги 2 и 3. У граждан, обозначенных буквами А и Г, одинаковые доходы (по 100 денежных единиц у каждого), значит, они имеют одинаковые (связанные) ранги. Их ранг получается сложением мест (4 + 5) и делением этой величины на 2. В результате получилась величина 4,5. Граждане с доходом, равным 90 единиц, занимают шестое, седьмое и восьмое места. Поэтому их ранг будет равен величине (6 + 7+ 8) / 3 = 7, что и зафиксировано в таблице. Последний ранг (9) присвоен гражданину, обозначенному буквой Ж.

Процедура ранжирования может оказаться весьма полезной в дальнейшем, если провести ранжирование указанных граждан по другим основаниям. После этого можно провести сравнение этих ранжированных рядов на степень их согласованности друг с другом. Если эта согласованность будет обнаружена, то можно говорить о взаимосвязанности между свойствами, по которым осуществляется ранжирование. Это может оказаться весьма важным обстоятельством для целей прогнозирования поведения потребителей.

В рассмотренном примере ранжирование служит целям упорядочения объективно существующей информации, и является очень важным приемом исследования. Бывает, что в ходе экономических исследований с помощью процедуры ранжирования решается и другая задача — измерения потребительских предпочтений.

Чаще всего процедура прямого ранжирования используется, когда потребитель присваивает ранги каждому объекту ранжирования. Понятно, что каждый потребитель в общем случае задаст собственный ранжированный ряд, который будет отличаться от аналогичных рядов, полученных другими потребителями. Анализ полученных ранжированных рядов может позволить исследователю выявить систему предпочтений данной группы потребителей в среднем. Как правило, при обработке ранжированных рядов используют элементарные процедуры с использованием таких простых статистических показателей, как мода и медиана.

Методы прямого ранжирования, как легко убедиться, просты. В этом их существенное преимущество. В то же время сама процедура ранжирования приводит к измерению свойств ранжируемого ряда в очень простой шкале — шкале отношений. Действия с этой шкалой измерения ограничены. Это тем более досадно, что исходный ряд, как правило, измерен в шкале более высокого уровня.

Более информативная шкала расстояний может быть получена при проведении ***процедуры парных сравнений Терстоуна***. Данный метод был разработан Л. Терстоуном и впервые использован им для ранжирования преступлений по степени серьезности. Он заключается в том, чтобы опрашиваемый осуществил попарное сравнение объектов ранжирования по заданному основанию. Легко подсчитать, что если для анализа выбрано N объектов, то число попарных сочетаний этих объектов будет равно $N(N-1)/2$. В том случае, когда число объектов равно, например 10, число всех возможных пар будет равно $10 \times 9 / 2 = 45$. В том случае, когда объектов будет 20, множество пар резко увеличится и станет равным 190! Поэтому необходимо иметь в виду, что число объектов для проведения успешного исследования должно быть невелико, иначе опрашиваемый может запутаться в своих предпочтениях, да и объемы вычислений так увеличатся, что получение необходимых результатов будет затруднительно.

При проведении процедуры попарных сравнений респонденту предлагается поочередно сравнить объекты, сгруппированные по парам, а результаты сравнения — свести в специальную таблицу (см., например, табл. 2.7). Объекту, которому отдается предпочтение, присваивается единица, объекту, который проигрывает это сравнение, ставится нуль. Если, на взгляд опрашиваемого, объекты равноценны, им выставляются одинаковые баллы (0,5). Таблица, в которую заносятся результаты сравнения, представляет собой прямоугольник в клетки которого, находящиеся на пересечении столбцов и строк, заносятся результаты сравнения (за исключением последнего столбца, куда заносятся итоговые значения). Диагональные элементы таблицы представляют собой результат сравнения объекта самого с собой. Такое сравнение бессмысленно, поэтому во всех диагональных элементах таблицы ставятся прочерки.

Таблица 2.7

Результаты парных сравнений одним респондентом

	Объект 1	Объект 2	Объект 3	Объект 4	Объект 5	Число предпочтений
Объект 1	—	1	1	1	0	3
Объект 2	0	—	1	0	0	1
Объект 3	0	0	—	1	1	2
Объект 4	0	1	0	—	0,5	1,5
Объект 5	1	1	0	0,5	—	2,5

Что означает единица, выставленная в табл. 2.7 на пересечении второго столбца с первой строчкой? Это означает, что первый объект предпочтительнее второго. Именно поэтому в клеточке, представляющей собой пересечение второй строчки с первым столбцом, выставлена нулевая оценка. Таблица включает два элемента с одинаковой оценкой (0,5). Это говорит о том, что данные объекты респондентом оцениваются как одинаковые.

Ранжирование проводится по результатам числа предпочтений, которое получается суммированием по строкам для каждого объекта и заносится в последний столбец таблицы. Для табл. 2.7 ранжирование объектов произведено так (по ходу уменьшения ранга): объект 1, объект 5, объект 3, объект 4, объект 2. При этом сами ранги в табл. 2.7 представляют информацию, измеренную в шкале расстояний. А это означает, что с полученными значениями работать удобнее, чем с простыми рангами. Действительно, ранги характеризуют здесь не только занимаемое место в системе упорядоченных предпочтений, но и степень, силу этого предпочтения.

В ходе экономических исследований интерес вызывает мнение не одного, а многих респондентов с тем, чтобы определить предпочтение группы потребителей, относящихся, например, к выбранному сегменту. Для этого необходимо, чтобы процедуру парного сравнения прошли и другие респонденты данного сегмента. Предметом анализа для экономиста является не содержание таблицы, а обобщенные результаты, которые сведены в последнем столбце каждой таблицы, названном «Число предпочтений». Обычно информация из этого столбца каждой индивидуальной таблицы выписывается и обобщается. Само обобщение осуществляется так.

По каждому объекту (в табл. 2.7 таких объектов пять) определяется число предпочтений, которое он получил от всех респондентов. Для этого складывают числа предпочтений по каждому из объектов, которые он получил от каждого опрашиваемого. Удобнее всего это сделать в форме, пример которой приведен в табл. 2.8. Всего было опрошено восемь потребителей, каждый из которых заполнил таблицу наподобие табл. 2.7. Последние столбцы каждой из восьми таблиц типа 2.7, в которых рассчитаны числа предпочтений, представляют собой столбцы новой таблицы —2.8. По каждому из объектов легко рассчитать итоговую сумму чисел предпочтений. Однако работать с суммой указанных чисел не очень удобно, поскольку итоговые цифры могут быть значительными, и дать им толкование будет сложно. Поэтому чаще всего полученную сумму делят на общее число ответивших и получают среднюю оценку.

В табл. 2.8 приведены результаты парных сравнений каждого респондента по каждому из объектов сравнения, а также суммарная оценка и средняя арифметическая оценка каждого объекта.

Таблица 2.8

Результаты парных сравнений

Респонденты	1	2	3	4	5	6	7	8	Сумма	Средняя сумма
Объект 1	3	4	2,5	4	4	6	2	3	28,5	3,56
Объект 2	1	1	1	0	0	1	2	1	7	0,87
Объект 3	2	1	2	1	2	1	2	1	12	1,5
Объект 4	1,5	2	2	2	2	1	2	2	14,5	1,81
Объект 5	2,5	2	2,5	3	2	1	2	3	18	2,25

Последний итоговый столбец дает общую среднюю оценку предпочтений группы респондентов (в рассматриваемом примере их восемь). Для простых исследований итоговой цифры вполне достаточно. Однако следует проверить и достоверность полученных значений по нескольким позициям.

Первая позиция определяется количеством респондентов. Чем больше респондентов, отнесенных к данному сегменту опрошенных, тем более достоверными оценками истинного отношения к объектам со стороны потребителей данного сегмента являются средние арифметические. При этом, конечно, необходимо помнить, что респонденты должны в обязательном порядке входить в выборку из одного и того же сегмента, иначе она будет неоднородной и полученные значения будут содержать в себе существенную ошибку (расчетные значения будут смещены относительно истинных значений).

Вторая позиция определяется статистическими характеристиками, вычисляемыми по результатам парных сравнений. В данном случае в первую очередь необходимо убедиться в том, что средняя арифметическая является лучшей оценкой изучаемого общего мнения. Для этого по данным таблицы типа 2.8 проверяется гипотеза о нормальном распределении ответов респондентов. После того, как гипотеза о нормальном характере распределения вероятностей подтверждена, рассчитываются и другие параметры, характеризующие степень доверия к полученным результатам. В самом начале рассчитывается дисперсия, которая отражает степень разброса мнений респондентов относительно средних значений по каждому из объектов. После получения ее численного значения определяются доверительные границы мнения респондентов относительно данного объекта. Затем они приводятся к относительным величинам и к процентам. Если разброс значений, находящихся в доверительных границах, высок, то необходимо провести дополнительные исследования. Вполне возможно, например, что в ходе обработки данных была допущена арифметическая ошибка.

Таким образом, многообразие задач социально-экономического прогнозирования предопределяет и многообразие методов прогнозирования, которое формируется не только различными периодами упреждения, разными типами характера динамики, но и тем, в какой шкале измерена соответствующая информация. Именно поэтому методы социально-экономического прогнозирования оригинальны и существенно отличаются от методов прогнозирования других объектов.

2.3. Предварительный анализ и обработка данных

Первая задача, которую необходимо решить при прогнозировании с помощью фактографических методов, — собрать необходимую информацию о прогнозируемом объекте и обработать ее так, чтобы она была пригодна для дальнейшего использования в моделировании и прогнозировании. Несмотря на кажущуюся простоту этой задачи, прогнозисту весьма часто приходится сталкиваться с ситуацией, когда он не может решить ее сразу. При постановке задачи прогнозирования формируется ряд показателей, отражающих основные характеристики объекта прогнозирования. Информацию об этих показателях за некоторый определенный задачами прогнозирования промежуток времени и должен собрать прогнозист.

Различают первичные (полевые) и вторичные (кабинетные) методы исследования и сбора необходимой информации[1].

Основными методами получения *первичных данных* являются опрос, наблюдение и эксперимент. Методы опроса, наблюдения, эксперимента и панельных исследований могут использоваться как отдельно, так и в сочетании с другими методами. Более подробно с ними знакомит дисциплина «Маркетинговые исследования». Первичная информация отличается особой достоверностью, поскольку исследователь сам собирает и обрабатывает эту информацию. Но следует отметить, что она касается исключительно микроуровня (поведение потребителей; изменение цен в группе магазинов или на локальном рынке (например, ММВБ); передвижение товаров по магистралям и т.п.). Способ получения первичной информации в ходе эмпирических исследований обладает существенными *недостатками*.

1. Стоимость информации оказывается весьма высокой, поскольку ее сбор осуществляется в результате «полевых» исследований, для которых привлекаются значительные силы; на подготовку, проведение и обработку полученных результатов затрачивается много времени.

2. Объем полученной информации незначителен и не всегда удовлетворяет прогнозиста. Например, он может получить информацию о динамике изменения потребительских предпочтений, но не узнать, почему происходит эта динамика — факторы, оказывающие на нее влияние, могут остаться невыявленными. В результате исследователь будет не в состоянии сделать верный прогноз.

3. Для того, чтобы в ходе проведения полевых исследований получить достоверную информацию, необходимо

---

[1] *Светуньков С. Г.* Методы маркетинговых исследований. СПб. : Изд-во ДНК, 2003.

проводить эти исследования на довольно высоком уровне — обязательно следует привлекать для этого специалистов соответствующей квалификации, чаще всего, когда речь идет о непосредственном контакте с некоторым социумом или его представителями для получения от них первичной информации, — социологов. Попытки экономиста самостоятельно решить эту задачу чаще всего заканчиваются неудачей.

Таким образом, сбор первичной информации для прогнозирования могут позволить себе только организации, имеющие для этого достаточное количество свободных средств, для которых точность прогноза имеет определяющее значение.

В связи с вышесказанным прогнозисты чаще всего удовлетворяются использованием вторичной информации.

Вторичные исследования, как правило, базируются на уже имеющейся информации и потому носят название кабинетных исследований. По своему содержанию они представляют собой анализ имеющихся источников информации об изучаемой проблеме.

Различают внешние и внутренние (по отношению к объекту социально-экономического прогнозирования) источники для вторичных исследований. В качестве внутренних источников могут использоваться:

- внутренняя статистика (характеристика товарооборота, объем сбыта, объем распродаж, импорт, экспорт, рекламации);
- данные о производственных затратах;
- прочие данные (о производительности установок, оборудования, прайс-листы на сырье и материалы, характеристика системы складирования, карты потребителей и др.).

В качестве внешних источников выступают:

- публикации национальных и международных официальных организаций;
- публикации государственных органов, министерств, муниципальных комитетов и организаций;
- публикации торгово-промышленных палат и объединений;
- ежегодники статистической информации;
- отчеты и издания отраслевых фирм и совместных предприятий;
- книги, сообщения в журналах и газетах;
- публикации учебных, научно-исследовательских, проектных институтов и общественно-научных организаций, симпозиумов, конгрессов, конференций;

• прайс-листы, каталоги, проспекты и другие фирменные публикации.

Значимость для вторичных исследований внутренней и внешней информации в каждом конкретном случае определяется исследователем. Основным достоинством вторичных исследований является то обстоятельство, что затраты на проведение кабинетных исследований существенно меньше, чем на получение такого же объема информации с помощью полевых исследований. Другим существенным преимуществом кабинетных исследований для сбора информации, пригодной для прогнозирования, является то, что по сравнению с полевыми исследованиями они затрачивают на порядок меньше времени на сбор и обработку информации и прогноз можно получить в оперативном порядке.

К сожалению, кабинетные исследования могут привести к получению недостоверной информации, поскольку опираются на вторичную информацию, которая была собрана другими исследователями и может быть недостаточно высокого уровня.

В любом случае полевые исследования всегда дороже кабинетных. Поэтому они применяются в случаях, когда:

• в результате вторичного исследования не достигнут требуемый результат и невозможно прогнозирование динамики;

• высокие затраты на полевые исследования могут быть компенсированы существенным повышением точности прогноза за счет получения в качестве базы достоверной и представительной информации.

На практике поступают следующим образом. Вначале проводят кабинетные исследования, собирают и анализируют максимально возможное число вторичной информации. Если такой информации оказывается недостаточно, проводят полевые исследования. Если взять объем информации, с которым работают современные прогнозисты, за 100%, то можно с уверенностью утверждать, что в 95% случаев прогнозисты работают с вторичной информацией, которую уже кто-то собрал и обработал. И здесь прогнозист может столкнуться с проблемой недостоверности имеющейся в его распоряжении информации либо с ее отсутствием. Обращаясь к опубликованным источникам, прогнозист, прежде всего, отметит, что информации много, но той, которая ему нужна, в наличии нет. Очень часто приходится сталкиваться с тем, что в имеющемся массиве данных нет того показателя, который необходим для осуществления прогноза. В таком случае прогнозисту приходится использовать другие показатели, имеющие похожий или аналогичный смысл. Например, среди доступных статистических данных может не оказаться данных по среднемесячной зарплате. Тогда прогнозист вынужден будет использовать близкий по смыслу показатель среднедушевых доходов, хотя по экономическому смыслу и содержанию этот показатель отличается от требуемого.

Но если предположить, что подготовленная к анализу информация полностью ясна и отвечает необходимым условиям с позиций ее смыслового содержания, то следует помнить, что она может содержать в себе другие ошибки, которые могут сильно повлиять на точность прогноза. От этих ошибок по возможности необходимо избавляться, но для того, чтобы это сделать, следует понять, какого рода ошибки могут содержаться в статистической информации.

По своим свойствам и характеру влияния на результаты наблюдений ошибки подразделяют на грубые, систематические и случайные[1].

Грубые ошибки возникают в случае невнимательности человека, эту информацию записывающего, передающего или получающего. В результате невнимательности измеренное число увеличивается или уменьшается на порядок и резко выделяется в общей совокупности данных. С такой ошибкой несложно бороться — она легко обнаруживается и устраняется. Такие значения в математической статистике часто называют «выбросами». Единого мнения относительно того, что делать с выбросами, нет, однако принято поступать одним из следующих способов: просто избавиться от наблюдения, интерполировать наблюдение (т.е. заменить не выбивающимся из динамики) либо каким-то образом учесть выброс в модели. Стоит обратить внимание, что при построении модели по ряду данных с такими выбросами, исследователь может получить искаженную модель, несущую в себе эту самую грубую ошибку, что впоследствии может привести к неточному прогнозу. Однако прежде чем принимать какие-либо меры относительно выбросов, нужно понять, чем вызвано полученное значение: является ли оно следствием внешних шоков, следствием изменяющейся структуры предприятия или же просто появилось из-за ошибки, допущенной при сборе данных.

---

[1] *Большаков В. Д.* Теория ошибок наблюдений. М. : Недра, 1983. С. 80—81.

Для того чтобы выявить сами выбросы, существует несколько возможных процедур.

I. **Визуальный анализ графического представления**. Статистические данные наносят на некоторый график. Если рассматривается статистическая взаимосвязь между двумя факторами, то изучается эта взаимосвязь. Если изучаются динамические ряды, то строится зависимость изменения показателей во времени. Данные, которые резко выделяются из совокупности наблюдений, отбрасываются. Конечно, надо быть полностью уверенным в том, что это наблюдение не является проявлением некоторых факторов, присущих анализируемому процессу. Ведь возможны ситуации, когда на объект начинают действовать некоторые новые факторы, влияние которых еще не известно, а в будущем это приведет к резким отклонениям от общей тенденции. Резкое отклонение, находящееся в начале или посередине довольно продолжительного отрезка времени, можно смело отбрасывать, а вот в случае, когда отклонение наблюдается в конце ряда наблюдений, необходимо выяснить, действительно ли такое отклонение является результатом грубой ошибки, или же это предвестник возможных серьезных изменений в прогнозируемом объекте?

При визуальном анализе принято обращаться к следующим типам графиков.

1. *Линейный график* показывает изменение показателя во времени. По оси абсцисс откладывается время наблюдения, а по оси ординат — значение показателя (рис. 2.1).

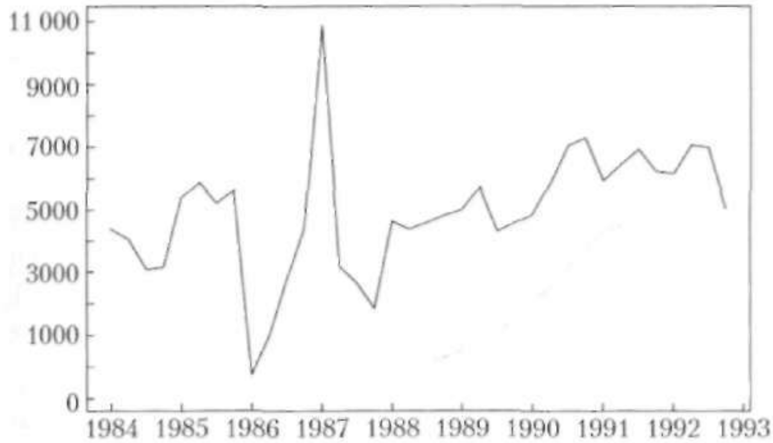

Рис. 2.1. **Линейный график по показателю**

На данном графике представлены квартальные данные некоторого показателя. Как видим, из общей динамики выбиваются значения за 1 квартал 1986 г. и 1 квартал 1987 г. Для дальнейшей работы с этим рядом данных требуется понять, чем вызваны эти всплески, соответственно, необходимо провести дополнительное исследование объекта прогнозирования.

2. *График «сезонных изменений»* отражает изменение показателя в данном сезоне из периода в период и обычно применяется для временных рядов с сезонностью. По оси абсцисс откладываются периоды (например, годы), а по оси ординат — значения показателя в данный сезон (например, только январские значения) (рис. 2.2).

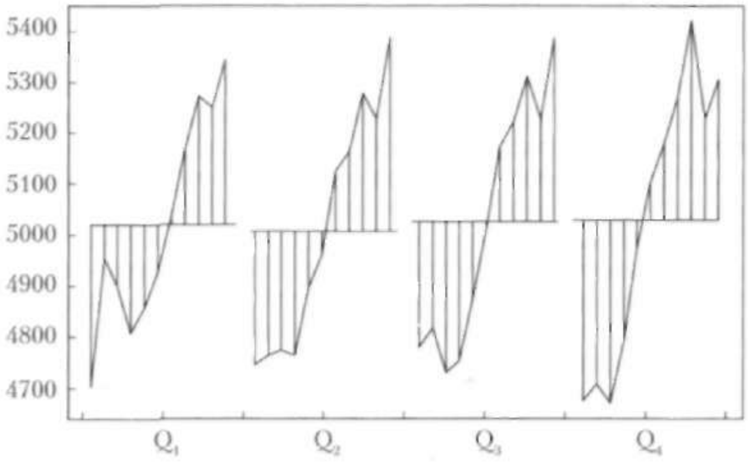

Рис. 2.2. **График сезонных изменений показателя**

Данный рисунок показывает, как из года в год меняется значение показателя по кварталам. Видим, что из общей динамики выбивается второе значение за первый квартал: по другим кварталам оно значительно ниже, а по первому кварталу за ним следует более плавное изменение.

3. *Точечная диаграмма* отражает изменение одного показателя при изменении другого. По оси абсцисс откладываются значения одного фактора, по оси ординат — другого в те же моменты наблюдений (рис. 2.3).

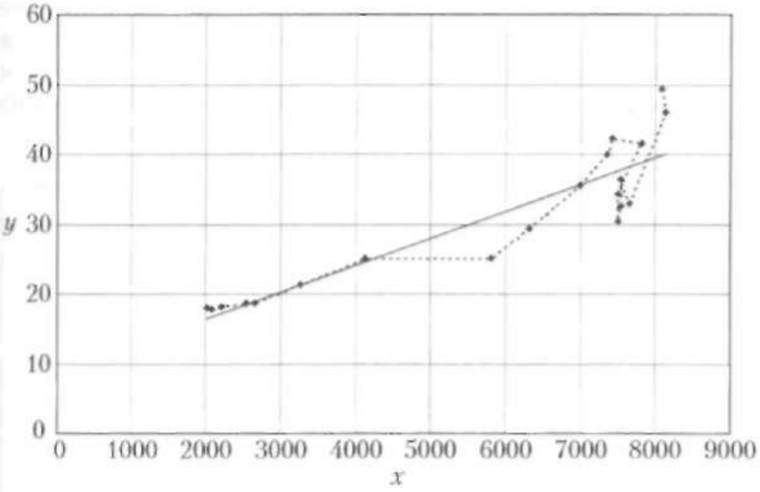

Рис. 2.3. **Точечная диаграмма**

В случае с пространственными данными получающиеся точки на такой диаграмме не соединяются никакими линиями, так как значения ряда данных не имеют естественного упорядочивания. В приведенном выше случае, с временным рядом, мы соединили точки для того, чтобы иметь представления о том, как изменялась связь между показателями во времени. Иногда это не имеет никакого смысла, так же, как и с пространственными данными. Однако в данном случае оказалось полезным. С чисто математической точки зрения выбросами на данной диаграмме можно было бы считать точки, лежащие дальше от прямой, проведенной через все облако точек. К таким значениям, возможно, следовало бы отнести последнее наблюдение. Но соединение точек, которое мы осуществили на графике, показывает, что первая часть ряда укладывается в одну кривую, которая хорошо описывает эту связь (со случайными отклонениями от нее), а последние 8 наблюдений (точки в правой части графика) формируют уже несколько иную зависимость между x и y: с изменением x на единицу происходит более резкое изменение y. Здесь мы сталкиваемся с эволюционным характером исследуемого процесса, а значит, признавать последнее наблюдение «грубой ошибкой» было бы неправильно.

4. *Гистограмма* – график, показывающий распределение исследуемого показателя. По оси абсцисс откладываются значения показателя, по оси ординат — частота появления тех или иных значений, которая может быть представлена как абсолютным, так и относительным числом (рис. 2.4).

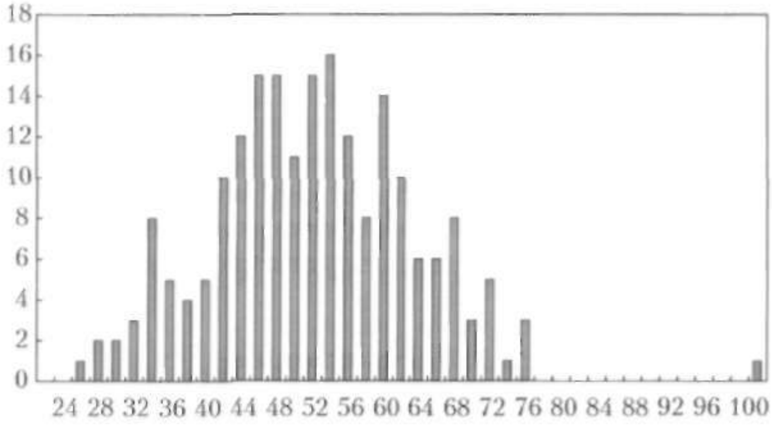

Рис. 2.4. **Гистограмма**

На себя обращает внимание значение, лежащее в правом «хвосте» распределения: среди наблюдений показателя оказалось какое-то одно, значительно отстоящее от всех остальных. Оно может быть признано выбросом.

Заметим, что гистограмма по временны́м рядам не всегда позволяет сделать однозначный вывод о наличии грубых ошибок. Выбивающееся значение может быть вызвано какими-то объективными причинами динамики показателя и может объясняться соответствующими значениями показателя, влияющего на изучаемый.

5. *Ящичковая диаграмма* (в английском — «boxplot») — специальный график, дающий визуальное представление об основных статистических характеристиках показателя (рис. 2.5).

Слева от ящичка представлены значения случайной величины. Они соответствуют значениям на гистограмме по плотности распределения. Тело ящичка ограничивает 50% всех наблюдений, нижняя его часть характеризует нижний квартиль Q_1 (т.е. ниже этого значения лежит 25% всех наблюдений), а верхняя — соответственно верхний квартиль Q_3.

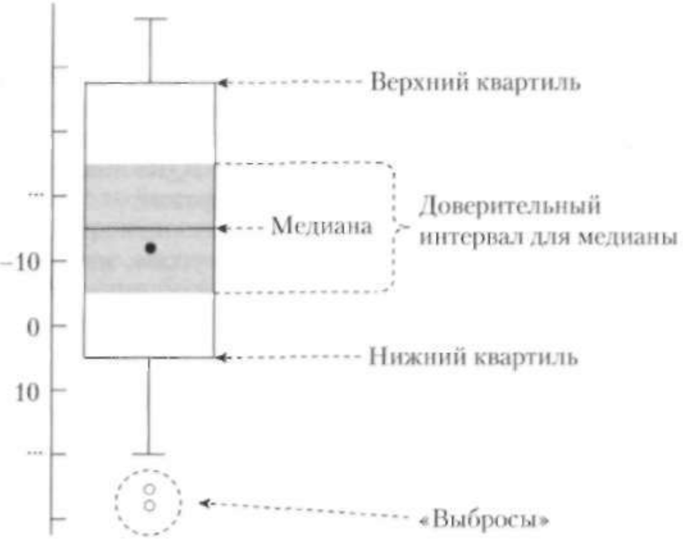

Рис. 2.5. **Схема ящичковой диаграммы**

Расстояние между верхним и нижним квартилем называется «интерквартильным расстоянием». В английском языке оно обозначается как IQR и рассчитывается как простая разность между верхним и нижним квартилями:

$$IQR = Q_3 - Q_1. \quad (2.2)$$

В середине ящика горизонтальной линией отмечена медиана — величина, ниже которой лежит 50% всех наблюдений.

Серой областью вокруг медианы обозначен доверительный интервал, рассчитываемый на основе все того же интерквартильного расстояния по формуле:

$$Me(x) \pm 1.57\frac{IQR}{\sqrt{T}}, \quad (2.3)$$

где $Me(x)$ — медиана по выборке исследуемой случайной величины; T — число наблюдений в выборке.

Иногда вместо темной области на ящичковой диаграмме изображают сужение к медиане (нечто похожее на женскую талию). Интерпретация области такого сужения идентична описанной выше.

Сплошной точкой в середине обозначена средняя величина по выборке. Если она совпадает с медианой, то можно заключить, что распределение случайной величины симметрично. Если точка находится в пределах медианного доверительного интервала, то нельзя с уверенностью говорить, что медиана и средняя величина статистически различаются.

«Усы» ящичка показывают значения, лежащие внутри интервала, рассчитываемого на основе IQR: нижний «ус» начинается от значения $Q_1 - 1{,}5 \cdot IQR$, а верхний — заканчивается значением $Q_3 + 1{,}5 \cdot IQR$.

Точками, выходящими за пределы «усов», изображены «выбросы».

Фактически ящичковая диаграмма в некотором роде заменяет гистограмму и дает более наглядное представление о распределении случайной величины.

II. **Визуальный анализ табличных данных**. Данные, которые невозможно или нежелательно отображать графически, анализируют непосредственно в таблицах. Например, если данные содержат в себе несколько тысяч наблюдений, их графическое изображение будет представлять собой хаотическое нагромождение точек из-за того, что их общее количество велико, а размеры графика ограничены.

Возможен и другой вариант — когда изучается многофакторная зависимость, графическое изображение которой осуществить довольно сложно. Тогда имеющиеся данные располагают в таблице в некотором порядке, присущем анализируемому процессу. Например, в порядке поступления информации (временные ряды) или возрастания какого-либо признака. После того, как произведено такое упорядочивание, проводится анализ полученной последовательности. Наблюдения, которые визуально выбиваются из этого порядка, отбрасываются. При этом необходимо иметь в виду, что такой анализ субъективен, и очень высока вероятность того, что исследователь может не заметить в общей совокупности информационных массивов наблюдение, содержащее грубую ошибку.

III. **Статистический анализ данных**. В том случае, когда визуальный анализ затруднен из-за большого количества данных, следует использовать статистические процедуры. Полученные данные проверяются на одну из статистических гипотез относительно характера распределения вероятностей появления анализируемой совокупности данных. Затем данные, противоречащие этому закону, отбрасываются. Чаще всего предполагают наличие нормального закона распределения. В самом простом случае, когда измеренные величины колеблются около некоторого значения, рассчитываются средняя арифметическая этого значения, дисперсия, доверительные интервалы. Данные, выходящие за эти интервалы, отбрасываются. Естественно, что доверительные границы на этапе предварительного анализа данных должны быть наиболее широкими, с очень высоким уровнем доверительной вероятности.

В случае наличия «выбросов» в исходных данных для получения более точной оценки связи между признаками требуется каким-то образом нивелировать их влияние на оценку модели. Прежде, чем что бы то ни было делать с идентифицированными «выбросами», надо разобраться в том, что они собой представляют и почему произошли. Так, прогнозисту стоит идентифицировать время появления «выброса» и найти экономическое обоснование его существования. В случае если такого обоснования нет, «выброс» можно признать грубой ошибкой. В таком случае с ним можно поступить следующим образом:

- удалить (если «выброс» пришелся на первые наблюдения в имеющихся данных, сделать это можно безболезненно. Однако если он произошел ближе к концу, то его удаление может повлечь за собой сложности при прогнозировании последних сложившихся тенденций и связей в ряде данных);
- интерполировать[1] (в этом случае значение выброса заменяется на некоторое усредненное либо значение соседнего наблюдения. Такая замена решает проблему отсутствия наблюдения, однако интерполяция может лишить прогнозиста важной информации относительно произошедшего наблюдения).

В случае, если прогнозист смог экономически обосновать существование «выброса», он может все так же на свой страх и риск удалить, интерполировать «выброс» либо включить его в модель с помощью фиктивных переменных. В ряде случаев может быть предпочтителен именно третий вариант, так как тогда «выброс» становится частью модели, а не инородным объектом. Подробней о фиктивных переменных речь пойдет в параграфе 4.6.

Источниками **систематической ошибки** могут являться как *инструмент сбора и обработки информации*, так и *чело-*

---

[1] Один из вариантов интерполяции — интерполяция с помощью формулы Ньютона, к которой мы обратимся позже.

веческий фактор (желание приукрасить ситуацию или скрыть часть неблагоприятной информации). К сожалению, при работе с вторичными данными социально-экономической динамики (официальной и неофициальной статистикой) часто приходится иметь дело с ошибками, вызванными влиянием человеческого фактора. Особенно это касается данных отечественной статистики и статистики стран — республик бывшего СССР. Дело в том, что бо́льшая часть экономических показателей отражает эффективность деятельности того или иного подразделения, той или иной системы, того или иного региона. Классическим примером ошибки информации такого рода являются записываемые год от года в статистические сборники данные о количественных показателях экономического развития бывшего СССР, которые отражали не столько реальные процессы, сколько желаемые результаты. Известно, например, что в Узбекской ССР долгие годы шли приписки о сборе невыращенного хлопка, которые попадали в статистические сборники. По отчетам о выполнении приписанных плановых заданий по сбору хлопка составлялись планы работы текстильной промышленности, которая, соответственно, из несобранного хлопка не могла выпустить несуществующую ткань. Но в статистические отчетности часть этой невыпущенной ткани попадала, а другая списывалась на разного рода потери и брак. Эта цепочка пронизывала весомую часть статистической отчетности, в результате чего ошибочно определялись и производные экономические инструменты — расход хлопка на единицу ткани, нормы электропотребления, производительность труда и т.п. Не секрет, что в экономике бывшего СССР осуществлялись и другие приписки. Таким образом, практически все обобщающие данные экономического развития (валовой продукт, национальный доход и т.п.) отдельных регионов, да и страны в целом оказывались год от года засоренными ошибками такого рода.

Сегодня причиной возникновения подобной ошибки может быть, например, желание уменьшить выплаты по платежам в бюджеты и внебюджетные фонды, искажение данных в ходе «информационной войны» с конкурентами и т.п. Многие предприятия в результате этого представляют в статистические органы «засоренную» информацию, и выявить в ней ошибки такого рода практически невозможно.

Есть и другой источник подобной информации. Например, если прогнозист в качестве одного из показателей собирается использовать величину стоимости основных производственных фондов России за некоторый промежуток времени, то он столкнется с рассматриваемой ошибкой. Например, в табл. 2.9 приведены данные по динамике основных фондов и инвестиций по Псковской области[1]. Из курса «Экономика предприятия» студент знает, что полная стоимость основных производственных фондов в текущем году определяется как остаточная стоимость фондов на конец предыдущего периода плюс инвестиции данного года в основные фонды минус выбытие фондов за год.

Пример

Данные табл. 2.9 свидетельствуют о том, что приведенные значения по основным фондам и инвестициям не вписываются в эту логику. Остаточная стоимость основных фондов, например, в 2006 г. (третий столбец), равная 81 641 млн руб., никак не коррелирует с полной стоимостью основных фондов 2006 г. – 164 095 млн руб. Инвестиции этого года (7603,3 млн руб.) еще больше запутывают суть того, что скрывается за этими цифрами, поскольку если к величине остаточной стоимости прибавить величину инвестиций, то будет получено такое число: 81 641 млн руб. + 7603,3 млн руб. = 89 244,3 млн руб.

Сравнив это полученное с помощью вычислений число с тем, что приводится во втором столбце, можно убедиться, что числа отличаются друг от друга почти в два раза! Объяснить этот двукратный рост с помощью других вычислений невозможно. На самом деле разбалансировка значений объясняется тем, что фактические значения основных фондов, которые собираются на местах, передаются в Росстат (Федеральную службу государственной статистики), где и осуществляется переоценка стоимости фондов по методике, разработанной для этого случая Росстатом.

Таблица 2.9

Некоторые данные по экономике Псковской области

Год	Полная стоимость основных фондов, млн руб.	Остаточная стоимость основных фондов в предыдущем периоде, млн руб.	Инвестиции в основной капитал в отчетном периоде, млн руб.
2001	90 537	49 234	2791
2002	111 128	52 604	3135
2003	122 694	60 463	5613,6

[1] Данные представлены в 2009 г. студентами О. И. Коровка и Г. К. Бомейко.

Окончание табл. 2.9

Год	Полная стоимость основных фондов, млн руб.	Остаточная стоимость основных фондов в предыдущем периоде, млн руб.	Инвестиции в основной капитал в отчетном периоде, млн руб.
2004	129 423	67 313	5904,9
2005	144 880	69 984	5546,9
2006	164 095	73 973	7603,3
2007	188 943	81 641	11 831,2

Эта ситуация с основными фондами, очевидно, характерна для всех областей России и для некоторых республик бывшего СССР, где также производится переоценка основных фондов по собственным методикам. Поэтому прогнозисту в подобных случаях необходимо проводить дополнительные исследования, чтобы понять, как определить ту величину стоимости основных фондов, которую он собирается применить в своих расчетах. Без такой дополнительной работы использовать показатели табл. 2.9 для построения прогнозных моделей нельзя. Понятно, что подобные же перерасчеты, агрегирование и трансформация органами Росстата относятся и к другим показателям социально-экономического прогнозирования.

Систематические ошибки, вызванные применением неисправного инструмента измерения, могут быть выявлены и исключены, так как имеют примерно одну и ту же величину, один и тот же знак. Поэтому исходные данные, содержащие этот тип ошибки, всегда несколько завышены или занижены. Объективным источником этой ошибки служат, в основном, измерительные устройства, приемы или приборы, вносящие одну и ту же погрешность при измерениях.

Случайные ошибки неизбежны. Причины их появления многообразны и связаны с действием множества случайных неконтролируемых факторов, а поэтому не поддаются анализу. В результате этого практически любое измерение содержит случайные ошибки, но так как источников их возникновения достаточно много, они, как правило, обладают следующими свойствами[1].

1. Для ряда результатов наблюдений с известным параметром распределения абсолютные величины случайных ошибок с заданной вероятностью P не превосходят определенного предела. Это значит, что влияние случайных ошибок на результат все-таки незначительно.

---

[1] *Большаков В. Д.* Теория ошибок наблюдений. М. : Недра, 1983.

2. Положительные и отрицательные случайные ошибки равновозможны, т.е. одинаково часто встречаются при наблюдениях. Из этого вытекает и следующее свойство.

3. Математическое ожидание случайной ошибки равно нулю.

4. Малые по абсолютной величине случайные ошибки встречаются при наблюдениях чаще, чем большие.

Следовательно, в большинстве случаев можно предполагать, что случайные ошибки подчиняются закону нормального распределения вероятностей и их математическое ожидание равно нулю. С учетом перечисленных свойств, создается ситуация, когда проявляются условия действия закона больших чисел, в соответствии с которым «совокупное действие большого числа случайных факторов приводит при некоторых, достаточно широких условиях к результату, почти не зависящему от случая»[1]. Таким образом, избежать влияния случайных ошибок можно, если увеличить объем выборки.

После того, как будут устранены или учтены ошибки информации, перед исследователем встает задача ее систематизации и обработки. В достаточно редких случаях необходимая исследователю информация представлена в систематизированном виде и в виде, пригодном для последующего анализа и обработки. Чаще всего информация представляет собой некоторую неупорядоченную совокупность, нуждающуюся в упорядочении и систематизации.

Систематизация информации заключается в ее представлении в виде таблиц, графиков, диаграмм и в других формах, удобных для исследователя и показывающих некоторые наиболее очевидные закономерности. В большинстве случаев исследователи предпочитают сведение информации в статистические таблицы (при этом возможен их последующий формализованный анализ с помощью математических методов).

Для того чтобы неупорядоченную совокупность данных можно было свести в таблицу, необходимо определить признак упорядочивания данных. Такими признаками могут являться:

- номер наблюдения;
- время наблюдения;

---

[1] *Дунин-Барковский И. В., Смирнов Н. В.* Теория вероятностей и математическая статистика в технике. М. : Гос. изд-во технико-теоретической литературы, 1955. С. 104.

• показатель, по мере увеличения (уменьшения) которого можно упорядочить данные;

• номера экспертов, дававших оценку объекту исследования;

• ранги, полученные для свойств товара по шкале отношений или интервалов;

• товарный ряд и т.п.

Зачастую на практике встречаются случаи, когда сведенная в таблицы информация оказывается неполной — часть данных отсутствует (например, когда при опросе один из респондентов ответил не на все вопросы или в статистическом сборнике не приводятся данные за какой-либо год). В результате возникает необходимость восстановления утерянной в процессе сбора и обработки наблюдений информации. Определить неизвестную величину внутри статистического ряда можно с помощью одного из методов интерполяции.

Теория интерполяции — один из старейших разделов математики, она начиналась работами И. Ньютона, Ж. Лагранжа, Н. Абеля, Ш. Эрмита и др. Интерполирование — это способ нахождения какой-либо величины по известным отдельным значениям этой же или других величин, связанных с ней[1].

Теория интерполяции является одним из наиболее разработанных разделов численных методов, и поэтому поставленная задача может быть решена с той или иной степенью точности.

Проще всего воспользоваться методом разностей (хотя это и не самый точный метод экстраполяции), суть которого заключается в следующем.

Первая производная функции, как известно, находится по формуле

$$y_t' = \lim_{x \to 0} \frac{\Delta y_t}{\Delta x_t}. \qquad (2.4)$$

Эта производная остается постоянной и не равной нулю, если между двумя переменными существует линейная функциональная зависимость. Если между двумя переменными существует линейная регрессионная зависимость, то в каждой конкретной точке наблюдения за переменными в момент t

[1] *Янович Л. А.* Интерполирование // Математическая энциклопедия. Т. 2. М. : Советская энциклопедия, 1979. С. 622.

первая производная будет иметь значения, в общем случае отличающиеся от значений первой производной в другие моменты времени. Эти отклонения вызваны действием множества случайных факторов и поэтому значения первой производной в разных точках будут колебаться вокруг своего математического ожидания, лучшей оценкой которой в данном случае является средняя арифметическая.

Как вычислить первую производную регрессионной зависимости, если не известны коэффициенты линейной функции, которая описывает эту зависимость? У исследователя имеются в распоряжении только эмпирические значения x_t и y_t. Для решения поставленной задачи первую производную заменяют отношением конечных разностей:

$$y'_t \approx \frac{\Delta y_t}{\Delta x_t}. \tag{2.5}$$

Такая замена возможна только в том случае, когда приращения Δx_t (конечные разности первого порядка) достаточно малы. Обычно с такой ситуацией и приходится иметь дело на практике. Поэтому легко найти первые разности Δx_t и Δy_t:

$$\Delta y_t = y_t - y_{t-1},\ \Delta x_t = x_t - x_{t-1}, \tag{2.6}$$

а затем — их отношение:

$$y'_t \approx \frac{\Delta y_t}{\Delta x_t} = \frac{y_t - y_{t-1}}{x_t - x_{t-1}}. \tag{2.7}$$

Если это отношение при разных значениях t действительно колеблется около некоторого значения, можно найти одну из оценок этого значения, а именно, среднюю арифметическую:

$$\bar{\Delta} = \frac{1}{T-1}\sum_{t=1}^{T}\frac{\Delta y_t}{\Delta x_t}, \tag{2.8}$$

которая будет в первом приближении характеризовать первую производную зависимости, а значит, будет являться одной из оценок коэффициента линейной регрессии y на x. С помощью этого коэффициента можно решать задачи интерполяции подобных линейных зависимостей. Однако точность такой интерполяции не очень высока, к тому же в практике прогнозирования социально-экономических процессов линейные зависимости встречаются крайне редко. Поэтому метод вычисления конечных разностей для целей интерполяции в настоящее время применяется в простых случаях, но именно конечные разности легли в основу интерполяции методом полинома Ньютона.

Интерполяционная формула Ньютона применяется в том случае, когда упорядоченные значения x_t находятся на равном расстоянии друг от друга, т.е., когда $\Delta x_t = x_{t+1} - x_t = h =$ $= \text{const}$ для всех t. Константа h получила название шага наблюдений (шаг таблицы наблюдений). С учетом этого свойства значения функции двух переменных x_t и y_t характеризуются только изменением переменной y_t. Эти изменения можно определить, вычислив значения конечных разностей. Сами разности можно осуществлять с шагом назад, как это было сделано в (2.6), а можно — и с шагом вперед, как это предусмотрено методом Ньютона. При этом формула для расчета первых разностей будет иметь вид $\Delta y_t = y_{t+1} - y_t$, вторых разностей — $\Delta^2 y_t = \Delta y_{t+1} - \Delta y_t$, третьих — $\Delta^3 y_t = \Delta^2 y_{t+1} - \Delta^2 y_t$, и так далее для других конечных разностей (табл. 2.10).

Таблица 2.10

Конечные разности различных порядков

x_t	y_t	Δy_t	$\Delta^2 y_t$	$\Delta^3 y_t$	$\Delta^4 y_t$	$\Delta^5 y_t$
x_1	y_1	Δy_1	$\Delta^2 y_1$	$\Delta^3 y_1$	$\Delta^4 y_1$	$\Delta^5 y_1$
x_2	y_2	Δy_2	$\Delta^2 y_2$	$\Delta^3 y_2$	$\Delta^4 y_2$	
x_3	y_3	Δy_3	$\Delta^2 y_3$	$\Delta^3 y_3$		
x_4	y_4	Δy_4	$\Delta^2 y_4$			
x_5	y_5	Δy_5				
x_6	y_6					

Легко убедиться в том, что если наблюдений будет не шесть, как это показано в таблице 2.10, а, например, восемь, то можно будет вычислить еще и разности шестого и седьмого порядков; если будет T наблюдений, то можно вычислить разности $(T - 1)$-го порядка. Разность каждого порядка некоторым образом характеризует производную степени, соответствующей данному порядку.

Так как значения x_t в рассматриваемом случае представляют собой арифметическую прогрессию, то, введя обозначение

$$q = \frac{x - x_1}{h}, \tag{2.9}$$

получим интерполяционную формулу Ньютона на основе вычисленных значений конечных разностей:

$$P_T(x) = y_1 + q\Delta y_1 + \frac{q(q-1)}{2!}\Delta^2 y_1 + \ldots + \frac{q(q-1)\ldots(q-T+1)}{T!}\Delta^T y_1. \quad (2.10)$$

Подставляя в (2.9) известное значение x_t, легко найти q, используя которое в (2.10), вычисляется интерполируемое значение y_t.

Формула (2.10) называется интерполяционной формулой Ньютона для интерполирования вперед[1]. Существует также формула Ньютона для интерполирования назад, которая использует разности, вычисленные по принципу (2.6).

В случаях, когда необходимо получить более точные результаты интерполяции, рекомендуется использовать и более сложные нелинейные интерполяционные формулы, в первую очередь, интерполяционные формулы Лагранжа. Методика интерполяций этим методом исходит из необходимости построения интерполирующей функции, проходящей через все точки (x_t, y_t). Подобной функцией является многочлен (T – 1)-й степени, который, очевидно, пройдет через все T точек:

$$L(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_{T-1}x^{T-1}. \quad (2.11)$$

Однако достаточно часто построение подобных функций оказывается излишним, поскольку с подобной задачей могут успешно справиться и функции с более низкими степенями. Именно эту задачу решает метод вычисления интерполяционного многочлена (полинома) Лагранжа, который рассчитывается по формуле[2]

$$L_T(x) = \sum_{t=1}^{T} \frac{(x-x_1)(x-x_2)\ldots(x-x_{t-1})(x-x_{t+1})\ldots(x-x_T)}{(x_t-x_1)(x_t-x_2)\ldots(x_t-x_{t-1})(x_t-x_{t+1})\ldots(x_t-x_T)} y_t. \quad (2.12)$$

Здесь x_t и y_t — фактические значения наблюдаемых показателей.

---

[1] *Самарин М. К.* Ньютона интерполяционная формула // Математическая энциклопедия. Т. 3. М. : Советская энциклопедия, 1982. С. 1092.

[2] *Кудрявцев Л. Д., Самарин М. К.* Лагранжа интерполяционная формула // Математическая энциклопедия. Т. 3. М. : Советская энциклопедия, 1982. С. 170—171.

Этот интерполяционный многочлен вычисляется для имеющихся пар значений и дает исследователю формулу зависимости интерполируемого значения $y(x) = L_T(x)$ от любого значения x. Определив вид этой функции по известному значению x_t $(t < T)$, интерполируют значение y_t, подставляя значение x_t в формулу (2.12).

Каждая из интерполяционных формул (2.8) и (2.10) дает при вычислении ошибки, которые при необходимости можно рассчитать и учесть в вычислениях[1]. Однако величина этих ошибок достаточно мала, поэтому их влиянием в практике прогнозирования пренебрегают, тем более что эмпирические данные загрязнены другими ошибками, которые мы рассматривали ранее.

К числу методов интерполирования относят также интерполирование методами математической статистики, чаще всего — с помощью МНК. Подробней данный метод будет рассмотрен нами в параграфе 3.5. С учетом того, что сегодня в распоряжении прогнозиста имеются разнообразные пакеты прикладных программ, реализованных применительно к персональным компьютерам, в которых вычисление оценок МНК представляет собой элементарную процедуру, этот способ используют довольно часто.

После того, как информация об объекте прогнозирования собрана, по возможности очищена от ошибок, систематизирована и, при необходимости, восстановлена, прогнозист становится обладателем данных, которые можно использовать для последующих вычислений и построения прогнозных моделей.

2.4. Математическая обработка и прогнозирование информации, измеренной в неметрических шкалах

Измерять информацию необходимо для того, чтобы ее можно было сравнивать, оценивать, выявлять закономерности и т.п. — без измерения невозможно познание. В то же время необходимо отметить, что указанные приемы не всегда непосредственно применимы к собранной и измеренной информации. Например, сравнивать информацию, собранную и измеренную в номинальной шкале, не имеет смысла и нет

[1] *Хемминг Р. В.* Численные методы для научных работников и инженеров. М. : Наука, 1968. С. 110–112.

такой возможности, поскольку номинальная шкала позволяет только отнести измеренный объект к тому или иному классу. Название, наименование класса не может быть предметом статистической обработки и уж тем более — сравнения.

Каждой шкале измерения присущ свой оригинальный набор методов обработки информации. Применение инструментария, рассчитанного на другой тип шкалы, в большинстве случаев приводит к ошибочной интерпретации результатов. При этом следует отметить, что методы обработки информации, измеренной в номинальной шкале, могут быть использованы для обработки информации, измеренной в шкале более высокого уровня, например, порядковой. Но методы, с помощью которых обрабатывается информация, измеренная в порядковой шкале, не всегда могут быть использованы для обработки информации, измеренной в шкале номинальной.

В простейшей **номинальной шкале** моделируются самые простые действия с информацией — отношения «равенства — неравенства». Эта шкала обладает только характеристикой описания — дается множество элементов, из которых следует указать один элемент, причем как результат идентификации, а не сравнения.

С информацией, измеренной по номинальной шкале, можно осуществлять только действия по отнесению объекта с измеренным признаком к тому или иному типу объектов. Это означает, что с помощью идентификации объектов при использовании номинальной шкалы можно осуществлять группировку объектов, и это имеет большое значение, ведь наиболее мощным инструментом современного научного анализа как раз и выступает метод типологизации — мысленное расчленение сложной совокупности объектов наблюдения на устойчивые типы, состоящие в свою очередь из классов, групп, родов и т.п. Чаще всего эту процедуру называют классификацией или группировкой, поскольку типологизация — высшая таксономическая категория, результат не только простой, пусть и многомерной группировки, но и аналитических исследований.

Классификацией называется распределение объектов по тому или иному существенному свойству, в результате чего каждый из них попадает в точно указанный класс, подмножество или группу[1].

---

[1] *Рузавин Г. И.* Логика и аргументация : учеб. пособие. М. : Культура и спорт ; ЮНИТИ, 1997. С. 64.

Одним из наиболее часто используемых методов классификации объектов, информация о которых приводится в номинальной шкале, является кластерный анализ. При этом структурирование группы осуществляется «снизу доверху»[1]: вначале предполагается, что каждый объект образует самостоятельный кластер, а затем производится слияние близко стоящих кластеров в более крупные кластеры (до получения кластера, объединяющего рассматриваемую группу). На каждом этапе слияния объединение происходит на основе нового признака классификации (например, вначале создаются кластеры по национальности респондентов, затем — по уровню полученного образования и т.п.).

Если обозначить измерение каждого объекта буквами алфавита (А, Б и т.д.), то символическая запись номинальной шкалы будет иметь вид

$$(А) \vee (Б) \vee (В) ... \vee (Я). \qquad (2.5)$$

Здесь знак $\vee$ означает дизъюнкцию, т.е. операцию «либо — либо».

Номинальная шкала позволяет выполнять следующие операции[2].

1. Нахождение частот распределения по пунктам шкалы с помощью процентирования или в натуральных единицах.

2. Поиск средней тенденции по модальной частоте. Модальной называют группу с наибольшей численностью.

3. Установление взаимосвязи между рядами свойств с помощью перекрестных таблиц.

Следует обратить внимание на то, что данные операции производятся не с самими числами, измеренными в этой шкале, а с количеством наблюдений объектов, обладающими теми или иными признаками и отнесенными в отдельные группы. Количество наблюдений или элементов, отнесенных к той или иной группе, — это уже информация, измеренная в метрической шкале, с которой можно работать всеми доступными и уместными математическими методами.

Для того чтобы определить частоту распределения каждой группы, вычисляется численность элементов каждой

---

[1] Типология и классификация в социологических исследованиях. М. : Наука, 1982. С. 96.

[2] *Ядов В. А.* Стратегия социологического исследования. Описание, объяснение, понимание социальной реальности. М. : Добросвет ; Книжный дом "Университет", 1998. С. 160–161.

из групп, а затем находится отношение этой численности к общему числу единиц признаков. Пусть, например, в группу с высшим образованием из числа респондентов в 500 человек были отнесены 133 человека. Тогда частота респондентов с высшим образованием в данной выборке составит 133 / 500 = 0,266. Эта величина дает исследователю дополнительную информацию (например, что число людей с высшим образованием составляет 26,6%). Частоту распределения удобно анализировать с помощью различного рода диаграмм — столбиковых, ленточных или круговых.

В статистике под «модой» понимают наиболее часто встречающееся значение. В номинальной шкале мода характеризует группу с наибольшей численностью. Вычисление модальной группы не вызывает затруднений, так как требует выполнения только операции сравнения типа «больше — меньше». Если в имеющейся выборке есть только одна модальная группа, то говорят об «унимодальной» группе. Значительно реже встречаются случаи, когда две и более группы включают одинаковое число единиц, которое для данной выборки является максимальным. В этом случае говорят о многомодальных случаях. Наличие нескольких мод говорит о своеобразии имеющегося распределения. Их анализ позволит получить дополнительную информацию.

Важным способом количественного анализа информации является установление взаимосвязи между рядами свойств с помощью таблиц сопряженности признаков. Здесь также используется информация о количестве наблюдаемых признаков. Эти количественные значения позволяют оценить возможность наличия взаимосвязи между наблюдаемыми признаками. Информацию о количестве наблюдаемых признаков сводят в таблицу сопряженности так, как это сделано в табл. 2.11.

Таблица 2.11

Общий вид таблицы сопряженности для анализа взаимосвязи между признаками

Значение признаков	0	1	Итог
0	a	b	$a + b$
1	c	d	$c + d$
Итог	$a + c$	$b + d$	$N = a + b + c + d$

Используя обозначения, приведенные в табл. 2.11, можно рассмотреть оценку взаимосвязи между наблюдаемыми признаками. Наиболее простым в использовании является коэффициент ассоциации Юла. С учетом введенных нами обозначений он имеет вид

$$Q = \frac{ad - bc}{ad + bc}. \tag{2.6}$$

Этот коэффициент изменяется в пределах от −1 до +1. Чем ближе коэффициент по модулю к единице, тем сильнее связь между измеренными признаками. Если он имеет отрицательное значение, то это свидетельствует о том, что увеличение значений одного признака приводит к уменьшению значений другого.

Пример

Любой студент российского вуза может обладать одним из признаков — мужским или женским полом. Можно выяснить еще одну характеристику — состоит или не состоит в браке. Здесь важно, что шкала наименований дает исчерпывающие состояния — ни в том, ни в другом случае нельзя предложить хотя бы один дополнительный класс, к которому может быть отнесен студент (при измерении свойства «принадлежность к полу» у студентов может быть только одно из двух состояний — принадлежность либо к мужскому, либо к женскому полу).

По признакам принадлежности к тому или иному полу, а также по отношению к браку можно выделить четыре сопряженные группы, условные данные о которых представлены в табл. 2.12.

Именно количество измеренных в номинальной шкале данных и подвергается статистической обработке. Это количество, очевидно, измеряется в метрической шкале.

Таблица 2.12

Условный пример таблицы сопряженности

Значение признаков	Мужской пол	Женский пол	Итог
Состоит в браке	34	66	100
Не состоит в браке	88	62	150
Итог	122	128	250

Для нашего условного примера коэффициент Юла будет равен $Q = \frac{34 \cdot 62 - 66 \cdot 88}{34 \cdot 62 + 66 \cdot 88} = -0,467.$

Это говорит о слабости возможной зависимости между признаками. Знак «минус» можно проинтерпретировать так: при переходе от признака «мужской пол» к «женскому полу» не следует ожидать перехода у объекта свойства «состоит в браке» к свойству «не состоит в браке», а наоборот, следует ожидать появления противоположного признака «состоит в браке».

Поскольку взаимосвязь между признаками все же имеется, можно получить дополнительную информацию о наблюдаемом явлении для получения элементарных прогнозов, например, для прогнозирования того, состоит ли студент в браке или нет.

Для этого преобразуем табл. 2.12 к значениям в процентах от общего числа наблюдаемых (табл. 2.13).

Таблица 2.13

Таблица сопряженности в процентах

Значение признаков	Мужской пол	Женский пол	Итог
Состоит в браке	14	26	40
Не состоит в браке	35	25	60
Итог	49	51	100

По данным этой таблицы можно утверждать следующее.

1. Если перед нами студент, то в 49 случаях из 100 — это мужчина, а в 51 случае — женщина.

2. Зная, что перед нами студент, у объекта с этой характеристикой мы можем ожидать, что в 40 случаях из 100 он будет состоять в браке, а в 60 случаях — нет.

3. Если перед нами студент мужского пола, то в 14 / 49 = 28 случаях из 100 он будет холост.

4. Если перед нами студентка, то в 26 / 51 = 52 случаях из 100 она окажется замужем.

Таким образом, измерение информации в номинальной шкале оказывается полезным для прогнозирования социально-экономических процессов, поскольку мы можем по количеству наблюдаемых признаков давать количественные характеристики взаимосвязи между показателями и получать некоторые прогнозные оценки.

Недостаток коэффициента ассоциации Юла заключается в том, что он не очень точно оценивает взаимосвязь между факторами. Если, например, хотя бы в одной клетке таблицы сопряженности будет иметься нуль, то коэффициент ассоциации Юла будет равен единице, но это не говорит об однозначной зависимости между признаками.

От этого недостатка свободен коэффициент хи-квадрат Пирсона:

$$\chi^2 = \frac{(ad-bc)^2}{(a+c)(b+d)(a+b)(c+d)}. \tag{2.7}$$

Поскольку он лежит в пределах от нуля до единицы, трудно дать интерпретацию его значениям, к тому же он не указывает на направление взаимосвязи между признаками. Поэтому значительно чаще на практике вместо него используют корень квадратный из этой величины, который получил название «коэффициент сопряженности Пирсона»:

$$\chi = \frac{ad-bc}{\sqrt{(a+c)(b+d)(a+b)(c+d)}}. \tag{2.8}$$

Из анализа этих формул следует, что коэффициенты Пирсона могут принимать значения от -1 до $+1$. При этом значение -1 коэффициент будет иметь только в том случае, когда $a = d = 0$, а значение $+1$ — когда $b = c = 0$.

Пример

Вычислим коэффициенты хи-квадрат и сопряженности Пирсона применительно к условным данным табл. 2.12. Получим для коэффициента хи-квадрат:

$$\chi^2 = \frac{(34 \cdot 62 - 66 \cdot 88)^2}{122 \cdot 128 \cdot 100 \cdot 150} = 0{,}058.$$

Как видно, дать интерпретацию этим значениям сложно. А близость коэффициента к нулю, казалось бы, свидетельствует о том, что взаимосвязи нет, хотя из анализа таблицы некое представление о наличии взаимосвязи все-таки остается. Рассчитаем коэффициент сопряженности Пирсона:

$$\chi = \frac{34 \cdot 62 - 66 \cdot 88}{\sqrt{122 \cdot 128 \cdot 100 \cdot 150}} = -0{,}242.$$

Значение коэффициента невелико, ближе по модулю к нулю, чем к единице. Это свидетельствует о том, что связь между группами признаков есть, но довольно слабая. Знак «минус» говорит об обратной зависимости между признаками и подтверждает выводы, которые были получены с помощью коэффициента Юла.

Как видно из приведенных примеров, степень связи между признаками довольно сложно выявить однозначно. Как и в случае с корреляциями между признаками, измеренными в метрической шкале, логика использования этих коэффициентов должна быть такой: вначале следует аналитическими методами обосновать наличие взаимосвязи между признаками, а затем подтвердить эту гипотезу с помощью коэффициентов Юла или Пирсона. Следует отметить, что коэффициент сопряженности Пирсона всегда меньше коэффициента Юла. Обычно рекомендуется придерживаться следующего эмпирического правила: можно говорить о том, что взаимосвязь подтверждается, если коэффициент Юла по модулю больше 0,5, а коэффициент сопряженности Пирсона — больше 0,3.

На основе статистики хи-квадрат (2.7) были разработаны и другие коэффициенты, отражающие некую связь между двумя величинами (например, коэффициент сопряженности и коэффициент Крамера). Но от рассмотренного нами коэффициента (2.8) они отличаются несущественно и имеют схожий смысл, поэтому подробно останавливаться на них мы не будем.

Выше мы проанализировали ситуации изучения связи между двумя бинарными переменными, но может возникнуть закономерный вопрос: как быть в случае с категориальными переменными, принимающими больше чем два значения? В этом случае строить таблицу сопряженности по самим переменным не имеет смысла, так как в номинальных переменных нет упорядочения, а любая рассчитанная характеристика показывает связь между тем, как упорядочены элементы в одном признаке и элементы — в другом. Поэтому прежде чем строить таблицу сопряженности по категориальной переменной, эту переменную надо разбить на ряд бинарных.

Пример

Прогнозист салона «Мега-моторс» собрал данные о том, какие машины пользуются спросом. В этих данных есть два интересующих его признака.

1. Цвет машины:
- желтый;
- красный;
- черный;
- другой.

2 Тип кузова:
- седан;
- хэтчбэк;
- универсал.

По этим данным прогнозист составил следующую таблицу сопряженности (табл. 2.14).

Таблица 2.14

Таблица сопряженности в частотах по типам автомобилей

Значение признаков	Желтый	Красный	Черный	Другой	Итог
Седан	29	26	101	51	207
Хэтчбэк	15	20	35	150	220
Универсал	15	108	30	62	215
Итог	59	154	166	263	642

Из таблицы следует, что «другой» цвет — самая многочисленная группа. Можно также сделать вывод, что черный седан и красный универсал пользуются популярностью среди покупателей. Однако рассчитывать по этой таблице коэффициенты Юла и Пирсона нельзя: они будут показывать, как при переходе от черного цвета к «другому» будут меняться предпочтения от хэтчбэка в пользу универсала, что абсолютно бессмысленно. Поэтому каждую из этих переменных надо разбить на бинарные:

1. Желтая машина:
- 1 — если выбрана желтая машина;
- 0 — если не выбрана желтая.

2. Красная машина:
- 1 — если выбрана красная машина;
- 0 — если не выбрана красная.

3. Черная машина:
- 1 — если выбрана черная машина;
- 0 — если не выбрана черная.

4. Другой цвет:
- 1 — если выбрана машина другого цвета;
- 0 — если не выбрана машина другого цвета.

и

1. Седан:
- 1 — если выбрана машина с кузовом седан;
- 0 — если не выбрана машина с кузовом седан.

2. Хэтчбэк:
- 1 — если выбрана машина с кузовом хэтчбэк;
- 0 — если не выбрана машина с кузовом хэтчбэк.

3. Универсал:
- 1 — если выбрана машина с кузовом универсал;
- 0 — если не выбрана машина с кузовом универсал.

На основе этих новых переменных можно составить таблицы сопряженности и рассчитать коэффициенты Юла и Пирсона. Пример одной из таких таблиц (черных машин с кузовом седан) — табл. 2.15.

Таблица 2.15

Таблица сопряженности в частотах по типам автомобилей для черных седанов

Значение признаков	Черный	Не черный	Итог
Седан	101	106	207
Не седан	65	370	435
Итог	166	476	642

По данным этой таблицы коэффициент Юла составил 0,689, а коэффициент Пирсона — 0,361. Эти значения указывают на то, что между выбором черного цвета и седана есть некоторая связь. Интерпретироваться она может следующим образом: люди, покупающие машины с кузовом седан, отдают предпочтение черному цвету кузова.

Кроме того, на основе данных табл. 2.14 можно оценить, какие машины пользуются наибольшей популярностью у потребителей. Например, среди красных автомобилей наибольшей популярностью пользуется универсал, поэтому при решении о том, какие автомобили салону закупать в большем количестве, имеет смысл сделать акцент на автомобили этого типа.

Для номинальной шкалы измерения информации имеется возможность вычисления важной статистической характеристики — моды. Например, для рассмотренных выше условных примеров совокупности признака «мужской пол» модой является состояние «не состоит в браке», а для признака «женский пол» модой является состояние «состоит в браке». Для признака «состоит в браке» модой является состояние «женский пол», а для признака «не состоит в браке» — «мужской». Все эти моды при использовании принципов индукции могут служить основаниями для выполнения элементарных прогнозов.

Итак, операции с информацией, измеренной в номинальной шкале, позволяют исследователю выполнить некоторые прогнозы. Значительно больше знаний и выводов об объекте прогнозирования можно получить, если собрана и измерена информация в шкалах более высокого уровня. Разнообразнее становятся и инструменты прогнозирования.

Шкала порядка (ранговая шкала) наряду с описанием имеет еще и порядок, в результате чего возможно установление приоритетов и сравнений информации. Помимо операции типа «равенство — неравенство» с величинами, измеренными в шкале порядка, выполняются действия типа «больше — меньше». Кроме операции отнесения объекта по информации о нем к тому или иному классу, появляется возможность сравнения этой информации по каким-либо признакам на предмет того, какой признак оказывается больше другого. При этом шкала порядка имеет тем или иным образом сформированные ранги. Поэтому вместе с методами обработки данных, применимыми для шкалы наименований (номинальной шкалы), выполняются и другие действия, поскольку здесь есть некоторые количественные признаки. Например, рост двух количественных признаков уже может помочь более точно оценить наличие возможной взаимосвязи между признаками, чем это получается сделать с информацией, измеренной в номинальной шкале.

Работая с информацией, измеренной в этой шкале, необходимо помнить, что интервалы в ней не равны друг другу и, вообще говоря, интервалами не являются, а числа означают лишь порядок следования признаков.

С числами в шкале порядков можно выполнять следующие действия[1].

1. Числа могут быть монотонно преобразованы в другие числа.

2. Возможно суммирование оценок по ряду упорядоченных шкал.

3. Помимо моды *Mo* появляется возможность рассчитать медиану *Me*.

4. Взаимосвязь между признаками может быть определена с помощью коэффициентов ранговой корреляции.

Поясним суть каждого действия. Прежде всего, рассмотрим монотонное преобразование чисел порядковой шкалы. С учетом того, что числа в ранговой шкале означают лишь порядок следования признаков, то сам порядок можно обозначать любыми числами. Поэтому и становится возможным монотонное преобразование одних чисел другими с сохранением прежнего порядка. Пусть, например, объекты с измеряемым признаком получили следующие ранги:

1 2 3 4 5 6.

Эти числа могут быть изменены так, что отношения между рангами останутся неизменными, например:

–1,5 –1 –0,5 0 0,5 1,0.

Это свойство важно в тех случаях, когда возникает необходимость получения интегрированной оценки, выраженной

---

[1] *Ядов В. А.* Стратегия социологического исследования. Описание, объяснение, понимание социальной реальности. С. 169–170.

в одной шкале с постоянной величиной заданных интервалов, — появляется возможность суммирования оценок в таких шкалах.

Действительно, когда оцениваемый объект имеет не одно свойство, а ряд свойств, то его общая интегрированная оценка будет представлять собой сумму оценок этих свойств. Так, например, при оценивании преподавателя студентами могут учитываться такие свойства, как: «внешний вид», «знание материала», «умение доходчиво преподнести материал лекции» и т.п. Если все эти свойства оцениваются в шкале порядков в одной и той же шкале с одним и тем же постоянным шагом между рангами и одним и тем же диапазоном, то полученные для каждого преподавателя оценки по отдельным свойствам могут быть просуммированы, а полученная суммарная оценка будет характеризовать преподавателя в целом.

Пример

Пусть по итогам сессии студент получил следующие оценки:

- высшая математика — «неудовлетворительно» (2 балла);
- история России — «отлично» (5 баллов);
- психология — «отлично» (5 баллов);
- линейная алгебра — «отлично» (5 баллов).

Получается, что в среднем студент сдал сессию на 4,25 балла ((5 + + 5 + 5 + 2) / 4 = 4,25).

Число 4 в данном случае измерено в метрической шкале (она показывает число оценок), поэтому деление на него вполне возможно и имеет смысл. Однако в порядковой шкале оценок нет числа 4,25! Дать толкование полученному результату невозможно. Что же произошло в результате вышеприведенных действий и как интерпретировать полученный результат? Шкала так и осталась порядковой, но изменился ее диапазон (масштаб). Если раньше в шкале были только числа 2, 3, 4 и 5, то теперь появилось значительно большее количество упорядоченных чисел, а именно:

- число 2 = (2 + 2 + 2 + 2) / 4;
- число 2,25 = (2 + 2 + 2 + 3) / 4;
- число 2,5 = (2 + 2 + 3 + 3) / 4;
- число 2,75 = (2 + 3 + 3 + 3) / 4 и т.д. до числа 5.

Вместо четырех элементов порядковой шкалы получены 12 упорядоченных элементов. В этой шкале число 4,25 больше, чем число 4 и меньше, чем число 5. Поэтому показатель «средняя оценка за сессию» имеет простую интерпретацию. Как видно, приведенным выше действием мы просто расширили диапазон шкалы.

Пусть по итогам сессии один студент сдал ее со средним баллом 4,75, а другой — со средним баллом 3,25. Поскольку и для первого, и для второго студента используется одна и та же шкала с одним и тем же количеством элементов в ней, сравнение полученных результатов имеет смысл. Мы можем сказать, что первый студент учится лучше, чем второй. Но можно ли получить ответ на вопрос: во сколько раз лучше сдал сессию первый студент, чем второй? Казалось бы, надо просто разделить 4,75 на 3,25. В результате получим: 4,75/3,25 = 1,461538.

Что означает полученный результат? Ничего! Числа 1,461538 в той шкале, с которой мы работаем, просто не существует! Шкала начинается с числа 2. Любые попытки интерпретации полученного результата будут неверными. Поэтому на поставленный вопрос получить ответ невозможно — порядковая шкала ограничивает наши познавательные возможности.

Часто для объективной оценки уровня знаний преподаватели используют систему тестов — формализованную систему оценки ответов на закрытые вопросы. При этом при ответе на вопрос необходимо подчеркнуть один из предлагаемых вопросов или же ответить «да» или «нет». Допустим, что имеется 10 вопросов. В случае правильного ответа выставляется единица, в случае неправильного — ноль. Число правильных ответов подсчитывается, и на их основе выставляется соответствующая оценка:

- «отлично», если получено 9 или 10 баллов;
- «хорошо», если получено 7 или 8 баллов;
- «удовлетворительно», если получено от 4 до 6 баллов;
- «неудовлетворительно», если получено до 3 баллов.

В данном случае также используются порядковые шкалы. Сначала используется шкала оценки ответа на вопрос, которая предусматривает два числа, — 1 или 0. Число 0,24 здесь, очевидно, не существует.

Затем полученные оценки суммируются и получается та же порядковая шкала, но ее диапазон (масштаб) увеличился — от 0 до 10. Увеличилось и количество возможных чисел. Их, как легко подсчитать, стало уже 11.

После этого шкала вновь преобразуется в другую порядковую шкалу — шкалу принятых в вузе оценок: «отлично», «хорошо», «удовлетворительно», «неудовлетворительно» или 5, 4, 3 и 2.

Одна из характеристик, используемых для оценки порядковой шкалы — это «медиана». По сути, она делит ранжированный упорядоченный ряд значений пополам, т.е. 50% оценок находятся до этого значения, а 50% — после него.

Пример

Пусть 10 экспертов оценили вкус мороженого на 1 балл; 15 — на 2 балла; 20 — на 3 балла и 5 экспертов — на 4. Выполним расчет медианы для этого случая (табл. 2.16).

Всего к экспертизе было привлечено 50 экспертов (10 + 20 + 15 + 5). Тогда относительная частота ответов по той или иной оценке будет находиться отношением числа экспертов, давших эту оценку, к общему числу экспертов (например, относительная частота экспертов, давших оценку в один балл, будет равна 10 / 50 × 100% = 20%). Эти цифры занесены в третью строку таблицы.

Таблица 2.16

Расчет медианы по условному примеру

Оценка	1	2	3	4
Число экспертов, давших оценку	10	15	20	5
Относительная частота, %	20	30	40	10
Накопленная частота, %	20	50	90	100

Накопленная частота представляет собой сумму всех предыдущих относительных частот. Она характеризует процент экспертов, давших оценку не выше данной. Очевидно, что накопленная частота последней оценки будет равна 100%. Накопленная частота, равная 90%, которая приведена для оценки в 3 балла, говорит о том, что 90% экспертов оценили свойства объекта на 3 балла и ниже. Медиана в данном случае будет равна 2 баллам, так как именно для этой оценки характерна граница в 50% ответов. В случае если медиана не выпадает на границу (как с примером в табл. 2.16), медианой будет считаться то значение, в котором происходит переход за 50%. Мода, как легко увидеть, приходится на оценку, равную трем, так как относительная частота этого ответа является наибольшей. Мода и медиана могут совпадать либо отличаться друг от друга, как в этом примере.

Так как оценки каждого из показателей, измеренных в ранговой шкале, упорядочены, то можно анализировать, насколько эти порядки показателей одного и того же объекта исследования совпадают или отличаются друг от друга, т.е. можно говорить о том, являются ли эти факторы взаимосвязанными. Взаимосвязь, очевидно, проявляется так: если с ростом ранга одного признака его количества меняются аналогично количеству другого признака при росте его ранга, данные признаки являются взаимосвязанными. Эту взаимосвязь вычисляют с помощью различных методов ранговой корреляции, чаще всего используя коэффициенты ранговой корреляции Спирмена и Кендалла.

Если проранжированные показатели свойств x_t и y_t t-го объекта имеют один и тот же диапазон рангов от 1 до T, то коэффициент ранговой корреляции Спирмена будет вычисляться по формуле[1]

[1] *Гласс Дж., Стэнли Дж.* Статистические методы в педагогике и психологии. М. : Прогресс, 1976. С. 160.

$$r_s = 1 - \frac{6\sum_{t=1}^{T}(x_t - y_t)^2}{T(T^2-1)}. \tag{2.9}$$

Если этот коэффициент по модулю близок к единице, то говорят о сильной линейной взаимосвязи между этими показателями, если он близок по модулю к нулю, то это свидетельствует об отсутствии линейной взаимосвязи.

Для работы с так называемыми «связанными рангами» (одинаково важными рангами) этот коэффициент оказался непригодным. Так, например, если некоторое свойство товаров Г и Д эксперт оценивает как равное, он придает им одинаковую оценку, и в шкале порядков товары будут занимать следующие друг за другом ранги, имеющие одинаковые оценки. Эта ситуация и получила название «связанных рангов». В этом случае рекомендуется использовать коэффициент ранговой корреляции Кендалла, который будет иметь вид[1]:

$$\tau = \frac{4P}{T(T-1)} - 1. \tag{2.10}$$

Здесь P — число совпадений для пар рангов. Под «совпадением» понимается одинаковый порядок оценок пар рангов. Например, первым экспертом товару А присвоен ранг 1, а товару Б — ранг 3. Необходимо посмотреть, сколько еще экспертов дают товару А более высокий ранг, чем товару Б. Это и есть число совпадений.

Пример

Покажем, как информацию, измеренную в шкале порядков, можно использовать для выявления взаимосвязи между показателями с помощью коэффициента ранговой корреляции Спирмена (2.9). В табл. 2.17 приведен пример результатов ранжирования 10 товаров по нескольким показателям.

Таблица 2.17

Ранги десяти товаров

Показатель	Товар, i									
	1	2	3	4	5	6	7	8	9	10
Качество	2	1	4	7	3	8	5	9	6	10
Цена	8	2	1	9	6	10	3	4	5	7

[1] Там же. С. 164.

Окончание табл. 2.17

Показатель	Товар, i									
	1	2	3	4	5	6	7	8	9	10
Упаковка	3	1	4	6	8	9	5	2	7	10
Гарантийное обслуживание	5	1	4	9	3	10	2	7	6	8
Престижность	5	2	3	7	4	8	1	6	10	9

Посмотрим, насколько, по мнению экспертов, взаимосвязаны два показателя — качество товара и его престижность. Расчеты коэффициента приведены в табл. 2.18.

Таблица 2.18

Расчет коэффициента ранговой корреляции Спирмена

Показатель	Товар, i										Сумма
	1	2	3	4	5	6	7	8	9	10	
Качество, x_t	2	1	4	7	3	8	5	9	6	10	
Престижность, y_t	5	2	3	7	4	8	1	6	10	9	
$(x_t - y_t)$	−3	−1	1	0	−1	0	4	3	−4	1	
$(x_t - y_t)^2$	9	1	1	0	1	0	16	9	16	1	54

Теперь можно рассчитать соответствующий коэффициент ранговой корреляции:

$$r_s = 1 - \frac{6 \cdot 54}{10(100-1)} = 0{,}67. \tag{2.11}$$

Если этот коэффициент по модулю близок к единице, то говорят о сильной линейной взаимосвязи между показателями, если он близок по модулю к нулю, то это свидетельствует об отсутствии линейной взаимосвязи. В нашем случае коэффициент приближается к величине 0,7, что ближе к единице. Поэтому можно говорить о том, что взаимосвязь между рассматриваемыми факторами существует, хотя и не очень тесная.

Операции с числами, измеренными в **интервальных шкалах**, помимо всех перечисленных выше действий, включают в себя следующие новые[1]:

1) числа могут быть линейно преобразованы в другие числа;

---

[1] *Ядов В. А.* Стратегия социологического исследования. Описание, объяснение, понимание социальной реальности. С. 172—173.

2) для определения степени взаимосвязи используется более чувствительный коэффициент парной корреляции Пирсона.

Если с числами порядковой шкалы можно было осуществлять только монотонные преобразования, то линейные преобразования более сложны и имеют вид:

$$y_t = a_0 + a_1 x_t. \tag{2.12}$$

Это означает, что, получив значение x в шкале интервалов, исследователь может его преобразовать в новое число y с помощью соотношения (2.12). Как видно, к числам теперь можно не только прибавлять некоторую константу, но и умножать их на число соотношения между интервалами, их расстояния при этом не изменятся. Числа в интервальной шкале можно складывать или вычитать, находить средние арифметические или средние взвешенные. В шкале порядков число, равное 5,345, не имеет никакого смысла, так как между числами 5 и 6 в этой шкале нет и не может быть никакого расстояния, поэтому остаток от целого числа не дает никакой информации. Более того, такой остаток совершенно бессмыслен. В шкале интервалов это число показывает, насколько оно близко к соседним целым числам — 5 и 6.

С учетом того, что в шкале интервалов нулевое значение не означает полного отсутствия свойства, то отношения чисел друг к другу не имеет особого смысла, хотя эти операции и могут выполняться. Покажем это на простом примере.

Одна из самых известных интервальных шкал — шкала температур по Цельсию. Если сегодня температура воздуха в городе составила +1° С, а завтра — +5° С, то это не означает, что температура воздуха выросла в пять раз. В этом легко убедиться, если перейти к новой шкале измерения температуры — по Фаренгейту. Ранги температур в обеих шкалах остаются неизменными, а вот отношения между ними получаются разными. Если такую же операцию провести с числами метрической шкалы, то мы увидим принципиальное отличие — отношения чисел в метрической шкале имеют смысл. Для того чтобы убедиться в этом, разделите среднемесячную заработную плату по России в настоящее время на ежемесячную стипендию студента российского вуза в то же время. А теперь переведите измеренные доходы в доллары, немецкие марки или японские йены и вновь найдите отношения среднемесячной зарплаты по России к стипендии российского студента. Вы убедитесь в том, что это отношение осталось неизменным.

Этот пример показывает, что не все математические операции с числами в шкале интервалов можно осуществлять так, как нам этого хотелось бы. Линейные преобразования имеют смысл, нелинейные — бессмысленны.

В то же время, наличие расстояния между числами значительно обогащает возможности вычисления степени взаимосвязи между показателями, измеренными в шкале интервалов. Конечно, в этом случае можно воспользоваться и всеми предыдущими способами вычисления степени взаимосвязи, но более точно это можно сделать с помощью коэффициента корреляции Пирсона, который будет нами изучен в параграфе 3.6.

Числа, полученные с помощью **метрической шкалы**, могут быть обработаны с помощью всего арсенала методов математической статистики, включая методы, применимые для шкал более низкого уровня.

2.5. Оценка адекватности прогнозных моделей

Когда статистические данные собраны, обработаны и являются пригодными для использования в прогнозировании, возникает задача построения прогнозной модели. В данном параграфе мы рассмотрим только одну часть этого процесса, которая является общей для прогнозирования всех обратимых и необратимых процессов — оценку адекватности модели.

При построении математических прогнозных моделей исследователю нужно каким-то образом выбрать модель, которая позволит получить наиболее точный прогноз. Однако оценить точность прогноза до получения фактических данных оказывается либо крайне сложным, либо практически невозможным процессом. Именно поэтому были придуманы различные процедуры и инструменты, позволяющие получить косвенное суждение о прогнозной точности модели.

Основная идея всех применяемых на практике процедур отражается в следующей индуктивной гипотезе: если модель хорошо описывает прошлое, значит, есть основания предполагать, что она будет хорошо описывать будущее.

Рекомендуемый подход к прогнозированию заключается в построении нескольких моделей, из которых исследователь на основе статистических данных и экспертного суждения выбрал бы наилучшую[1]. Рассмотрим инструменты и показатели, которые позволяют исследователю принять такое решение.

2.5.1. Графический анализ модели

Одним из базовых методов оценки адекватности построенной модели является простой линейный график, на котором показаны фактические значения y_t, расчетные (полученные по модели) $\hat{y}_t$ и иногда ошибки модели ($e_t = y_t - \hat{y}_t$), все упорядоченные во времени. На такой график бывает полезно добавить и прогнозные значения — они позволяют понять, насколько модель дала прогноз, вписывающийся в общую динамику. Кроме того, по оси абсцисс имеет смысл откладывать не просто номера наблюдений, а даты, на которые имеются те или иные наблюдения — это позволяет понять, какие события описаны моделью, а какие — нет, что дает дальнейшую пищу для анализа (рис. 2.6).

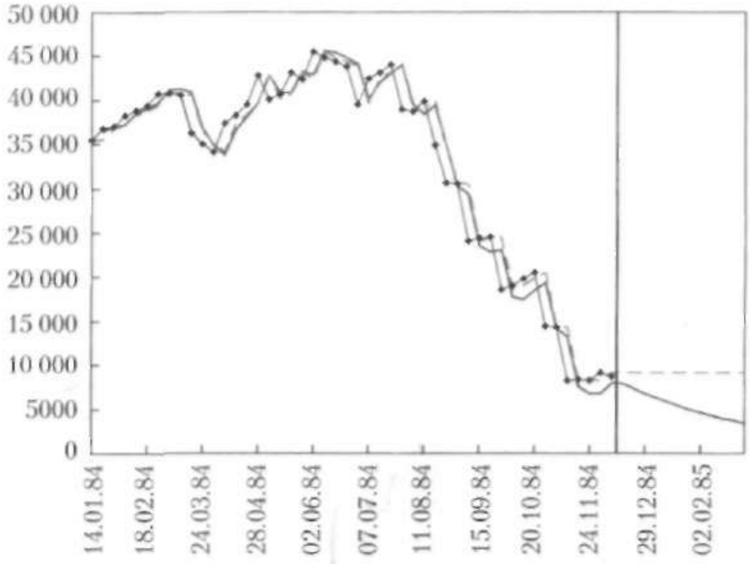

Рис. 2.6. **Графический анализ модели:**

сплошная линия с точками — условный ряд данных; *сплошная линия до вертикальной* (*модель № 1*), *пунктирная* (*модель № 2*) — расчетные значения; *сплошная линия после вертикальной* (*модель № 1*), *пунктирная* (*модель № 2*) — прогноз по некоторым моделям

---

[1] *Armstrong J. Scott, Fildes Robert.* Making progress in forecasting // International Journal of Forecasting. 2006. Vol. 22. P. 433–441.

По рисунку видно, что построенная модель № 1 как будто представляет собой исходный ряд данных, сдвинутый на 1 наблюдение вперед. Создается ощущение, будто модель «запаздывает», однако заметно, что на участках, на которых существенных изменений не происходит (например, первые восемь наблюдений), модель достаточно точно аппроксимирует ряд данных. В целом такая аппроксимация может быть признана вполне приемлемой. Главный вывод, который можно сделать из данного графического анализа, заключается в том, что модель не имеет систематических завышений либо занижений, явных огрехов в аппроксимации не наблюдается. Прогнозируемая моделью тенденция может быть признана допустимой хотя бы потому, что в целом, начиная с 02.06.1984, ряд данных демонстрирует тенденцию к снижению. Конечно, мы ни в коем случае не можем делать каких-то выводов о динамике показателя только на основе графического анализа — мы лишь диагностируем общую описательную способность модели и констатируем адекватность данного ею прогноза с технической точки зрения.

Модель № 2 во многом напоминает первую, единственное существенное различие заключается в самом прогнозе: в соответствии с этой моделью значения в будущем не будут сильно изменяться, а, скорее всего, будут располагаться примерно на одном уровне.

Остатки идеальной модели не должны демонстрировать тенденции к росту либо снижению. Наличие таких тенденций говорит о том, что выбранная модель плохо аппроксимирует ряд данных. Кроме того, дисперсия остатков должна быть постоянной, как в начале ряда, так и в конце: колебания относительно нуля должны быть примерно одинаковыми на всем промежутке. В случае нарушения этого правила можно диагностировать наличие гетероскедастичности — свойства, при котором с ростом какого-то фактора растет дисперсия остатков. Если остатки носят в себе перечисленные недостатки, имеет смысл обратиться к другой модели, которая даст остатки без них.

Графики остатков по первой и по второй моделям изображены на рис. 2.7.

Как видим, в остатках первой модели сильных нарушений не наблюдается. Можно лишь обратить внимание на то, что отрицательные значения к концу ряда имеют по модулю бóльшую величину, чем положительные, что сигнализирует о том, что модель не поспевает за фактическими значениями и не одинаково прогнозирует значения выше и ниже некоторого уровня. Дисперсия же остатков чисто внешне выглядит постоянной.

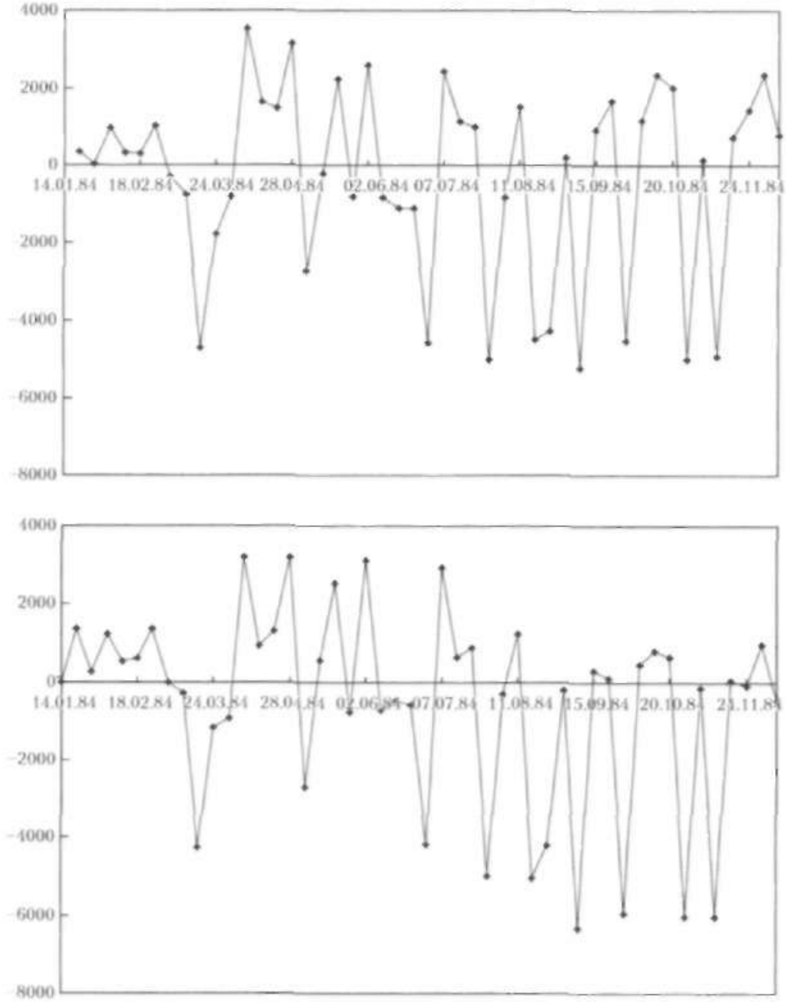

Рис. 2.7. **Остатки модели № 1 (сверху) и модели № 2 (снизу), построенных по условному ряду данных**

Однако во второй модели явно наблюдается какая-то тенденция в остатках к концу ряда: дисперсия увеличилась за счет сдвига остатков вниз, в отрицательную область. Это может сигнализировать о наличии гетероскедастичности в модели, которая возможна из-за неправильной спецификации.

Для дальнейшего графического изучения модели и ее ошибок можно рассмотреть гистограмму распределения ошибок, в которой по оси абсцисс откладываются значения ошибок, а по оси ординат — частота их появления (рис. 2.8).

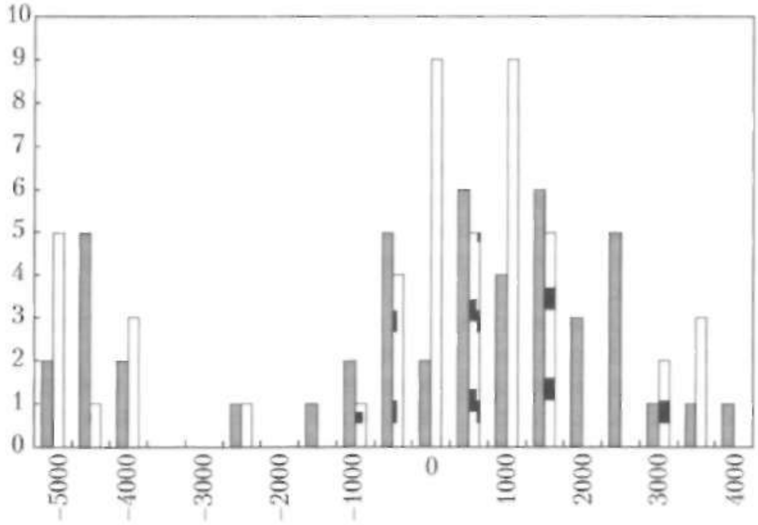

Рис. 2.8. **Гистограмма распределения остатков по модели № 1 (сплошные столбики) и модели № 2 (заштрихованные столбики)**

Считается, что в идеале остатки, полученные по модели, должны быть распределены в соответствии с нормальным законом распределения случайных величин, с нулевым математическим ожиданием — это указывает на то, что модель описала влияние всех остальных факторов и лишь мелкие неучтенные, влияние каждого из которых ничтожно, остались не описанными. Соответственно, в идеальной модели распределение остатков должно напоминать колокол. Это требование крайне желательно, но не обязательно при прогнозировании. На него опираются вся дальнейшая проверка статистических гипотез, а также точность построенных доверительных интервалов, однако и без него можно дать достаточно точный прогноз. Тем не менее, в случае, если это правило не соблюдается, имеет смысл рассмотреть альтернативные модели.

Кроме того, по распределению остатков иногда можно диагностировать наличие различных подвыборок в исследуемом ряде данных, что может указывать на то, что в исходных данных не были учтены какие-то важные факторы (например, сезонность в ряде данных).

Кроме того, по данному графику можно обнаружить «выбросы» — наблюдения, сильно выбивающиеся из общей динамики.

По рисунку можно заметить, что остатки как в одной, так и в другой модели разделены на две группы: до –3000 и после –3000. Это указывает на то, что обе модели дают систематически заниженные расчетные значения по сравнению с имеющимися фактическими значениями. Кроме того, даже на глаз видно, что распределение остатков в обоих случаях не нормально.

Пока что графики на рис. 2.7 и 2.8 не дают сделать однозначного выбора в пользу первой или второй модели.

Следующий график, также дающий полезную информацию о распределении случайной величины — это ящичковая диаграмма (см. параграф 2.3).

На рис. 2.9 представлен пример ящичковой диаграммы, построенной по остаткам модели. В данном случае ящичковая диаграмма представлена в горизонтальном, а не вертикальном виде, но это не меняет всех ее свойств, рассмотренных ранее.

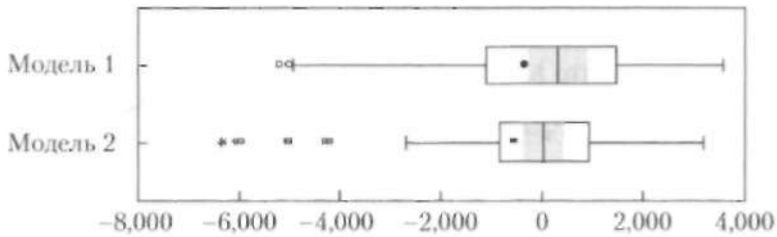

Рис. 2.9. **Ящичковая диаграмма по остаткам модели № 1 и модели № 2**

По данной диаграмме можно сделать два вывода.

Первый заключается в том, что в ряде данных ошибок наблюдается несколько «выбросов» (которые для модели № 1 сами по себе незначительны, так как располагаются близко к нижнему усу, а вот в модели № 2 — существенны). Это те самые значения отрицательных остатков, которые мы видели на гистограмме. На основе этого вывода можно попытаться провести дальнейший анализ этих значений: получились ли «выбросы» потому, что модель не смогла их объяснить, или же они являются результатами каких-то шоков, которые стандартными методами не выявляются. К таким шокам, например, могут относиться продажи продукции в магазине в одном месяце в результате закрытия магазина-конкурента на ремонт. Если «выбросы» относятся именно к таким непрогнозируемым ситуациям, их можно либо исключить, либо заменить интерполированными значениями, после чего перейти к дальнейшему изучению остатков модели. В нашем случае «выбросы» вызваны скорее неточной аппроксимацией ряда соответствующей моделью.

Второй вывод по диаграмме заключается в том, что распределение остатков как в первом, так и во втором случае, скорее всего, несимметрично: средняя величина находится ниже нижней границы медианного доверительного интервала.

2.5.2. Основные коэффициенты оценки качества модели

Очевидно, что, используя только графический анализ, адекватно оценить модель затруднительно. Именно поэтому были разработаны различные показатели, характеризующие точность аппроксимации и «качество» модели. Все эти показатели рассчитываются на основе остатков модели.

Базовым показателем, используемым в эконометрике, является *сумма квадратов отклонений фактических значений от расчетных* (т.е. сумма квадратов остатков):

$$RSS = \sum_{t=1}^{T} e_t^2. \tag{2.13}$$

Возведение в квадрат нужно для того, чтобы при сложении остатков положительные и отрицательные значения не перекрывали друг друга.

Для первой модели *RSS* оказался равным 295 281 628,8, что не дает нам практически никакой ценной информации. Мы просто видим, что сумма квадратов велика, но оценить эту величину не представляется возможным. Во второй модели *RSS* оказался выше: 322 067 656,0, что можно записать как очко в пользу первой модели.

Как видим, по величине *RSS* оценивать аппроксимацию модели либо крайне сложно, либо вообще невозможно. Поэтому вместо нее можно использовать *дисперсию остатков*, которая рассчитывается на основе *RSS*:

$$\sigma_e^2 = \frac{RSS}{T-k} = \frac{1}{T-k}\sum_{t=1}^{T} e_t^2, \qquad (2.14)$$

где T — число наблюдений; k — число коэффициентов модели.

Величина дисперсии остатков обычно используется как характеристика колеблемости остатков и нужна, скорее, для дальнейших расчетов доверительных интервалов, но сравнивать модели по дисперсиям остатков все так же затруднительно.

Вместо дисперсии остатков можно рассчитать более простую величину — *среднюю квадратическую ошибку*, которая отличается от (2.14) лишь знаменателем — деление в ней происходит не на число степеней свободы модели, а просто на число наблюдений:

$$MSE = \frac{1}{T}\sum_{t=1}^{T} e_t^2. \qquad (2.15)$$

По смыслу *MSE* аналогична дисперсии, но незначительно отличается от нее.

Для модели № 1 получились следующие значения: $\sigma_e^2 = 6\,710\,946{,}1$, $MSE = 6\,282\,587{,}8$. У модели № 2 — $\sigma_e^2 = 7\,001\,470{,}8$ и $MSE = 6\,852\,503{,}3$. Как видим, значения дисперсии и *MSE* отличаются друг от друга, хотя и достаточно близки по величине. Кроме того, соотношения, которые наблюдались между *RSS* первой и второй моделей, сохранились.

Заметим, как дисперсия, так и *MSE* сложно интерпретируемы: если, например, исходные величины измерялись в рублях, то остатки также будут измеряться в рублях, а вот дисперсия и *MSE* — уже в квадратных рублях, что, конечно же, не имеет смысла. Извлечение корня из (2.14) и (2.15) отчасти решает проблему: мы фактически получаем стандартные отклонения — величины, измеренные в рублях. Однако в интерпретации это совершенно не помогает, потому что величины стандартных отклонений не являются средними, а представляют собой некие величины, деленные на корень из числа наблюдений. Так, для первой модели стандартные отклонения составили соответственно 2561,6 и 2506,5, что позволяет сравнить модели, однако не дает нам практически никакой полезной информации о точности модели.

Именно по этим причинам иногда для выявления меры колеблемости используют *среднее относительное отклонение ошибок*[1]:

$$MAE = \frac{1}{T}\sum_{t=1}^{T}|e_t|. \tag{2.16}$$

Величина MAE уже измеряется в рублях и показывает, на какую величину остатки модели в среднем отклоняются от нуля.

Для первой модели MAE оказалось равным 1947,5, а для второй — 1805,2. Эти значения уже имеют смысл: в среднем остатки отклоняются от нуля на эту величину. По данному показателю вторая модель оказалась лучше.

Стоит отметить, что величины (2.13) — (2.16) позволяют сравнивать прогнозные модели между собой, но, как мы видим, не позволяют судить о точности аппроксимации ряда данных. Для получения некой интерпретируемой величины, позволяющей судить об этом, используют относительные ошибки аппроксимации, общая идея которых заключается в том, чтобы привести полученную величину MSE или MAE к относительным величинам. Достичь этого можно различными способами. Один из них заключается в делении MSE и MAE на среднюю величину по ряду. В связи с тем, что MSE измеряется в квадратах, для получения финальной величины из него нужно извлечь корень. Таким образом, выявляются *средние относительные ошибки аппроксимации*:

$$MAPE_1 = \frac{\sqrt{MSE}}{\bar{y}}; \tag{2.17}$$

$$MAPE_2 = \frac{MAE}{\bar{y}}, \tag{2.18}$$

где $\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$ — средняя величина по исходному ряду.

---

[1] *Makridakis Spyros G., Wheelwright Steven C., Hyndman Rob J.* Forecasting: Methods and Applications. Wiley, 1998. P. 43.

Альтернативой ошибок (2.17) и (2.18) представляется расчет на основе относительных ошибок: каждая из ошибок делится на фактическое значение ряда, после чего либо возводится в квадрат, либо берется по модулю и суммируется. Таким образом, получаются следующие относительные ошибки аппроксимации:

$$MAPE_3 = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{e_t}{y_t}\right|; \tag{2.19}$$

$$MAPE_4 = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\frac{e_t}{y_t}\right)^2}. \tag{2.20}$$

Все эти варианты средних относительных ошибок аппроксимации для наглядности можно измерять в процентах. Тогда ошибки будут показывать, на сколько процентов в среднем модель ошибается при аппроксимации ряда данных.

Для модели № 1 получились следующие средние относительные ошибки аппроксимации:

$MAPE_1 = 7{,}71\%$;

$MAPE_2 = 5{,}99\%$;

$MAPE_3 = 8{,}34\%$;

$MAPE_4 = 13{,}47\%$.

Для модели № 2 получены:

$MAPE_1 = 8{,}05\%$;

$MAPE_2 = 5{,}55\%$;

$MAPE_3 = 7{,}45\%$;

$MAPE_4 = 14{,}75\%$.

По каждому из полученных значений можно заключить, что соответствующая модель имеет представленные проценты ошибок. Правда, интерпретация ошибок (2.17) и (2.20) все так же затруднена из-за корня квадратного из числа наблюдений — эти ошибки так же на самом деле не являются средними.

Сравнивая модели по *MAPE*, видим, что в половине случаев ошибки оказались меньше в первой модели, чем во второй, а в другой половине — наоборот. Это и не удивительно: первая и четвертая ошибки рассчитаны на основе квадратов, в которых вторая модель проигрывала первой. Все из-за того, что квадраты ошибок сильнее «наказывают модель» — при наличии больших ошибок те значительно увеличиваются во время расчетов *RSS*.

Стоит отметить важный недостаток всех перечисленных ошибок аппроксимации: они очень чувствительны к масштабу. Так, если данные измеряются в миллионах единиц, то знаменатели (2.17)–(2.20) будут велики, в результате чего величины будут получаться маленькими даже при больших ошибках. Если же Y_t оказывается близким к нулю (например, прибыль на предприятии может быть как положительной, так и отрицательной), то ошибки значительно увеличиваются без каких бы то ни было оснований.

Именно поэтому, получив ошибку аппроксимации, нельзя сделать однозначный вывод о том, точная получена модель или нет. Однако сравнивать разные модели, построенные по одному и тому же ряду данных с помощью этих ошибок, значительно удобней, чем с помощью показателей (2.13)–(2.16). Кроме того, получение очень больших значений *MAPE* может послужить сигналом: если в ряде данных нет значений, близких к нулю, значит, модель неточно аппроксимировала ряд данных.

Еще одним недостатком *MAPE* является заложенное в них предположение о наличии в ряде данных естественной точки отсчета, т.е. они применимы лишь для данных, измеренных в абсолютных шкалах. Например, в случае измерения температуры в градусах по Цельсию или по Кельвину, по одной и той же модели с одной и той же объясняющей способностью, будут получаться совершенно разные средние относительные ошибки аппроксимации.

И, наконец, *MAPE* не ограничены никаким пределом, поэтому можно получить среднюю относительную ошибку аппроксимации, измеряемую в тысячах процентов, что сложно для восприятия.

Для того чтобы решить частично обозначенные проблемы, была предложена *симметричная средняя относительная ошибка аппроксимации*[1]:

$$sMAPE = \frac{2}{T}\sum_{t=1}^{T}\left|\frac{y_t - \hat{y}_t}{y_t + \hat{y}_t}\right|. \qquad (2.21)$$

Стоит пояснить появление в этой формуле числа 2 в числителе. Вызвано это тем, что в отличие от других формул ошибок аппроксимаций в данной формуле фактически происходит деление модуля отклонения на модуль средней величины между фактическим и расчетным значениями: $\left|\frac{y_t + \hat{y}_t}{2}\right|$. В результате этого деления дробь переворачивается и 2 оказывается в числителе, после чего выносится за знак суммы. Взятие суммы в знаменателе по модулю позволяет ограничить ошибку только положительными значениями.

С одной стороны, деление на среднее значение позволяет получить более адекватную оценку ошибки в случаях с фактическими значениями, близкими к нулю. Однако это не до конца решает проблему, так как ситуации, в которых и фактические, и расчетные значения близки к нулю, остаются вполне возможными на практике.

Достоинством *sMAPE* является также то, что ошибки, рассчитанные по формуле (2.21), будут лежать в пределах от 0 до 200%, что легче воспринимается. Однако сохраняется проблема с оценкой данных, измеренных не в абсолютных шкалах.

Кроме того, симметричная ошибка аппроксимации на самом деле не является до конца симметричной. Так, если по одной модели расчетное значение оказалось завышенным на величину x, а по другой — заниженным на ту же величину x, то *sMAPE* будет выше у второй модели (несмотря на одинаковую величину отклонения от факта), в связи с тем, что знаменатель в формуле (2.21) у нее будет меньше. Однако это свойство коэффициента соответствует общему представлению о том, что «переоценка лучше, чем недооценка». Например, при прогнозировании спроса, заниженный про-

[1] *Hyndman Rob J., Koehler Anne B., Ord J. Keith, Snyder Ralph D.* Forecasting with Exponential Smoothing: The State Space Approach. Springer-Verlag Berlin Heidelberg, 2008. P. 26.

гноз оказывается хуже для предприятия, нежели завышенный.

Для модели № 1 получилось следующее значение симметричной средней относительной ошибки аппроксимации: *sMAPE* = 7,95%. Для модели № 2 *sMAPE* оказалась меньше — 6,67%, так как в основе (2.21) лежит сумма модулей, а не сумма квадратов, по которой вторая модель проигрывает. Сама по себе полученная величина как в первом, так и во втором случае невелика, что указывает на отсутствие серьезных нарушений в моделях.

Последняя величина, дающая оценку ошибки, фактически основана на сравнении прогнозных свойств модели с прогнозными свойствами элементарного метода под названием Naïve (мы к нему еще обратимся в параграфе 5.2), суть которого заключается в том, что в качестве прогноза будущего используется предыдущее фактическое значение. Рассчитываемая величина называется *«U-статистика»* и рассчитывается по формуле:

$$U = \sqrt{\frac{\sum_{t=1}^{T-1}\left(\frac{y_{t+1}-\hat{y}_{t+1}}{y_t}\right)^2}{\sum_{t=1}^{T-1}\left(\frac{y_{t+1}-y_t}{y_t}\right)^2}}. \tag{2.22}$$

Значение U-статистики может быть равным нулю только в случае, если все фактические значения оказались равными расчетным. В остальных случаях она будет больше нуля. Можно выделить следующие области для интерпретации U-статистики:

1. $U = 1$ — метод Naïve по описательной способности оказался идентичен используемому методу;
2. $U < 1$ — используемый метод оказался лучше Naïve по описательной способности;
3. $U > 1$ — используемый метод оказался хуже метода Naïve по описательной способности;

Из-за того, что сравнение в данном случае идет с другим методом, данная статистика имеет смысл и при оценке величин, измеренных в интервальных шкалах. Кроме того, масштаб не сильно влияет на данную величину все из-за того же сравнения с методом Naïve. Конечно, значение U-статистики не ограничено сверху, однако получение значения больше 1 уже сигнализирует о том, что, возможно, имеет смысл обратиться к другому методу прогнозирования.

Для модели № 1 U оказалась равной 1,02, что указывает на то, что метод Naïve лучше справляется с аппроксимацией ряда данных, чем используемый нами метод. Для модели № 2 U оказалась незначительно выше 1 (отличие на седьмом знаке после запятой), что говорит о том, что модель аппроксимирует ряд на уровне модели «Naïve».

Помимо рассмотренных нами здесь оценок ошибок аппроксимации используют некоторые расчетные коэффициенты, которые дают полезную информацию об аппроксимационных свойствах модели. Один из них — это *коэффициент детерминации*, который рассчитывается по формуле:

$$R^2 = \frac{ESS}{TSS}, \tag{2.23}$$

где $ESS = \sum_{t=1}^{T} \left(\hat{y}_t - \overline{y}\right)^2$ — объясненная сумма квадратов отклонений, а $TSS = \sum_{t=1}^{T} \left(y_t - \overline{y}\right)^2$ — общая сумма квадратов отклонений.

Суммы квадратов отклонений в случае с линейными регрессионными уравнениями связаны следующим равенством:

$$TSS = RSS + ESS, \tag{2.24}$$

откуда следует еще одна формула для расчета R^2:

$$R^2 = 1 - \frac{RSS}{TSS}. \tag{2.25}$$

Фактически, R^2, рассчитанный по формулам (2.23) и (2.25) имеет смысл, схожий со смыслом U-статистики. В нем также сравниваются значения по используемой модели со значениями по более простой модели. Только в данном случае в качестве такой простой модели выступает средняя величина по ряду, кроме того в основе коэффициента лежит не ошибка, а отклонение от средней величины. Соответственно тут также возможны три ситуации с разными интерпретациями.

1. $R^2 = 1$ — модель идеально описала ряд данных. Такая ситуация возможна только в том случае, если все расчетные значения оказались равными всем фактическим.

2. $R^2 < 1$ — модель описала соответствующий процент дисперсии фактических значений.

3. $R^2 \notin (0; 1)$ — модель имеет слишком высокие значения по сравнению с фактическими.

Последняя ситуация возможна чисто технически при использовании нелинейных моделей. Причем в этом случае равенство (2.24) не выполняется, что приводит к тому, что R^2, рассчитанные по формулам (2.23) и (2.24), не будут совпадать: в первом случае коэффициент будет больше единицы, а во втором — меньше нуля.

Во всех остальных случаях коэффициент детерминации лежит в пределах от 0 до 1.

Помимо коэффициента детерминации по модели можно рассчитать *квадрат коэффициента корреляции между фактическими и расчетными значениями*. Этот показатель в случае с линейными регрессионными моделями будет эквивалентен коэффициенту детерминации. Рассчитывается он как

$$c^2 = \operatorname{corr}\left(y_t; \hat{y}_t\right)^2 = \frac{\operatorname{cov}\left(y_t; \hat{y}_t\right)^2}{\sqrt{D\left(y_t\right) D\left(\hat{y}_t\right)}}, \tag{2.26}$$

однако в случае со сложными нелинейными моделями показатели, рассчитанные по (2.23) и (2.26), будут различаться.

В нашем примере были построены более сложные нелинейные модели, поэтому коэффициенты детерминации для них, рассчитанные по формуле (2.23), получились равными 1,0033 и 0,9104, по формуле (2.24) — 0,9545 и 0,9504, а квадрат коэффициента корреляции получился равным 0,9559 и 0,9531 соответственно. Как интерпретировать полученные значения, не совсем понятно. Даже сравнить полученные модели по этим значениям затруднительно. Поэтому эти коэффициенты и не используются в таких ситуациях.

У коэффициента детерминации есть один недостаток, который заключается в том, что он никак не позволяет определить систематические занижение или завышение модели в случаях, вызванных какими-то техническими причинами.

Еще один коэффициент, правда, отличающийся от всех предыдущих, называется *коэффициентом «сбалансированности»*[1] и рассчитывается по формуле

---

[1] *Светуньков И. С.* Новые коэффициенты оценки качества эконометрических моделей // Прикладная эконометрика. 2011. № 4 (24). С. 85–99.

$$B = \frac{\sum_{t=1}^{T} \sqrt{y_t^2 - \hat{y}_t^2}}{\sum_{t=1}^{T} \sqrt{y_t^2 + \hat{y}_t^2}}. \qquad (2.27)$$

Коэффициент сбалансированности основан на идее представления чисел на псевдоевклидовой плоскости и использует теорию функций комплексных переменных. По формуле (2.27) видно, что B может быть:

1) действительным числом в случае, если у модели имеются систематические отклонения, такие что $|y_t| > |\hat{y}_t|$, причем, если все $\hat{y}_t = 0$, то $B = 1$;

2) мнимым числом, в случае, когда у модели имеются систематические отклонения от реальных данных другого характера: $|y_t| < |\hat{y}_t|$, причем, если все $y_t = 0$, то $B = i$;

3) комплексным числом, если систематических отклонений нет;

4) равным нулю, если модель идеально описывает ряд данных.

Для удобства коэффициент можно представить в процентах. По полученным действительному и мнимому значениям коэффициента (2.27) можно судить о симметричности остатков. Так, если в остатках наблюдается отрицательная асимметрия (т.е. у распределения «левый хвост» длиннее правого), то мнимая часть коэффициента будет больше действительной. Если же наблюдается положительная асимметрия, то действительная часть будет выше мнимой.

Кроме того, коэффициент может принимать только такие значения, при которых $|B| \leq 1$. Графически область допустимых значений коэффициента может быть представлена в виде треугольника в первом квадранте, показанном на рисунке 2.10. Чем ближе в таком случае величина коэффициента B к границе, тем большие отклонения имеет модель от фактических значений.

На основе коэффициента сбалансированности можно рассчитать еще два полезных коэффициента.

1. *Модуль коэффициента сбалансированности*:

$$|B| = \sqrt{\left(\mathrm{Re}(B)\right)^2 + \left(\mathrm{Im}(B)\right)^2}. \qquad (2.28)$$

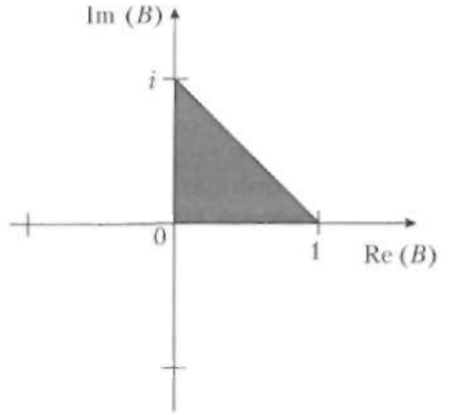

Рис. 2.10. **Область допустимых значений коэффициента сбалансированности**

Этот показатель по смыслу близок к *sMAPE* и может принимать значения от 0 до 100%. В случае, если модель не описывает изучаемый процесс, его значение будет близко к 100%. В обратной ситуации он будет близок к 0. Коэффициент менее чувствителен к масштабу данных, нежели, например, простые ошибки аппроксимации *MAPE*. Однако этот коэффициент может быть ошибочным в случае, если фактические и расчетные значения оказываются противоположными по знаку. Такие ситуации в прогнозировании встречаются редко, однако это неприятное свойство коэффициента сбалансированности надо иметь в виду, если приходится работать с отрицательными значениями.

2. *Тангенс угла наклона коэффициента сбалансированности*:

$$b = \frac{\mathrm{Re}(B) - \mathrm{Im}(B)}{\mathrm{Im}(B) + \mathrm{Re}(B)}. \tag{2.29}$$

В общем случае тангенс угла наклона вектора рассчитывается относительно оси абсцисс. Однако в нашем случае это была бы неудобная характеристика, так как тангенс $\frac{\pi}{2}$ – это бесконечность, и полученный коэффициент имел бы область значений от нуля до бесконечности, что сложно трактовать. Именно поэтому для коэффициента сбалансированности имеет смысл рассчитывать угол наклона относительно линии $\text{Re}(B)=\text{Im}(B)$, характеризующей полный баланс модели. Именно этого и можно достичь, рассчитав угол наклона по формуле (2.29).

B в таком случае лежит в промежутке от -1 до 1. Чем ближе его значение к -1, тем более завышенный прогноз дает модель (наблюдается отрицательная асимметрия). Чем ближе значение коэффициента (2.29) к 1, тем более заниженным будет прогноз. Значение b можно измерять в процентах, тогда полученная величина будет показывать процент соответствующих отклонений.

Для первой модели коэффициент сбалансированности оказался равным 10,3% + i11,2%, его модуль — 15,15%, а b = 4,67%. Это говорит о незначительной отрицательной асимметрии (модель дает незначительное завышение относительно фактических значений). Кроме того, число 15,15% говорит о невысокой описательной способности модели. Во второй модели B = 9,3% + i10,8%, $|B|$ = 14,25%, b = 7,46%. Все так же видим отрицательную асимметрию, превышающую асимметрию в первой модели, и то, что описательная способность второй модели оказалась выше.

Помимо перечисленных показателей существуют еще и *стандартные статистические*, с помощью которых можно изучить и оценить остатки модели. К таким показателям относятся асимметрия и эксцесс. *Асимметрия* рассчитывается по формуле

$$Sk=\frac{\frac{1}{T}\sum_{t=1}^{T}e_t^3}{\sigma_e^3}, \tag{2.30}$$

где σ_e — стандартное отклонение ошибок, рассчитываемое как корень из дисперсии (2.13).

Смысл данного показателя прост: если в ряде ошибок имеются большие (по модулю) отрицательные значения, не компенсирующиеся такими же большими положительными, то их куб в (2.30) будет большим отрицательным числом, а значит и итоговый показатель будет отрицательным. И наоборот. Деление на куб в (2.30) проведено для того, чтобы нормировать полученные кубы ошибок. Однако как-то иначе интерпретировать показатель асимметрии, кроме как «положительный» или «отрицательный», а также «близко к нулю» или «не близко к нулю» не представляется возможным. Как известно из теории вероятностей, получить асимметрию, равную нулю, практически невозможно[1].

Для модели № 1 коэффициент асимметрии составил −0,98, а для модели № 2 — −1,44, что вполне соответствует распределению ошибок, представленному на рис. 2.6. Асимметрия распределения ошибок во второй модели сильнее, чем в первой.

Помимо асимметрии важна еще и «острота» распределения, которая характеризуется таким показателем, как *эксцесс*, рассчитываемый по формуле

$$Kur = \frac{\frac{1}{T}\sum_{t=1}^{T} e_t^4}{\sigma_e^4}. \qquad (2.31)$$

В связи с тем, что эталонным распределением случайных величин является нормальное распределение, у которого эксцесс равен 3, формулу (2.31) иногда модифицируют: отнимают от полученного числа 3. В таком случае показатель будет характеризовать эксцесс по сравнению с нормальным законом распределения.

У данного показателя есть две ценные интерпретации.

1. $Kur < 3$ — в распределении ошибок наблюдается отрицательный эксцесс, т.е. значения, лежащие близко к 0, встречаются реже, чем это должно быть по сравнению с нормальным законом.

2. $Kur > 3$ — в распределении ошибок наблюдается положительный эксцесс, т.е. значения, близкие к нулю, встречаются чаще, чем в нормальном законе распределения.

Ситуация, в которой $Kur = 3$, принципиально невозможна по тем же причинам, по которым асимметрия не может быть равной нулю.

По нашим остаткам эксцесс оказался равным 2,406 и 3,566 — для первой и для второй моделей соответственно. Это говорит о том, что остатки, близкие к 0, в первой модели встречаются реже, чем должны были бы в случае с нормально распределенными остатками с такой же дисперсией, как в нашем ряде данных, а во второй модели — чаще, чем это предполагает нормальное распределение.

---

[1] *Вентцель Е. С.* Теория вероятностей : учебник. 11-е изд. М. : КноРус, 2010. С. 81.

Последний инструмент, использующийся при оценке аппроксимационных свойств моделей — это *статистические гипотезы*. Подробней мы их рассмотрим в параграфе 3.4. Пока же приведем наиболее распространенную формулу, с помощью которой проверяется гипотеза о статистической значимости модели:

$$F = \frac{ESS/(k-1)}{RSS/(T-k)} = \frac{(T-k)ESS}{(k-1)RSS}. \qquad (2.32)$$

Данная формула позволяет рассчитать значение F-статистики, которое по результатам расчетов сравнивается с F-табличным с заданным уровнем вероятности и степенями свободы $(k - 1)$ и $(T - k)$. В случае если значение, рассчитанное по формуле (2.32), оказалось больше критического, у исследователя появляются основания отклонить нулевую гипотезу и сделать вывод о статистической значимости модели. В наше время F-статистика стала важнейшим показателем адекватности модели. Можно обратить внимание, что, если ошибки модели невелики (RSS имеет небольшое, а ESS — большое значение), то F оказывается достаточно большим числом, в результате чего гипотеза, скорее всего, будет отклонена, а значит и модель получилась статистически значимой. Это достаточно важный показатель качества модели. Однако обратим внимание на то, что в числителе (2.30) приведено умножение на число наблюдений. В данном случае идея заключается в том, что с увеличением числа наблюдений даже незначительная описательная способность модели будет приводить к высокому значению F-статистики.

Здесь реализован основной механизм выборочного метода: чем больше собрано наблюдений, тем более точную оценку величины можно получить. Но на самом деле в этом же кроется и некая опасность: если модель имеет очень слабую объясняющую способность, но при этом построена по большому числу наблюдений, то она чисто технически может оказаться статистически значимой. Так, даже простая прямая линия при достаточном числе наблюдений становится статистически значимой — начинает работать закон больших чисел, который дает достаточно хорошие, устойчивые статистически характеристики, но, к сожалению, во многих случаях точность прогнозов не повышает.

Отметим, что расчет F-статистики имеет смысл только в случае с линейными регрессионными моделями, в которых остатки распределены нормально. В связи с тем, что мы имеем дело с нелинейной моделью, остатки которой распределены явно ненормально, рассчитывать F-статистику не имеет смысла.

Собственно говоря, один из критериев адекватности модели заключается в проверке статистических гипотез относительно полученного распределения остатков. Нулевая гипотеза в таких тестах обычно формулируется как «случайная величина распределена нормально», а альтернативная — «ненормально». Для проверки такой гипотезы разработано несколько тестов, некоторые из которых мы рассмотрим в параграфе 3.4. Пока же обратим внимание на то, что в результате проведенных тестов исследователь может лишь сделать вывод относительно остатков каждой отдельной модели и решить, считать ли полученные остатки распределенными нормально или нет. Сравнивать же между собой модели по полученным статистикам (как будто в данной модели они «более нормальные», чем в другой) некорректно.

Помимо описанных показателей можно упомянуть и о классе показателей под названием «*информационные критерии*». Их суть сводится к тому, что при большом числе регрессоров, включенных в модель, многие из рассмотренных нами показателей оказываются заниженными (в случае с ошибками) или завышенными (в случае с коэффициентом детерминации). При этом вклад в описательную способность некоторых регрессоров может быть совершенно незначительным. Поэтому для того, чтобы учесть увеличение описательной способности за счет добавления лишних регрессоров, используются такие информационные критерии, как «*AIC*», «*BIC*», «*BICc*» и пр.[1] Во всех них используется один и тот же принцип расчета:

$$IC = \alpha f(k) - \beta \ln L, \tag{2.33}$$

где $\ln L$ — это логарифм функции правдоподобия (см. параграф 3.5); k — число коэффициентов в модели; α, β — коэффициенты, отличающие один информационный критерий от другого; f — функция от числа коэффициентов, использующаяся в критерии.

---

[1] *Makridakis Spyros G., Wheelwright Steven C., Hyndman Rob J.* Forecasting: Methods and Applications. Wiley, 1998. P. 280.

Например, в случае с AIC $\alpha = \beta = 2$, $f(k) = k$. В моделях с хорошей описательной способностью $\ln L$ будет больше, чем в моделях с плохой, а значит и AIC будет меньше. При этом число коэффициентов играет роль «штрафа»: при большом числе коэффициентов AIC будет увеличиваться. Таким образом, с помощью AIC можно выбрать среди моделей ту, которая будет наилучшим образом описывать ряд данных при наименьшем числе коэффициентов. Выбор модели осуществляется в пользу той, у которой AIC меньше.

Фактически в информационных критериях заложен принцип бритвы Оккама: не стоит плодить сущность сверх необходимого (переводя на язык прогнозирования, — зачем включать в модель массу переменных, не дающих значительного увеличения в точности модели, если можно использовать более компактную и удобную модель).

Для первой модели (у которой было три коэффициента) $AIC = 946{,}58$, в то время как для второй (у которой был один коэффициент) он составил 944,33. Таким образом, по информационному критерию стоит отдать предпочтение второй модели.

Для того чтобы понять, какая модель в итоге дала более точный прогноз, рассмотрим график с фактическими значениями на периоде прогноза (рис. 2.11).

Как видим, на первых 10 значениях точным оказался прогноз второй модели, в то время как на последних трех — прогноз по первой модели, т.е. однозначного вывода о точности прогноза на основе только аппроксимации ряда сделать, в общем-то, нельзя. Для того чтобы сравнить эту точность численно, можно воспользоваться любым из показателей, рассмотренных нами в данном параграфе. Учитывая, что вторая модель точнее спрогнозировала большее число наблюдений, ошибки прогноза (*MAPE*, *sMAPE* и т.п.) по ней будут явно меньше.

В целом стоит заметить, что точность аппроксимации гарантирует получение точного прогноза лишь в случае с обратимыми процессами. При этом модель, заведомо плохо аппроксимирующая ряд данных, навряд ли даст точный прогноз. Здесь мы имеем дело с двумя противоположностями, с которыми прогнозисту нужно научиться обращаться: с одной стороны, не стоит погружаться в поиски модели с идеальной объясняющей способностью, с другой, — не стоит полагаться на неадекватную модель (какой бы статистически значимой она ни была). Тут можно в качестве пояснения привести два графических примера (рис. 2.12).

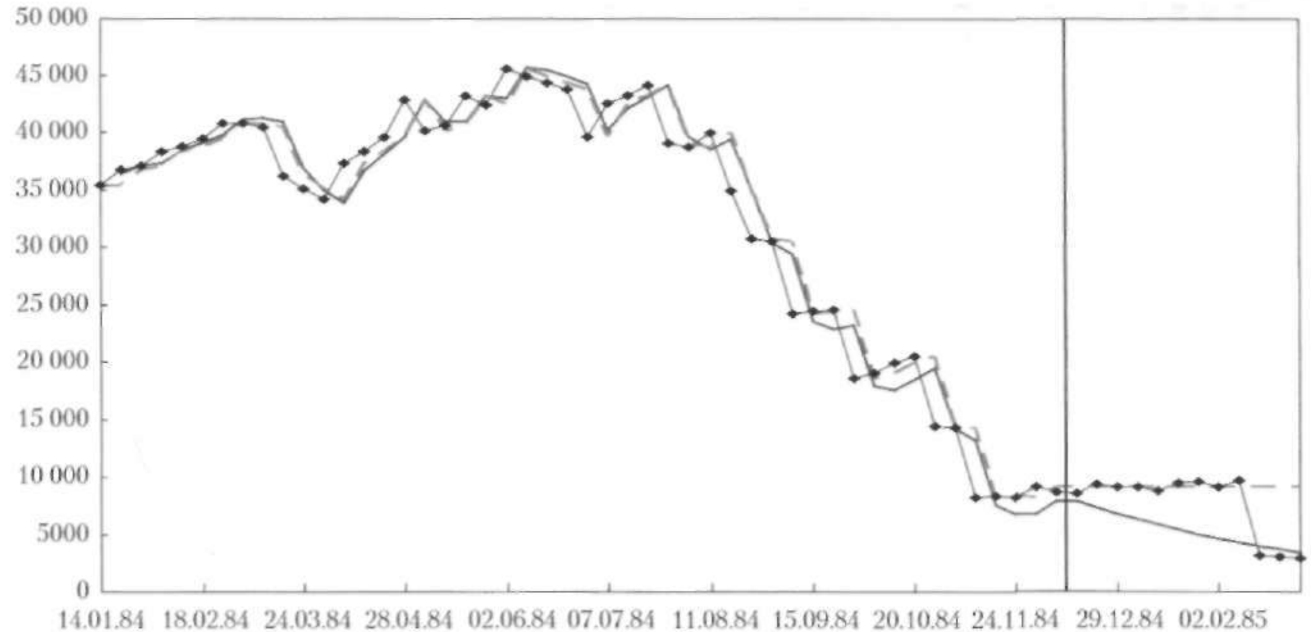

Рис. 2.11. **Фактические значения на период прогноза:**
сплошная линия с точками — условный ряд данных; *сплошная линия до вертикальной* (*модель 1*), *пунктирная* (*модель 2*) — расчетные значения; *сплошная линия после вертикальной* (*модель 1*), *пунктирная* (*модель 2*) — прогноз по некоторым моделям

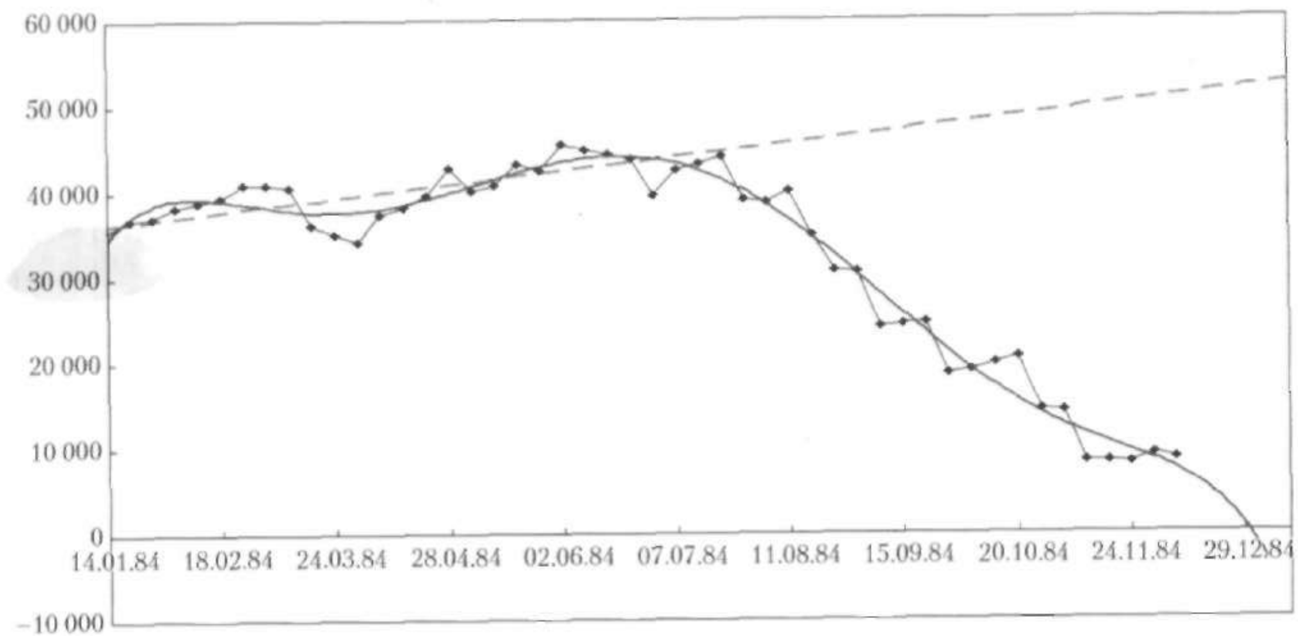

Рис. 2.12. **Условный ряд данных и прогноз по двум моделям:**
пунктирная линия — модель № 1, *сплошная линия* — № 2

На рисунке показаны две ситуации.

Первая модель дает неадекватный прогноз, потому что не учитывает весь промежуток, начиная с изменения тенденции на снижение. Ошибки аппроксимации у модели, очевидно, достаточно высокие, R^2 низкий, но при этом модель статистически значима (F-статистика составляет 18,31, при критическом в 4,05).

Вторая модель также дает неадекватный прогноз, но при этом имеет очень хорошие аппроксимационные способности: ошибки в ней значительно ниже, чем в первой модели, R^2, скорее всего, достаточно высок. Модель также будет, скорее всего, статистически значима.

Данный пример демонстрирует простую истину: при выборе прогнозирующей модели нужно следовать срединному пути — нельзя опираться лишь на графики, либо показатели аппроксимации, либо статистические данные. Главными показателями качества модели для прогнозиста должны быть его собственный опыт и интуиция.

2.5.3. Процедура ретропрогноза

В связи с тем, что точность аппроксимации не гарантирует точность прогнозирования, попытки получить хоть какую-то более адекватную оценку точности предлагаемого прогноза привели к разработке процедуры ретропрогноза, в соответствии с которой имеющаяся выборка делится на две части.

1. Обучающая выборка. Она больше по объему и включает в себя наблюдения от 1 до $T - h$, где h выбирается исследователем, но чаще всего так, чтобы не быть больше $T / 3$.

2. Тестовая выборка меньше по объему и включает в себя наблюдения от $T - h + 1$ до T.

По первой части строится модель и оцениваются значения ее коэффициентов. Затем по построенной модели дается прогноз на $h - 1$ наблюдение вперед. Полученный прогноз сравнивается с фактическими значениями прогнозируемого показателя. Далее ошибка ретропрогноза используется для расчета показателей точности прогноза (*MAPE* и пр.), после чего делается вывод о точности модели.

В процедуре ретропрогноза лучше использовать несколько моделей, тогда на их основе можно будет выбрать ту модель, которая дает более точный прогноз на тестовой выборке. Такая процедура, с одной стороны, позволяет оценить точность прогноза в будущем (при условии, что намеченные тенденции и связи сохранятся), а с другой, — вроде бы гарантирует, что будет выбрана модель, которая дает более точный прогноз для исследуемого ряда данных.

Практикум

Вопросы для самоконтроля

1. Что такое «данные» и что такое «информация», как они взаимосвязаны?

2. Почему самая простая шкала измерения информации — номинальная — имеет столько других названий?

3. Чем порядковая шкала отличается от номинальной?

4. Интервальная шкала уже имеет такую характеристику, как «расстояние». Что под этим понимается?

5. Что такое «аттитюд»?

6. Почему для изучения поведения потребителя и прогнозирования его реакции на товар и его маркетинговое сопровождение необходимо измерять не один, а несколько показателей?

7. Какие типы ошибок могут встретиться при предварительной обработке информации, собранной для выполнения прогноза?

8. Как устранить грубую ошибку в информации?

9. Что делать с систематической ошибкой в исходной информации для того, чтобы в дальнейшем сделать прогноз?

10. Можно ли устранить из собранной информации случайную ошибку?

11. Информация, измеренная в метрической шкале, не дает возможности осуществлять сравнение измеренных в этой шкале чисел друг с другом. Над ней тем более невозможно выполнение никаких математических действий. Как можно использовать эту информацию для прогнозирования?

12. Почему при измерении степени взаимосвязи между показателями, измеренными в неметрических шкалах, не удается ее точно определить?

13. Что представляет собой графический анализ прогнозной модели?

14. Для чего используют расчетные показатели оценки качества прогнозной модели?

15. Что собой представляет процедура ретропрогноза? Какова ее основная цель? Какой из общенаучных методов вывода здесь используется — дедуктивный или индуктивный?

Задания

Задание 1. В вашем распоряжении имеются данные о средней стоимости докторской колбасы в Санкт-Петербурге по месяцам (руб. за кг):

Месяц	Цена	Месяц	Цена
Январь	301	Июль	337
Февраль	307	Август	341
Март	313	Сентябрь	346
Апрель	319	Октябрь	353
Май	325	Ноябрь	359
Июнь	331	Декабрь	366

1. Переведите данные этой таблицы в интервальную шкалу на основе максимальной, минимальной и средней цен.

2. Переведите данные таблицы в порядковую шкалу, состоящую из четырех групп: «до 315 руб. за кг», «от 315 до 330 руб. за кг», «от 330 до 345 руб. за кг», «от 345 руб. за кг». Рассчитайте медиану по полученным данным. Попытайтесь оценить среднюю величину по эти же данным. Какой они имеют смысл?

3. Переведите данные таблицы в номинальную шкалу и создайте соответствующую переменную «Колбаса стоит выше 350 руб.». Рассчитайте моду.

Задание 2. На сайте «Росстата» (www.gks.ru) найдите данные по базовым индексам потребительских цен за последние 10 лет.

1. Импортируйте данные в MS Excel, используя функцию «Вставить — Линейный график», постройте график изменения индекса цен. Что можно сказать о динамике индекса цен?

2. Используя функцию MS Excel «ЧАСТОТА», сформируйте таблицу частот появления индекса цен (за основу возьмите значения от 1 до 1,2 с шагом 0,05). На основе полученной таблицы постройте гистограмму. Какое значение встречается чаще остальных?

Есть ли в ваших данных выбросы? Чем они могут быть вызваны? Можно ли их убрать? Если да, то чем их лучше заменить?

Задание 3. Исследователь изучает работу курьерской службы. Для этого он собрал следующие данные:

Хрупкое содержимое	Доставка в черте города	Класс посылки	Температура на улице (°C)	Время доставки (в минутах)
1	1	3	10	61
0	0	2	12	54
1	0	2	5	41
0	1	1	7	13
0	0	1	8	15

Хрупкое содержимое	Доставка в черте города	Класс посылки	Температура на улице (°C)	Время доставки (в минутах)
1	1	3	10	55
0	1	3	13	71
0	0	2	11	53
0	0	3	5	67
1	1	2	3	51
1	0	3	5	23
1	1	3	2	120

Хрупкое содержимое — это переменная, которая принимает значение «1», если посылка содержит хрупкие предметы, и «0» — в противном случае.

Доставка в черте города — «1», если доставка осуществляется в черте города, «0» — в противном случае.

Класс посылки — чем ниже значение, тем более важная посылка с более приоритетной доставкой.

Температура на улице — средняя температура воздуха днем во время доставки посылки.

Время доставки — время, прошедшее с момента получения посылки до момента ее доставки.

1. Определите, в каких шкалах измерены данные.
2. Исследователь считает, что между некоторыми переменными может существовать связь, поэтому ему нужно рассчитать коэффициенты, которые покажут наличие либо отсутствие этих связей. Определите, какие коэффициенты нужно использовать для расчета силы связи между разными переменными.
3. Проведите расчеты коэффициентов, которые вы выбрали в п. 2.

Имеется ли какая-нибудь зависимость между «хрупкостью содержимого» и «доставкой в черте города»? Если есть, то какая?

Есть ли связь между «классом посылки» и «временем доставки»? Если есть, то какая именно?

Связаны ли между собой «температура на улице» и «время доставки»?

Задание 4. Аналитик попытался дать прогноз по продажам автомобилей своей компании на год вперед на основе исторических данных. Для этого он использовал процедуру ретропрогноза, построил несколько моделей и выбрал из них две наилучшие. В вашем распоряжении имеются прогнозы по этим двум моделям, а также ряд фактических значений, которые аналитик сохранил для сравнения точности моделей (см. таблицу).

Месяц	Факт	Модель 1	Модель 2
Январь 2012	6002,7	6459,8	6408,9
Февраль 2012	6370,5	6340,4	6202,8
Март 2012	7664,1	7352,9	7236,4
Апрель 2012	6969,6	7346,4	7241,8
Май 2012	6963,9	7191,6	7522,7
Июнь 2012	7576,8	7350,0	8012,3
Июль 2012	6618,6	6893,6	7114,9
Август 2012	7473,6	7118,6	7867,3
Сентябрь 2012	7668,0	7598,0	8111,4
Октябрь 2012	7725,0	7727,7	8321,6
Ноябрь 2012	7223,1	7092,9	7066,8
Декабрь 2012	6461,4	6768,1	6398,4

1. На основе данных таблицы постройте линейные графики и сравните точность прогнозов по этим двум моделям. Какие особенности можно выделить в том, как каждая из моделей прогнозирует динамику продаж?
2. Рассчитайте все коэффициенты, рассмотренные в параграфе 2.5, чтобы получить наиболее полную информацию о том, насколько точный каждая из моделей в среднем дает прогноз.
3. Какой из моделей стоит отдать предпочтение и почему?

Задание 5. Прогнозист по 40 наблюдениям построил две модели: в одной было четыре коэффициента, а сумма квадратов ошибок оказалась равной 4537, в другой было три коэффициента, а сумма квадратов ошибок составила 4497. Дисперсия изучаемого показателя составила 262,675, а средняя величина — 512.

1. На основе этих данных рассчитайте ошибки аппроксимации (все, какие сможете), R^2, F-критерий;
2. Сделайте выводы относительно точности аппроксимации данных этими двумя моделями. Можно ли выбрать модель на основе этих данных, и если можно, то какой бы из них вы отдали предпочтение?

Глава 3
ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ПРОГНОЗИРОВАНИЯ СТАЦИОНАРНЫХ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ПРОЦЕССОВ

В результате освоения данной главы студент должен:

знать

- основные понятия, методы и инструменты количественного и качественного анализа стационарных социально-экономических процессов;
- основные информационные технологии, используемые в прогнозировании стационарных процессов;
- свойства средних величин, дисперсий и других мер колеблемости;

уметь

- различать генеральную и выборочную совокупности;
- оценивать степень приемлемости использования выборочного метода для прогнозирования социально-экономических процессов;
- давать интервальную оценку прогнозируемому показателю;
- осуществлять статистическую проверку гипотез;
- строить регрессионные модели;
- оценивать наличие взаимосвязи между случайными факторами;

владеть

- навыками получения точечных оценок характеристик выборочных совокупностей;
- методами построения доверительных интервалов для точечных оценок характеристик выборочных совокупностей;
- методами регрессионного и корреляционного анализа применительно к задачам социально-экономического прогнозирования;
- навыками самостоятельной научной и исследовательской работы в части выявления причинно-следственных связей и моделирования их статистическими методами;
- методикой построения регрессионных моделей социально-экономического прогнозирования.

3.1. Генеральная совокупность, выборка и выборочный метод

Поскольку охватить всю совокупность наблюдений за социально-экономическими явлениями невозможно, прогнозисту приходится использовать только некоторую их часть. Из всего множества имеющихся наблюдений он использует только те, которые некоторым образом выбраны им из всего множества возможных значений для того, чтобы по нескольким значениям составить правильное представление об изучаемом явлении в целом. Поскольку всю возможную совокупность наблюдений обычно представляют как генеральную совокупность, а часть из нее — как выборочную, следует разобраться в том, какая именно совокупность наблюдений может быть названа генеральной в строго статистическом смысле, а какая, опять же в строго математическом понимании, — выборочной. Это очень важно, поскольку под понятиями «генеральная совокупность» и «выборочная совокупность» в теории вероятностей и математической статистики понимаются вполне определенные и четко очерченные в границах определения характеристики соответствующих объектов. Поэтому, прежде чем использовать их, необходимо уяснить, что же они собой представляют. Это позволит прогнозисту избежать терминологической путаницы и не ошибиться в выборе инструмента прогнозирования.

О том, что простого ответа на поставленный вопрос нет, может свидетельствовать следующая цитата: «Анализируемые ряды динамики являются *почти всегда* выборками из более длинных рядов»[1]. Она говорит о том, что еще в 1977 г. ученые считали, что имеющиеся в их распоряжении ряды не всегда являются выборками. Как различить случаи, когда анализируемые ряды являются выборками из более длинных рядов, а когда они таковыми не являются? Ответ на этот вопрос найти необходимо, ведь для обработки выборочных значений используется выборочный метод, а он подразумевает использование методов математической статистики. Но, как было показано, в предыдущей главе, например, для необратимых процессов методы математической статистики неприемлемы — свойства рядов необратимой динамики не соответствуют посылкам выборочного метода.

[1] *Вайну Я. Я.-Ф.* Корреляция рядов динамики. М. : Статистика, 1977. С. 90.

Итак, прежде всего, рассмотрим понятие **«генеральная совокупность»**. Мы не будем рассматривать определения типа: «та совокупность, из которой проводится отбор, называется генеральной совокупностью; отобранные данные составляют выборочную совокупность»[1], поскольку такие определения не отражают сущностных свойств определяемых понятий. Они позволяют включить в число генеральных совокупностей практически все ряды наблюдений, например, социально-экономические показатели Руси и России за всю историю существования государства. Но ни один нормальный прогнозист не станет рассматривать статистические данные такого ряда с 1000 по 2012 г. для того, чтобы по этой «генеральной совокупности» сделать прогноз на последующие за 2012 годы. Значит, генеральная совокупность — не просто некоторая база данных. Она включает те из них, которые обладают определенными свойствами, знание которых нужно прогнозисту, чтобы легко определить, генеральная совокупность перед ним или нет.

Для понимания свойств генеральной совокупности обратимся к одной из фундаментальных отечественных работ по теории вероятностей и математической статистики и посмотрим, как ее авторы определяли интересующие нас понятия. «Пусть имеется многочисленная совокупность однородных элементов (объектов), каждый из которых может обладать или не обладать каким-либо признаком; неизвестная нам доля тех из них, которые обладают этим признаком, и подлежит определению. Наше испытание заключается в том, что мы выбираем наугад один элемент из множества элементов, отмечаем, обладает или нет этот элемент данным признаком и возвращаем его обратно в совокупность. При выборе элемента из совокупности принимаются меры к тому, чтобы вероятность быть выбранным была одинакова для всех элементов. Тогда имеющееся множество элементов называется *генеральной* совокупностью. Группа из *n* элементов, наблюдаемых при повторных испытаниях, называется *случайной выборкой*, число отобранных элементов — объемом выборки, а описанный процесс отбора элементов — простым

[1] *Елисеева И. И., Юзбашев М. М.* Общая теория статистики : учебник / под ред. И. И. Елисеевой. М. : Финансы и статистика, 2004. С. 216; Справочник по математике для экономистов : учеб. пособие / под ред. проф. В. И. Ермакова. М. : ИНФРА-М, 2007. С. 372; *Доугерти К.* Введение в эконометрику. М. : ИНФРА-М, 1997. С. 4. и др.

случайным выбором. Определив частоту признака среди отобранных в выборке объектов, мы можем по ней, опираясь на теорему Лапласа, приближенно оценить долю признака в генеральной совокупности: в самом деле, эта доля в данных условиях играет роль неизменной вероятности появления объекта, обладающего интересующим нас признаком, а уклонения частоты от вероятности приближенно следуют нормальному закону. При правильной организации отбора выборки мы должны обеспечить каждому объекту генеральной совокупности равную вероятность попадания в выборку.

На практике выбор из генеральной совокупности производится различными способами, в частности, выбор подразделяется на следующие разновидности.

1. Выбор с возвратом или повторением, или, иначе говоря, простой случайный выбор, о котором говорилось выше.

2. Выбор без возврата или без повторения, когда каждый отобранный индивидуум перед выбором следующего индивидуума обратно в генеральную совокупность не возвращается.

Полученная выборка называется репрезентативной (представительной), если она достаточно хорошо представляет пропорции генеральной совокупности»[1].

В этом отрывке приведены характерные свойства, присущие генеральной совокупности и выборке из нее, поэтому, воспользовавшись им, конкретизируем эти свойства. Относительно генеральной совокупности ясно, что это совокупность всех *однородных* элементов. Что касается выборки, то для нее характерны *случайность* выбора, *одинаковая вероятность* для каждого элемента быть выбранным, а также способность выборки *характеризовать свойства* генеральной совокупности.

Генеральная совокупность может быть конечной или бесконечной. В первом случае множество однородных элементов, хотя и очень велико, но все же конечно (например, количество подберезовиков в лесах Поволжья осенью текущего года). Во втором — множество этих элементов не ограничено, поскольку новые наблюдения за ними дают все новые и новые значения этой совокупности (например, количество

---

[1] *Дунин-Барковский И. В., Смирнов Н. В.* Теория вероятностей и математическая статистика в технике. М. : Гос. изд-во технико-теоретической литературы, 1955. С. 179.

подберезовиков в Поволжье с момента их появления в этом регионе и до настоящего времени. Поскольку «настоящее время», как верхняя временна́я граница этого множества со временем устремляется к бесконечности, то и генеральная совокупность элементов является бесконечной).

Важно, что генеральная совокупность представляет собой именно совокупность однородных элементов — элементов одного рода. Это ключ к толкованию того, какую совокупность элементов можно отнести к генеральной, а какую — нельзя.

Четких определений понятия «однородность» в научной литературе не встречается. Говорят об однородной функции, об упруго-однородном теле и т.п. Поэтому, отмечая сложность попытки дать однозначное толкование термину «однородная выборка», некоторые ученые вводят собственные определения, удобные для практического применения. Например: «однородной называется такая совокупность, элементы которой формируются под воздействием общих основных причин и условий, а их законы распределения имеют простую структуру»[1]. Так, например, если динамика некоторого показателя x описывается моделью постоянного прироста со случайными отклонениями

$$x_t = x_{t-1} + a + e_t, \tag{3.1}$$

где a — постоянный прирост; e_t — не зависящие друг от друга случайные отклонения, имеющие нормальный закон распределения с нулевым математическим ожиданием и дисперсией, не зависящей или зависящей от t, то все элементы совокупности x_1, x_2, ..., x_T имеют различные законы распределения, но сама совокупность в соответствии с указанным определением будет однородной.

Сложность использования такого «рабочего» определения однородности совокупности заключается в том, что понятия «общие причины и условия» недостаточно конкретны. Например, рассматривая динамику развития промышленности России на протяжении 30 лет, можно говорить, что основной причиной формирования этого динамического ряда является спрос на товары промышленности, а условием — работа поставщиков. На основе этого, воспользовав-

---

[1] *Смоляк С. А., Титаренко Б. П.* Устойчивые методы оценивания (статистическая обработка неоднородных совокупностей). М. : Статистика, 1980. С. 14.

шись приведенным выше определением, кажется, что можно говорить об однородности совокупности. Но учитывая, что сама причина — спрос на продукцию — непрерывно меняется и качественно, и количественно, да и структура промышленного производства претерпела кардинальные изменения, то необходимо будет признать, что совокупность все же является неоднородной[1]. Принципиально важной характеристикой однородности процесса является ее инвариантность времени или порядку наблюдения. Поэтому *совокупность элементов будет являться однородной только в том случае, когда ее элементы y_t, $t = 1, 2, 3, ..., m, ..., T$ формируются под воздействием общих основных неизменных причин и условий x_t так, чтобы при возникновении для $t = k$ условий и причин, равных $x_k = x_m$, элемент y_k был равен*

$$y_k = y_m + \varepsilon_k, \tag{3.2}$$

где ε_k — не зависящие друг от друга случайные отклонения, имеющие нормальный закон распределения с нулевым математическим ожиданием и конечной дисперсией.

Используя наше определение для случая (3.1), можно убедиться в том, что все элементы совокупности x_1, x_2, ..., однородны. Действительно, при достижении $x_t = x_\tau$, величина x будет определяться как

$$x_\tau = x_{\tau-1} + a + e\tau. \tag{3.3}$$

Теперь можно дать определение понятия генеральной совокупности, которое мы будем использовать в дальнейшем в качестве основного: *генеральная совокупность — это множество всех возможных элементов однородной совокупности.* Из этого определения с учетом введенного понятия однородности следует, что каждый элемент генеральной совокупности обладает характеристиками генеральной совокупности, но в силу воздействия на него множества случайных факторов эти характеристики в каждом элементе наблюдаются с некоторым отклонением.

Теперь можно ответить на проверочный вопрос, сформулированный ранее: социально-экономические показатели за всю многовековую историю существования Руси, а затем и России представляют собой все мысленно возможные состояния; это

---

[1] *Светуньков С. Г., Литвинов А. А.* Конкуренция и предпринимательские решения. Ульяновск : Изд-во Корпорации технологий продвижения, 2000. С. 186.

генеральная совокупность или нет? Если предложить российской промышленности, например, выпустить продукцию в количестве и качестве, соответствующих 1960-му г., то, например, количество потребленной при этом электроэнергии будет значительно меньше электропотребления, соответствовавшего 1960-му г., поскольку технологии промышленного производства существенно изменились. Если в 1960 г. на промышленных предприятиях наряду с электрическим приводом машин встречались паровые, газовые и гидравлические приводы, то к настоящему времени в подавляющем большинстве используются электрические приводы, да еще и с программным управлением. Но, что еще более важно, сегодня промышленность России просто не в состоянии выпустить ту номенклатуру изделий, которая была в 1960 г. — разрушены оснастка и технологические линии, уничтожена проектная документация, потеряны навыки производства продукции, которая давно не выпускается. Если по каким-то причинам будет все же необходимо в точности повторить эту номенклатуру и количество, то условия производства, затраты сырья, труда, денежных средств и т.п. будут существенно отличаться от имевших место в 1960 г.

Это значит, что социально-экономические показатели Руси и России в разные периоды времени отражали разные условия неоднородной динамики этой системы. Поэтому их ни в коем случае нельзя назвать генеральной совокупностью! Множество наблюдений социально-экономического развития России за многие века, пусть даже упорядоченное во времени, остается только множеством и никогда не превратится в генеральную совокупность.

Теперь следует определить понятие «**выборочной совокупности**». Понятно, что оно связано с генеральной совокупностью, и выборочная совокупность представляет собой ее часть. Обратимся к существующим определениям этого понятия.

«Если ограничиться числом наблюдений, которое в той или иной степени меньше общего числа элементов исследуемой совокупности, то будем говорить, что используется выборка... В широком смысле выборка — это любая часть всех элементов совокупности. Обычно эта часть извлекается таким образом, чтобы она «представляла» совокупность»[1].

---

[1] *Джессен Р.* Методы статистических обоснований. М. : Финансы и статистика, 1985. С. 21–22.

Из этого определения не совсем ясно, о какой совокупности идет речь — о генеральной или вообще о любом множестве каких-то элементов, представленном в виде некоторой совокупности. Но в нем есть одна важная характеристика, которая обязательно встречается при констатации основных характеристик выборочной совокупности — она должна отражать свойства генеральной совокупности. Если выборка не будет обладать таким свойством, то ее исследование бессмысленно, и называть ее выборочной нельзя.

«Случайная выборка есть совокупность единиц (выборочная совокупность), отобранных из генеральной совокупности случайно, т.е. таким образом, что все единицы генеральной совокупности имеют равные шансы быть отобранными (попасть в выборку)»[1].

В этом определении указывается на наличие источника выборки (генеральная совокупность), способ получения выборки (случайная выборка), но нет указания на то, что выборка должна представлять основные характеристики генеральной совокупности.

Этот же недостаток содержится в другом определении: «Выборкой размера n из распределения F называется случайный вектор $\{\xi_1, ..., \xi_n\}$, компоненты которого независимы и одинаково распределены с функцией распределения $F(x)$»[2]. Впрочем, необходимо отметить, что здесь появилась еще одна сущностная характеристика выборки — независимость элементов выборки.

«Выборка из данной генеральной совокупности — это результаты ограниченного ряда наблюдений $x_1, x_2, ..., x_n$ случайной величины ξ... Если ... ряд наблюдений $x_1, x_2, ..., x_n$ образует последовательность независимых и одинаково распределенных случайных величин, то выборка называется случайной»[3].

Итак, можно утверждать, что выборочная совокупность обладает следующими характеристиками, отличающими ее от других подмножеств:

1) элементы выборочной совокупности формируются из генеральной совокупности;

---

[1] *Головач А. В., Ерина А. М., Трофимов В. П.* Критерии математической статистики в экономических исследованиях. М. : Статистика, 1973. С. 5.

[2] *Лагутин М. В.* Наглядная математическая статистика : учебное пособие. М. : БИНОМ, Лаборатория знаний, 2009. С. 42.

[3] *Айвазян С. А., Мхитарян В. С.* Прикладная статистика и основы эконометрики : учебник для вузов. М. : ЮНИТИ, 1998. С. 197.

2) они формируются случайным образом — так, чтобы каждый элемент генеральной совокупности имел одинаковую вероятность попасть в выборку;

3) способ формирования выборки должен обеспечивать независимость элементов выборки друг от друга;

4) элементы выборки должны давать представление о свойствах генеральной совокупности.

Зная эти свойства случайной выборки, дадим ее определение, которое будем использовать в дальнейшем.

Выборочная совокупность — это случайно выбранные элементы генеральной совокупности, дающие представление о ней и выбранные так, что каждый элемент генеральной совокупности имеет одинаковую вероятность попасть в выборку не зависимо от других элементов.

Определив понятия генеральной и выборочной совокупностей, можно понять суть **выборочного метода**, который является основой математической статистики и эконометрии, а также одним из основных методов реализации индуктивного вывода, когда по изученным свойствам нескольких элементов всего множества делается вывод о свойствах самого множества в целом. Следует напомнить, что метод индуктивного вывода пришел на смену методу дедуктивного вывода, который более 1000 лет оставался единственным методом обобщения научных результатов и получения достоверных знаний. Именно применение индуктивного метода, как метода на порядок более дешевого и требующего значительно меньшего времени, послужило бурному росту науки с середины II тыс. н.э. Выборочный метод является математическим обоснованием этого индуктивного вывода, дает возможность математически корректно сформулировать его, и, что немаловажно, оценить вероятность его правильности.

Процесс выборочного метода заключается в следующем. Пусть имеется многочисленная генеральная совокупность однородных элементов, каждый из которых обладает каким-либо признаком. Из этой совокупности выбирается наугад один из элементов, причем при выборе элемента из совокупности принимаются все меры к тому, чтобы вероятность быть выбранным была одинаковой для всех элементов. Затем таким же образом выбирается другой элемент и т.д. до тех пор, пока сформированная выборка не будет удовлетворять некоторым заранее сформулированным требованиям исследователя (размер выборки, стоимость получения выборки и т.п.).

Полученные выборочные значения обрабатываются с помощью методов математической статистики на предмет выявления основных закономерностей генеральной совокупности и оценки ее основных характеристик. Чаще всего это выявление вида и формы некоторого закона изменения элементов генеральной совокупности, оценка параметров этого закона и достоверности полученных значений.

Если процессы, в ходе которых формируется совокупность элементов, не являются стационарными, то это означает, что информация о них, которую несут в себе фиксируемые в такие моменты времени наблюдения, не дает знания о совокупности в целом — они отражают лишь поведение объекта в этот конкретно-исторический промежуток времени, поведение объекта до этого промежутка и после него происходило в других условиях с другими характеристиками, которые, чаще всего, не вытекают друг из друга. В этом случае использовать выборочный метод нельзя. В лучшем случае при этом будут получены некоторые характеристики текущего момента, но знания о прошлом и будущем из них не вытекают.

Как определить, что имеющийся в распоряжении прогнозиста ряд статистических значений о некотором социально-экономическом объекте является выборочной совокупностью в статистическом смысле? Это сделать довольно просто. Надо только понять: безграничное расширение числа этих наблюдений позволит уточнить характеристики изучаемого объекта или, наоборот, помешают? Если увеличение числа наблюдений действительно улучшит понимание изучаемого процесса и позволит построить более точные модели, то перед нами выборочная совокупность, для обработки которой необходимо использовать методы математической статистики. Любое увеличение этой выборки из числа наблюдений, относимых к генеральной совокупности, только улучшит понимание характеристик генеральной совокупности.

Например, если прогнозист строит модель зависимости производительности труда швей-мотористок пятого разряда от уровня оплаты их труда, то новые дополнительные данные позволят улучшить результаты его работы по выявлению такой зависимости и построению соответствующей модели этой работы.

Если же включение в имеющуюся совокупность новых элементов только ухудшит свойства модели, то очевидно, что прогнозист имеет дело не с выборочной совокупностью. Например, если прогнозист собирается построить модель производственной функции Кобба—Дугласа для промышленности современной России и сделать с ее помощью прогнозы о том, как в ближайшие годы будут меняться объемы промышленного производства, то для этого ему нужны данные, в лучшем случае, с 1999 г. по настоящее время. Включение в имеющуюся совокупность данных значения прошлых лет (1990, 1980, 1960 и т.п.) бессмысленно — тогда была другая социально-экономическая система и эти данные, которые на первый взгляд расширяют статистическую базу, при их использовании для построения прогнозной модели производственной функции на самом деле только ухудшат ее характеристики.

Поскольку в данной главе изучаются методы прогнозирования однородных стационарных процессов, мы будем иметь дело исключительно с выборочными совокупностями, характеристики которых мы определили выше.

3.2. Средние величины в прогнозировании однородных стационарных процессов

Важнейшими из показателей, характеризующих генеральные совокупности, являются средние величины, вычисляемые на основе выборочных значений. Довольно часто при прогнозировании однородных стационарных социально-экономических процессов средняя величина отражает типичные черты процесса и может выступать как прогнозное значение изучаемого показателя.

Существует несколько типов средних величин, но при прогнозировании в экономике используются только две средние — средняя арифметическая и средняя геометрическая.

Для всех средних величин справедливо следующее свойство: средняя всегда меньше максимального значения ряда и всегда больше минимального значения ряда за исключением ситуации, когда все переменные равны друг другу.

В числе средних величин, используемых в прогнозировании, на первом месте стоит *средняя арифметическая*, представляющая собой частное от деления суммы значений показателя на число элементов выборочной совокупности:

$$\bar{y} = \frac{y_1 + y_2 + \ldots + y_T}{T} = \frac{1}{T}\sum_{t=1}^{T} y_t. \qquad (3.4)$$

Далее в некоторых случаях мы будем опускать пределы суммирования, понимая, что оно осуществляется по t в пределах от 1 до T.

Средняя арифметическая обладает рядом математических свойств, которые активно используют в математической статистике.

1. Средняя арифметическая суммы двух и более величин равна сумме средних арифметических этих двух и более величин:

$$\overline{y} = \frac{1}{T}\sum_{t=1}^{T} y_t = \frac{1}{T}\sum_{t=1}^{T}\left(x_t + z_t\right) = \overline{x} + \overline{z}. \tag{3.5}$$

2. При наличии в переменной постоянного слагаемого его можно вынести за пределы суммирования:

$$\overline{y} = \frac{1}{T}\sum_{t=1}^{T}\left(a + x_t\right) = a + \overline{x}. \tag{3.6}$$

3. Постоянный множитель или делитель можно вынести за знак средней:

$$\overline{y} = \frac{1}{T}\sum_{t=1}^{T}\left(ax_t\right) = a\overline{x}. \tag{3.7}$$

4. Сумма отклонений переменной от ее средней арифметической равна нулю:

$$\sum_{t=1}^{T}\left(y_t - \overline{y}\right) = \sum_{t=1}^{T} y_t - \sum_{t=1}^{T}\overline{y} = \sum_{t=1}^{T} y_t - T\frac{1}{T}\sum_{t=1}^{T} y_t = 0. \tag{3.8}$$

5. Сумма квадратов отклонений от средней арифметической всегда не меньше, чем сумма квадратов отклонений этого же показателя от любого другого числа:

$$\sum_{t=1}^{T}\left(y_t - \overline{y}\right)^2 \le \sum_{t=1}^{T}\left(y_t - a\right)^2. \tag{3.9}$$

Действительно:

$$\sum_{t=1}^{T}(y_t - a)^2 = \sum_{t=1}^{T}\left(y_t - \overline{y} - (a - \overline{y})\right)^2 =$$
$$= \sum_{t=1}^{T}(y_t - \overline{y})^2 + \sum_{t=1}^{T}(a - \overline{y})^2 - 2\sum_{t=1}^{T}(y_t - \overline{y})(a - \overline{y}),$$

раскрывая последнюю сумму, получим:

$$2\sum_{t=1}^{T}(y_t-\overline{y})(a-\overline{y})=2(a-\overline{y})\sum_{t=1}^{T}(y_t-\overline{y}),$$

а так как сумма отклонений в соответствии с (3.8) равна нулю, то и полученная сумма будет равна нулю. Следовательно:

$$\sum_{t=1}^{T}(y_t-a)^2=\sum_{t=1}^{T}(y_t-\overline{y})^2+\sum_{t=1}^{T}(a-\overline{y})^2.$$

Эти математические свойства средней арифметической используются в математической статистике при выводе других обобщающих характеристик совокупности изучаемых элементов.

Значительно реже и на практике, и в теории используется *средняя геометрическая*, которая представляет собой корень T-й степени из произведения n переменных:

$$\tilde{y}=\sqrt[T]{y_1y_2...y_T}. \tag{3.10}$$

Ее смысл легко понять, если рассмотреть вначале случай усреднения двух величин. Их произведение, которое находится в подкоренном выражении, представляет собой площадь прямоугольника со сторонами, соответствующими величине каждого из элементов. Вычисляя квадратный корень из этого произведения, мы тем самым определяем величину стороны квадрата, площадь которого равна площади прямоугольника. Теперь, если рассмотреть случай с тремя величинами, то средняя геометрическая позволяет вычислить сторону равностороннего куба и т.п. Поскольку эта средняя, как видно, имеет геометрическое содержание, она и получила такое название.

Средняя геометрическая также *обладает рядом* интересных *свойств*.

1. Если хотя бы одна из переменных равна нулю, средняя геометрическая также равна нулю. Для средней арифметической это свойство не выполняется.

2. Общий множитель переменных можно вынести за знак средней:

$$\tilde{y}=\sqrt[T]{(ay_1)(ay_2)...(ay_T)}=a\sqrt[T]{y_1y_2...y_T}. \tag{3.11}$$

3. Средняя геометрическая всегда не больше среднего арифметического тех же чисел:

$$\tilde{y} \le \bar{y}. \tag{3.12}$$

Это неравенство превращается в равенство только тогда, когда переменные являются константами, равными друг другу.

Существенным *отличием* средней геометрической от средней арифметической является то, что средняя арифметическая выступает как самостоятельная оценка изучаемой совокупности, а средняя геометрическая — только как одна из характеристик совокупности.

Действительно, для того, чтобы иметь представление о генеральной совокупности, необходимо знать такой «центр группирования» случайной величины, вокруг которого и находится основная часть всех наблюдений. Этот «центр группирования» называется, как известно, математическим ожиданием случайной величины. Математическое ожидание дискретной случайной величины Y при конечном числе ее возможных значений есть сумма произведений каждого из этих значений на его вероятности, а потому лучшей оценкой математического ожидания на выборке в большинстве случаев как раз и выступает средняя арифметическая (3.4).

Другой важной характеристикой генеральной совокупности является дисперсия дискретной случайной величины, которая представляет собой сумму квадратов отклонения каждого значения случайной величины от ее математического ожидания, умноженную на вероятность этого значения.

Переходя к выборочным значениям (подразумевая для начала, что мы имеем дело с большим числом наблюдений), получим, опираясь на это определение, такую формулу для вычисления дисперсии:

$$D(y) = M\left(y_t - \bar{y}\right)^2 = \frac{1}{T}\sum_{t=1}^{T}\left(y_t - \bar{y}\right)^2. \tag{3.13}$$

Для моделирования случайных величин широко используются моменты, введенные в научный оборот П. Л. Чебышевым.

Моментом r-го порядка случайной величины называется математическое ожидание величины $(y - a)^r$, где a — любое вещественное число, а r — вещественное целое неотрицательное число. Применительно к выборочным значениям момент r-го порядка запишется так:

$$v_r = \frac{1}{T}\sum_{t=1}^{T}\left(y_t - a\right)^r. \qquad (3.14)$$

Если $a = M(y)$, то момент называется *центральным*. А поскольку для выборочных значений средняя арифметическая зачастую выступает лучшей оценкой математического ожидания, то выборочные значения центральных моментов можно вычислить так:

$$\mu_{y,r} = \frac{1}{T}\sum_{t=1}^{T}\left(y_t - \overline{y}\right)^r. \qquad (3.15)$$

Легко заметить, что дисперсия представляет собой центральный момент второго порядка.

Как видно из вышенаписанного, средняя арифметическая является важной величиной в математической статистике, а поскольку прогнозирование социально-экономических процессов использует в своей основе аппарат математической статистики, без средней арифметической представить себе методы и модели прогнозирования в полной мере невозможно.

Чаще всего в практике прогнозирования стационарных однородных процессов приходится встречаться с нормальным законом распределения вероятностей. В этом случае роль средней арифметической является центральной.

Действительно, если попытаться найти такую оценку случайной дискретной величины, для которой выполняется требование сходимости по вероятности оценки к оцениваемому параметру при неограниченном возрастании объема наблюдения, то такой оценкой будет выступать именно средняя арифметическая (3.4). Это свойство полученных оценок называется *состоятельностью*.

Второе требование, предъявляемое к статистической оценке, — отсутствие в ней систематической погрешности. Это требование соответствует *несмещенной* оценке. Средняя арифметическая в рассматриваемом случае удовлетворяет именно этому требованию.

Помимо состоятельности и несмещенности средняя арифметическая характеризуется еще одним важным свойством. Дисперсии случайной величины относительно ее возможных оценок принимают разные значения. Та оценка, для которой указанная дисперсия является минимальной, называется *эффективной*. Дисперсия случайной нормально распределенной величины относительно ее средней арифметической как раз и обладает этим свойством.

Итак, средняя арифметическая при нормальном распределении обладает важными свойствами состоятельности, несмещенности и эффективности, а, следовательно, является лучшей оценкой генеральной совокупности в случае нормального распределения случайных величин и может выступать как прогнозная величина.

В то же время следует помнить о том, что средняя арифметическая представляет собой лишь одну из оценок математического ожидания генеральной совокупности. Если в распоряжении исследователя появится новое значение из выборки, то средняя арифметическая будет уточнена и изменит свое значение. Поэтому использовать просто величину средней арифметической как прогнозную величину показателя будет неверным — следует определить те границы, в которых будет находиться математическое ожидание. Иначе говоря, средняя арифметическая в рассматриваемом случае будет наилучшей прогнозной оценкой математического ожидания, но прогноз будет заключаться еще и в том, чтобы определить те доверительные границы, в которых может находиться прогнозная величина с учетом воздействия на нее множества случайных факторов. Для этого необходимо использовать выборочное значение дисперсии. Формула для его вычисления (3.13) не является лучшей оценкой дисперсии генеральной совокупности, поскольку она является смещенной. Дисперсия, как известно, представляет собой меру отклонений случайной величины от его математического ожидания. Применительно к дискретному случаю (а именно его мы рассматриваем в нашей дисциплине), генеральная дисперсия может быть записана так:

$$\sigma_y^2 = M\left(y - \mu_y\right)^2, \tag{3.16}$$

где μ_y — математическое ожидание случайной величины в генеральной совокупности.

Для того чтобы вычислить дисперсию, необходимо иметь все значения случайной переменной y и точно знать величину ее математического ожидания μ_y. На практике эти значения неизвестны — известны лишь некоторые выборочные значения y, которые обычно считаются независимыми, одинаково распределенными, и их средняя арифметическая равна $\overline{y}$. Если дисперсию рассчитать по выборке, используя вместо неизвестного математического ожидания генеральной совокупности среднюю величину (по формуле (3.13)), то мы получим смещенную оценку дисперсии. Для того чтобы показать это, оценим математическое ожидание такой дисперсии:

$$M\left(D(y)\right)=M\left(\frac{1}{T}\sum_t\left(y_t-\overline{y}\right)^2\right)=\frac{1}{T}\sum_t M\left(\left(y_t-\mu_y\right)-\left(\overline{y}-\mu_y\right)\right)^2.$$

Сумма квадратов в правой части данного равенства раскладывается на следующие составляющие:

$$M\left(D(y)\right)=\frac{1}{T}\sum_t\left(M\left(y_t-\mu_y\right)^2+M\left(\overline{y}-\mu_y\right)^2-2M\left(y_t-\mu_y\right)\left(\overline{y}-\mu_y\right)\right),$$

откуда следует, что

$$M\left(D(y)\right)=\frac{1}{T}\sum_t\left(\sigma_y^2+D(\overline{y})-2\operatorname{cov}\left(y_t-\mu_y;\overline{y}-\mu_y\right)\right). \qquad (3.17)$$

где σ_y^2 — генеральная дисперсия; $D(\overline{y})$ — дисперсия средней арифметической относительно генерального математического ожидания.

Дисперсия случайных величин относительно их средней арифметической вычисляется легко, а что представляет собой $\sigma(\overline{y})^2$ — дисперсия средней арифметической относительно математического ожидания? Это квадрат отклонения средней арифметической от математического ожидания случайной величины. Для нормального распределения вероятностей средняя арифметическая является лучшей оценкой математического ожидания, причем, чем большее число наблюдений включается в расчет средней арифметической, тем ближе находится средняя арифметическая к математическому ожиданию. Из этого следует, что с ростом числа наблюдений n дисперсия средней арифметической относительно математического ожидания стремится к нулю. Чтобы понять, как это происходит, вычислим дисперсию средней арифметической n одинаково распределенных взаимно независимых случайных величин:

$$D(\overline{y}) = M\left(\overline{y} - \mu_y\right)^2 = M\left(\frac{\sum_t y_t}{T} - \frac{T\mu_y}{T}\right)^2 = \frac{1}{T^2} M\left(\sum_t \left(y_t - \mu_y\right)\right)^2.$$

Математическое ожидание суммы дисперсий в данном случае будет равно сумме математических ожиданий дисперсий, так как μ не случайно. Следовательно, дисперсию средней величины можно записать в виде:

$$D(\overline{y}) = \frac{1}{T^2} \sum_t M\left(y_t - \mu_y\right)^2 = \frac{1}{T^2} \sum_t \sigma_y^2 = \frac{T}{T^2} \sigma_y^2 = \frac{1}{T} \sigma_y^2. \qquad (3.18)$$

Таким образом, дисперсия средней арифметической относительно математического ожидания в n раз меньше дисперсии всей генеральной совокупности.

Теперь оценим ковариацию в (3.17):

$$\operatorname{cov}\left(y_t - \mu_y; \overline{y} - \mu_y\right) = \operatorname{cov}\left(y_t - \mu_y; \frac{1}{T} \sum_j y_j - \frac{1}{T} T\mu_y\right) =$$

$$= \frac{1}{T} \operatorname{cov}\left(y_t - \mu_y; \sum_j \left(y_j - \mu_y\right)\right).$$

Ковариация между случайной величиной и суммой случайных величин равна сумме ковариаций между соответствующими случайными величинами. Однако, учитывая то, что значения y независимы друг от друга, все ковариации кроме одной будут равны нулю:

$$\operatorname{cov}\left(y_t - \mu_y; \overline{y} - \mu_y\right) = \frac{1}{T} \operatorname{cov}\left(y_t - \mu_y; y_t - \mu_y\right) = \frac{1}{T} \sigma_y^2. \qquad (3.19)$$

Из равенств (3.17), (3.18) и (3.19) получим:

$$M\left(D(y)\right) = \frac{1}{T} \sum_t \left(\sigma_y^2 + \frac{1}{T}\sigma_y^2 - 2\frac{1}{T}\sigma_y^2\right) = \frac{1}{T} \sum_t \left(\sigma_y^2 - \frac{1}{T}\sigma_y^2\right) = \frac{T-1}{T}\sigma_y^2. \qquad (3.20)$$

Как видим, дисперсия, рассчитанная по формуле (3.13), будет смещена относительно дисперсии в генеральной совокупности. Только на больших выборках (с ростом T) эта дисперсия будет приближаться к истинной оценке. Чтобы избежать этого смещения, рассчитывают исправленную дисперсию по формуле:

$$s_y^2 = \frac{T}{T-1} D(y) = \frac{1}{T-1} \sum_{t=1}^{T} (y_t - \overline{y})^2. \qquad (3.21)$$

Теперь у прогнозиста имеются все величины для осуществления прогноза случайного стационарного процесса, протекающего в условиях однородности. Прежде всего, он оценивает саму прогнозируемую величину по средней арифметической, после чего определяет величину дисперсии для данной выборки. Средняя арифметическая характеризует наиболее вероятное значение прогнозной величины, а дисперсия — меру колеблемости относительно этой величины.

Пример

В табл. 3.1 приведены данные об изменении цены за 1 кг картофеля в некоторых магазинах и ларьках Выборгского района Санкт-Петербурга. Необходимо спрогнозировать стоимость килограмма картофеля в ближайшем магазине или ларьке, расположенном ближе всего к дому проживания авторов учебника.

Таблица 3.1

Данные о выборочных значениях цены за 1 кг картофеля

Номер наблюдения	1	2	3	4	5	6	7	8	9	10	11
Цена за килограмм картофеля, руб/кг	13,4	14,1	15,0	12,5	12,5	11,0	13,4	13,1	11,9	13,0	12,0

Поскольку данные значения были собраны в короткий промежуток времени, они представляют собой реализацию случайного стационарного процесса, протекающего в условиях однородности. Поэтому лучшей прогнозной оценкой стоимости 1 кг картофеля в Выборгском районе Санкт-Петербурга будет являться средняя арифметическая. Вычисляя ее по формуле (3.4), получим:

$$\overline{y} = \frac{13{,}4 + 14{,}1 + 15{,}0 + \ldots + 12{,}0}{11} = 12{,}9.$$

Следовательно, при принятии решения о покупке картофеля нам следует ориентироваться на его цену в 12,9 руб/кг. В то же время это значение является выборочным значением оценки математического ожидания цены, поэтому оно найдено с некоторым случайным отклонением от этой величины. Мерой этого отклонения служит дисперсия, которая в силу малой выборки должна оцениваться по формуле (3.21). В рассматриваемом случае эта дисперсия будет равна:

$$s_y^2 = \frac{1}{11-1}\left[(13{,}4-12{,}9)^2 + (14{,}1-12{,}9)^2 + \ldots + (12{,}0-12{,}9)^2\right] = 1{,}214.$$

В этом простом случае для оценки доверительных границ полученного прогнозного значения можно воспользоваться правилом трех сигм, в соответствии с которым интервал прогнозируемой величины стоимости килограмма картофеля будет определяться так:

$$\bar{y} - 3s_y < y < \bar{y} + 3s_y.$$

Поскольку средняя арифметическая равна 12,9 руб/кг, а среднеквадратическое отклонение, как квадратный корень из дисперсии, будет равно 1,1, то прогнозируемая величина с высокой вероятностью будет лежать в пределах:

$$12{,}9 - 3{,}3 < y < 12{,}9 + 3{,}3;\ 9{,}6 < y < 16{,}1.$$

Здесь, однако, надо учесть следующее обстоятельство. Мы имеем дело со случайной величиной, поэтому она с той или иной степенью вероятности может отклоняться от своего математического ожидания. Эти вероятности можно учесть при прогнозировании. При этом следует ставить вопрос следующим образом: в каких пределах при заданной вероятности p может находиться прогнозная величина, если за основу прогноза взять ее среднюю арифметическую и знать при этом величину дисперсии для выборочных значений? Ответ на этот вопрос получен в теории вероятностей и математической статистике. Рассмотрим его ниже.

3.3. Общие принципы определения доверительных границ для выборочных значений из генеральной совокупности

Из всего многообразия возможных проявлений случайности, в экономической практике она чаще всего проявляется в форме нормального закона распределения вероятностей. Именно это положение и лежит в основе рассматриваемого мой процедуры определения доверительных границ выборочных оценок. Прежде всего, необходимо вспомнить, что функция распределения вероятностей случайной величины $F(y)$ определяет вероятность того, что случайная величина y_t при испытании примет значение, меньшее произвольно изменяемого действительного числа y ($-\infty < y < +\infty$):

$$F(y) = P(y_t < y). \tag{3.22}$$

Эта функция положительна и меньше единицы. Для непрерывно возрастающего y график этой функции является возрастающим. На рис. 3.1 построен график функции нормального распределения вероятностей. Если взять на оси y этого графика две произвольные точки y_0 и $y_0 + \Delta y$, то их ординатами соответственно будут $F(y_0)$ и $F(y_0 + \Delta y)$. Для приращения функции распределения вероятностей на этом участке имеем:

$$F(y_0 + \Delta y) - F(y_0) = P(y_0 < y_t < y_0 + \Delta y). \tag{3.23}$$

Первая производная функции распределения вероятностей получила название плотности вероятности и имеет вид

$$\phi(y) = \lim_{\Delta y \to 0} \frac{F(y_0 + \Delta y) - F(y_0)}{\Delta y} = \lim_{\Delta y \to 0} \frac{P(y_0 < y_t < y_0 + \Delta y)}{\Delta y}, \tag{3.24}$$

из чего следует, что плотность вероятности представляет собой предел отношения вероятности того, что случайная величина y_t примет значение, лежащее в границах от y_0 до $y_0 + \Delta y$, к величине интервала Δy, когда этот интервал стремится к нулю.

Функция распределения вероятностей является первообразной функцией по отношению к функции плотности вероятности, поэтому вероятность (3.23) того, что случайная величина Y примет значение, лежащее в границах от y_0 до $y_0 + \Delta y$, может быть найдена так:

$$P(y_0 < y_t < y_0 + \Delta y) = F(y_0 + \Delta y) - F(y_0) = \int_{y_0}^{y_0 + \Delta y} \phi(y)dy. \tag{3.25}$$

На графике плотности вероятности полученная вероятность (3.25) будет представлять собой площадь криволиней-

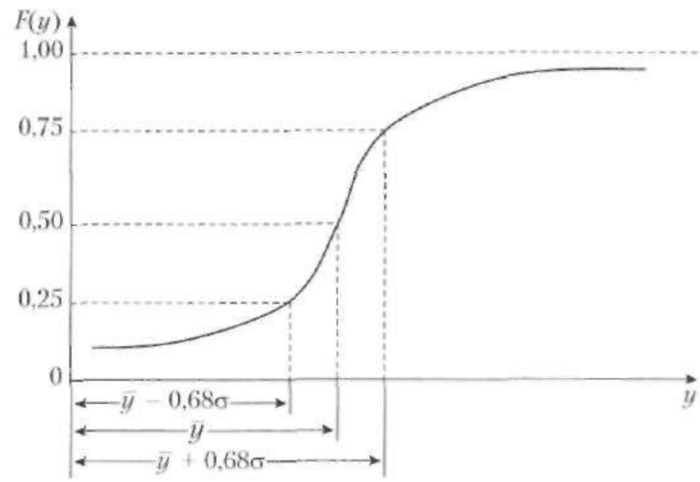

Рис. 3.1. **График функции нормального распределения**

ной трапеции с основанием от y_0 до $y_0 + \Delta y$, ограниченной сверху кривой плотности вероятности (рис. 3.2).

Иногда функцию распределения вероятностей $F(y)$ называют «интегральной функцией распределения», а функцию плотности вероятности $\phi(y)$ — «дифференциальной кривой распределения», исходя из их математического смысла.

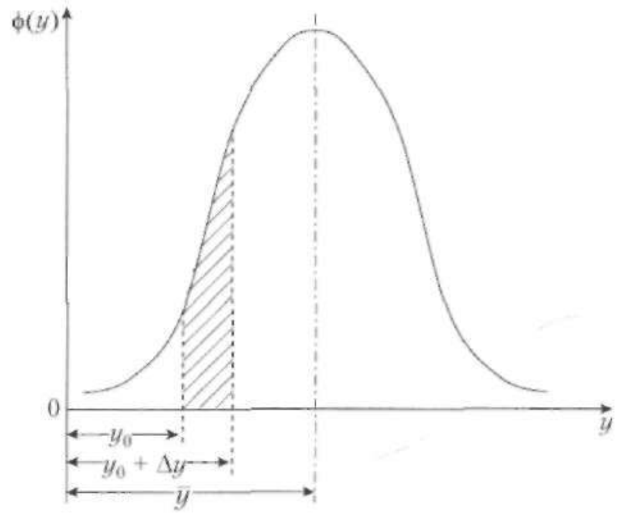

Рис. 3.2. **График функции плотности нормального распределения вероятностей**

В математике среди элементарных функций известна функция Гаусса, которая имеет вид

$$e^{-y^2}. \tag{3.26}$$

Эта функция симметрична относительно нулевого значения y, всегда положительна, а кроме того — принимает свое максимальное значение, равное единице, в том случае, когда $y = 0$. По своему виду она как нельзя лучше подходит для описания графика плотности вероятности нормального распределения (рис. 3.2), что и дало возможность Гауссу предложить функцию, аппроксимирующую нормальный закон распределения вероятностей и носящую его имя:

$$\phi(y) = \frac{1}{\sigma_y\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2}. \tag{3.27}$$

Часто для того, чтобы показать, что случайная величина распределена нормально, в соответствии с формулой (3.27) прибегают к записи вида:

$$y \sim N\left(\mu_y; \sigma_y^2\right).$$

Характер функции (3.27) определяется двумя характеристиками — дисперсией $\sigma_y{}^2$ и математическим ожиданием μ_y. Увеличение математического ожидания μ_y приводит к сдвигу кривой вправо вдоль оси $0y$, а ее уменьшение — к сдвигу влево. С возрастанием дисперсии максимальная ордината нормальной кривой убывает, а сама кривая становится более пологой.

Так как, варьируя эти параметры, можно получить любое семейство кривых, то следует взять за основу функцию плотности вероятности при каких-то фиксированных стандартных значениях, а затем — подставлять в нее эти две характеристики. Именно так и поступил в свое время Лаплас. Для этого он принял $\mu_y = 0$, т.е. получил ситуацию, когда график рис. 3.2 симметричен относительно нулевого значения на оси y, и приравнял дисперсии распределения единице. Так была получена нормированная кривая плотности распределения, или «кривая плотности стандартного нормального распределения»:

$$\varphi(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}. \tag{3.28}$$

Случайная величина, распределенная в соответствии со стандартным нормальным законом распределения, обозначается соответственно: $y \sim N(0,1)$.

К этому виду можно привести любой ряд y, для чего следует от каждого значения ряда отнять его математическое ожидание μ_y, а затем полученные значения разделить на среднее квадратичное отклонение σ_y:

$$z = \frac{y - \mu_y}{\sigma_y}. \tag{3.29}$$

Такой ряд будет называться стандартизированным.

Больший интерес, чем функция (3.28), представляет ее первообразная, которая характеризует вероятность того, что y_t лежит в интервале от нуля до некоторого значения Z:

$$\Phi(y) = \frac{1}{\sqrt{2\pi}} \int_0^Z e^{-\frac{1}{2}z^2} dz. \tag{3.30}$$

Эта функция получила название нормированной функции Лапласа. Подставляя в нее различные значения y, можно получить разные значения вероятностей. Эта работа и была в свое время выполнена, а расчетные значения вероятности сведены в соответствующие таблицы, которые можно встретить в любом учебнике по теории вероятностей и математической статистике. Очевидно, что для условия $y = 0$, т.е. для ситуации, когда выборочное значение y_t точно соответствует математическому ожиданию процесса, вероятность будет равна нулю. А вот уже вероятность того, что y_t лежит в интервале от нуля до значения $Z = 0{,}01$, не равна нулю. Подставляя это значение в функцию (3.30) или просто заглянув в соответствующую строку таблицы, получим вероятность, равную 0,0040. А вот вероятность того, что для нормированной величины ее значение окажется в интервале от нуля до значения $Z = 5{,}00$, равна 0,4999997, т.е. очень высока.

Таким образом, можно увидеть, что имеется возможность оценить то, с какой вероятностью выборочное значение попадет в заданный интервал от нуля до Z.

На практике можно каждый имеющийся ряд y отцентрировать относительно его математического ожидания μ_y, а затем полученные значения разделить на среднее квадратичное отклонение σ_y. Но значительно удобнее, воспользовавшись функцией Лапласа, привести формулу (3.25) к такому виду, чтобы при нормальном распределении можно было сразу определить вероятность того, что случайная величина y_t примет значение, лежащее в границах от y_0 до $y_0 + \Delta y$. Для этого воспользуемся очевидным равенством:

$$P(y_0 < y_t < y_0 + \Delta y) = \int_{y_0}^{y_0+\Delta y} \phi(y)dy = \int_{0}^{y_0+\Delta y} \phi(y)dy - \int_{0}^{y_0} \phi(y)dy. \quad (3.31)$$

Для того чтобы применить функцию Лапласа, определим из (3.29) исходную переменную y: $y = \sigma_y z + \mu_y$, а $dy = \sigma_y dz$. Теперь можно найти новые пределы интегрирования: если $y = y_0 + \Delta y$, то в соответствии с (3.29) $z = (y_0 + \Delta y - \mu_y) / \sigma_y$, а если $y = y_0$, то $z = (y_0 - \mu_y) / \sigma_y$. Теперь, подставляя в (3.31) функцию Лапласа с этими пределами интегрирования, получим

$$P(y_0 < y_t < y_0 + \Delta y) = \frac{1}{\sqrt{2\pi}} \int_{0}^{y_0+\Delta y-\mu_y} e^{-\frac{z^2}{2}} dz - \frac{1}{\sqrt{2\pi}} \int_{0}^{y_0-\mu_y} e^{-\frac{z^2}{2}} dz =$$
$$= \Phi\left(\frac{y_0 + \Delta y - \mu_y}{\sigma_y}\right) - \Phi\left(\frac{y_0 - \mu_y}{\sigma_y}\right). \quad (3.32)$$

Пусть теперь необходимо решить задачу нахождения вероятности того, что выполняется неравенство: $|y_t - \mu_y| < \delta$. Это неравенство равносильно двойному неравенству: $\mu_y - \delta < y_t < \mu_y + \delta$, что, как легко заметить, дает формулировку в терминах задачи (3.32), поэтому решение этой задачи легко найти с помощью (3.32):

$$P(|y_t - \mu_y| < \delta) = \Phi\left(\frac{\mu_y + \delta - \mu_y}{\sigma_y}\right) - \Phi\left(\frac{\mu_y - \delta - \mu_y}{\sigma_y}\right) =$$
$$= \Phi\left(\frac{\delta}{\sigma_y}\right) + \Phi\left(\frac{\delta}{\sigma_y}\right) = 2\Phi\left(\frac{\delta}{\sigma_y}\right). \quad (3.33)$$

Так как δ — некоторое наперед заданное число, его можно задавать различными способами, в частности, как некоторую линейную функцию от среднеквадратичного отклонения, например, так: $\delta = \sigma_y t$. Откуда $t = \delta / \sigma_y$. Пусть, например, $t = 3$. Тогда вероятность того, что отклонение случайной величины y_t от его математического ожидания μ по абсолютной величине будет меньше $\delta = 3\sigma_y$ равна

$P(|y_t - \mu| < 3\sigma_y) = 2\Phi(t) = 2\Phi(3) = 0{,}9973$, что известно в математической статистике под правилом трех сигм.

Неравенство $|y_t - \mu| < \sigma_y t$ равносильно не только двустороннему неравенству: $\mu_y - \sigma_y t < y_t < \mu_y + \sigma_y t$, но и другому двустороннему неравенству:

$$y_t - \sigma_y t < \mu_y < y_t + \sigma_y t. \qquad (3.34)$$

Здесь следует обратить внимание на то, что в последнем двойном неравенстве неизвестно математическое ожидание.

Воспользовавшись (3.33), заменив Y на среднюю арифметическую и общую дисперсию σ_y^2 на дисперсию средней арифметической относительно математического ожидания, получим

$$P(|\bar{y} - \mu| < \delta) = \Phi\left(\frac{\delta}{\sigma_{\bar{y}}}\right). \qquad (3.35)$$

Так как в соответствии с (3.18)

$$\sigma_{\bar{y}}^2 = \frac{\sigma^2}{n},$$

а ранее мы рассмотрели замену $\delta = \sigma_y t$, которая для случая средней арифметической примет вид:

$$\delta = \frac{t_m \sigma_y}{\sqrt{T}},$$

то (3.35) можно записать в другой форме

$$P\left(|\bar{y} - \mu_y| < \frac{t_m \sigma_y}{\sqrt{T}}\right) = 2\Phi\left(\frac{t_m \sigma_y}{\sqrt{T}} \Big/ \frac{\sigma_y}{\sqrt{T}}\right) = 2\Phi(t_m) = \alpha. \qquad (3.36)$$

Как следует из (3.38), при достаточно большом числе наблюдений n выборочная дисперсия практически равна генеральной дисперсии σ_y^2, поэтому можно утверждать, что с заданной доверительной вероятностью α математическое ожидание случайной величины лежит в пределах:

$$\bar{y} - \frac{t_m \sigma_y}{\sqrt{T}} < \mu_y < \bar{y} + \frac{t_m \sigma_y}{\sqrt{T}}. \qquad (3.37)$$

Но при малых выборках ($T < 30$), с которыми в основном и приходится иметь дело в прогнозировании социально-экономических процессов, выборочная дисперсия отличается от общей, поэтому значение σ_y требуется заменить на исправленную дисперсию s_y^2. Тем не менее, введение поправочного коэффициента (3.38) не меняет ситуацию, потому что при малых выборках выборочное значение дисперсии ведет себя иначе, чем это следовало бы из нормального закона распределения. В результате и нормированный ряд (3.29), в котором вместо дисперсии подставляется ее выборочное значение, а вместо случайной переменной — средняя арифметическая:

$$\frac{\overline{y} - \mu_y}{s_y}\sqrt{T} = t_m, \tag{3.38}$$

не будет распределен нормально; функция Лапласа к нему неприменима.

Английский статистик В. Госсет предложил описывать распределение величины (3.38) близким по форме к нормальному. Оно получило название «распределение Стьюдента», поскольку под этим псевдонимом В. Госсет опубликовал соответствующие материалы. Это распределение также симметрично, как и функция Лапласа, имеет такую же форму, но ее максимум несколько меньше, а с увеличением величины t_m функция более пологая, чем функция нормального распределения Лапласа. С увеличением числа наблюдений T распределение случайной переменной (3.38) стремится к нормальному и при $T > 30$ практически совпадает с ним.

Плотность распределения величины t (3.38) определяется только одним параметром — количеством наблюдений $T = m + 1$:

$$t_m = \frac{\Gamma\left(\frac{m+1}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\sqrt{m\pi}}\left(1 + \frac{t^2}{m}\right)^{-\frac{m+1}{2}}, \tag{3.39}$$

где $\Gamma(x)$ — Гамма функция Эйлера в точке x; $m = T - 1$ — величина, получившая название «число степеней свободы».

С помощью этой плотности распределения случайной величины t можно рассчитать вероятность $1 - \alpha$ того, что истинное значение нормированной переменной t лежит в пределах от минус $t_{\alpha,\,m}$ до плюс $t_{\alpha,\,m}$:

$$P\left(-t_{\alpha,m} < t_m < t_{\alpha,m}\right) = 1 - \alpha. \tag{3.40}$$

Эта вероятность получила название «доверительной вероятности».

Подставляя в это равенство значения t_m, взятые из (3.38), получим, что для выборочных значений средней арифметической и дисперсии с доверительной вероятностью $1 - \alpha$ математическое ожидание случайной величины лежит в пределах:

$$\overline{y} - t_{\alpha,m}\frac{s_y}{\sqrt{T}} < \mu_y < \overline{y} + t_{\alpha,m}\frac{s_y}{\sqrt{T}}. \tag{3.41}$$

Здесь множитель $t_{\alpha,\,m}$ рассчитывается по функции (3.39) или берется из таблиц t-статистики Стьюдента. В таблицах эта величина выбирается, исходя из числа степеней свободы m и остаточной вероятности α.

В общем случае, если перед исследователем стоит задача построить доверительные границы для некоторой зависимой переменной y при известных параметрах и значениях независимых переменных $x_1, x_2, ..., x_k$, исследователю достаточно оценить условное математическое ожидание $M\left(y \mid x_1, x_2, ..., x_k\right)$ (оно соответствует расчетному значению $\hat{y} = f\left(x_1, x_2, ..., x_k\right)$) и условную дисперсию $D\left(y \mid x_1, x_2, ..., x_k\right)$, а далее, предполагая *совместное нормальное распределение случайных величин*, воспользоваться формулой для построения доверительных границ на основе t-статистики с $T - k$ числом степеней свободы:

$$M\left(y \mid x_1, ..., x_k\right) - t_{\alpha,T-k}\sqrt{D\left(y \mid x_1, ..., x_k\right)} < y < M\left(y \mid x_1, ..., x_k\right) + \\ + t_{\alpha,T-k}\sqrt{D\left(y \mid x_1, ..., x_k\right)}.$$

3.4. Статистическая проверка гипотез

Для того чтобы оценить, насколько те или иные выявленные закономерности оказались действительно существенными, а не случайными, в теории вероятностей и математической статистике используется инструментарий проверки статистических гипотез. Сразу же стоит заметить, что этот инструментарий привнесен в общественные науки из технических дисциплин, а потому применим исключительно для обратимых процессов.

Вообще статистической гипотезой называется предположение о свойстве генеральной совокупности, которое можно проверить, опираясь на данные выборки[1], т.е. в основе статистической проверки гипотез лежит выборочный метод, который мы уже рассматривали ранее.

Любая статистическая гипотеза обычно состоит из двух частей: H_0 —так называемой «нулевой гипотезы», некоего базового утверждения, которое требует проверки, и H_1 — альтернативной гипотезы, утверждения, противоречащего базовому.

Стоит заметить, что стандартный научный подход заключается в проверке того, противоречит ли нулевая гипотеза статистическим данным или нет. Нулевая, так же, как и альтернативная гипотеза, никогда не может быть принята. У исследователя либо может быть недостаточно доказательств для того, чтобы отклонить нулевую гипотезу, либо ему удается собрать такие доказательства в нужном количестве, в результате чего он с некой степенью уверенности может отклонить нулевую гипотезу.

В основе этого подхода лежит индуктивный метод, предложенный Ф. Бэконом. Необходимость использования индуктивного метода в науке вызвана тем, что в реальной жизни нет возможности изучить все элементы генеральной совокупности, но можно сделать выводы о том, что в ней происходит на основе имеющихся наблюдений, т.е. мы в любом случае вынуждены делать умозаключения о генеральной совокупности на основе выборки, имеющейся в нашем распоряжении.

Индуктивный подход применительно к статистической проверке гипотез не идеален и не очень хорошо применим к некоторым сферам жизни, однако альтернативного подхода в науке на данный момент не существует.

3.4.1. Общие принципы проверки статистических гипотез

Проверка статистических гипотез позволяет лишь получить один из двух результатов: либо по данным есть осно-

---

[1] *Нестеров В. П., Дмитриева В. С., Гришакина Н. И., Манова Н. В.* Проверка статистических гипотез : монография. Великий Новгород : ООО Типография "Виконт", 2009. С. 4.

вания отклонить нулевую гипотезу, либо таких оснований нет. Соответственно, для того, чтобы получить какой-нибудь полезный результат, нулевую гипотезу всегда формулируют в терминах «утверждения», что в математической статистике реализуется через равенство. Например, в качестве нулевой гипотезы мы можем выдвинуть предположение о том, что средняя заработная плата (*salary*) по России составляет 50 000 руб.: H_0: *salary* = 50 000.

Отклонение этой гипотезы в пользу альтернативной дает нам некоторую полезную информацию. Однако и сама альтернативная гипотеза также требует четкой формулировки. Она всегда формулируется в терминах неравенства, и, соответственно может быть сформулирована как:

1) H_1: *salary* ≠ 50 000;
2) H_1: *salary* < 50 000;
3) H_1: *salary* > 50 000.

Выбор того, как именно сформулировать альтернативную гипотезу, осуществляется самим исследователем в зависимости от поставленных целей. Например, если нас интересует, меньше ли средняя заработная плата по России 50 000 руб., то мы воспользуемся вторым вариантом формулирования альтернативной гипотезы, но нулевую гипотезу все равно будем формулировать в терминах равенства.

Для проверки гипотезы по некоторой выборке нужно иметь представления о том, как распределена исследуемая случайная величина в генеральной совокупности. Естественно, никаких точных знаний об этом нет и быть не может, поэтому обычно исследователем делается некое предположение о законе распределения в генеральной совокупности, из которой случайным образом были получены имеющиеся в его распоряжении наблюдения. Для подкрепления своих предположений какими-нибудь расчетами, исследователь обычно также предварительно проверяет гипотезу о соответствии распределения случайной величины предполагаемому закону.

В общем случае проверка гипотезы сводится к сравнению некоторой расчетной статистики $\hat{x}$ (полученной по фактическим данным по определенной формуле) с теоретической статистикой x_1 (характеризующей нижнюю границу интервала) и x_2 (характеризующей верхнюю границу интервала), которые вычисляются на основе математической функции предполагаемого закона распределения случайной величины и заданной исследователем доверительной вероятности $(1 - \alpha)$. Конкретные значения соответствующих теоретических статистик могут быть найдены в статистических таблицах.

Здесь стоит вспомнить о таких понятиях, как доверительная и остаточная вероятности.

Доверительная вероятность — это вероятность, с которой интервал заданной ширины накроет случайную величину. Иногда встречается и другое определение: доверительная вероятность — это степень уверенности в том, что полученный интервал накроет изучаемую случайную величину. В качестве примера применения доверительной вероятности можно привести суждение: «я на 95% уверен, что полученный интервал будет содержать изучаемую случайную величину».

Остаточная вероятность — это вероятность, с которой границы интервала не накроют случайную величину. Остаточная вероятность обычно обозначается через α, в то время как доверительная — через $(1 - \alpha)$. Очевидно, что чем выше исследователь задает значение доверительной вероятности, тем более широкий доверительный интервал он получает. Типичными являются 90%, 95% и 99% доверительные вероятности, которым соответствуют значения $\alpha = 0{,}1$, $\alpha = 0{,}05$ и $\alpha = 0{,}01$.

Значения статистик и результат сравнения зависят не только от остаточной вероятности, но и от того, как была сформулирована альтернативная гипотеза. Возможны три ситуации:

1. H_1: *неравенство*:

a. В случае если $\hat{x} \in (x_1; x_2)$, у исследователя нет оснований отклонить нулевую гипотезу;

b. В случае если $\hat{x} \notin (x_1; x_2)$, у исследователя на уровне значимости α появляются основания отклонить нулевую гипотезу.

2. H_1: *величина меньше некоего значения*:

a. Нулевую гипотезу нет оснований отклонить в случае, когда $\hat{x} > x_1$;

b. Нулевую гипотезу есть основания отклонить на уровне значимости α в случае, когда $\hat{x} < x_1$.

3. H_1: *величина больше некоего значения*:

a. Нулевую гипотезу нет оснований отклонить в случае, когда $\hat{x} < x_2$;

b. Нулевую гипотезу есть основания отклонить на уровне значимости α в случае, когда $\hat{x} > x_2$.

Ситуации (a) могут быть объяснены следующим образом. Из генеральной совокупности, распределенной по предполагаемому закону, могли быть получены различные выборки, по которым, в свою очередь, могут быть получены разные интересующие нас статистики (например, средняя величина). При этом в $(1 - \alpha)$ процентах случаев интервалы, построенные для этих выборочных статистик, должны содержать истинное значение из генеральной совокупности. Если наша расчетная величина вошла в этот интервал, то у нас нет оснований утверждать, что она не получена из предполагаемого нами закона распределения. Но это, конечно же, не значит, что она действительно получена из него.

Результаты, полученные в ситуациях (b), конечно же, не говорят о том, что случайная величина не распределена по предполагаемому закону распределения. Однако они говорят о том, что доказательств в пользу такого распределения в имеющейся выборке обнаружено не было.

Возможно, если исследователю удастся собрать больше данных, результат проверки гипотезы изменится на противоположный.

Для большей наглядности рассмотрим приведенные ситуации на примере рис. 3.3.

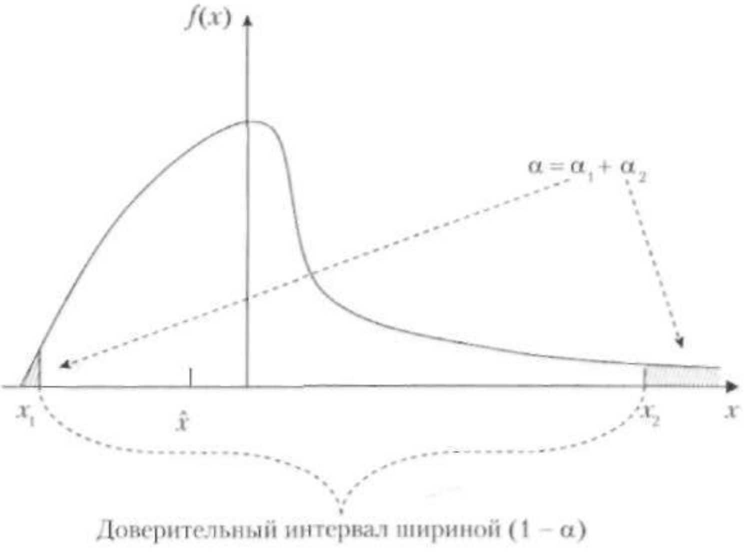

Рис. 3.3. **Условный пример плотности распределения случайной величины x**

На рис. 3.3 показана некоторая условная несимметричная функция распределения, использующаяся при проверке некой статистической гипотезы. Если случайная величина подчинена этому закону распределения вероятностей, то ее значения должны в $(1 - \alpha)$ процентах случаев лежать в выбранном исследователем интервале (на рисунке ограничен значениями x_1 и x_2).

Если альтернативная гипотеза представляет собой неравенство, то нам нужно узнать, попало ли наше наблюдаемое значение в интервал $(x_1; x_2)$: значения, выпадающие из него, мы просто отсекаем из рассмотрения как редко встречающиеся (с вероятностью меньше α). На рис. 3.3 изображена ситуация именно с такой альтернативной гипотезой. Остаточная вероятность α на нем разделена на две части, так, чтобы одна часть — α_1 была слева, а α_2 — справа. В ряде случаев исследователь может быть заинтересован в том, чтобы $\alpha_1 = \alpha_2$, что, например, в случае с симметричными распределениями обеспечивает одинаковое отсечение с разных сторон. В приведенном графическом примере нет оснований отклонить нулевую гипотезу, так как расчетное значение вошло в доверительный интервал.

Если альтернативная гипотеза меньше некоторого значения, то нужно оценить, как расчетная величина соотносится с выбранным критическим значением x_1. Если величина оказывается меньше, то мы можем наблюдать ситуацию, когда случайная величина не вписывается в предполагаемое распределение, а значит, есть основания отклонить нулевую гипотезу в пользу альтернативной.

Аналогична логика рассуждений в отношении оставшейся третьей альтернативной гипотезы.

Многие статистические программы значительно упрощают проверку статистических гипотез: наряду с расчетными значениями статистик, они выдают и значение остаточной вероятности, *p-value*. В таком случае исследователю достаточно знать нулевую гипотезу для того, чтобы понять результат теста. Если полученное значение *p-value* оказывается выше установленного критического α, то это оказывается эквивалентным условию $\hat{x} \in (x_1; x_2)$, т.е., у исследователя нет оснований отклонить нулевую гипотезу. В обратном случае такие основания появляются.

На рис. 3.4 графически отображен механизм проверки односторонней гипотезы, в которой альтернативная гипотеза описывается в терминах «больше». Заштрихованной

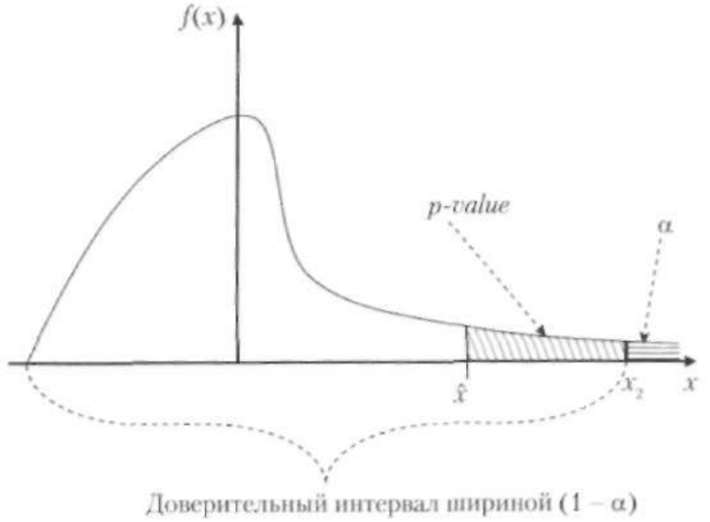

Рис. 3.4. **Условный пример плотности распределения случайной величины x**

областью показана остаточная вероятность, соответствующая расчетному значению $\hat{x}$. С одной стороны, мы видим, что расчетное значение оказалось меньше табличного x_2, выбранного для ширины интервала $(1 - \alpha)$. С другой, — мы видим, что значение *p-value* оказалось по величине больше α (отсекает бóльшую площадь с правой стороны, нежели площадь, соответствующая α). Оба результата говорят о том, что в данном случае оснований отклонить нулевую гипотезу нет.

Обычно при выборе значений статистик слева (x_1) и справа (x_2) стараются их выбрать таким образом, чтобы полученный интервал между этими точками оказался минимальным по ширине. Достигается это за счет выбора таких значений, чтобы функция плотности $f(x_1) = f(x_2)$. В случае с симметричными функциями распределения это сводится к достаточно простому условию: $|x_1| = |x_2|$.

3.4.2. Проверка гипотез с помощью нормального распределения

Самое популярное среди экономистов (да еще и симметричное) распределение, вероятностей — это нормальное распределение, или «распределение Гаусса». Для того чтобы проверить какую-либо гипотезу с помощью нормального распределения, нужно имеющуюся случайную величину центрировать и нормировать для того, чтобы получить стандартную нормально распределенную величину, т.е. с нулевым математическим ожиданием и единичной дисперсией. Так, если:

$$y_t \sim N\left(\mu_y; \sigma_y^2\right), \tag{3.42}$$

то для того, чтобы получить стандартно нормально распределенную величину, надо из y_t вычесть генеральное среднее μ_y, после чего полученное значение разделить на стандартное отклонение генеральной совокупности σ_y:

$$z_t = \frac{y_t - \mu_y}{\sigma_y} \sim N\left(0;1\right). \tag{3.43}$$

Полученное расчетное значение можно сравнить с теоретической z-статистикой по распределению Гаусса. Далее весь алгоритм проверки статистической гипотезы сводится к тому, который был описан выше.

Стоит, однако, заметить, что проверка гипотез с помощью нормального распределения допустима лишь на больших выборках (с числом наблюдений в несколько сотен), в случае, если известны параметры μ_y и σ_y, и, конечно же, она имеет смысл только в том случае, если выполняется предположение (3.42).

Пример

В технологическом процессе компании, выпускающей шампуни, предусмотрено, что в среднем в каждой бутылке с шампунем должно быть 200 мл жидкости. Однако точного значения достичь невозможно, поэтому технологическим процессом предусмотрено и стандартное отклонение от этих 200 мл — 8 мл. Кроме того, допустимой считается остаточная вероятность величиной не больше 5%. Несоблюдение установленных норм сигнализирует о том, что получен бракованный товар, подлежащий списанию. В некоторый момент времени в одну из бутылок было налито 216 мл. Требуется оценить, вписывается данное значение в установленные нормы или нет.

Сформулируем нулевую и альтернативную гипотезы: H_0: $y_t - \mu = 0$, H_1: $y_t - \mu \neq 0$.

Для лучшего понимания нулевую гипотезу можно сформулировать как «получено статистически незначимое отклонение от установленного значения в генеральной совокупности».

Для того чтобы проверить сформулированную гипотезу, нужно получить расчетное значение статистики, что легко сделать, подставив в формулу (3.43) все имеющиеся в нашем распоряжении значения:

$$z_t = \frac{216-200}{8} = 2.$$

Учитывая, что альтернативная гипотеза — неравенство, мы должны рассмотреть оба «хвоста» нормального распределения. Критическое значение для левого «хвоста» соответствует 2,5% квантилю, а для правого — 97,5% квантилю. По модулю z-статистика для нашего случая составляет примерно 1,96, что дает границы, в которых должна лежать нормально распределенная случайная величина: (−1,96; 1,96). Как видим, расчетное значение $z_t = 2$ оказалось выше критического для 97,5% квантиля, а это значит, что нулевая гипотеза отклоняется: превышение установленного значения оказалось статистически значимым, что говорит о том, что был получен бракованный товар и бутылку нужно списать.

На практике прогнозисту приходится работать с малыми выборками, а параметры μ_y и σ_y крайне редко (если вообще) оказываются известными, поэтому для проведения тестов в этих условиях используется распределение Стьюдента, разработанное специально для работы с малыми выборками.

3.4.3. Проверка гипотез с помощью распределения Стьюдента

Для того чтобы понять, как проверить какую-либо гипотезу с помощью распределения Стьюдента, нужно понять, как связаны эти распределения. А связь у них достаточно простая. Если случайная величина z имеет стандартное нормальное распределение, т.е.:

$$z_t \sim N(0;1), \tag{3.44}$$

где $i = 0, 1, \ldots, T$, причем все z_t независимы друг от друга, то величина, рассчитанная по формуле

$$t = \frac{z_t}{\sqrt{\frac{1}{T}\sum_{t=1}^{T} z_t^{\,2}}}, \tag{3.45}$$

будет распределена по Стьюденту с T степенями свободы[1]:

$$t \sim t(T).$$

Самой простой гипотезой, которую можно проверить с помощью распределения Стьюдента, является гипотеза о равенстве

[1] *Вентцель Е. С.* Теория вероятностей : учебник. С. 372.

математического ожидания в генеральной совокупности некоторой величине. Мы уже обращались к такой гипотезе.

Пример

Пусть по некоторым расчетам средняя месячная заработная плата по выборке из 16 регионов России оказалась равной 40 000 руб. По этим же данным стандартное отклонение оказалось равным 18 000 руб. Нам необходимо проверить по этим данным гипотезу о том, что в России средняя заработная плата ниже уровня 50 000 руб. Сформулируем необходимую гипотезу в статистических терминах: H_0: $\mu = 50\ 000$, H_1: $\mu < 50\ 000$.

В основе всех дальнейших рассуждений лежит предположение о том, что средняя величина заработной платы в генеральной совокупности распределена нормально с некоторым средним μ_y и дисперсией $\sigma_{\bar{y}}^2$ (причем так, что наблюдения независимы):

$$\bar{y} \sim N\left(\mu_y; \sigma_{\bar{y}}^2\right). \tag{3.46}$$

Для того чтобы прийти к формуле (3.45), нам нужно из (3.46) получить стандартную нормально распределенную величину (3.44). Достичь этого можно путем центрирования и нормирования $\bar{y}$:

$$\bar{z} = \frac{\bar{y} - \mu_y}{\sigma_{\bar{y}}}. \tag{3.47}$$

Если теперь подставить (3.47) в (3.45), то мы получим *значение t, которое будет распределено по Стьюденту:*

$$t = \frac{\bar{z}}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}\bar{z}^2}}. \tag{3.48}$$

После следующих элементарных преобразований получим

$$t = \frac{\frac{\bar{y} - \mu_y}{\sigma_{\bar{y}}}}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\bar{y} - \mu_y}{\sigma_{\bar{y}}}\right)^2}} = \frac{\bar{y} - \mu_y}{\sigma_{\bar{y}} \cdot \sqrt{\frac{\frac{1}{T}\sum_{t=1}^{T}\left(\bar{y} - \mu\right)^2}{\sigma_{\bar{y}}^2}}} =$$

$$= \frac{\bar{y} - \mu_y}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\bar{y} - \mu_y\right)^2}} = \frac{\bar{y} - \mu_y}{\sigma_{\bar{y}}}. \tag{3.49}$$

В знаменателе (3.49), как видим, представлена формула для расчета дисперсии средней величины относительно математического ожидания в генеральной совокупности. Чему равно это значение, мы уже рассматривали в параграфе 3.2. Подставляя значение (3.18) в формулу (3.49), получим итоговую формулу для получения расчетного значения по t:

$$t = \frac{\bar{y} - \mu}{\sigma_y}\sqrt{T}. \tag{3.50}$$

Теперь мы можем подставить в формулу (3.50) имеющиеся в нашем распоряжении выборочные данные:

$$t = \frac{40\,000 - 50\,000}{18\,000}\sqrt{16} = \frac{-10\,000}{4500} \approx -2{,}22. \tag{3.51}$$

Так как мы оцениваем среднее значение, у нас появляется условие: $\frac{1}{T}\sum_{t=1}^{T} y_t = \bar{y}$, которое отнимает одну степень свободы, в результате чего величина, рассчитанная по формуле (3.51), будет распределена по Стьюденту с $T-1$ степенью свободы.

Проверим гипотезу на общепринятом 5%-ном уровне остаточной вероятности. Теоретическое значение t для правого хвоста распределения, 15 степеней свободы и $\alpha = 0{,}05$ составляет:

$$t_{tab} = t(0{,}05;15) \approx 1{,}75. \tag{3.52}$$

В связи с тем, что распределение Стьюдента симметрично, критическое значение в левом «хвосте» распределения составит $-1{,}75$.

Как видим, расчетное значение (3.51) оказалось меньше табличного $-1{,}75$, следовательно, мы можем отклонить нулевую гипотезу на 5%-ном уровне и сделать вывод о том, что заработная плата по России, скорее всего, меньше 50 000 руб. Однако, если бы в качестве остаточной вероятности мы выбрали не 0,05, а 0,01, то отклонить гипотезу у нас не было бы оснований (так как табличное значение для этого уровня примерно равно 2,60), а значит, и такой вывод о заработной плате сделать было нельзя бы.

Стоит отметить, что определение уровня остаточной вероятности целиком и полностью ложится на плечи исследователя и зависит от тех целей, которые он хочет достичь. Однако получение значения, которое при одном уровне приводит к отклонению нулевой гипотезы, а при другом (близком к первому) — не приводит, может сигнализировать о том, что для более достоверного результата стоит провести дополнительные исследования и собрать больше данных.

Обратим внимание на то, что в основе проверки гипотезы по статистике Стьюдента лежит предположение о нормальности распределения исходной случайной величины. Если это предположение не выполняется, то все дальнейшие шаги по проверке гипотезы не имеют смысла.

3.4.4. Проверка гипотез с помощью распределения хи–квадрат

Распределение хи-квадрат (χ^2) также известно под названием «распределение Пирсона». По аналогии с применением распределения Стьюдента в проверке статистических гипотез рассмотрим вначале, как связаны нормальное распределение с распределением χ^2. Если случайная величина z имеет стандартное нормальное распределение, т.е.

$$z_t \sim N(0;1), \tag{3.53}$$

где $t = 0, 1, ..., T$, причем все z_t независимы друг от друга, то величина, рассчитанная по формуле

$$\chi^2 = \sum_{t=1}^{T} z_t^{\ 2}, \tag{3.54}$$

будет распределена по χ^2 с T степенями свободы[1]:

$$\chi^2 \sim \chi^2(T).$$

Как видим, в основе распределения χ^2 лежит сумма квадратов нормального распределения. Типичная гипотеза, проверяемая с помощью χ^2, — это гипотеза о дисперсии.

Пример

Некая фирма разработала математическую модель для оценки продаж продукции по дням. Считается, что если стандартное отклонение фактических значений от расчетных за прошедший месяц не превысило 50 единиц, то продажи продукции не претерпевают значительных изменений, модель дает все такой же точный прогноз. Если же стандартное отклонение превышает это значение, то модель требует переоценки. За прошедший месяц (в котором был 31 день) стандартное отклонение составило 55 единиц. Стоит ли пересчитывать коэффициенты модели?

---

[1] *Вентцель Е. С.* Теория вероятностей : учебник. 11-е изд. С. 171.

Сформулируем нулевую и альтернативную гипотезы: H_0: $\sigma^2 = 2500$, H_1: $\sigma^2 > 2500$.

Альтернативная гипотеза сформулирована как «больше» в связи с тем, что в тексте примера указывалось на важность превышения значения критического уровня.

Предположим, что ошибки модели распределены нормально с некоторым математическим ожиданием μ и дисперсией σ^2. Стандартное отклонение, рассчитанное по выборке, фактически уже представляет собой сумму квадратов центрированной случайной величины, поэтому для получения величины, распределенной в соответствии с (3.53), нам достаточно каждое значение в выборке разделить на генеральную дисперсию. Однако сами значения нам не даны, но они, в принципе, и не нужны. Из формулы выборочной дисперсии:

$$s^2 = \frac{1}{T-1}\sum_{t=1}^{T}\left(y_t - \mu\right)^2 \tag{3.55}$$

следует, что

$$\sum_{t=1}^{T}\left(y_t - \mu\right)^2 = s^2(T-1). \tag{3.56}$$

Разделив левую и правую части (3.56) на генеральную дисперсию, получим

$$\frac{\sum_{t=1}^{T}\left(y_t - \mu\right)^2}{\sigma^2} = \frac{s^2}{\sigma^2}(T-1). \tag{3.57}$$

Внесем в левой части формулы (3.57) дисперсию под знак суммы, тогда получим

$$\sum_{t=1}^{T}\left(\frac{y_t - \mu}{\sigma}\right)^2 = \frac{s^2}{\sigma^2}(T-1). \tag{3.58}$$

В формуле (3.58) в левой части под знаком суммы мы видим стандартную нормально распределенную случайную величину, рассчитываемую по формуле (3.43). Подставим значение из формулы (3.43) в (3.58), чтобы получить

$$\chi^2 = \sum_{t=1}^{T} {z_t}^2 = \frac{s^2}{\sigma^2}(T-1). \tag{3.59}$$

Величина, рассчитанная по формуле (3.59) будет распределена по $\chi^2(T-1)$ в том случае, если исходная величина x распределена нормально. Число степеней свободы получилось равным $T-1$ из-за наложения условия (3.55).

Подставим в формулу (3.59) значения из задачи:

$$\chi^2 = \frac{55^2}{50^2}(31-1) = 36,3.$$

В связи с тем, что в задаче не был указан уровень остаточной вероятности, примем стандартный: $\alpha = 0,05$. В таком случае табличное значение для $\chi^2(0,05; 30)$ составляет примерно 43,8 — это значение соответствует правому хвосту распределения. Сравнивая его с расчетным, приходим к выводу о том, что у нас нет оснований отклонить нулевую гипотезу. Судя по всему, модель можно не переоценивать.

Заметим, что распределение χ^2 в отличие от распределений Гаусса и Стьюдента не симметрично. Поэтому в случае, если потребуется оценить статистики с двух сторон, нам нужно будет отдельно оценить $\chi^2\left(\frac{\alpha}{2}; df\right)$ и $\chi^2\left(1-\frac{\alpha}{2}; df\right)$, где df — число степеней свободы.

В очередной раз обращаем внимание на предположения, лежащие в основе теста: y_t должны быть одинаково и независимо распределенными по нормальному закону величинами.

3.4.5. Проверка гипотез с помощью распределения Фишера

Последний типичный вид статистических гипотез — это гипотезы на основе распределения Фишера (или «распределения Снедекора»). Оно напрямую связано уже не с нормальным распределением, а с распределением хи-квадрат.

Если V_1 и V_2 — независимые случайные величины, распределенные по χ^2 со степенями свободы k_1 и k_2, то величина

$$F = \frac{V_1/k_1}{V_2/k_2} \tag{3.60}$$

будет распределена по Фишеру со степенями свободы k_1 и k_2[1]:

$$F \sim F(k_1; k_2).$$

---

[1] *Гмурман В. Е.* Теория вероятностей и математическая статистика : учеб. пособие для вузов. 7-е изд. М. : Высшая школа, 2001. С. 147.

Это распределение в основном используется для проведения *анализа дисперсий* (ANOVA — «Analysis of Variance»). Так, с помощью распределения Фишера можно проверить гипотезы о равенстве дисперсий в разных выборках или гипотезу о совместном равенстве нескольких случайных величин заданному числу. Это распределение активно используется в эконометрике при сравнении различных регрессионных моделей.

Рассмотрим один из типичных примеров, в котором с помощью этого распределения проверяется гипотеза.

Прогнозист построил две модели по 50 наблюдениям и рассчитал по ним дисперсии ошибок. В первой модели дисперсия ошибок составила $s_1^2 = 470$, во второй — $s_2^2 = 500$. У первой модели семь коэффициентов, у второй — четыре. На 1%-ном уровне прогнозисту нужно выбрать ту модель, которая дает наименьшую дисперсию ошибок при прогнозировании.

Сформулируем гипотезы. Стандартная формулировка выглядит следующим образом: H_0: $\sigma_1^2 = \sigma_2^2$, H_1: $\sigma_1^2 < \sigma_2^2$.

Альтернативная гипотеза могла бы быть сформулирована и иначе (дисперсия первой модели больше дисперсии второй), но мы видим, что выборочная дисперсия первой модели меньше, чем второй, поэтому логично предполагать, что в «генеральной совокупности» такое соотношение может сохраниться.

Обычно при проверке гипотез все случайные величины переносятся в одну сторону, а неслучайные — в другую. Разделим в наших гипотезах левую и правую части на дисперсию второй модели. Тогда получим гипотезу в более привычном виде:

$$H_0\text{: } \frac{\sigma_1^2}{\sigma_2^2} = 1;$$

$$H_1\text{: } \frac{\sigma_1^2}{\sigma_2^2} < 1.$$

Кроме того, такая формулировка гипотез подсказывает нам, как рассчитать требуемую статистику. Если предположить, что ошибки в моделях распределены нормально, то из этого следует, что дисперсии в моделях будут распределены по χ^2 с $T - k_1$ и $T - k_2$ степенями свободы соответственно (где k_j — число коэффициентов в j-й модели). Отношение дисперсий, соответственно, будет в таком случае распределено по Фишеру. Рассчитаем это отношение:

$$F = \frac{470}{500} = 0{,}94. \tag{3.61}$$

В связи с тем, что альтернативная гипотеза заключается в том, что первая дисперсия меньше второй, мы должны отсечь левый «хвост» распределения Фишера, т.е. 99%-ный квантиль. Статистические таблицы обычно содержат статистики для общепринятых вероятностей: 1%, 5% и 10%. Для того, чтобы получить 99%-ный квантиль можно воспользоваться соотношением

$$F(1-\alpha; df_1; df_2) = \frac{1}{F(\alpha; df_2; df_1)}.$$

Число степеней свободы в числителе — 50 – 7 = 43. В знаменателе — 50 – 4 = 46. $F(0{,}01; 46; 43) \approx 2{,}04$, значит искомое критическое значение для 99%:

$$F(0{,}99; 43; 46) = \frac{1}{2{,}04} \approx 0{,}49. \tag{3.62}$$

Сравнивая теоретическое значение (3.61) с расчетным (3.62), видим, что расчетное попало в область неотклонения нулевой гипотезы (в промежуток от 0,49 до бесконечности). Значит, по полученным данным нет оснований утверждать, что в первой модели дисперсия ошибок статистически значимо меньше, чем во второй, и вопрос выбора лучшей прогнозной модели так и не решен.

3.4.6. Другие способы проверки гипотез

С помощью рассмотренных нами в параграфе 3.4 распределений можно проверить значительно больше различных гипотез, в том числе и составных (например, вначале с помощью F-теста проверить на основе двух выборок, равны ли дисперсии генеральных совокупностей, а затем — с помощью t-теста проверить, равны ли в них средние величины). Однако механизм проверки во всех них один и тот же: на основе предположения о нормальности распределения y_t и на основе связи распределений рассчитывается некоторая статистика, которая затем сравнивается с теоретической.

Конечно же, в теории вероятностей и математической статистике существует значительно больше распределений, чем те, которые были рассмотрены нами выше. Однако они пользуются меньшей популярностью, чем распределения Гаусса, Стьюдента, Пирсона и Фишера.

Выше мы уже упомянули о том, что в ряде случаев, прежде чем проверять более простую гипотезу, исследователю следует убедиться в том, что его случайная величина распределена в соответствии с выдвинутыми предположениями. Наиболее важным, как мы уже убедились, является нормальное распределение, для проверки гипотез относительного которого было разработано множество различных тестов. Нулевая гипотеза во всех них заключается в том, что случайная величина распределена в соответствии с нормальным законом, а альтернативная — не в соответствии с ним. Соответственно, после получения результатов этих тестов, появляется некоторый элемент неопределенности: если данные указывают на то, что у исследователя нет оснований отклонить нулевую гипотезу, то это, как мы уже знаем, не говорит о том, что величина распределена нормально. Похожие результаты, например, могут быть получены и при проверке соответствия эмпирического распределения равномерному распределению случайной величины. Однако никаких других инструментов, дающих бо́льшую уверенность относительно закона распределения случайной величины, у исследователя в распоряжении нет, а значит, в итоге принимаемое им решение в достаточной степени основывается на желании или нежелании исследователя отклонить нулевую гипотезу.

Мы не ставим перед собой цель рассказать обо всех статистических тестах (их разработано очень много), поэтому в данном параграфе упомянем только о трех популярных тестах на нормальное распределение. Это критерий согласия Пирсона (он же «критерий χ^2»), тест Шапиро — Уилка (Shapiro — Wilk test) и тест Харке — Бера (Jarque — Bera test).

Критерий χ^2 является достаточно общим и позволяет проверить гипотезу о соответствии распределения выбранному исследователем. Суть теста сводится к тому, чтобы разделить эмпирическое и теоретическое распределения на одинаковые интервалы и оценить, какой процент наблюдений должен был бы попасть в эти интервалы в теории и какой попал на самом деле. Критерий требователен к размеру выборки: на малых выборках эмпирические частоты могут оказаться заниженными, в результате чего будет получен некорректный результат.

Тест Шапиро — Уилка считается одним из самых эффективных и основан на оценке дисперсии случайной величины, на поквантильном сравнении эмпирического и теоретического (нормального) распределений.

Тест Харке — Бера основывается на сравнении эксцесса и асимметрии эмпирического распределения со значениями в нормальном распределении. Он не очень точен по сравнению с первыми двумя тестами, однако в случае, если при его проведении гипотеза отклоняется, более сложные тесты можно не проводить — это уже указывает на сильные отличия основных параметров распределения от нормальных.

Исследователи также замечают, что на больших выборках тесты становятся более требовательными и нулевая гипотеза в них отклоняется даже в тех случаях, в которых по многим признакам (различные графики, эксцесс и асимметрия) создается впечатление, что распределение случайной величины соответствует распределению Гаусса. Дело в том, что на малых выборках при построении моделей можно допустить различные ошибки, которые незначительно скажутся на эмпирическом распределении случайной величины, однако в случае с большими выборками любые мелкие ошибки увеличивают свое влияние в разы. Кроме того, на малых выборках может быть просто недостаточно данных для точной оценки, в результате чего гипотеза чисто технически не будет отклоняться в тех случаях, в которых должна бы отклониться.

Помимо рассмотренных нами тестов и критериев, существуют «непараметрические» тесты, не требующие нормального распределения случайной величины. Обычно они основываются на оценке медианы (а не средней величины) либо на ранжировании данных, что делает результаты тестов более устойчивыми к мелким ошибкам и «выбросам». Благодаря отсутствию предположений относительно закона распределения, они применимы в большем ряде случаев. Однако с помощью этих тестов можно получить и результат, не соответствующий реальности, так как мощность таких тестов невысока: в случае, если гипотеза на самом деле должна быть отвергнута, непараметрические тесты не будут ее отвергать в связи с большей робастностью (нечувствительностью к ошибкам).

Как видим, инструменты проверки статистических гипотез достаточно разнообразны и позволяют проверять совершенно разные гипотезы относительно различных случайных величин. Стоит, однако, отметить, что этот инструментарий применим исключительно в случаях с обратимыми процессами, в которых работает выборочный метод. Его применимость для пространственных данных неоспорима, в то время как в ситуациях с проверкой гипотез по временным рядам необратимых процессов он вызывает некоторые сомнения. Проблема заключается в том, что получаемое выборочное значение для таких процессов может не являться репрезентативным, а значит, даже в тех случаях, когда полученное значение изучаемой величины можно признать нормально распределенным, утверждать, что какие-то закономерности сохраняются в «генеральной совокупности», может быть не корректно.

3.5. Основы регрессионного анализа в прогнозировании

Задача прогнозирования стационарных однородных процессов довольно проста. Значительно более сложной представляется задача прогнозирования стационарных процессов, протекающих в условиях неоднородности внешней среды. Стационарные процессы, как об этом было сказано в первой главе учебника, характеризуются тем, что они не претерпевают качественных изменений, т.е. не меняют свою структуру, силу и направление взаимосвязей между элементами. Но поскольку эти процессы наблюдаются в условиях неоднородной среды, эта неоднородность оказывает влияние на количественные характеристики процесса — помимо случайных факторов на него воздействуют факторы неслучайной природы. Рассмотрим вначале простой случай, когда на прогнозируемый показатель y оказывает влияние только один фактор x, и это влияние является линейным, т.е. между переменными имеется следующая регрессионная зависимость:

$$y_t = a_0 + a_1 x_t + \varepsilon_t.$$

Пусть функция плотности распределения случайной величину y имеет вид $f(y, \theta)$, где y — это изучаемые случайные величины ($y = y_1, y_2, ...$), θ — некоторый параметр. Тогда событие, заключающееся в независимом отборе из этой совокупности T элементов, в соответствии с теоремой умножения вероятностей независимых событий будет описываться следующей общей функцией плотности вероятностей:

$$f(y_1, y_2, ..., y_T, \theta) = f(y_1, \theta) f(y_2, \theta) ... f(y_T, \theta). \quad (3.63)$$

Так как случайные величины ($y = y_1, y_2, ..., y_T$) уже имеются в распоряжении исследователя, то неизвестной переменной выступает параметр θ. В рамках заданной функции плотности вероятности различные значения этого параметра приводят к тому, что случайные величины ($y = y_1, y_2, ..., y_T$) принимают разные значения. Но в рассматриваемом случае эти величины нам заданы, следовательно, есть некоторое значение параметра θ, которому и соответствует имеющаяся выборка. Наша задача — найти это неизвестное значение параметра.

Теперь можно сформулировать правило нахождения данного параметра. Надо найти оценки θ так, чтобы придать максимальное правдоподобие высказанному предположению о характере распределения вероятностей, в соответствии с которым получена данная выборка. Иначе говоря, нам надо найти такое значение параметра θ, при котором функция плотности вероятностей (3.63) принимала бы максимальные значения. Это правило, предопределяющее направление поиска наилучших оценок параметров модели, получило название *метода максимального правдоподобия*. Так как нами априорно задан закон распределения вероятностей f, то, подставляя его математическое выражение в формулу (3.63), сводим тем самым задачу к поиску максимального значения этой функции по неизвестному параметру θ. Можно применить этот подход непосредственно к функции (3.63), но значительно проще свести задачу от мультипликативной формы (3.63) к линейной аддитивной форме с помощью логарифмирования, например, по натуральному основанию:

$$\ln f(y_1, y_2, ..., y_T, \theta) = \ln f(y_1, \theta) + \ln f(y_2, \theta) + ... + \ln f(y_T, \theta). \quad (3.64)$$

Правомерно ли это? Так как $\ln f(y)$ есть монотонная функция от самой $f(y)$, то ее максимум достигается в той же самой точке, что и у исходной функции. Возьмем первую производную (3.64) по параметру θ:

$$\frac{\partial \ln f(y_1, y_2, ..., y_T, \theta)}{\partial \theta} = \frac{1}{f(y_1, y_2, ..., y_T, \theta)} \left(\frac{\partial f(y_1, y_2, ..., y_T, \theta)}{\partial \theta} \right). \quad (3.65)$$

Так как функция (3.63) по определению не равна нулю в точке своего максимума, то в этой точке из равенства $\frac{\partial \ln f(y_1, y_2, ..., y_T, \theta)}{\partial \theta} = 0$ следует $\frac{\partial f(y_1, y_2, ..., y_T, \theta)}{\partial \theta} = 0$.

Покажем, как, используя метод максимального правдоподобия, получить искомые оценки коэффициентов линейной зависимости. Пусть закон плотности распределения вероятностей соответствует нормальному:

$$f(y, \theta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\theta}{\sigma}\right)^2}. \quad (3.66)$$

Так как моделью процесса выступает линейная регрессия, то параметр θ представляет собой уравнение этой модели, т.е.

$$\theta = a_0 + a_1 x. \tag{3.67}$$

Подставим (3.66) и (3.67) в (3.64). Получим

$$\ln f\left(y, a_0 + a_1 x\right) = T \ln \frac{1}{\sigma\sqrt{2\pi}} - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\left(y_t - a_0 - a_1 x_t\right)^2. \tag{3.68}$$

Наша задача — нахождение максимума этой функции по параметрам a_0 и a_1. Легко убедиться в том, что первое слагаемое полученного выражения не зависит от этих параметров, поэтому максимум функции (3.68) равнозначен нахождению максимума такой функции:

$$F'\left(a_0, a_1\right) = -\frac{1}{2\sigma^2}\sum_{t=1}^{T}\left(y_t - a_0 - a_1 x_t\right)^2.$$

В свою очередь, так как сомножитель перед знаком суммы есть величина постоянная (неизвестны только a_0 и a_1) и отрицательная, то максимум этой функции соответствует минимуму другой функции:

$$F\left(a_0, a_1\right) = \sum_{t=1}^{T}\left(y_t - a_0 - a_1 x_t\right)^2 = \sum_{t=1}^{T}\varepsilon_t^2. \tag{3.69}$$

Полученная функция соответствует известной в математической статистике сумме квадратов ошибок, минимизация которой позволяет найти оценки методом наименьших квадратов (МНК), ведь теперь нам нужно минимизировать сумму квадратов отклонений расчетных значений от фактических. Следовательно, принцип максимального правдоподобия в случае, когда есть основания утверждать наличие нормального закона распределения вероятностей, в качестве лучшего метода оценки параметров модели характеризует метод наименьших квадратов. А с учетом того, что для стационарных процессов чаще всего наблюдается именно нормальный закон вероятностей, то МНК получил максимальное распространение на практике.

Обратим внимание на одно следствие из полученных результатов: в случае, если остатки модели не распределены нормально, вместо МНК имеет смысл использовать метод максимального правдоподобия с другим заданным априори законом распределения случайной величины. Например, в ситуации, когда в качестве зависимой переменной выступает целочисленная величина (такая как количество посетителей кафе), в качестве такого закона можно использовать Пуассоновское распределение.

Применительно к нашей модели линейной регрессии условие минимума квадратов отклонений сводится к нахождению первых производных функции (3.69) по неизвестным параметрам a_0 и a_1 и приравниванию их к нулю, т.е.

$$\begin{cases} \dfrac{\partial F(a_0, a_1)}{\partial a_0} = \dfrac{\partial \sum\limits_{t=1}^{T}(y_t - a_0 - a_1 x_t)^2}{\partial a_0} = -2\sum\limits_{t=1}^{T}(y_t - a_0 - a_1 x_t) = 0, \\ \dfrac{\partial F(a_0, a_1)}{\partial a_1} = \dfrac{\partial \sum\limits_{t=1}^{T}(y_t - a_0 - a_1 x_t)^2}{\partial a_1} = -2\sum\limits_{t=1}^{T}x_t(y_t - a_0 - a_1 x_t) = 0. \end{cases} \tag{3.70}$$

Отсюда легко найти систему уравнений, которая в данном случае будет называться «системой нормальных уравнений», поскольку соответствует нормальному закону распределения:

$$\begin{cases} \sum\limits_{t=1}^{T} y_t = Ta_0 + a_1 \sum\limits_{t=1}^{T} x_t, \\ \sum\limits_{t=1}^{T} y_t x_t = a_0 \sum\limits_{t=1}^{T} x_t + a_1 \sum\limits_{t=1}^{T} x_t^2. \end{cases} \tag{3.71}$$

Пример

В табл. 3.2 приведены значения изменения производительности труда сдельщика в зависимости от величины расценки за единицу труда. Необходимо оценить значения коэффициентов линейной регрессионной модели с тем, чтобы в последующем использовать эту модель для прогнозирования.

Таблица 3.2

Изменение производительности труда при увеличении сдельной оплаты

Номер наблюдения	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Расценка за единицу продукции, руб/шт.	0,25	0,26	0,27	0,28	0,29	0,30	0,32	0,34	0,35	0,38	0,40	0,42	0,45	0,50

Окончание табл. 3.2

Номер наблюдения	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Количество произведенной продукции, шт.	15	17	18	18	19	21	22	26	26	28	29	30	32	36

Используя данные табл. 3.2, найдем необходимые суммы и подставим их в систему нормальных уравнений (3.71) $\begin{cases} 337 = 14a_0 + 4{,}81a_1, \\ 122{,}1 = 4{,}81a_0 + 1{,}7293a_1. \end{cases}$

Решая эту систему, получим такое уравнение регрессии: $\hat{y}_i = 82{,}3x_i - 4{,}2$. Графически эту модель мы представили на рис. 3.5.

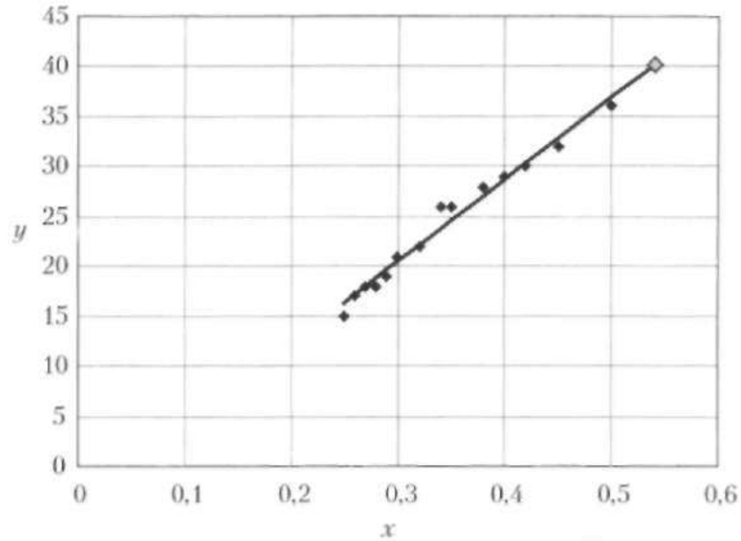

Рис. 3.5. **Точечная диаграмма зависимости между расценками за единицу продукции (ось $0x$) и количеством произведенной продукции (ось $0y$), а так же построенная модель парной регрессии (прямая линия)**

Как видим, построенная модель проходит посередине «облака», образованного из точек по нашим данным.

Эта модель описывает исходные данные с ошибкой, дисперсия которой равна $\sigma_\varepsilon^2 = 0{,}992$. Средняя относительная ошибка аппроксимации $sMAPE = 3{,}30\%$.

Теперь уравнение регрессии можно использовать для прогнозирования. Например, с помощью этой модели можно спрогнозировать производительность труда рабочего, если поднять расценку до 0,54 руб/шт. Для этого в полученную прогнозную модель надо подставить $x = 0{,}54$ и выполнить расчет:

$$\hat{y}_t = 82{,}3 \times 0{,}54 - 4{,}2 = 40{,}24.$$

Это значение на графике изображено ромбиком в правой верхней части.

Зададим доверительную вероятность, равную 0,95. Для 14 наблюдений при этой доверительной вероятности из таблиц t-статистики Стьюдента получим значение $t = 2{,}18$.

Тогда можно с вероятностью 0,95 ожидать, что при расценке в 0,54 руб/шт. рабочий будет вырабатывать:

$$40{,}24 - 2{,}18 \times 0{,}996 < y < 40{,}24 + 2{,}18 \times 0{,}996, \text{ или } 38{,}0 < y < 42{,}4.$$

Для получения более точных границ необходимо строить доверительный интервал, исходя из расчета условного математического ожидания и дисперсии, т.е. рассмотреть $M(y|x)$ и $D(y|x)$ на основе коэффициентов модели и дисперсии ошибки модели.

Коэффициенты модели парной регрессии вида $y_t = a_0 + a_1 x_t + \varepsilon_t$ могут быть найдены и по более простым формулам. Чтобы вывести их, центрируем предварительно значения x и y. Так мы избавимся от константы в модели. Получим: $y'_t = a_1 x'_t + \varepsilon_t$, где $y'_t = y_t - \overline{y}$, $x'_t = x_t - \overline{x}$.

Применяя МНК к такой модели, получим уравнение с одной неизвестной $\sum_t y'_t x'_t = a_1 \sum_t x'^2_t$, решая которое, получим формулу

$$a_1 = \frac{\sum_t y'_t x'_t}{\sum_t x'^2_t} = \frac{\sum_t (y_t - \overline{y})(x_t - \overline{x})}{\sum_t (x_t - \overline{x})^2} = \frac{\operatorname{cov}(y_t; x_t)}{D(x_t)}. \qquad (3.72)$$

Формула для константы выводится из первого уравнения системы нормальных уравнений (3.71): $\sum_t y_t = Ta_0 + a_1 \sum_t x_t$.

Разделим левую и правую части этого уравнения на число наблюдений n и перегруппируем составляющие в формуле:

$$a_0 = \frac{1}{T}\sum_t y_t - a_1 \frac{1}{T}\sum_t x_t. \qquad (3.73)$$

В (3.73) мы видим средние значения y и x, т.е. формулу можно представить в простом виде:

$$a_0 = \overline{y} - a_1 \overline{x}. \qquad (3.74)$$

С линейными моделями прогнозисты встречаются повсеместно. Но не менее часто встречаются случаи, когда изменения показателя в зависимости от изменений фактора носят нелинейный характер и использование линейных моделей в прогнозировании ошибочно и возникает задача нахождения коэффициентов таких нелинейных моделей.

С позиций того, насколько просто удается применить МНК к оценке параметров таких нелинейных моделей, выделяют два вида моделей:

- линейных по параметрам;
- нелинейным по параметрам.

К первому типу моделей относят те из них, оценки МНК которых получаются либо непосредственно при применении МНК, либо при осуществлении линеаризации модели. Вторые модели невозможно привести к линейному виду с помощью подобных преобразований.

Рассмотрим сначала линейные по параметрам нелинейные модели и способы нахождения выборочных значений коэффициентов этих моделей.

Довольно просто найти оценки МНК коэффициентов моделей полиномов разных степеней, например, для квадратичной регрессии:

$$\hat{y}_t = a_0 + a_1 x_t + a_2 x_t^2. \tag{3.75}$$

Применяя МНК к этой модели, получим такую систему нормальных уравнений:

$$\begin{cases} \sum\limits_{t=1}^{T} y_t = Ta_0 + a_1 \sum\limits_{t=1}^{T} x_t + a_2 \sum\limits_{t=1}^{T} x_t^2, \\ \sum\limits_{t=1}^{T} y_t x_t = a_0 \sum\limits_{t=1}^{T} x_t + a_1 \sum\limits_{t=1}^{T} x_t^2 + a_2 \sum\limits_{t=1}^{T} x_t^3, \\ \sum\limits_{t=1}^{T} y_t x_t^2 = a_0 \sum\limits_{t=1}^{T} x_t^2 + a_1 \sum\limits_{t=1}^{T} x_t^3 + a_2 \sum\limits_{t=1}^{T} x_t^4. \end{cases} \tag{3.76}$$

Решая ее относительно неизвестных коэффициентов, можно получить оценку МНК их выборочных значений.

Известно, что полиномы высоких степеней очень неустойчивы, поэтому на практике моделей в форме полиномов со степенью больше двух стремятся избегать, хотя системы нормальных уравнений для этих моделей вычисляются довольно просто.

Также просто найти систему нормальных уравнений для других аддитивных нелинейных моделей. Например, для модели

$$\hat{y}_t = a_0 + a_1/x_t \qquad (3.77)$$

получим систему уравнений МНК:

$$\begin{cases} \sum_{t=1}^{T} y_t = Ta_0 + a_1 \sum_{t=1}^{T} 1/x_t, \\ \sum_{t=1}^{T} y_t/x_t = a_0 \sum_{t=1}^{T} 1/x_t + a_1 \sum_{t=1}^{T} 1/x_t^2, \end{cases} \qquad (3.78)$$

решая которую, найдем оценки коэффициентов модели (3.77), соответствующие оценкам МНК.

Применение МНК к этим аддитивным моделям будет давать несмещенные оценки коэффициентов, когда сумма отклонений расчетных значений показателя от его фактических величин будет равна нулю. Это свидетельствует о том, что модель в среднем проходит через все точки так, что количество точек, расположенных выше модели, примерно равно количеству точек, расположенных ниже нее.

Но в моделях, в которых неизвестные коэффициенты представлены в мультипликативной форме, ситуация изменяется. Рассмотрим в качестве примера задачу оценивания коэффициентов одной из простых прогнозных экспоненциальных моделей:

$$\hat{y}_t = a_0 e^{a_1 x_t}. \qquad (3.79)$$

Как следует из логики МНК, необходимо найти такие значения коэффициентов a_0 и a_1 этой модели, для которых выполняется условие

$$\sum_t \left(y_t - \hat{y}_t\right)^2 = \sum_t \left(y_t - a_0 e^{a_1 x_t}\right)^2 \to \min. \qquad (3.80)$$

Поскольку сумма квадратов отклонений расчетных значений от фактических представляет собой функцию от двух переменных a_0 и a_1, то для нахождения этого минимума необходимо вычислить первые производные функции (3.80) от каждой из переменных и приравнять их к нулю. Решая полученную систему уравнений, можно найти искомые значения коэффициентов.

Применительно к модели (3.79) это условие запишется так:

$$\begin{cases} \dfrac{\partial \sum\limits_t \left(y_t - a_0 e^{a_1 x_t}\right)^2}{\partial a_0} = 0, \\ \dfrac{\partial \sum\limits_t \left(y_t - a_0 e^{a_1 x_t}\right)^2}{\partial a_1} = 0. \end{cases} \qquad (3.81)$$

Вычисляя первые производные каждого равенства системы (3.81) и подставляя их, получим искомую систему нормальных уравнений:

$$\begin{cases} \sum\limits_{t=1}^{T} y_t e^{a_1 x_t} - a_0 \sum\limits_{t=1}^{T} e^{2a_1 x_t} = 0, \\ a_0 \sum\limits_{t=1}^{T} y_t x_t e^{a_1 x_t} - a_0^2 \sum\limits_{t=1}^{T} x_t e^{2a_1 x_t} = 0. \end{cases}$$

Получена система нелинейных уравнений, которую невозможно решить известными методами линейной алгебры. Для решения этой системы необходимо использовать многоитеративную процедуру одного из численных методов. Для специалиста в области экономико-математического моделирования это не составит труда, проблемы могут быть чисто методическими — необходимо выбрать один из алгоритмов нахождения оптимума численных методов, реализовывать его в той или иной программной среде и т.п., т.е. затратить относительно большой промежуток времени на решение довольно тривиальной задачи. Если прогнозист не обладает необходимыми знаниями, он эту задачу решить не сможет.

Несколько десятков лет назад, когда ученые не были вооружены вычислительной техникой так, как это имеет место сегодня, использование численных методов для решения этой задачи становилось чрезвычайно трудоемким, поэтому задача нахождения коэффициентов мультипликативных моделей решалась с помощью линеаризации нелинейных моделей. Применительно к модели (3.79) это делается так.

Прологарифмируем левую и правые части равенства (3.79) по натуральному основанию:

$$\ln \hat{y}_t = \ln a_0 + a_1 x_t. \qquad (3.82)$$

Как видно, в результате логарифмирования модель из мультипликативной формы превратилась в модель аддитивной формы, да к тому же еще и линейного вида. Именно поэтому подобная процедура и получила название «линеаризация».

После линеаризации задача нахождения коэффициентов модели с помощью МНК ставится следующим образом. Логарифмы фактических наблюдений y_t должны описываться логарифмами (3.82) так, чтобы сумма квадратов отклонений между ними была минимальна: $\sum_t (\ln y_t - \ln \hat{y}_t)^2 =$
$= \sum_t (y_t - \ln a_0 - a_1 x_t)^2 \to \min$.

Находя первые производные этой функции по ее параметрам $\ln a_0$ и a_1 и приравнивая их к нулю, получим систему нормальных уравнений:

$$\begin{cases} \sum_{t=1}^{T} \ln y_t = T \ln a_0 + a_1 \sum_{t=1}^{T} x_t, \\ \sum_{t=1}^{T} x_t \ln y_t = \ln a_0 \sum_{t=1}^{T} x_t + a_1 \sum_{t=1}^{T} x_t^2. \end{cases}$$

Эта система имеет довольно простое алгебраическое решение, что позволяет легко найти коэффициенты модели. Но поскольку, решая систему, находятся значения $\ln a_0$, а не a_0, как это необходимо для построения прогнозной модели (3.79), следует из логарифма коэффициента найти его исходное значение: $a_0 = e^{\ln a_0}$.

Теперь полученные с помощью МНК оценки коэффициентов можно подставлять в модель и выполнять с ее помощью прогнозы.

Таким же образом поступают и с другими нелинейными моделями. Покажем, как это делается, на примере основных из них.

Если прогнозист считает необходимым использовать в качестве модели тренда степенную функцию $\hat{y}_t = a_0 x_t^{a_1}$, то для ее линеаризации надо вновь прологарифмировать по любому основанию левую и правую части равенства. Получим $\ln \hat{y}_t = \ln a_0 + a_1 \ln x_t$.

Тогда, применяя для линеаризированной модели МНК, получим систему нормальных уравнений:

$$\begin{cases} \sum_{t=1}^{T} \ln y_t = T \ln a_0 + a_1 \sum_{t=1}^{T} \ln x_t, \\ \sum_{t=1}^{T} \ln x_t \ln y_t = \ln a_0 \sum_{t=1}^{T} \ln x_t + a_1 \sum_{t=1}^{T} \ln^2 x_t. \end{cases}$$

Если в основании показательной степени лежит число 10, то показательная функция имеет вид

$$\hat{y}_t = a_0 10^{a_1 x_t}. \tag{3.83}$$

Чтобы найти коэффициенты этой модели, необходимо ее линеаризовать. Поскольку основанием показательной степени является число 10, то и логарифмировать левую и правую части равенства (3.83) надо по десятичному основанию: $\lg \hat{y}_t = \lg a_0 + a_1 x_t$.

Тогда задача ставится так — минимизировать сумму квадратов отклонений логарифмов расчетных значений от фактических. Для этого формируется система нормальных уравнений:

$$\begin{cases} \sum_{t=1}^{T} \lg y_t = T \lg a_0 + a_1 \sum_{t=1}^{T} x_t, \\ \sum_{t=1}^{T} x_t \lg y_t = \lg a_0 \sum_{t=1}^{T} x_t + a_1 \sum_{t=1}^{T} x_t^2. \end{cases}$$

Решение задачи дает расчетные значения $\lg a_0$ и a_1. Чтобы использовать модель (3.83) в прогнозировании, находим исходное значение коэффициента a_0: $a_0 = 10^{\lg a_0}$.

При использовании в прогнозировании показательной модели с любым основанием

$$\hat{y}_t = a_0 k^{a_1 x_t} \tag{3.84}$$

поступают также. Но поскольку основание этой модели не равно e и не равно 10, то непонятно, по какому же основанию логарифмировать левую и правые части модели для того, чтобы линеаризовать ее? На самом деле, никакой особенной разницы здесь нет, поэтому логарифмировать можно по любому основанию. Для определенности прологарифмируем левую и правую части (3.84) по натуральному основанию: $\ln \hat{y}_t = \ln a_0 + a_1 x_t \ln k$.

Система нормальных уравнений для МНК применительно к этой линеаризованной модели будет иметь вид

$$\begin{cases} \sum_{t=1}^{T} \ln y_t = T \ln a_0 + a_1 \ln k \sum_{t=1}^{T} x_t, \\ \sum_{t=1}^{T} x_t \ln y_t = \ln a_0 \sum_{t=1}^{T} x_t + a_1 \ln k \sum_{t=1}^{T} x_t^2. \end{cases}$$

Из решения данной системы нормальных уравнений легко перейти к исходному виду модели (вычисляя коэффициент a_0) так, как это делается для других моделей: $a_0 = e^{\ln a_0}$.

Казалось бы, найдены оценки этих нелинейных моделей, и усложнять больше ничего не нужно. Надо брать полученные расчетные значения и, подставляя их в модель, вычислять для заданного прогнозного периода значения прогнозируемого показателя. Но поскольку мы вычисляем коэффициенты модели на выборочном множестве вероятного процесса, необходимо оценить доверительные границы этих коэффициентов. Тогда на основе этих границ вычисляются интервальные значения прогнозируемой величины. Эта процедура рассмотрена в параграфе 3.3.

Однако простота указанных рассуждений скрывает за собой одну непростую проблему. Дело в том, что МНК во всех случаях линеаризации нелинейных моделей применяется не для исходной, а для линеаризованной модели. Поэтому все замечательные свойства оценок коэффициентов, полученных для случайных нормально распределенных процессов с помощью МНК (состоятельность, несмещенность и эффективность), соответствуют не исходным, а линеаризованным моделям.

Напомним, что оценки коэффициентов модели будут являться состоятельными, если они по вероятности сходятся к оцениваемому параметру при неограниченном возрастании объема выборки. Чем большее число наблюдений учитывается при оценивании коэффициентов модели, тем точнее становится оценка выборочного значения коэффициента. Несмещенными называются оценки выборочного значения параметра, в которых отсутствуют систематические отклонения от оцениваемого параметра. Поскольку по выборочным значениям можно оценить коэффициенты модели разными способами, то каждый из них даст различную величину дисперсии отклонения расчетных значений от фактических. Тот метод оценивания, который даст наименьшую дисперсию, будет являться эффективным.

Доказано, что в случае нормально распределенной случайной величины, использование МНК приводит к тому, что полученные выборочные оценки будут являться состоятельными, несмещенными и эффективными. Но, эти оценки характерны для линеаризованных, а не для исходных моделей. Покажем, какими будут оценки для исходных нелинейных мультипликативных моделей. Воспользуемся для этого экспоненциальной моделью: $\hat{y}_t = a_0 e^{a_1 x_t}$.

Линеаризованная модель будет описывать логарифм исходного ряда с некоторой ошибкой ε_t: $\varepsilon_t = \ln y_t - \ln \hat{y}_t$.

Поэтому критерий МНК

$$\sum_t \left(\ln y_t - \ln \hat{y}_t\right)^2 = \sum_t \left(\ln y_t - \ln a_0 - a_1 x_t\right)^2 \to \min$$

по сути, сводится к минимизации суммы квадратов этой ошибки:

$$\sum_t \left(\ln y_t - \ln \hat{y}_t\right)^2 = \sum_t \varepsilon_t^2 \to \min. \qquad (3.85)$$

Несмещенность оценок МНК коэффициентов линеаризованной модели означает, что на рассматриваемом выборочном множестве сумма отклонений ε_t будет равна нулю:

$$\sum_t \varepsilon_t = 0. \qquad (3.86)$$

Из (3.85) легко получить следующее равенство: $\ln y_t = \ln \hat{y}_t + \varepsilon_t$, откуда

$$y_t = \hat{y}_t e^{\varepsilon_t}. \qquad (3.87)$$

Ошибка ε_t является мультипликативной по отношению к исходной модели. Получается, что при линеаризации исследователь автоматически делает предположение о том, что ошибка в модели имеет мультипликативный вид. А это, в свою очередь, будет означать, что с ростом значения $\hat{y}$ будет расти и величина ошибки, что, конечно же, не всегда выполняется и не всегда имеет смысл.

Обозначим аддитивную ошибку отклонений расчетных значений модели от фактических через ξ_t: $\xi_t = y_t - \hat{y}_t$, откуда

$$y_t = \xi_t + \hat{y}_t. \qquad (3.88)$$

Левые части (3.87) и (3.88) равны друг другу, следовательно, равны друг другу и правые части этих равенств, т.е. $\hat{y}_t e^{\varepsilon_t} = \xi_t + \hat{y}_t$.

Тогда можно вывести аддитивную ошибку в зависимости от расчетных значений показателя и мультипликативной ошибки, характерной для оценок МНК линеаризованной модели:

$$\xi_t = \hat{y}_t\left(e^{\varepsilon_t} - 1\right). \tag{3.89}$$

Поскольку исходный ряд значений y_t положителен и не равен нулю — он изменяется по тенденции, близкой к экспоненте, — постольку и его расчетные значения положительны. Тогда из полученного равенства видно, что аддитивная ошибка ξ_t будет равна нулю только в одном случае, — когда мультипликативная ошибка ε_t также будет равна нулю. Относительно мультипликативных ошибок ε_t, соответствующих оценке МНК, известно, что они в сумме дают нуль (3.86), а поскольку оценки МНК состоятельны и эффективны, то дисперсия этой ошибки мала. Размахи отрицательных ($\varepsilon_{\min} = \min\limits_t \varepsilon_t$) и положительных величин ($\varepsilon_{\max} = \max\limits_t \varepsilon_t$) этих ошибок по вероятности равны друг другу: $|\varepsilon_{\min}| = \varepsilon_{\max}$.

Множитель при расчетных значениях показателя представляет собой функцию, изображенную на графике рис. 3.6.

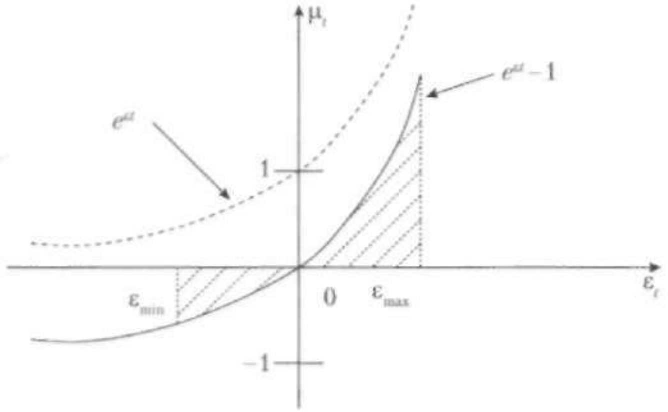

Рис. 3.6. **Функция изменения множителя ($e^{\varepsilon t}$ – 1) в зависимости от εt**

Зная, как ведет себя этот множитель в зависимости от мультипликативной ошибки, можно ответить и на вопрос: чему в среднем будет равна сумма отклонений ξ_t, т.е. как пройдет модель с коэффициентами, оцененными с помощью МНК, по линеаризованной модели?

Исходная модель будет несмещенной, если сумма

$$\sum_t \xi_t = \sum_t \left(y_t - \hat{y}_t\right) \tag{3.90}$$

будет равна нулю.

Равенство (3.90) с учетом (3.89) можно записать так:

$$\sum_t \xi_t = \sum_t \hat{y}_t \left(e^{\varepsilon_t} - 1\right). \tag{3.91}$$

Чтобы ответить на поставленный вопрос, вначале примем для простоты, что $\hat{y}_t = \text{const} > 0$.

Тогда знак суммы (3.91) определяется знаком суммы $\sum\limits_t \hat{y}_t \left(e^{\varepsilon_t} - 1\right)$.

Его можно понять, если обратиться к графику рис. 3.6. Эта сумма будет представлять собой площадь фигуры, изображенной на графике двумя заштрихованными областями. Левее нулевой точки расположена отрицательная часть этой суммы, правее — положительная часть. На рисунке видно, что площадь отрицательной части меньше, чем площадь положительной. Это означает, что знак суммы будет положительным:

$$\sum_t \xi_t > 0. \tag{3.92}$$

Поскольку $y_t \neq \text{const}$, но при этом является положительным, знак меняться не будет.

Таким образом, линеаризация исходных моделей и применение к ним МНК приводит к смещению оценок исходных моделей. Применение МНК непосредственно к исходной модели, хотя и приводит к серьезным вычислительным сложностям, поскольку необходимо решать систему нелинейных уравнений типа (3.80), но позволяет получить несмещенные оценки модели так, что дисперсия ошибки аппроксимации при этом будет минимальной.

Кстати, из вывода о том, что сумма аддитивной ошибки модели положительна, следует вывод о том, как опишет исходный ряд модель нелинейного тренда со смещенными оценками. Подставим в (3.92) значения ошибки из (3.88): $\sum\limits_t \xi_t = \sum\limits_t \left(y_t - \hat{y}_t\right) > 0$.

Это означает, что модель в среднем пройдет ниже исходных точек: $\sum\limits_t y_t > \sum\limits_t \hat{y}_t$.

Следовательно, и прогноз, который будет выполняться с помощью такой модели, будет давать прогнозисту искаженные оценки.

Поэтому, хотя процедура линеаризации и используется повсеместно на практике для применения МНК, необходимо помнить, что при прогнозе такие модели будут неточными и содержащими в себе систематическую ошибку.

Пример

Пусть исходный ряд представляет собой некоторую возрастающую последовательность, подчиняющуюся экспоненциальному закону изменения: $y_t = 2e^{0{,}1x_t} + \varepsilon_t$, где ε_t — случайная нормально распределенная величина с нулевым математическим ожиданием (табл. 3.3).

Логарифмируя исходный ряд, тем самым его линеаризуя, мы приводим его к виду, удобному для применения МНК. Третий столбец табл. 3.3 содержит эти значения. МНК, примененный к этим логарифмам так, как это показано в системе нормальных уравнений (3.83), позволяет найти значения коэффициентов линеаризованной модели: $\ln \hat{y}_t = 0{,}0992x_t + 0{,}6933$.

Экспонируя полученную функцию, найдем исходную модель:

$$\hat{y}_t = 2{,}0003e^{0{,}0992x_t}. \qquad (3.93)$$

В четвертом столбце табл. 3.3 приведены эти расчетные величины. Теперь можно вычислить ошибку аппроксимации модели — она приведена в этой же таблице в последнем столбце. Просуммировав значения этого столбца по всем наблюдениям, найдем сумму ошибок аппроксимации: $\sum_{t=1}^{T} \varepsilon_t = 1{,}0860 > 0$.

Таблица 3.3

Условный пример, демонстрирующий смещенность оценок МНК для линеаризованной модели

x_t	y_t	$\ln y_t$	$\hat{y}_t$	ε_t
1	2,0086	0,6974	2,2088	−0,2003
2	3,2403	1,1757	2,4391	0,8012
3	2,6568	0,9771	2,6935	−0,0366
4	3,3981	1,2232	2,9743	0,4238
5	2,7140	0,9984	3,2844	−0,5704
6	3,7978	1,3344	3,6269	0,1710
7	3,0772	1,1240	4,0050	−0,9279

Окончание табл. 3.3

x_t	y_t	$\ln y_t$	$\hat{y}_t$	ε_t
8	4,7934	1,5672	4,4226	0,3708
9	3,9584	1,3758	4,8837	–0,9253
10	6,2936	1,8395	5,3929	0,9007
11	5,2232	1,6531	5,9552	–0,7320
12	7,6326	2,0324	6,5761	1,0564
13	7,3203	1,9907	7,2618	0,0585
14	8,6179	2,1538	8,0189	0,5989
15	8,1964	2,1037	8,8550	–0,6586
16	10,6042	2,3613	9,7783	0,8259
17	10,0786	2,3104	10,7978	–0,7192
18	12,2277	2,5037	11,9237	0,3040
19	12,6084	2,5344	13,1669	–0,5585
20	15,4433	2,7372	14,5397	0,9036

Как и ожидалось, эта сумма не равна нулю, а это значит, что построенная модель смещена, т.е. она в среднем проходит ниже, чем фактические наблюдения. На рис. 3.7 графически представлены фактические значения и расчетные значения, полученные по модели (3.93).

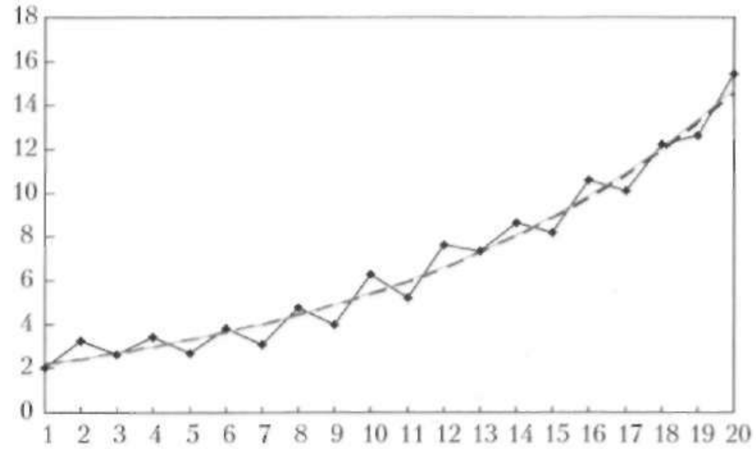

Рис. 3.7. **Условный пример (сплошная линия с точками) и его аппроксимация линеаризованной моделью (пунктирная линия) и не линеаризованной моделью (сплошная линия)**

Расчетные значения по модели (3.93) на рис. 3.7 изображены сплошной линией. По рисунку видно, что модель в целом достаточно хорошо усреднила исходный ряд данных, сильных проблем в модели невооруженным глазом не видно.

Однако если каждое расчетное значение модели сдвинуть вверх на величину $\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t = \frac{1{,}0860}{20} = 0{,}0543$, то модель проходит через среднюю арифметическую точку.

Применительно к математической форме записи модели (3.93) это означает ее модификацию к следующему виду:

$$\hat{y}_t = 2{,}0003e^{0{,}0992x_t} + 0{,}0543\,. \qquad (3.94)$$

Сумма ошибок аппроксимации такой модели тогда действительно станет равной нулю, но эта модель все равно не будет обладать свойствами оценок МНК (несмещенность, состоятельность и эффективность). Модели типа (3.94) иногда называют моделями модифицированной экспоненты, оценку коэффициентов которой осуществляют обычно так, как это было показано выше на условном примере.

Нормальные оценки нелинейных моделей в наше время можно найти численными методами, реализованными во многих статистических программах. Например, в MS Excel для этих целей можно воспользоваться надстройкой «Поиск решения» («Solver»). Для нашего случая надстройка позволяет минимизировать сумму квадратов отклонений y_t от $\hat{y}_t$, рассчитанных непосредственно по формуле без линеаризации $\hat{y}_t = a_0e^{a_1x_t}$.

Чтобы реализовать такой расчет, на листе MS Excel нужно выделить ячейки для коэффициентов и создать столбец с расчетными значениями, которые ссылались бы на эти ячейки. После этого нужно в отдельной ячейке оценить сумму квадратов отклонений и запустить «Поиск решения». В качестве целевой ячейки задается ячейка с суммой, в качестве «равенства» указывается «минимум», в качестве изменяющихся ячеек — выделенные ячейки с коэффициентами.

После таких расчетов мы получили модель вида

$$\hat{y}_t = 1{,}9873e^{0{,}1003x_t}. \qquad (3.95)$$

Как видим, коэффициенты этой модели незначительно отличаются от коэффициентов модели (3.92). Однако сумма ошибок в этой модели уже равна нулю и модель оказывается несмещенной. На рис. 3.7 расчетные значения по модели изображены сплошной линией. Видно, что с ростом значения y_t усиливается расхождение между значениями по модели (3.92) и (3.95). Это расхождение иногда списывается экономистами на «научно-технический прогресс», когда они с помощью мультипликативных моделей описывают экономическую динамику, хотя, как видим, оно вызвано чисто техническими проблемами.

3.6. Корреляционный анализ при построении однофакторных прогнозных моделей

В реальной экономике практически не встречаются процессы, изолированные от других процессов, объектов и явлений. Любой прогнозируемый экономический показатель отражает некоторое состояние экономического объекта и является результатом сложной взаимосвязи прогнозируемого объекта и других объектов, факторов и условий. Еще Г. В. Лейбниц писал: «Аксиома, что ничего не бывает без основания, должна считаться одной из самых важных и плодотворных аксиом во всем человеческом познании; на ней основывается большая часть метафизики, физики и нравственного учения, и без нее нельзя ни доказать ни существования Бога из творений, ни построить доказательство от причин к следствиям или от следствий к причинам, ни сделать какие-либо выводы в делах гражданских»[1]. Если попытаться выявить все факторы и условия, оказывающие воздействие на прогнозируемый социально-экономический объект, то таких факторов будет очень много, и учесть влияние каждого из них практически невозможно. Поэтому возникает задача абстрагирования — выбора факторов, оказывающих на прогнозируемый объект наибольшее влияние. Зачастую на практике эту задачу упрощают — необходимо найти главный фактор, изменения которого оказывают влияние на прогнозируемый объект, т.е. использовать для прогнозирования однофакторную зависимость изменения прогнозируемого показателя y_t от основного фактора x_t.

Поскольку в данной главе мы анализируем процессы, представляющие собой реализацию некоторых случайных причин, а имеющиеся в распоряжении прогнозиста статистические данные об этих процессах рассматриваем как выборочные значения из некоторой генеральной совокупности, то к ним в полном объеме применим инструментарий математической статистики. Из всего многообразия инструментария современной математической статистики к задачам социально-экономического прогнозирования чаще всего применяют аппарат *корреляционно-регрессионного анализа*, базирующийся на выборочном методе.

Под корреляционным анализом, как известно, понимается раздел математической статистики, нацеленный на выявле-

[1] *Лейбниц Г. В.* Соч. : в 4 т. Т. 3. М., 1984. С. 141.

ние взаимосвязи между случайными факторами и оценки степени этой взаимосвязи.

Под регрессионным анализом понимается раздел математической статистики, посвященный выбору формы регрессионной зависимости, оценке ее коэффициентов и достоверности полученных результатов.

Сразу же уточним. Ученые часто имеют дело с функциональными зависимостями, когда одному x соответствует только одно y. При обработке эмпирических данных ученые сталкиваются с другого рода зависимостью, когда одна случайная величина зависит от другой случайной величины (когда одному значению x по вероятности соответствует другое значение y). Она называется корреляционной — «такая зависимость между двумя случайными переменными, при которой одна из них реагирует на изменения другой изменениями своего математического ожидания»[1]. Математическое описание корреляционной зависимости представляет собой линию регрессии.

Рассмотрим вначале корреляционный анализ применительно к задаче социально-экономического прогнозирования, понимая, что все приводимые расчетные коэффициенты будут являться выборочными значениями. Основной задачей корреляционного анализа является выявление из множества случайных факторов, связанных с прогнозируемым показателем, тех из них, влияние которых оказывается наиболее значимым. В однофакторном случае из множества факторов необходимо выбрать один, чье влияние на прогнозируемый показатель оказывается решающим. В корреляционном анализе не предполагается определять направление этой взаимосвязи, т.е. искать ответ на вопрос: y_t влияет на x_t или наоборот x_t влияет на y_t. Необходимо выявить наличие взаимосвязи между двумя (в случае однофакторной зависимости) случайными величинами и оценить тесноту этой взаимосвязи.

Корреляционный анализ оперирует довольно обширным арсеналом различных коэффициентов и показателей, но о наличии взаимосвязи в итоге предлагается судить в основном по трем расчетным коэффициентам:

- корреляционному отношению;
- коэффициенту парной корреляции;
- коэффициенту детерминации.

---

[1] *Дунин-Барковский И. В., Смирнов Н. В.* Теория вероятностей и математическая статистика. М. : Гос. изд-во технико-теоретической литературы, 1955. С. 370.

Корреляционное отношение используется в ситуации, когда рассматриваются зависимости нескольких групп одного фактора от нескольких групп других факторов. Например, когда рабочие разных разрядов получают зарплаты разного уровня и результаты их работы в зависимости от этого уровня изменяются. Для решения подобной задачи можно разбить рабочих на несколько групп в зависимости от уровня их зарплаты, соответствующего разряду квалификации рабочего, например:

- крайне маленькая зарплата (2—5 денежных единиц);
- низкий уровень (5,1—8 денежных единиц);
- ниже среднего (8,1—11 денежных единиц);
- средний уровень (11,1—14 денежных единиц);
- чуть выше среднего (14,1—17 денежных единиц);
- выше среднего (17,1—20 денежных единиц);
- значительно выше среднего (20,1—23 денежных единиц);
- очень высокий (от 23,1 денежных единиц и выше).

В каждой из этих i групп зарплата конкретных рабочих варьируется вокруг некоторой средней $\bar{x}_i$, соответствующей каждому из разрядов, т.е. внутри каждой i-й группы есть некоторая внутригрупповая дисперсия $\sigma^2_{x,\,i}$. Точно так же и производительность труда этих рабочих, измеренная в некоторых единицах, варьируется. Эту производительность разного уровня также можно сгруппировать. И вовсе не обязательно, что число j групп будет совпадать с числом i групп рабочих, разбитых по уровню их зарплаты. Внутри каждой группы, соответствующей разным уровням производительности труда, также есть некоторый средний уровень $\bar{y}$ и отклонения от него, т.е. имеется собственная внутригрупповая дисперсия $\sigma^2_{y,\,j}$. Такая ситуация типична для применения к ее анализу корреляционного отношения. В этом случае можно вычислить выборочное корреляционное отношение y к x. Оно будет рассчитываться так. Для всех значений y рассчитывается общая средняя арифметическая $\bar{y}$ и общее среднее квадратическое отклонение σ_y. Затем для каждой группы j считаются внутригрупповые средние квадратические отклонения σ_{yj} и их средняя по всем группам $\bar{\sigma}_y$. Искомое корреляционное отношение η_{yx} будет иметь вид

$$\eta_{yx} = 1 - \frac{\bar{\sigma}_y}{\sigma_y}. \tag{3.96}$$

Если в какой-либо j-й группе будет всего лишь одно наблюдение, или все наблюдения в группе будут равны одному и тому же значению, то очевидно, что внутригрупповая дисперсия в такой группе будет равна нулю. Если это наблюдается в каждой группе, т.е. одному x соответствует одно и только одно y, то числитель дроби (3.96) будет равен нулю и корреляционное отношение η_{yx} будет равно единице, характеризуя ситуацию, когда между y и x имеется функциональная зависимость. Другой крайний случай, когда переменные независимы друг от друга и вне зависимости от того, какие значения принимает один фактор, разброс точек в каждой группе другого фактора равномерный, а средние каждой группы одинаковы. Тогда средняя $\bar{\sigma}_y$ будет равна общему среднему квадратическому отклонению σ_y, а это значит, что числитель дроби (3.96) равен ее знаменателю и что корреляционное отношение η_{yx} будет равно нулю, характеризуя отсутствие какой-либо зависимости.

Точно так же можно рассчитать и корреляционное отношение x к y:

$$\eta_{xy} = 1 - \frac{\bar{\sigma}_x}{\sigma_x}. \tag{3.97}$$

Однако чаще всего на практике встречаются ситуации, когда имеются пары случайных значений y_t; x_t, т.е. каждому y_t соответствует x_t и наоборот — каждому x_t соответствует y_t. Но поскольку каждая из переменных является случайной величиной, то переменной y одного уровня может соответствовать несколько переменных x, отличающихся друг от друга на случайную величину. Здесь нет групп и соответственно нет внутригрупповых дисперсий. Если формально применить к этой ситуации корреляционное отношение, то будет получена единица. Но это не говорит о том, что между переменными имеется строгая функциональная зависимость, ведь по n точкам можно построить функциональную зависимость — полином $(n - 1)$-й степени, а именно об этом в данном случае и сигнализирует корреляционное отношение. Просто для данного случая корреляционное отношение неприменимо. Надо использовать другой инструмент измерения случайной взаимосвязи между факторами, или, иначе говоря, корреляционной зависимости между ними.

Здесь основным инструментом выступает коэффициент парной корреляции. Рассмотрим его свойства.

Чаще всего в современной математической статистике коэффициент корреляции случайных величин y и x рассматривают как отношение корреляционного момента μ_{xy} к произведению средних квадратических отклонений этих величин:

$$r_{yx} = \frac{\mu_{xy}}{\sigma_x \sigma_y}. \tag{3.98}$$

Корреляционный момент случайных величин определяют как математическое ожидание произведения отклонений этих величин:

$$\mu_{xy} = M\{[x - M(x)][y - M(y)]\}. \tag{3.99}$$

Зная свойства математических ожиданий, а именно, что математическое ожидание произведения независимых случайных величин равно произведению математических ожиданий этих сомножителей, получим

$$\mu_{xy} = M\{[x - M(x)][y - M(y)]\} = \\ = M[x - M(x)]M[y - M(y)] = 0. \tag{3.100}$$

Из этого делается вывод о том, что для двух независимых случайных величин коэффициент корреляции будет равен нулю, поскольку из (3.100) следует, что числитель формулы (3.98) равен нулю, а значит и сам коэффициент равен нулю. Поэтому если на практике прогнозист сталкивается с двумя независимыми случайными переменными, он может быть уверен, что коэффициент корреляции между ними будет равен нулю.

Необходимо обратить внимание на одностороннюю направленность этого утверждения, а именно на то, что истинным является только вывод о том, что *для независимых случайных величин коэффициент корреляции равен нулю.* Обратное утверждение о том, что при наличии коэффициента парной корреляции между двумя переменными, равным нулю, можно говорить о независимости этих переменных, истинным не является.

Действительно, прямой вывод имеет схему классического дедуктивного вывода, и если нам известно, что две случайные переменные независимы друг от друга, то мы будем утверждать, что для них коэффициент парной корреляции равен нулю. Вот логическая схема этого вывода.

Все случайные переменные (все A), являющиеся независимыми (при условии B), имеют коэффициент парной корреляции, равный нулю (имеют $r = 0$). Поэтому если перед нами некоторые другие случайные переменные (D принадлежит к множеству A), которые являются независимыми (удовлетворяют условию B), то для них коэффициент парной корреляции равен нулю (имеют $r = 0$).

Или: *для всех A, удовлетворяющих условию B, выполняется $r = 0$. D принадлежит к A и удовлетворяет условию B, следовательно, для него выполняется $r = 0$.*

Вывод дедуктивный и является истинным.

А теперь рассмотрим структуру обратного вывода. Чаще всего рассуждают так: «две случайные переменные имеют коэффициент парной корреляции, равный нулю. Поскольку известно, что этот коэффициент равен нулю для независимых переменных, то это означает, что рассматриваемые переменные не зависят друг от друга». Кажется, что так оно и есть, но на самом деле полученный вывод носит не утвердительный, а предположительный характер. Для того чтобы понять это, рассмотрим логическую структуру данного вывода.

Для всех случайных переменных (для всего множества A), являющихся независимыми (которые удовлетворяют условию B), коэффициент парной корреляции равен нулю (обладают $r = 0$). Случайные переменные (некоторая часть D множества A) имеют коэффициент парной корреляции, равный нулю (имеет свойство $r = 0$). Следовательно, они также являются независимыми (они удовлетворяют условию B).

Или: *для всех A, удовлетворяющих условию B, выполняется $r = 0$. D принадлежит к A и имеет $r = 0$. Значит, для D выполняется условие B.*

Как видно из структуры этого вывода, он является не дедуктивным, а индуктивным. Действительно, рассматриваемую ситуацию можно расширить: *для всех A, удовлетворяющих условию C, также может выполняться равенство $r = 0$. Из чего следует, что если D принадлежит A и для него $r = 0$, то D может удовлетворять и условию C, и условию B.*

Поясним сказанное на простом примере первого дедуктивного вывода и обратного ему индуктивного вывода.

Первый дедуктивный вывод: все члены партии «Единая Россия» являются гражданами России. Этот человек — член партии «Единая Россия», следовательно, он гражданин России. Это дедуктивный вывод, и он является истинным. А вот обратный вывод: все члены партии «Единая Россия» являются гражданами России. Авторы этих строк — граждане России, следовательно, они члены партии «Единая Россия». Он является индуктивным, носит предположительный характер, а потому не является истинным.

Точно так же и обратный индуктивный вывод о том, что равенство нулю коэффициента парной корреляции свидетельствует о независимости двух случайных переменных, не является истинным, он дает некоторую предположительную оценку, которая может оказаться и ложной. Надо всегда иметь это в виду.

В математической статистике доказывается, что корреляционный момент μ_{xy} лежит в пределах

$$-\sigma_x\sigma_y \le \mu_{xy} \le \sigma_x\sigma_y. \qquad (3.101)$$

Это двойное неравенство, как легко заметить, может быть заменено на такое:

$$|\mu_{xy}| \le \sigma_x\sigma_y. \qquad (3.102)$$

Если теперь обратиться к формуле коэффициента парной корреляции (3.98), то, воспользовавшись неравенством (3.102), получаем еще одно свойство этого коэффициента:

$$\frac{|\mu_{xy}|}{\sigma_x\sigma_y} = |r| \le 1, \qquad (3.103)$$

причем строгое равенство достигается в том случае, когда между исходными переменными имеется функциональная линейная зависимость.

Отсюда следует дедуктивный вывод о том, что для двух случайных переменных, между которыми имеется линейная зависимость, коэффициент парной корреляции по модулю будет равен единице. Обратный вывод о том, что если между двумя переменными коэффициент парной корреляции равен единице, то между ними существует линейная функциональная зависимость, является индуктивным, и поэтому не является истинным. Приведем подтверждающий это пример.

Пусть в некоторой больнице одному из пациентов поставлена капельница и количество влитых в его организм капель меняется строго линейно функционально относительно времени: $x_t = at$, а в некотором банке кассир, принимающий у клиента деньги на счет, получает их так, что сумма счета меняется линейно функционально относительно времени как $y_t = bt$.

Если кому-то придет в голову идея рассчитать выборочное значение коэффициента корреляции между этими двумя рядами x_t и y_t, то оно будет равно единице. Но коэффициент парной корреляции, равный единице для этого случая, очевидно, вовсе не означает, что между этими двумя рядами имеется строгая линейная функциональная зависимость. Никто из здравомыслящих людей не заявит о том, что чем больше капель лекарства получает пациент в больнице, тем больше денег появится на счете у клиента банка. Между этими двумя рядами, как это ясно из сути процессов, вообще взаимосвязи нет и быть не может.

Итак, по значениям коэффициента парной корреляции нельзя однозначно судить о степени корреляционной взаимосвязи между факторами, поскольку такой вывод является индуктивным и носит предположительный характер.

Для корректного применения этого коэффициента необходимо предварительно обосновать гипотезу о наличии корреляции между двумя случайными переменными, и только после этого подтвердить или опровергнуть эту гипотезу с помощью коэффициента корреляции.

Поскольку экономисты работают не с генеральными совокупностями, а только с выборочными значениями из них, то непосредственное применение формулы (3.98) для расчета коэффициента парной корреляции невозможно. Необходимо использовать выборочное значение корреляционного момента и выборочные значения средних квадратических.

Центральный выборочный корреляционный момент рассчитывается по формуле

$$\mu_{xy} = \frac{1}{T}\sum_{t=1}^{T}\left(x_t - \overline{x}\right)\left(y_t - \overline{y}\right). \tag{3.104}$$

Выборочные значения среднеквадратических отклонений рассчитываются так:

$$\sigma_x = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(x_t - \overline{x}\right)^2}; \tag{3.105}$$

$$\sigma_y = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(y_t - \overline{y}\right)^2}. \qquad (3.106)$$

Заменив показатели в формуле (3.98) на их выборочные значения (3.104) – (3.106), получим выборочное значение коэффициента парной корреляции:

$$r = \frac{\sum_{t=1}^{T}\left(x_t - \overline{x}\right)\left(y_t - \overline{y}\right)}{\sqrt{\sum_{t=1}^{T}\left(x_t - \overline{x}\right)^2 \sum_{t=1}^{T}\left(y_t - \overline{y}\right)^2}}. \qquad (3.107)$$

Понятно, что значения этого коэффициента лежат в пределах от минус единицы до плюс единицы, но какую им дать интерпретацию? Ведь любой коэффициент тем и хорош, что с его помощью можно с той или иной степенью успешности диагностировать некоторую анализируемую ситуацию.

Коэффициент парной корреляции имеет четкую статистическую интерпретацию, но на практике приходится сталкиваться с тем, что его смысл существенно искажается, особенно в экономической практике. Вот, например, как рекомендуется понимать его значения при проведении маркетинговых исследований: «Абсолютная величина коэффициента корреляции характеризует тесноту связи, а знак указывает на ее направление»[1]. И далее рекомендуется так интерпретировать абсолютные значения коэффициента парной корреляции: если он лежит в пределах от 1 до 0,81, – считать, что связь сильная; если от 0,61 до 0,80, – что сила связи умеренная; от 0,41 до 0,60, – слабая; от 0,21 до 0,40 – очень слабая; от 0,00 до 0,19 – считать, что связь между переменными отсутствует. На эту ошибочную интерпретацию результатов ориентируются очень многие экономисты, а не только те, кто занимается маркетинговыми исследованиями. Поэтому остановимся более подробно на свойствах коэффициента парной корреляции с тем, чтобы прогнозист не допустил ошибки на этом важном этапе прогнозирования.

Прежде всего, следует заметить, что изначально при обосновании коэффициента парной корреляции,

[1] *Голубков Е. П.* Маркетинговые исследования: теория, методология и практика. М. : Финпресс, 1998. С. 244–245.

использовалась прямолинейная регрессия, поскольку, как отмечал еще в 1912 г. в монографии «Теория корреляции и элементы учения о кривых распределения» Е. Е. Слуцкий: «Теория криволинейной регрессии до сих пор разработана недостаточно, и мы в дальнейшем будем иметь дела главным образом с первым типом (*линейной регрессией*), к счастью, по крайней мере, в известном приближении, преобладающим в ряде случаев, с которыми имеет дело статистическая практика»[1]. Далее, имея в виду исключительно линейную регрессию, исследователь писал: «Корреляционная зависимость между явлениями может быть, ... и более, и менее тесной. Начиная от полной независимости и переходя через ряд градаций, она в пределе превращается в строго функциональную связь двух величин»[2]. При этом он предупреждал, что «равенство коэффициента корреляции нулю может свидетельствовать об отсутствии корреляции только лишь при условии линейности»[3]. Если коэффициент парной корреляции равен нулю, это вовсе не означает отсутствие корреляции между случайными переменными вообще, это свидетельствует только об отсутствии линейной корреляции.

Чтобы понять суть этих предупреждений, рассмотрим другой способ получения формулы (3.104)[4], тем более что именно так этот коэффициент и был выведен Пирсоном в конце XIX в.

Запишем уравнение линейной регрессии:

$$y_t = a_0 + a_1 x_t + \varepsilon_t. \tag{3.108}$$

Коэффициенты данной модели можно найти с помощью МНК. Но для упрощения процесса нахождения выборочных значений этих коэффициентов произведем центрирование исходных переменных относительно их средних:

$$\left(y_t - \overline{y}\right),\ \left(x_t - \overline{x}\right).$$

[1] *Слуцкий Е. Е.* Экономические и статистические произведения : избранное. М. : Эксмо, 2010. С. 691.

[2] Там же. С. 695.

[3] Там же. С. 731.

[4] *Венецкий И. Г., Венецкая В. И.* Основные математико-статистические понятия и формулы в экономическом анализе : справочник. М. : Статистика, 1979.

Тогда легко найти значение коэффициента регрессии:

$$a_1 = \frac{\sum_t (y_t - \overline{y})(x_t - \overline{x})}{\sum_t (x_t - \overline{x})^2}. \tag{3.109}$$

Значение свободного члена при этом будет равно нулю. Вспомним, что коэффициент a_1 характеризует тангенс угла α наклона линии регрессии к оси $0x$ (рис. 3.8).

Поскольку на этом этапе исследования нам неизвестно направление зависимости, а мы изучаем сам факт ее существования, то вполне возможно существование и обратного направления зависимости — зависимости x от y:

$$x_t = b_0 + b_1 y_t + \xi_t. \tag{3.110}$$

Для центрированных значений исходных переменных относительно их средних арифметических легко найти значение коэффициента регрессии этой зависимости с помощью МНК. Формула вычисления коэффициента регрессии для этого случая будет записана так:

$$b_1 = \frac{\sum_t (y_t - \overline{y})(x_t - \overline{x})}{\sum_t (y_t - \overline{y})^2}. \tag{3.111}$$

Этот коэффициент, как известно, характеризует угол β наклона прямой линии к оси $0y$, он равен тангенсу этого угла (рис. 3.8). Из рисунка легко заметить, что сумма углов α и β меньше $\pi / 2$:

$$\alpha + \beta \le \pi / 2. \tag{3.112}$$

Откуда со всей очевидностью имеем:

$$\operatorname{tg} \alpha \times \operatorname{tg} \beta \le 1. \tag{3.113}$$

Причем строгое равенство достигается только в том случае, когда сумма углов будет равна $\pi / 2$, т.е., когда две линии регрессии совпадают. Чем дальше эти две линии регрессии друг от друга, тем дальше сумма углов от $\pi / 2$ и дальше от единицы произведение тангенсов этих углов (3.113).

Значит, по величине (3.113) можно судить, насколько линии регрессии близки друг к другу и насколько, в конечном итоге, зависимость между двумя случайными величинами приближается к линейной.

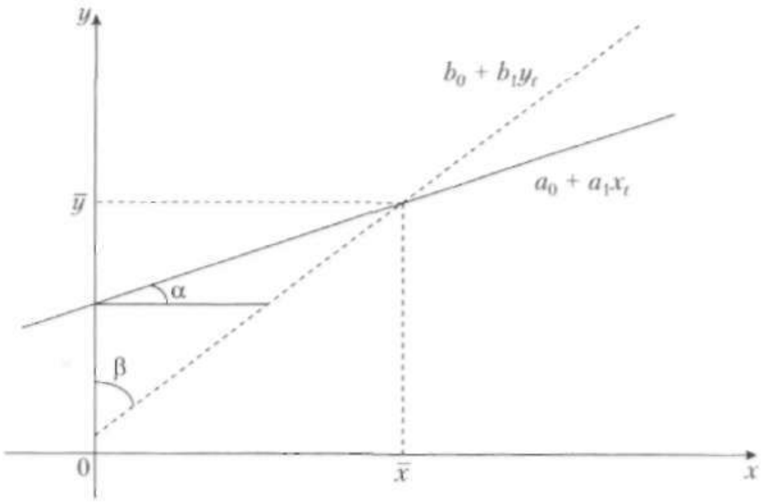

Рис. 3.8. **Графическое изображение линий регрессии *x* на *y* и *y* на *x***

Подставив в (3.113) вместо тангенсов углов их числовые выражения (a_1 и b_1), найденные ранее по формулам (3.109) и (3.111), и исчислив квадратный корень из этого произведения, получим:

$$r = \pm\sqrt{\operatorname{tg}\alpha \cdot \operatorname{tg}\beta} = \pm\sqrt{a_1 b_1} = \frac{\sum_{t=1}^{T}(x_t - \overline{x})(y_t - \overline{y})}{\sqrt{\sum_{t=1}^{T}(x_t - \overline{x})^2 \sum_{t=1}^{T}(y_t - \overline{y})^2}}. \quad (3.114)$$

Сравнив с (3.107), можно сделать вывод о том, что мы получили формулу выборочного значения коэффициента парной корреляции. Поэтому можно рассматривать *выборочное значение коэффициента парной корреляции как среднегеометрическое из коэффициентов регрессии, оцененных с помощью МНК.*

Что же характеризует выборочное значение коэффициента парной корреляции? Оно показывает, насколько зависимость между случайными переменными, если она есть, приближается к линейной. И только! Как было показано выше, из тех или иных значений коэффициента парной корреляции вовсе не следует вывод о наличии или отсутствии зависимости между случайными переменными. Первична гипотеза о наличии такой взаимосвязи, затем следует ее инструментальное подтверждение с помощью коэффициента парной корреляции.

Если поступать наоборот, то можно столкнуться с явлением *ложной корреляции, когда два совершенно независимых случайных показателя имеют высокое значение коэффициента парной корреляции*. Если два независимых показателя имеют монотонную тенденцию своего изменения, например, оба возрастают, то, нанеся эти точки на график рис. 3.8, можно убедиться в том, что они лягут на линию, с той или иной степенью приближающуюся к прямой. Коэффициент парной корреляции в этом случае будет показывать степень приближения этих точек к прямой линии и не будет как-то свидетельствовать о наличии или отсутствии корреляционной взаимосвязи между переменными. Если, например, рассчитать коэффициент парной корреляции между доходами на душу населения в Ненецком автономном округе с 1998 г. по настоящее время и численностью населения Индии за этот же промежуток времени, то можно убедиться в том, что этот коэффициент будет близок к единице. Но это высокое значение коэффициента парной корреляции, очевидно, совсем не говорит о наличии взаимозависимости между этими двумя переменными. Корреляция между ними будет ложной.

Таким образом, для того, чтобы избежать явления ложной корреляции, следует вначале исследования с помощью экономического анализа обосновать наличие зависимости между случайными переменными, а затем с помощью корреляционного анализа подтвердить приближение этой зависимости к линейной форме.

Если коэффициент парной корреляции по модулю близок к единице, то это подтверждает гипотезу о том, что между переменными может существовать линейная взаимосвязь.

Если коэффициент парной корреляции по модулю меньше, чем 0,8, но выше, чем 0,6, то можно говорить о том, что зависимость между переменными может быть представлена как линейная, но с очень большой погрешностью.

Если модуль коэффициента парной корреляции меньше, чем 0,6, то следует искать другую нелинейную форму зависимости.

Низкое значение коэффициента корреляции ни в коем случае не свидетельствует об отсутствии взаимосвязи вообще. Оно свидетельствует только об отсутствии линейной взаимосвязи. Например, для нелинейной функциональной зависимости $y_t = \sin x_t$ коэффициент парной корреляции между x и y будет почти равен нулю. Но равенство нулю коэффициента парной корреляции в этом случае, очевидно, не говорит об отсутствии взаимосвязи между y_t и x_t вообще, а свидетельствует именно об отсутствии *линейной* взаимосвязи между ними, поскольку функциональная нелинейная взаимосвязь очевидна.

Но что делать исследователю, если он убежден в наличии взаимосвязи между случайными факторами, а коэффициент парной корреляции по модулю близок к нулю? Как подтвердить наличие или отсутствие взаимосвязи вообще? Для этого можно использовать графический анализ взаимосвязи. Возможно, что точки на графике лягут так, что исследователь сможет предполагать наличие, например, экспоненциальной зависимости: $\hat{y}_t = a_0 e^{a_1 x_t}$.

Тогда для подтверждения этой гипотезы следует прологарифмировать левую и правую части равенства. Получим $\ln \hat{y}_t = \ln a_0 + a_1 x_t$.

Коэффициент парной корреляции необходимо находить между $\ln y_t$ и x_t. Близость его к единице по модулю будет подтверждать высказанную гипотезу о наличии корреляции между случайными факторами, корреляции, описываемой экспоненциальной зависимостью.

Так же можно поступать и при выдвижении гипотез о наличии нелинейных зависимостей других видов (показательной, степенной, квадратичной, логарифмической зависимости и т.п.).

Расчет и смысл коэффициента детерминации мы рассмотрели в параграфе 2.4. В случае с парной линейной регрессионной моделью он представляет собой квадрат коэффициента парной корреляции. Покажем это. Как мы помним, коэффициент детерминации рассчитывается по формуле

$$R^2 = \frac{ESS}{TSS} = \frac{\sum_t (\hat{y}_t - \bar{y})^2}{\sum_t (y_t - \bar{y})^2} = \frac{\sum_t (\hat{y}_t - \bar{y})(\hat{y}_t - \bar{y})}{\sum_t (y_t - \bar{y})^2}. \qquad (3.115)$$

Подставим в числитель (3.115) вместо одного расчетного значения значение из (3.108) и заменим второе расчетное значение фактическим с ошибкой:

$$R^2 = \frac{\sum_t (a_0 + a_1 x_t - \overline{y})(y_t - \varepsilon_t - \overline{y})}{\sum_t (y_t - \overline{y})^2} =$$

$$= \frac{\sum_t (a_0 + a_1 x_t - \overline{y})(y_t - \overline{y}) - \sum_t \varepsilon_t (a_0 + a_1 x_t - \overline{y})}{\sum_t (y_t - \overline{y})^2}. \quad (3.116)$$

Учитывая, что в регрессионных моделях, оцененных МНК, сумма ошибок и сумма произведений ошибок на независимую переменную равны нулю, вторую часть в числителе в (3.116) мы можем безболезненно отбросить. Одновременно с этим, так как регрессионная линия проходит через среднюю точку, мы можем заменить $\overline{y}$ на среднее расчетное значение: $\overline{y} = a_0 + a_1\overline{x}$.

В итоге получим:

$$R^2 = \frac{\sum_t (a_0 + a_1 x_t - a_0 + a_1\overline{x})(y_t - \overline{y})}{\sum_t (y_t - \overline{y})^2} = \frac{a_1 \sum_t (x_t - \overline{x})(y_t - \overline{y})}{\sum_t (y_t - \overline{y})^2}. \quad (3.117)$$

Теперь вместо a_1 подставим в (3.117) формулу (3.109) для того, чтобы получить окончательный результат:

$$R^2 = \frac{\sum_t (x_t - \overline{x})(y_t - \overline{y})}{\sum_t (x_t - \overline{x})^2} \frac{\sum_t (x_t - \overline{x})(y_t - \overline{y})}{\sum_t (y_t - \overline{y})^2} =$$

$$= \frac{\left(\sum_t (x_t - \overline{x})(y_t - \overline{y})\right)^2}{\sum_t (x_t - \overline{x})^2 \sum_t (y_t - \overline{y})^2} = r^2. \quad (3.118)$$

Коэффициент детерминации в этом случае имеет простой смысл — его значения показывают, на сколько процентов линейная динамика одного показателя соответствует линейной динамике другого показателя. Например, если коэффициент парной корреляции равен 0,9, то коэффициент детерминации в соответствии с (3.118) будет равен квадрату коэффициента парной корреляции, т.е. 0,81. Это говорит о том, что изменение каждого из показателей на 81% может быть объяснено линейным влиянием другого показателя, а 19% этого изменения может быть отнесено к влиянию иных факторов.

Данный вывод носит предположительный характер, поскольку является индуктивным. Для того чтобы оценить, насколько можно доверять подобному высказыванию, необходимо оценить доверительные границы для расчетных выборочных значений как коэффициента парной корреляции, так и коэффициента детерминации.

Для расчета доверительного интервала для коэффициента корреляции существует формальная процедура проверки статистической гипотезы. Нулевая гипотеза в данном случае заключается в том, что коэффициент корреляции равен нулю, а альтернативная обычно — в том, что не равен:

$$H_0: r = 0;$$
$$H_1: r \neq 0.$$

Проверяется эта гипотеза путем расчета t-статистики по формуле

$$t = \frac{r\sqrt{n-1}}{\sqrt{1-r^2}}. \tag{3.119}$$

Процедура проверки гипотезы стандартна: расчетное значение сравнивается с табличным $t(\alpha, n-1)$. Если расчетное значение оказалось по модулю меньше табличного, делается вывод о том, что нет оснований отклонить нулевую гипотезу. Этот результат может либо сигнализировать о недостаточности данных, либо об отсутствии линейной связи между исследуемыми показателями.

Основной вывод, который следует из материалов данного параграфа, заключается в том, что корреляционный анализ так и не дает ответа на главный вопрос о наличии или отсутствии взаимосвязи между случайными факторами. Его основные коэффициенты (парной корреляции и коэффициент детерминации) позволяют лишь оценить степень приближения случайной зависимости к линейной, если эта зависимость предварительно обоснована аналитически.

3.7. Коэффициент согласия в динамике

Ранее говорилось, что выводы математической статистики — типичные выводы по индукции. Структурно различают четыре вида научной индукции, в логике традиционно называемых методами[1].

1. Метод единственного сходства.
2. Метод единственного различия.
3. Метод сопутствующих изменений.
4. Метод остатков.

Тщательное изучение каждого из этих методов дается в учебниках по логике, поэтому здесь мы не будем останавливаться на особенностях каждого из них. Из этих методов нас интересуют два — метод единственного сходства и метод единственного различия, поскольку первый из них является методологической основой корреляционного анализа, а второй — дает базу для вычисления коэффициента согласия в динамике[2]. Знакомство с указанными методами даст возможность понять различие между подходами к выявлению корреляции и суть каждого из них.

Метод единственного сходства — это умозаключение о причине наблюдаемого явления, основанное на сравнении нескольких случаев, влекущих за собой это явление.

Если два или более случая исследуемого (наблюдаемого) явления связаны только с одним (из нескольких) общим, предшествующим явлению обстоятельством, то оно и есть причина или часть причины исследуемого (наблюдаемого) явления.

Схематически структура данного метода может быть выражена следующим образом:

случай 1: $Б_1$**В**$Г_1$ — обстоятельства, предшествующие явлению «**а**»;
случай 2: $Б_2Г_2$**В** — обстоятельства, предшествующие явлению «**а**»;
случай 3: **В**$Б_3Г_3$ — обстоятельства, предшествующие явлению «**а**»;
случай 4: $Б_4Г_4$**В** — обстоятельства, предшествующие явлению «**а**».

«**В**» является причиной или частью причины явления «**а**».

Именно этот вид индуктивного вывода и лежит в основе вычисления коэффициента парной корреляции. Действительно, логику этой части корреляционного анализа можно сформулировать так: имеется ряд значений y_t и ряд значений x_t, которые можно рассматривать, в соответствии со

[1] *Брюшинкин В. Н.* Логика : учебник. М. : Гардарики, 2001. С. 262–269.
[2] *Светуньков С. Г.* Количественные методы прогнозирования эволюционных процессов. Ульяновск : Изд-во УлГУ, 1997. С. 62–74.

схематической структурой метода единственного сходства как обстоятельства $Б_1$, $Б_2$, $Б_3$ ..., $Г_1$, $Г_2$, $Г_3$... Между ними имеется линейная зависимость (обстоятельство «**В**»). Для таких обстоятельств коэффициент корреляции близок к единице (явление «**а**»). Тогда и следует вывод: линейная зависимость («**В**») и является причиной высокого значения коэффициента парной корреляции («**а**»).

Как было показано в предыдущем параграфе, если все точки двух переменных, расположенных на одной плоскости, выстраиваются в фигуру, напоминающую прямую линию, коэффициент парной корреляции между ними всегда будет близок к единице (структура метода единственного сходства дает в данном случае правильный вывод). Другое дело, что например, возможно появление третьего фактора z_t, который и является истиной причиной высокого значения коэффициента парной корреляции, а он в этой схеме не рассматривается. Но на то он и индуктивный вывод, что дает вывод не однозначно истинный, а возможный.

Такой же вывод можно получить с помощью метода единственного различия, который представляет собой умозаключение о причине наблюдаемого явления, основанное на сравнении всего лишь двух случаев: когда интересующее нас явление имеет место и когда его нет. Если случай, в котором явление присутствует, отличается от случая, в котором его нет, только одним предшествующим явлению обстоятельством, то именно это обстоятельство и является причиной или частью причины данного явления:

случай 1: **В**СД — обстоятельства, повлекшие явление «**а**»;
случай 2: СД — обстоятельства, не повлекшие явление, т.е. «—»;

«**В**» — причина или часть причины явления «**а**».

Корреляционный анализ базируется на априорном предположении о том, что изменения случайных факторов могут быть какими угодно, но они линейны относительно друг друга. Если же эти изменения относительно друг друга носят более сложный нелинейный характер, корреляционный анализ оказывается бессильным для выявления этой сложной взаимосвязи.

Взаимосвязь факторов, как известно, проявляется не только в наличии некоторой тенденции взаимного изменения, которую корреляционный анализ и пытается оценить, но и в характере отклонений от тенденций во времени каждого из них. Если, например, между двумя факторами имеется некоторая корреляционная взаимосвязь (не обязательно линейная), то это означает, что резкое изменение одного фактора вызовет аналогичное изменение другого фактора.

Зная это, можно предложить подход, базирующийся на методе единственного различия. Имеется ряд значений y_t и ряд значений x_t, которые можно рассматривать, в соответствии со схематической структурой метода единственного различия как обстоятельства С и Д. Корреляционная взаимосвязь между факторами (обстоятельство «**В**») приводит к тому, что при изменении в тенденциях одного фактора наблюдаются изменения в тенденциях другого фактора (явление «**а**»). Когда между факторами нет взаимосвязи (обстоятельство «В» отсутствует), то динамика факторов «рассогласованна», т.е. резкий рост одного фактора никак не влияет на динамику другого (отсутствует явление «**а**»).

Следовательно, если удастся найти коэффициент, который будет характеризовать степень согласованности динамики показателей, по его значению можно будет судить о наличии или отсутствии взаимосвязи между факторами. Очевидно, что и в этом случае будет иметь место вывод по индукции, носящий предположительный, а не истинный характер.

Покажем, как можно получить коэффициент, обладающий подобными свойствами.

Известно, что по всем T наблюдениям о показателе y_t ($t = 1, 2, 3, ..., T$), имеющимся в распоряжении исследователя, можно построить степенную функцию ($T - 1$)-й степени:

$$\hat{y}_t = a_0 + a_1 t + a_2 t^2 + ... + a_{T-1} t^{T-1}. \qquad (3.120)$$

Причем этот полином так опишет исходный ряд наблюдений y_t, что его расчетные значения $\hat{y}_t$ в каждой t-й точке в точности будут соответствовать фактическим значениям y_t:

$$y_t = \hat{y}_t. \qquad (3.121)$$

Коэффициенты полинома (3.120) определяют динамические свойства анализируемого ряда. Поэтому по их значениям можно судить об исходном ряде и его свойствах. Очевидно, что каждый ряд, отличающийся в своей динамике от другого ряда, будет отличаться от него и своими коэффициентами, если оба ряда представить в виде модели (3.120). А если ряды в динамике меняются в среднем одинаково, то и коэффициенты степенной функции, в которую можно разложить любой ряд, будут во многом совпадать. Поэтому сравнительный анализ коэффициентов при одинаковых степенях может позволить получить искомую информацию о взаимосвязи факторов.

Но для того, чтобы осуществить такой сравнительный анализ, необходимо вычислить коэффициенты модели разложения функции в степенной ряд (3.120). Для этого есть несколько возможных вариантов:

1) построить степенные функции (T – 1)-й степени (3.120), решая системы из T уравнений с T неизвестными коэффициентами a;

2) воспользоваться разложением функций в степенные ряды (ряды Тейлора) и анализировать соответствующие члены степенного ряда;

3) оценить параметры степенных функций с помощью аппарата конечных разностей;

4) найти различные производные степенной функции (3.120) с помощью метода конечных разностей, поскольку они вполне характеризуют соответствующие коэффициенты.

Первый вариант предполагает решение системы T уравнений с T неизвестными, что в ряде случаев при большом числе наблюдений T представляет собой непростую вычислительную задачу, которую неразумно рекомендовать к широкому практическому использованию. К тому же полиномы высоких степеней очень неустойчивы к ошибкам округлений.

Второй вариант предполагает представление в явном виде самой функции, разложение которой необходимо осуществить, а она не известна. Существует большое количество различных методов, позволяющих найти приближенные коэффициенты рядов Тейлора. Однако их множество вызвано именно различной степенью неточности такого приближения и с удовлетворительной точностью этот вариант решен быть не может.

Третий вариант — оценить параметры степенных функций с помощью аппарата конечных разностей — значительно проще, чем все рассмотренные выше варианты, однако и он достаточно громоздок. Рассмотрим его более подробно, так как на основе этого рассмотрения можно будет увидеть особенности простого использования четвертого варианта.

Пусть нам дана линейная однофакторная модель

$$\hat{y}_t = a_0 + a_1 t, \tag{3.122}$$

параметры которой a_0 и a_1 следует найти на множестве наблюдений T.

Метод конечных разностей исходит из того факта, что первая производная линейной функции (3.122) будет равна a_1, а одной из оценок этой производной может быть средняя первая разность.

Обозначим через $\Delta^1 y_t$ первую разность ряда в точке t. Эта разность вычисляется элементарно:

$$\Delta^1 y_t = y_t - y_{t-1}. \qquad (3.123)$$

Она характеризует первую производную функции в точке t. Общей характеристикой первой производной на всем множестве значений может служить средняя этих разностей. Но поскольку первых разностей на единицу меньше, чем количество исходных членов ряда, то эта средняя будет вычисляться со второго члена ряда до последнего T-го:

$$\overline{\Delta}^1 y = \frac{1}{T-1} \sum_{t=2}^{T} \Delta^1 y_t. \qquad (3.124)$$

Подставляя в (3.124) значения разностей, полученные с помощью (3.123), и осуществив элементарные сокращения, получим окончательно:

$$\overline{\Delta}^1 y = \frac{1}{T-1} \left(y_T - y_1 \right). \qquad (3.125)$$

Эта средняя разность характеризует первую производную функции и дает некоторую оценку коэффициента a_1 линейной функции:

$$a_1 \approx \overline{\Delta}^1 y. \qquad (3.126)$$

Для нахождения параметра a_0 модели (3.122) по известному значению коэффициента a_1 необходимо найти средние значения y и t, подставить их в (3.122) и, решив уравнение, определить значение a_0.

Такую же процедуру можно провести с рядом значений другого показателя, который обозначим через x_t: $\hat{x}_t = b_0 + b_1 t$.

После этого можно сравнить друг с другом коэффициенты a_0 и b_0, a_1 и b_1. Очевидно, что они должны быть либо безразмерными, либо одной размерности.

Если теперь предположить наличие тенденции роста, описываемой с помощью параболы второй степени

$$\hat{y}_t = a_0 + a_1 t + a_2 t^2, \qquad (3.127)$$

то для определения параметра a_2 этой модели необходимо рассчитать вторые разности:

$$\Delta^2 y_t = \Delta^1 y_t - \Delta^1 y_{t-1} = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = y_t - 2y_{t-1} + y_{t-2}, \qquad (3.128)$$

а затем найти среднюю вторых разностей. Поскольку вторых разностей на единицу меньше, чем первых, средняя вторых разностей вычисляется так:

$$\overline{\Delta}^2 y = \frac{1}{T-2} \sum_{t=3}^{T} \Delta^2 y_t. \qquad (3.129)$$

Известно, что вторая разность характеризует вторую производную любой функции, а применительно к функции (3.127) она будет характеризовать значение коэффициента a_2, поскольку вторая производная указанной функции равна $2a_2$.

Подставляя в (3.129) значения вторых разностей из (3.128), которые, в свою очередь, определяются через первые разности (3.123), после ряда элементарных сокращений получим

$$\overline{\Delta}^2 y = \frac{1}{T-2}\left(\Delta^1 y_T - \Delta^1 y_2\right). \qquad (3.130)$$

Отсюда легко найти оценку коэффициента a_2:

$$a_2 = \frac{\overline{\Delta}^2 y}{2}. \qquad (3.131)$$

Теперь для нахождения неизвестных значений коэффициентов a_0 и a_1 следует воспользоваться процедурой, описанной для линейной модели, и последовательно найти a_1, а затем a_0.

Продолжая подобным образом, можно найти коэффициенты полинома третьей, четвертой и вообще любой степени N. Так, например, для полинома третьей степени третья разность будет находиться так:

$$\Delta^3 y_t = \Delta^2 y_t - \Delta^2 y_{t-1} = (y_t - 2y_{t-1} + y_{t-2}) - (y_{t-1} - 2y_{t-2} + y_{t-3}) = \\ = y_t - 3y_{t-1} + 3y_{t-2} - y_{t-3}. \qquad (3.132)$$

Средняя третьих разностей равна

$$\bar{\Delta}^3 y = \frac{1}{T-3}\sum_{t=4}^{T}\Delta^3 y_t = \frac{1}{T-3}\left(\Delta^2 y_T - \Delta^2 y_2\right). \qquad (3.133)$$

Эта разность в свою очередь характеризует третью производную функции, равную для параболы третьей степени $6a_3$. Тогда коэффициент a_3 можно найти так:

$$a_3 = \frac{\bar{\Delta}^3 y}{6}. \qquad (3.134)$$

Очевидно, что трудоемкость такого способа нахождения параметров параболы (T – 1)-й степени меньше, чем трудоемкость построения степенных функций (T – 1)-й степени, при решении системы из T уравнений с T неизвестными коэффициентами, или трудоемкость разложения функций в степенные ряды Тейлора. Однако и его можно упростить, вычисляя лишь конечные разности, которые характеризуют конечные производные функций.

Таким образом, для оценки коэффициентов полиномов a_1, a_2 и a_3 найденные оценки производных с помощью конечных разностей делились для коэффициента a_1 на 1 (3.126), для коэффициента a_2 – на 2 (3.131), для коэффициента a_3 – на 6 (3.133) и т.п. В случае расчета оценок производных надобность в подобной процедуре умножения коэффициентов при показателях, возводимых в одинаковую степень, отпадает, что значительно упрощает расчеты. Кроме того, при расчете средних разностей сумму разностей необходимо было делить на число этих разностей. Без этой процедуры также можно обойтись. Действительно, если необходимо сравнить вторую производную данного ряда со второй производной другого ряда, то от умножения или деления каждой производной на одну и ту же константу, не равную нулю, результат отношения не изменится. Поэтому процедура еще более упрощается – следует находить не среднюю разностей, а просто их сумму. Сравнив формулы для вычисления конечных разностей (3.123), (3.128) и (3.132), легко обнаружить закономерность, с помощью которой можно сгенерировать формулу для вычисления конечной разности любого порядка. Но поскольку подобная задача уже решена применительно к интерполяционной формуле Ньютона, воспользовавшись этим, запишем формулу вычисления конечной разности n-го порядка в точке t:

$$\Delta^n y_t = \sum_{k=0}^{n} (-1)^k C_n^k y_{t-k}, \qquad (3.135)$$

где k — порядковый номер члена разности (3.133); k = 1, 2, 3, ..., n; C — число всех сочетаний из n различных элементов по k, которое, как известно, находится по формуле:

$$C_n^k = \frac{n!}{k!(n-k)!},$$

в которой знак «!»— это знак факториала. Напомним, что по определению 0! = 1.

Сумма n-х конечных разностей, т.е. не усредненная оценка n-й производной данного ряда, определяется по формуле:

$$\overline{\Delta}^n y = \left(\Delta^{n-1} y_T - \Delta^{n-1} y_n\right), \qquad (3.136)$$

где $\Delta^{n-1} y_T$ — последняя n-я конечная разность в ряду вычисляемых n-х конечных разностей для исходных значений y_t, т.е. n-я разность при $t = T$; $\Delta^{n-1} y_n$ — первая в ряду вычисляемых n-х конечных разностей, когда наблюдение $t = n$.

Поскольку любой прогнозист без труда может вычислить конечные разности (3.135), а на их основе — сумму этих конечных разностей (3.136), то для любых двух рядов y_t и x_t, состоящих из T наблюдений, можно получить $(T - 1)$ пар чисел:

$$\left(\overline{\Delta}^1 y; \overline{\Delta}^1 x\right), \left(\overline{\Delta}^2 y; \overline{\Delta}^2 x\right), \left(\overline{\Delta}^3 y; \overline{\Delta}^3 x\right), ..., \left(\overline{\Delta}^{T-1} y; \overline{\Delta}^{T-1} x\right), \qquad (3.137)$$

которые характеризуют первые, вторые, третьи и т.д. до $(T - 1)$-й производные указанных рядов. С учетом того, что производные наилучшим образом характеризуют динамику процессов, сравнение указанных пар значений может характеризовать степень согласия в динамике рядов y_t и x_t. Очевидно, что сравнение пар значений производных следует производить так, чтобы итогом данного сравнения стал некоторый расчетный коэффициент, имеющий ясное смысловое толкование. Сам коэффициент будет характеризовать согласие двух исходных рядов в динамике, поэтому его имеет смысл назвать именно так.

Чтобы предложить формулу вычисления коэффициента согласия в динамике, воспользуемся неравенством Коши, которое для конечных сумм имеет вид

$$\left(\sum_t y_t x_t\right)^2 \le \sum_t y_t^2 \sum_t x_t^2. \tag{3.138}$$

Если левую часть указанного неравенства разделить на его правую часть, будет получено выражение, которое никогда не будет больше единицы:

$$\frac{\left(\sum_t y_t x_t\right)^2}{\sum_t y_t^2 \sum_t x_t^2} \le 1. \tag{3.139}$$

Равенство единице может быть только в том случае, когда ряд значений y_t представляет собой линейное преобразование ряда x_t.

Если вместо пар значений y_t и x_t подставить в предлагаемую формулу пары значений сумм конечных разностей (3.137) и извлечь квадратный корень из полученного выражения, то получим формулу, внешне напоминающую формулу для расчета коэффициента парной корреляции, который и назовем «коэффициент согласия в динамике»:

$$k = \frac{\sum_t \bar{\Delta}^t y \bar{\Delta}^t x}{\sqrt{\sum_t \left(\bar{\Delta}^t y\right)^2 \sum_t \left(\bar{\Delta}^t x\right)^2}}. \tag{3.140}$$

К этому коэффициенту можно будет применить толкование, аналогичное толкованию значений коэффициента парной корреляции.

Поскольку сумма разностей может быть как положительной, так и отрицательной, то и итоговое значение коэффициента согласия в динамике может быть как положительным, так и отрицательным. Для того чтобы понять, что именно отражает коэффициент согласия в динамике, представим, что ряды y_t и x_t находятся в линейной функциональной зависимости друг от друга, но под воздействием внешних по отношению к ним факторов меняются от наблюдения к наблюдению по сложному нелинейному закону. Поскольку они связаны друг с другом линейной функциональной зависимостью, их производные будут также линейно зависеть друг от друга, а неравенство Коши превращается в таком случае в равенство, поэтому коэффициент (3.140) будет равен единице.

Тогда можно утверждать, что в случае если значение коэффициента (3.140) близко по модулю к единице, динамика двух исходных рядов y_t и x_t находится в очень сильной степени согласия друг с другом, поскольку значения их производных разного порядка соответствуют друг другу. Согласованность их динамики свидетельствует о том, что между ними весьма вероятна взаимосвязь.

Если же вычисленное значение модуля коэффициента (3.140) близко к нулю, следует признать, что динамика двух рядов различна и наличие взаимосвязи между ними вряд ли возможно, ведь производные высших порядков рассогласованны, а это значит, что динамика одного ряда y_t не похожа на динамику другого ряда x_t и его изменение вряд ли вызвано действием фактора, отражаемого рядом значений x_t. В отличие от коэффициента парной корреляции, значение коэффициента согласия в динамике, близкое к нулю, однозначно указывает на то, что между двумя рядами взаимосвязи нет.

Так о чем же свидетельствуют значения коэффициента согласия в динамике, лежащие по модулю в пределах от 0 до 1? Для ответа на этот вопрос продемонстрируем использование данного коэффициента на условных данных.

Если ряд y_t представляет собой синусоиду, изменяющуюся во времени t, а ряд x_t — косинусоиду с той же амплитудой и фазой — $y_t = \sin t$; $x_t = \cos t$, то для 100 точек, снятых в течение одного периода этих гармонических функций, коэффициент парной корреляции составит величину, равную $r = -0{,}0013674$, что свидетельствует об очевидном отсутствии линейной взаимосвязи. При этом, кстати, многие практикующие экономисты сделают вывод об отсутствии взаимосвязи между y_t и x_t вообще.

Коэффициент согласия в динамике дает величину, равную $k = 0{,}9796672$, что свидетельствует о наличии очень сильного соответствия в динамике показателей. Это означает, что у исследователя, получившего такое значение коэффициента согласия в динамике, есть все основания предполагать наличие взаимосвязи между факторами.

Из данного примера можно сделать следующий вывод: если для изучаемых рядов коэффициент парной корреляции по модулю близок к нулю, а модуль коэффициента согласия при этом близок к единице, это свидетельствует о наличии тесной нелинейной взаимосвязи между исследуемыми факторами.

Для подтверждения свойств коэффициента согласия в динамике для выявления возможной взаимосвязи между факторами проведем расчеты еще на одном примере, условные данные которого приведены в табл. 3.4.

Таблица 3.4

Условные данные для расчета коэффициента согласия в динамике

t	1	2	3	4	5	6	7	8	9
y_t	0,0	1,0	2,5	4,0	7,0	7,0	6,0	7,0	8,5
x_t	1,0	1,5	2,0	3,0	2,0	1,5	2,0	2,5	3,0

Данные в этой таблице подобраны таким образом, чтобы резкий рост ряда x_t на четвертом наблюдении и его последующее падение вплоть до седьмого наблюдения могли бы быть причиной аналогичной динамики показателя y_t со сдвижкой во времени — показатель y_t растет на пятом наблюдении и начинает падение на следующем шаге вплоть до восьмого наблюдения (иначе говоря, между ними имеется лаг в единицу). Это отражено на графике рис. 3.9, который построен по данным табл. 3.4.

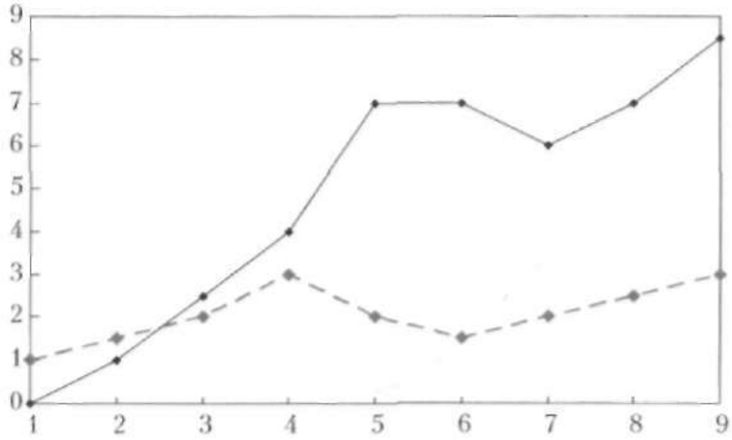

Рис. 3.9. **Динамика условных рядов данных:**
сплошная линия — ряд y; *прерывистая линия* — ряд x

Расчеты коэффициента парной корреляции и коэффициента согласия в динамике для этих двух рядов показывают следующее. Коэффициент парной корреляции между анализируемыми рядами равен $r = 0{,}5828$, а коэффициент согласия в динамике — $k = 0{,}9918$.

Различие между этими коэффициентами довольно велико. Чтобы понять причину такого различия, обратимся к графику зависимости y_t от x_t (рис. 3.10).

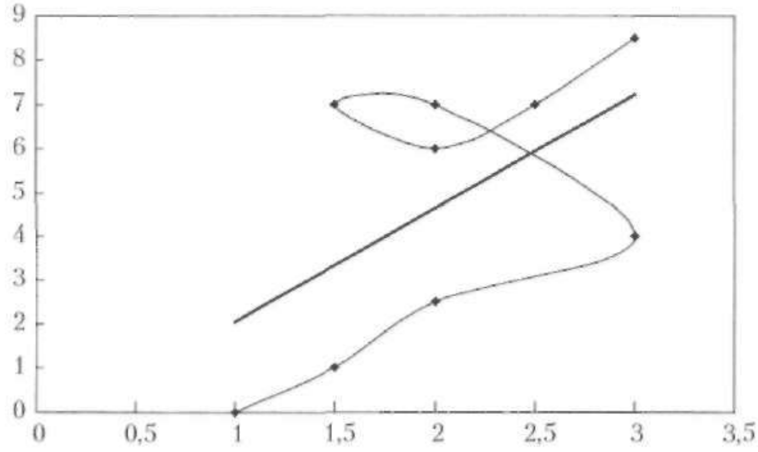

Рис. 3.10. **Совместное изменение x и y во времени:** *жирная линия по центру* — прямая парной регрессии

Как видно из графика, предположить наличие линейной зависимости между этими двумя рядами сложно, а именно наличие линейной корреляционной связи между случайными факторами и ищет коэффициент парной корреляции. Кроме того, если рассматривать эти два ряда так, как будто они засорены некоторыми случайными ошибками, и попытаться построить прямую линию, проходящую через них (она изображена на рисунке 3.10), можно убедиться в том, что ошибка аппроксимации такой линии будет очень высокой, что и демонстрирует значение коэффициента парной корреляции, равное 0,5828. Коэффициент парной корреляции не «уловил» возможную взаимосвязь между двумя рядами, а коэффициент согласия в динамике это сделал. Высокое значение коэффициента согласия в динамике по модулю выявляет наличие сложной взаимосвязи между рядами, которое формальный корреляционный анализ не подтверждает. Коэффициент согласия в динамике достаточно высок и свидетельствует о возможности обратной взаимосвязи между факторами, так как характеры динамики факторов близки друг к другу.

Для того чтобы установить вид взаимосвязи между факторами, необходимо осуществить дополнительные исследования, в том числе и расчет лагов между рядами посредством расчета кросс-корреляций. Проведем такое исследование. В табл. 3.5 приведены значения коэффициентов кросс-корреляций между значениями x_t и y_{t+1}.

Таблица 3.5

Условные данные для расчета коэффициента согласия в динамике

1	$r(x_t; y_{t+1})$
0	0,5828
1	0,7655
2	0,6449
3	0,1629

В нашем условном примере было мало наблюдений, тем не менее, по значениям, показанным в табл. 3.5, можно заметить, что между рядом x_t и рядом y_t, сдвинутым на 1, существует корреляция выше средней. Именно это мы и наблюдали по графику, это нам показал и коэффициент согласия в динамике.

Отсюда следует вывод: если для изучаемых рядов коэффициент парной корреляции мал, а модуль коэффициента согласия при этом близок к единице, это свидетельствует о наличии тесной взаимосвязи между исследуемыми факторами, возможно, нелинейной, а возможно — линейной с лагами.

Указанные примеры показывают, что предлагаемый коэффициент согласия в динамике действительно дает исследователю новую информацию, весьма важную для того, чтобы решить задачу определения степени взаимосвязи между факторами.

Таким образом, можно сделать вывод о том, что коэффициент согласия в динамике можно использовать в качестве дополнительного инструмента, позволяющего уточнить наличие взаимосвязи между факторами. Совместное использование сравниваемых коэффициентов дает прогнозисту значительно больше информации, чем простое использование каждого из них в отдельности. Коэффициент парной корреляции показывает исследователю, насколько тенденции двух изучаемых рядов совпадают в их линейном росте, а коэффициент согласия в динамике — насколько вариации одного показателя соответствуют вариациям другого.

Высокое значение модуля коэффициента согласия в динамике свидетельствует о том, что между изучаемыми факторами возможна тесная взаимосвязь. Если при этом коэффициент парной корреляции близок к единице, можно утверждать, что зависимость между факторами приближается к линейной. Если же коэффициент парной корреляции по модулю далек от единицы, прогнозисту необходимо осуществить дополнительные исследования для выявления формы взаимосвязи между изучаемыми случайными переменными.

Практикум

Вопросы для самоконтроля

1. Что такое «генеральная совокупность»?
2. Что такое «выборочная совокупность»?
3. Как связаны друг с другом выборочная совокупность и генеральная совокупность?
4. Почему выборочный метод не всегда может использоваться в социально-экономическом прогнозировании?
5. Что характеризует средняя величина? Какие виды средних величин могут использоваться для оценки свойств генеральной совокупности?
6. Дисперсия является важной статистической характеристикой. Как она вычисляется и что она характеризует? Чем отличаются друг от друга дисперсия и выборочная дисперсия?
7. Почему при работе с выборочными значениями необходимо находить интервальные оценки величин? Что собой представляют точечные оценки?
8. Приведите общую схему оценки доверительных границ средней арифметической для случая, когда распределение приближается к нормальному.
9. Как логика оценки доверительных границ распространяется на другие точечные оценки — коэффициентов регрессионных моделей, коэффициента парной корреляции и т.п.?
10. Что называется «статистической гипотезой»?
11. Что такое «нулевая гипотеза»? Что представляет собой «альтернативная гипотеза»?

12. Как можно определить понятия «доверительная вероятность» и «остаточная вероятность»?

13. В каком случае используется распределение Стьюдента?

14. Как друг с другом связаны нормальное распределение и распределение «хи-квадрат»?

15. Когда используется распределение Фишера?

16. Почему МНК является лучшим методом оценки коэффициентов регрессионных моделей по выборочным наблюдениям в случае нормального распределения?

17. Как использовать МНК для оценки мультипликативных моделей? Что представляет собой процедура линеаризации?

18. В каких случаях использования МНК могут быть получены смещенные оценки?

19. Что понимается под корреляционным анализом?

20. Что характеризует корреляционное отношение?

21. Что характеризует коэффициент парной корреляции?

22. Что собой представляет коэффициент детерминации?

23. В каких пределах изменяется коэффициент парной корреляции? Что означают те или иные значения коэффициента парной корреляции?

24. Как выявить нелинейную взаимосвязь между случайными факторами?

25. В чем логика вычисления коэффициента согласия в динамике?

26. Что дает прогнозисту совместное использование коэффициента парной корреляции и коэффициента согласия в динамике?

Задания

Задание 1. По 15 наблюдениям о стоимости чипсов в Санкт-Петербурге студент рассчитал среднюю величину и простую (неисправленную) дисперсию, которые получились соответственно 48,5 и 225 руб.

1. Рассчитайте исправленную дисперсию для этой ситуации.
2. Рассчитайте дисперсию средней величины.
3. Проверьте гипотезу о том, что средняя стоимость чипсов в генеральной совокупности равна 50.
4. Проверьте гипотезу о том, что средняя стоимость чипсов в генеральной совокупности равна 45. Противоречат ли результаты проверки данной гипотезы результатам проверки в п. 3?
5. На основе t-статистики Стьюдента постройте 90% доверительный интервал для средней величины.
6. Проверьте гипотезу о том, что дисперсия средней величины равна 10.

Студент также собрал данные по стоимости чипсов в Москве. По 13 наблюдениям средняя величина составила 53 руб., а дисперсия — 144. Проверьте гипотезу о том, что средняя стоимость чипсов в Санкт-Петербурге и Москве одинакова.

Задание 2. Студент решил собирать данные о стоимости своих обедов по дням и о количестве пар в день (см. данные таблицы).

Затраты на обед в столовой (в руб.)	150	120	125	130	150	145	135	120	100	150
Количество пар в день	5	2	2	2	4	3	3	2	1	4

1. По этим данным рассчитайте значение коэффициента корреляции. О чем говорит полученное значение?
2. Постройте точечную диаграмму зависимости стоимости обеда от количества пар.
3. Постройте модель линейной парной регрессии зависимости стоимости обеда от количества пар в день. Нанесите расчетные значения на построенную ранее точечную диаграмму.
4. Как вы считаете, насколько хорошо линейная модель описывает эту зависимость? Проведите расчеты, которые подкрепляли бы ваш вывод.
5. Студент решил, что, возможно, зависимость может носить нелинейный характер и построил степенную модель вида: $\hat{y}_t = a_0 x_t^{a_1}$. Рассчитайте коэффициенты этой модели и нанесите расчетные значения на график.
6. Проведите корректировку указанной в п. 5 модели так, чтобы избежать систематического завышения (чтобы сумма ошибок была равна нулю).
7. Оцените точность второй модели. Проведите расчеты, которые помогли бы вам это сделать.
8. Сравните модель из п. 3 с моделью из п. 5. Какой из них вы бы отдали предпочтение и почему?

Глава 4
МЕТОДЫ И МЕТОДИКИ ПРОГНОЗИРОВАНИЯ СТАЦИОНАРНЫХ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ПРОЦЕССОВ

В результате освоения данной главы студент должен:

знать

- типовые прогнозные модели;
- суть задачи построения прогнозной модели, ограниченность их практического применения;
- основные результаты новейших исследований в области многофакторного моделирования;
- основные понятия, методы и инструменты количественного и качественного анализа сложных многофакторных социально-экономических процессов;

уметь

- применить адекватную модель социально-экономического прогнозирования;
- строить многофакторные модели в условиях мультиколлинеарности;
- обрабатывать эмпирические и экспериментальные данные, а также проводить количественное прогнозирование сложных многофакторных социально-экономических явлений;
- учитывать при прогнозировании информацию о факторах, измеренных в неметрических шкалах;

владеть

- навыками самостоятельной научной и исследовательской работы в части статистической обработки данных о сложных многофакторных явлениях;
- прикладными методами построения однофакторных и многофакторных моделей в условиях мультиколлинеарности.

4.1. Типовые прогнозные модели

В том случае, когда внешнее окружение стационарного процесса меняется, условия становятся неоднородными. Эти неоднородные изменения внешней среды оказывают воздействие на прогнозируемый социально-экономический процесс, являясь причиной изменения его количественных показателей. В то же время, внутренняя среда объекта социально-экономического процесса не претерпевает качественных изменений — она только меняет свои количественные характеристики, отвечая на изменившееся влияние внешней среды. Для изучаемого процесса появляется новая по сравнению с однородным процессом характеристика — его математическое ожидание представляет собой функцию от изменяющихся внешних факторов (что в математической статистике называется условным математическим ожиданием):

$$M\left(y_t \mid x_{j,t}\right) = f\left(x_{j,t}\right). \tag{4.1}$$

Здесь $M(y_t \mid x_{j,t})$ — условное математическое ожидание значений прогнозируемого процесса; x_j — значения факторов внешней среды, оказывающих влияние на прогнозируемый процесс; j — номер фактора внешней среды ($j = 1, 2, 3, ..., k$).

Эта реакция социально-экономического объекта, процесс развития которого и необходимо прогнозировать, находится под влиянием множества случайных факторов, в результате чего ряд значений y_t прогнозируемого показателя на наблюдении t измеряется с некоторой случайной ошибкой ε_t. Поэтому вместо функциональной зависимости (4.1) прогнозисту приходится иметь дело с регрессионной зависимостью:

$$y_t = f\left(x_{j,t}\right) + \varepsilon_t. \tag{4.2}$$

Самый простой случай для прогнозирования — когда число изменяющихся факторов равно единице. В этом случае прогнозисту приходится иметь дело с однофакторной моделью:

$$y_t = f\left(x_t\right) + \varepsilon_t, \tag{4.3}$$

форму которой ему необходимо выявить, после чего следует оценить коэффициенты этой модели.

Для этого используется аппарат регрессионного и корреляционного анализа.

Задача проста, если фактор, который оказывает влияние на прогнозируемый показатель, заранее известен. Если это не так и прогнозист имеет дело с ситуацией, когда меняются несколько факторов окружающей среды и ему необходимо отобрать тот из них, который и является причиной изменения прогнозируемого показателя, следует использовать аппарат корреляционного анализа так, как это было показано в третьей главе данного учебника.

После того, как этот влияющий фактор выявлен, перед прогнозистом возникает задача подбора функции, наилучшим образом описывающей взаимосвязь между фактором x (причиной изменений) и прогнозируемым показателем y.

Поскольку характер изменения рядов социально-экономических показателей является многообразным, то и описывающие его модели могут иметь самые различные формы. Чаще всего в практике социально-экономического прогнозирования в качестве моделей однофакторных зависимостей используют несколько элементарных функций. Рассмотрим их.

Линейная функция:

$$\hat{y}_t = a_0 + a_1 x_t. \tag{4.4}$$

График этой функции приведен на рис. 4.1 Свободный член функции обозначает отрезок, который отсекает прямая линия на оси $0y_t$, а коэффициент пропорциональности a_1 характеризует тангенс угла наклона этой прямой к оси $0x_t$.

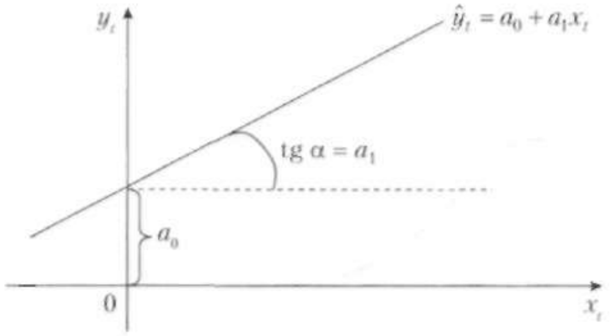

Рис. 4.1. **Линейная модель и геометрическая интерпретация значений ее коэффициентов**

Модель линейной функции часто используют в прогнозировании. По крайней мере, исходя из общенаучного принципа «от простого — к сложному», изучают свойства этой модели, разрабатывают различные методы оценивания ее коэффициентов, а также их пересчета при появлении новой информации, либо адаптации модели; выполняют прогнозы и считают доверительные интервалы, а затем на основе полученных знаний и навыков переходят к изучению более сложных моделей. На практике эту модель довольно часто предпочитают другим более сложным моделям, поскольку другой общенаучный принцип «простоты» гласит о том, что если сложная модель незначительно улучшает понимание процесса, то ей надо предпочесть более простую модель — нет смысла усложнять задачу, если она имеет простое решение.

В ряде случаев линейной зависимости может быть вполне достаточно для целей прогнозирования. В конце концов, многие нелинейные процессы в краткосрочной перспективе имеют линейную динамику. А значит, и аппроксимация нелинейной зависимости линейной функцией может быть вполне достаточной для решения некоторого ряда задач.

К линейной функции также обращаются в тех случаях, в которых сложно подобрать какую-либо более адекватную модель для описания зависимости.

Значительно реже, чем линейную модель, используют модель *квадратичной функции*, которая может быть записана так:

$$\hat{y}_t = a_0 + a_1 x_t + a_2 x_t^2. \qquad (4.5)$$

Графиком этой функции является парабола с осью симметрии, параллельной оси ординат. Характер функции определяется ее коэффициентами. Если коэффициент $a_2 > 0$, то ветви параболы направлены вверх (рис. 4.2); если $a_2 < 0$, — вниз. Ось ординат пересекается в точке $y_0 = a_0$. Поэтому значение свободного члена этой модели также характеризует начальный уровень. Эта модель используется в случаях, когда прогнозируемый ряд имеет нелинейную тенденцию, но, используя ее, необходимо помнить о том, что при довольно больших значениях x_t сам показатель y_t, который вычисляется по формуле (4.5), значительно меняет свои значения.

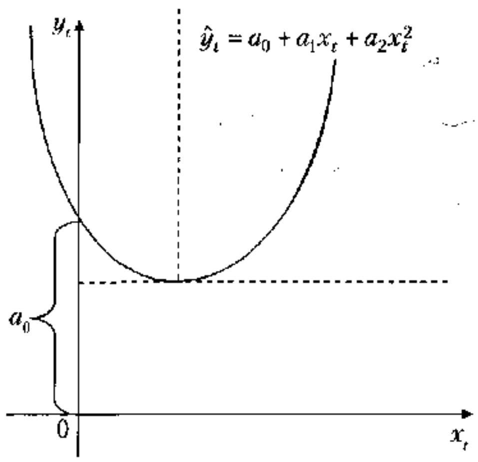

Рис. 4.2. **Квадратичная функция ($a_2 > 0$)**

Использование такой функции в регрессионном анализе позволяет моделировать ситуации, когда с ростом x y растет либо снижается с замедлением. При этом используется только одна ветвь параболы. Примером такой зависимости может быть объем спроса на бананы в зависимости от доходов потребителей: с ростом доходов объем спроса будет расти, но с некоторым замедлением. А в какой-то момент потребители и вовсе могут переключиться, например, на минибананы, что вызовет снижение спроса на простые бананы.

При включении в регрессионную модель квадрата зависимой переменной нельзя забывать про переменную x_t без квадрата — если ее опустить, то экстремум параболы будет приходиться на точку $x_t = 0$, а значит и сама модель уже будет моделировать совершенно другую зависимость.

Крайне редко в практике социально-экономического прогнозирования используются модели ***многочлена третьей степени***. Вообще-то многочлены выше третьей степени в практике прогнозирования не используются, поскольку они являются неустойчивыми. Да и многочлены третьей степени имеют непростую форму, поскольку при определенном сочетании значений коэффициентов они имеют локальный минимум и локальный максимум и, соответственно, точку перегиба. В общем виде эта модель будет записана так:

$$\hat{y}_t = a_0 + a_1 x_t + a_2 x_t^2 + a_3 x_t^3. \tag{4.6}$$

Графиком такой функции является кубическая парабола. Одной из ее характеристик выступает такая расчетная величина:

$$\Delta = 3a_3a_1 - a_2^1. \qquad (4.7)$$

Если, например, $\Delta > 0$, то кубическая парабола имеет возрастающий характер при $a_3 > 0$ и убывает при $a_3 < 0$. А если $\Delta < 0$ и $a_3 > 0$, то она имеет вид, представленный на рис. 4.3.

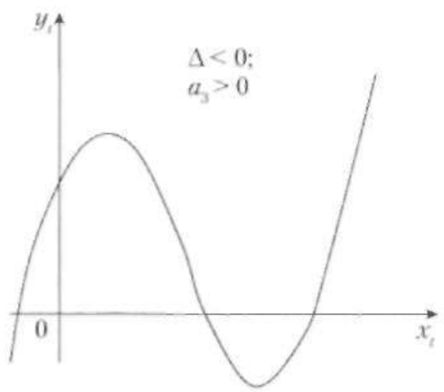

Рис. 4.3. **Кубическая парабола (4.6)**

Ветви такой параболы устремляются резко вниз в левой части графика и резко вверх — в правой части. Поэтому кубическая парабола может хорошо описать внутреннюю часть некоторой изменяющейся динамики, а вот для прогноза будет моделировать либо резкий рост, либо резкое падение в зависимости от коэффициентов модели, что, конечно же, не имеет особого смысла. Именно поэтому такие модели очень редко встречаются в прогнозировании.

Значительно чаще в практике прогнозирования используется *степенная функция*, которая имеет вид:

$$\hat{y}_t = a_0 x_t^{\,a_1}. \qquad (4.8)$$

Графиком этой функции является парабола порядка a_1. Она проходит через точки O (0; 0) и $A(1;\ a_0)$ и касается оси $0x_t$ в начале координат. Если показатель степени — целое положительное число, то при его четном значении график функции симметричен относительно оси $0y_t$ и в начале координат имеет минимум при $a_0 > 0$ и максимум при $a_0 < 0$.

Поскольку параметры модели в виде степенной функции оцениваются по статистическим данным, вероятность того, что показатель степени будет числом целым да еще четным, практически равна нулю. Маловероятно также, что показатель степени будет целым нечетным числом. К тому же, в подавляющем большинстве случаев точки социально-экономических показателей располагаются в первом квадранте плоскости. Поэтому на практике можно встретиться с двумя случаями (рис. 4.4).

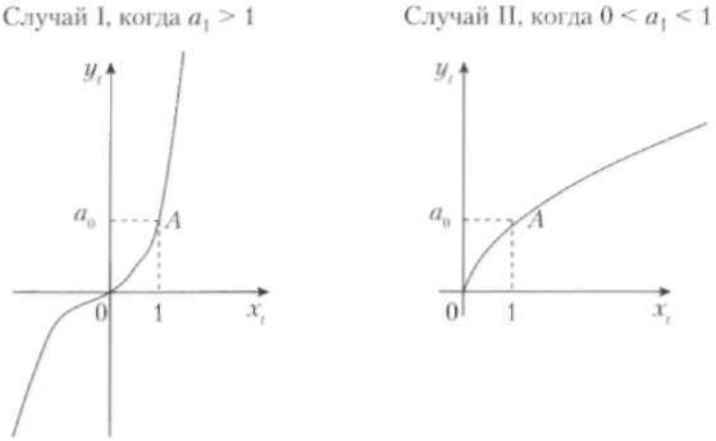

Рис. 4.4. **Модель в виде степенной функции**

Первый случай соответствует ситуации, когда показатель степени больше единицы. Тогда это возрастающая функция, причем, чем больше показатель степени, тем более резко возрастает функция (случай I рис. 4.4). Впрочем, в реальных социально-экономических процессах такой резкий рост показателей встречается довольно редко. А если он и встречается, то такая тенденция длится не очень долго. Чаще всего приходится иметь дело с более пологим характером степенной функции, и вполне может быть так, что применение МНК приведет к тому, что показатель степени будет лежать в пределах $0 < a_1 < 1$. Тогда функция будет возрастать не так быстро и ее характер можно определить как пологий (случай II рис. 4.4). Подобные ситуации встречаются довольно часто. В качестве примера можно привести известную производственную функцию Кобба — Дугласа, которая представляет собой многофакторную зависимость, но зависимость эта как раз степенная, а показатели степени у каждого фактора такой модели по определению лежит именно в пределах от нуля до единицы. Степенная функция во втором случае позволяет моделировать ситуацию роста y_t с замедлением при увеличении x_t. Однако в отличие от параболы этот рост никогда не останавливается — у степенной функции с $a_1 \in (0; 1)$ нет экстремума.

Интересен случай степенной функции, когда показатель степени отрицателен ($a_1 < 0$). Тогда степенная функция (4.8) имеет график равносторонней гиперболы, ветви которой симметричны относительно начала координат. Но поскольку модели чаще всего рассматривают применительно к первому квадранту, то в практике прогнозирования используется только одна ветвь гиперболы (рис. 4.5). Очевидно, что при малых значениях x_t моделируемый показатель устремлен в бесконечность.

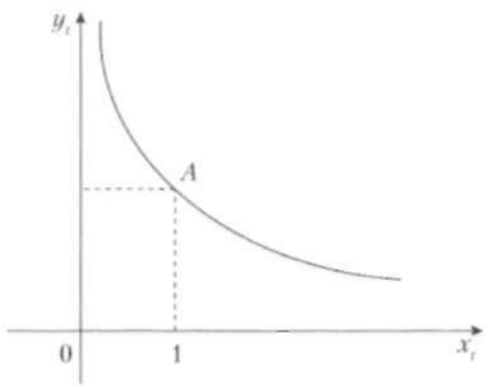

Рис. 4.5. **Модель в виде степенной функции при $a_1 < 0$**

Чаще на практике для моделирования гиперболической зависимости между y и x используют простую модель гиперболы вида

$$\hat{y}_t = a_0 + a_1 \frac{1}{x_t}, \tag{4.9}$$

в которой константа a_0 характеризует сдвиг вверх вдоль оси $0y_t$, а коэффициент пропорциональности a_1 — пологость функции. Графически модель выглядит аналогично изображенной на рис. 4.5, однако для ее оценки можно применить МНК непосредственно, не приводя ее к линейному виду.

Примером такой зависимости может выступать объем спроса типографии на бумагу в зависимости от цены на нее: для того, чтобы типография работала, ей нужна бумага, поэтому какое-то количество будет закупаться даже при достаточно высокой цене. Однако надо иметь в виду, что такой вид зависимости может носить локальный характер, в каких-нибудь границах изменения x_t, что, конечно же, может быть вполне достаточным для целей прогнозирования.

К числу элементарных функций, которые используют в социально-экономическом прогнозировании, относится и модель *показательной функции*, которая в общем виде может быть представлена так:

$$\hat{y}_t = a_0 k^{a_1 x_t}. \tag{4.10}$$

Здесь k — основание показательной функции, $k > 0$. Основанием показательной функции может быть любое положительное число. Как легко убедиться из уравнения показательной функции, при любом ее основании функция будет проходить через точку (0; a_0) и асимптотически приближаться к оси $0x_t$. Если основание показательной функции лежит в пределах $0 < k < 1$, то с ростом x_t функция убывает и стремится к оси $0x_t$. Если же основание показательной функции больше единицы, то функция возрастает с ростом x_t. Основание показательной функции, равное единице, исключается, потому что это означает постоянство результата, т.е. отсутствие функции как таковой.

Если коэффициент пропорциональности $a_0 > 0$, то значения функции всегда положительны, а отрицательные значения этого коэффициента обычно не рассматривают.

Важный частный случай показательной функции — *экспоненциальная функция*, когда ее основанием служит число Эйлера (которое приближенно равно 2,72). Для этого случая модель записывается так:

$$\hat{y}_t = a_0 e^{a_1 x_t}. \tag{4.11}$$

График экспоненциальной функции (экспоненты) приведен на рис. 4.6. Если показатель степени этой функции $a_1 < 0$, то функция убывающая; если $a_1 > 0$, то функция возрастает, причем этот рост имеет взрывной характер — изменение x приводит к все бо́льшим изменениям y. В связи с тем, что такие «взрывные» процессы в прогнозировании встречаются достаточно редко, ситуаций с $a_1 > 0$ стараются избегать.

Значительно реже, чем основание по числу e, применяют *показательную функцию с десятичным основанием*:

$$\hat{y}_t = a_0 10^{a_1 x_t}. \qquad (4.12)$$

Ее графическое изображение подобно экспоненте, только в случае, когда $a_1 x_t = 1$, функция пройдет через значение $y = a_0 10$, а не через $y = a_0 e$, как в случае с экспонентой.

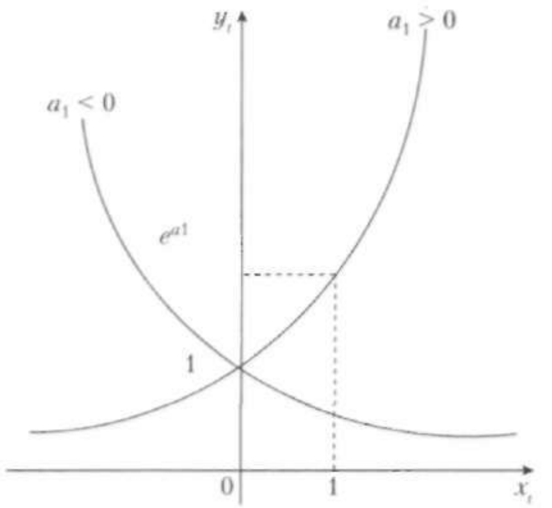

Рис. 4.6. **График экспоненты при $a_0 = 1$ и различных значениях a_1**

В принципе, любая показательная функция может быть приведена к другой показательной функции. Например, функцию (4.12) можно привести к (4.11) следующим образом: $\hat{y}_t = a_0 e^{a_1 x_t \ln 10}$.

В регрессионном анализе в связи с этим обычно используется экспонента, поскольку ее легче интерпретировать с экономической точки зрения. Увидеть эту интерпретацию можно в том случае, если прологарифмировать левую и правую части (4.11): $\ln \hat{y}_t = \ln a_0 + a_1 x_t$.

В данном виде изменение x_t на единицу будет увеличивать логарифм y_t на a_1:

$$\Delta \ln \hat{y}_t = a_1 = \ln \hat{y}_t - \ln \hat{y}_{t-1} = \ln\left(\frac{\hat{y}_t}{\hat{y}_{t-1}}\right) =$$
$$= \ln\left(\frac{\hat{y}_{t-1} + \Delta y_t}{\hat{y}_{t-1}}\right) = \ln\left(1 + \frac{\Delta y_t}{\hat{y}_{t-1}}\right). \quad (4.13)$$

В правой части (4.13) представлен логарифм единицы *плюс изменения y в процентах (по отношению к предыдущему значению y)*. При малых значениях a_1 логарифм этой величины может быть интерпретирован как процентное изменение y при изменении x на единицу.

Еще одной популярной функцией является *функция логарифма*:

$$\hat{y}_t = a_0 + a_1 \ln x_t. \quad (4.14)$$

Здесь a_0 характеризует уровень ряда в точке, в которой $x_t = 1$, а a_1 — изменения y при изменении логарифма x.

Так же, как и в случае с показательной функцией, основание данной функции может быть любым, но из тех же самых соображений удобности интерпретации обычно пользуются числом Эйлера: при малых значениях a_1 коэффициент будет показывать, на сколько единиц изменится y при изменении x на 1%.

Графически эта функция представлена на рис. 4.7.

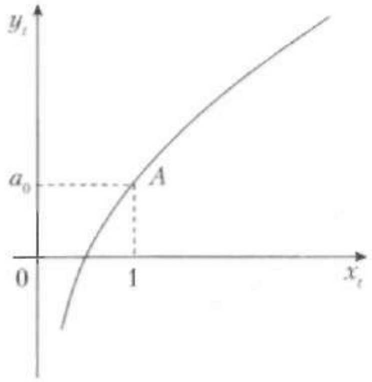

Рис. 4.7. **Модель в виде логарифмической функции**

Функция, как и степенная, моделирует процессы роста с замедлением. Ее особенностью является то, что при крайне малых значениях x она устремляется в $-\infty$, а при больших значениях — в $+\infty$. У этой функции нет экстремумов, и она возрастает менее резко, нежели, например, степенная функция.

Иногда в социально-экономическом прогнозировании цикличных процессов используют элементарные тригонометрические функции — синусоиды, обычно в сочетании с другими элементарными функциями. Функции, описывающие окружности и эллипсы, и другие функции более сложных форм, как правило, не используются.

Объясняется это тем, что для социально-экономических процессов не характерны резкие изменения характеристик. Даже динамические ряды стоимости ценных бумаг на фондовых рынках, которые имеют сложные графические изображения, стараются при прогнозировании «сгладить» для того, чтобы определить общую тенденцию. Эта тенденция отображается в виде некоторой гладкой функции, чьи коэффициенты и необходимо найти для того, чтобы спрогнозировать уровень изучаемого показателя.

Ни один социально-экономический объект не описывается моделью, точно совпадающей с той или иной функцией — всегда будут иметь место отклонения от нее. Если отклонения носят случайный характер, то подобранная функция может выступать моделью некоторой зависимости, освобожденной от влияния случайных факторов. Если отклонения являются систематическими, значит, необходимо подобрать другую модель.

Зная изображения этих элементарных функций, прогнозист может, изучив график прогнозируемого процесса, оценить, какая функция лучше всего описывает процесс, и принять ее в качестве основной.

Конечно, утверждать, что сам моделируемый процесс подчиняется выбранному характеру функции, нельзя. Можно лишь утверждать, что на рассматриваемом промежутке социально-экономический процесс лучше всего может быть описан этой моделью, поскольку другие элементарные модели имеют другой характер. В силу инерционности большей части социально-экономических процессов можно предполагать, что сложившиеся в рассматриваемый промежуток времени закономерности, проявляющиеся в характере зависимости, и дальше будут вести себя также. При этом если прогнозист колеблется в выборе и считает, что две отличающиеся друг от друга модели, на его взгляд, в одинаковой степени подходят для целей прогнозирования, то следует отдать предпочтение более простой модели, так как именно этого требует общенаучный принцип простоты.

Но, конечно же, доверять выбор модели одному лишь визуальному анализу не стоит. Поэтому очень часто прогнозист априорно задает некоторый набор элементарных или несколько более сложных моделей, а затем запускает процесс нахождения коэффициентов каждой из них, чаще всего, с помощью МНК. После этого построенные модели сравниваются по каким-нибудь критериям. Одним из простейших критериев является минимум дисперсии ошибок каждой модели, поскольку минимальная дисперсия свидетельствует о том, что модель лучше всего описала имеющийся исходный ряд значений. При этом надо иметь в виду, что не всегда самая точная в аппроксимации модель действительно даст точный прогноз. Например, известно, что по двум точкам можно найти коэффициенты линейной модели, так, что модель пройдет через все две точки, и дисперсия ошибки аппроксимации будет равна нулю. По трем точкам можно построить и квадратичную функцию, и дисперсия ошибки аппроксимации также будет равна нулю. В общем случае по T точкам можно построить полином $(T - 1)$-й степени. Его дисперсия ошибки будет равна нулю. Но разве эта модель даст точный прогноз по происходящему процессу или каким-нибудь образом отразит сложившиеся связи между факторами? Конечно, нет!

В любом случае окончательный выбор вида модели должен оставаться за прогнозистом и основываться на его опыте, интуиции и использовании формальных методов.

4.2. Метод наименьших квадратов и уравнения в отрезках

После выбора вида модели возникает задача оценки коэффициентов этой модели применительно к тем статистическим данным, которые есть в распоряжении прогнозиста, т.е. наполнения модели конкретным статистическим смыслом. Эта задача имеет довольно простое решение, о котором говорилось во второй главе нашего учебника, где рассматривался метод наименьших квадратов. Но для понимания сути процесса оценивания коэффициентов прогнозной модели с помощью того же МНК рассмотрим графическую интерпретацию этой задачи.

Рассмотрим вначале простую линейную модель:

$$\hat{y}_t = a_0 + a_1 x_t. \qquad (4.15)$$

Она описывает исходные значения показателя y_t с некоторой ошибкой аппроксимации ε_t: $\varepsilon_t = y_t - \hat{y}_t$.

Вычисляемое отклонение ε_t, если рассмотреть задачу на плоскости факторов, означает, что из точки y_τ $(\tau \in t)$ фактического наблюдения показателя проводится линия, перпендикулярная оси $0x_t$, до пересечения с прямой линией, которая соответствует графическому расположению модели и расчетному значению показателя $\hat{y}_t$. На рис. 4.8 эта точка на плоскости обозначена A_τ.

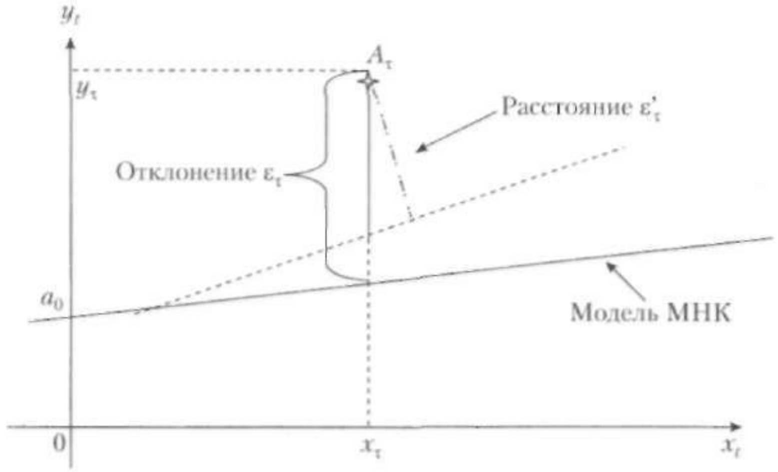

Рис. 4.8. **Фактическая точка, модель и отклонения между ними**

Метод наименьших квадратов, предусматривающий нахождение таких коэффициентов модели, для которых сумма квадратов отклонений ε_t будет минимальной:

$$\sum_t \varepsilon_t^2 = \sum_t \left(y_t - \hat{y}_t\right)^2 \to \min, \qquad (4.16)$$

означает построение такой модели, которая проходит через «облако» фактических точек так, чтобы дисперсия отклонений расчетных значений от фактических была бы минимально возможной.

Но, очевидно, что МНК — не единственный метод оценивания коэффициентов прогнозных моделей, а лишь один из многих. Прямая линия может пройти через фактические точки на плоскости и по-другому. Выбор того, как должна пройти через исходные точки модель, чтобы с ее помощью и получить наиболее точный прогноз, и составляет основную задачу прогнозирования. Использование МНК имеет очевидное преимущество в том, что прогнозируемый показатель y_t моделируется так, что дисперсия описания прошлого была минимальной из всех возможных, а для стационарных рядов это гарантия того, что и в будущем при прогнозе именно такая модель даст наименьшее отклонение прогнозируемого показателя от реального значения.

Но если задача заключается в том, чтобы наилучшим образом описать взаимосвязь между переменными y_t и x_t, с тем, чтобы прогнозировать не один показатель, а пару, то МНК будет не лучшим способом оценивания. В этом случае в качестве критерия оценивания коэффициентов линейной модели можно принять минимизацию квадратов расстояний ε'_t между прямой и фактическими наблюдениями (на рис. 4.8 это расстояние показано штрих-пунктирной линией). Модель, коэффициенты которой оцениваются с помощью такого критерия, в общем случае пройдет иначе, чем модель с оценками МНК. На рисунке эта модель показана пунктирной линией.

В любом случае, при разных способах оценивания модели получаются разные значения коэффициентов, что приводит к тому, что каждая прямая на плоскости факторов будет иметь свой оригинальный угол наклона и отсекать на оси $0y_t$ отличный от других моделей отрезок. Иначе говоря, на плоскость рис. 4.8 можно нанести бесконечное множество прямых, каждая из которых будет тем или иным образом описывать фактические значения и характеризоваться различными значениями дисперсии, суммами квадратов расстояний, средними абсолютными отклонениями и т.п., а также, что принципиально важно, — разными ошибками прогнозирования.

Поскольку процесс оценивания коэффициентов модели по сути сводится к поиску единственной пары оптимальных значений a_0 и a_1 из бесконечного множества возможных значений этих параметров, то имеет смысл рассмотреть задачу на плоскости этих неизвестных[1]. Действительно, варьируя, например, значения свободного члена a_0 модели (4.15), мы тем самым приводим к тому, что прямая линия МНК будет пересекать ось $0y_t$ в разных точках (рис. 4.8), модель будет отстоять от фактических точек по-разному, а значит, и сумма (4.16) будет различной. Именно неизвестные значения a_0 и a_1 являются параметрами задачи, а не значения y_t и x_t, как это может показаться на первый взгляд. Поэтому логично рассматривать задачу оценивания коэффициентов модели не на плоскости исходных переменных, а на плоскости неизвестных параметров, т.е. на плоскости с осями $0a_0$ и $0a_1$ (рис. 4.9).

Обозначим функцию, соответствующую критерию оптимизации коэффициентов модели (минимум суммы квадратов отклонений, минимум суммы абсолютных отклонений и т.п.), буквой Q. Применительно к задаче МНК это означает:

$$Q = \sum_t \varepsilon_t^2 = \sum_t \left(y_t - \hat{y}_t\right)^2. \qquad (4.17)$$

В случае нахождения оценок коэффициентов линейной модели это функция двух переменных — $Q(a_0; a_1)$.

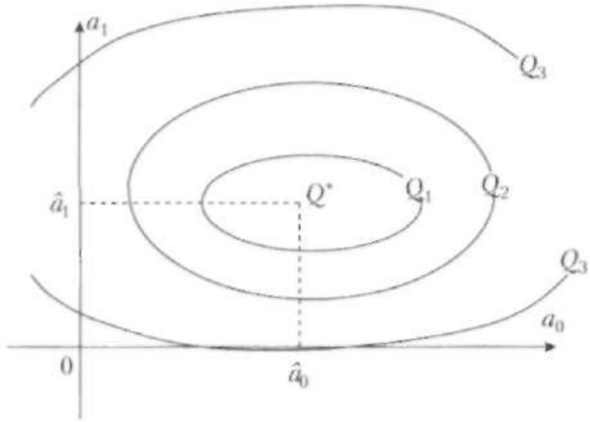

Рис. 4.9. **Плоскость коэффициентов линейной модели и значения функции**

На рис. 4.9 оценке МНК соответствует одна точка с координатами $(\hat{a}_0; \hat{a}_1)$. В этой точке значение функции Q достигает своей минимальной величины: $Q = Q^*$.

[1] *Светуньков С. Г.* Количественные методы прогнозирования эволюционных составляющих экономической динамики. Ульяновск : Изд-во УлГУ, 1999.

Все остальные точки этой плоскости будут давать в качестве своих координат такие значения коэффициентов a_0 и a_1, подставляя которые в (4.15) и вычисляя (4.17), будут получены значения функции Q, которые всегда будут больше Q^*. Множество точек, в каждой из которых функция $Q > Q^*$ имеет одно и то же значение, можно представить как замкнутые гладкие выпуклые кривые. По мере приближения точек к их оптимальным с позиций критерия МНК значениям, размер этих кривых уменьшается, как уменьшается и значение функции ($Q_3 > Q_2 > Q_1 > Q^*$). При достижении оптимума кривая превращается в точку, а значение критерия отбора принимает оптимальное значение Q^*. Именно это и происходит в ситуации, когда прогнозист оценивает значения коэффициентов нелинейной по параметрам прогнозной модели с помощью численных методов. Задавая на плоскости рис. 4.9 начальную точку, определяется значение функции Q в этой точке и определенным каждым численным методом подходом выбирается направление оптимизации — определяется следующая точка на плоскости коэффициентов модели. Для новой точки вновь высчитывается значение функции Q, после чего это значение функции сравнивается с предыдущим и определяется направление следующего шага. Так продолжается до тех пор, пока очередная точка на плоскости не окажется в заранее определенных допустимых окрестностях оптимальной точки Q^*.

Однако в простом случае оценивания коэффициентов линейной однофакторной модели с помощью МНК никаких многоитеративных процедур проводить не нужно, — довольно лишь решить систему нормальных уравнений:

$$\begin{cases} \sum_t y_t = Ta_0 + a_1 \sum_t x_t, \\ \sum_t y_t x_t = a_0 \sum_t x_t + a_1 \sum_t x_t^2, \end{cases} \tag{4.18}$$

где T — число наблюдений ($t = 1, 2, 3, \dots T$).

Но что будет означать эта система уравнений, если рассмотреть ее как задачу на плоскости коэффициентов рисунка 4.9? Первое и второе уравнения системы МНК представляют собой уравнения прямых линий, лежащих на плоскости коэффициентов a_0 и a_1. Эти уравнения можно представить для удобства в виде, называющемся «уравнением в отрезках». Для первого уравнения системы (4.18) оно будет иметь вид

$$\frac{a_0}{\frac{\sum_t y_t}{T}} + \frac{a_1}{\frac{\sum_t y_t}{\sum_t x_t}} = 1. \tag{4.19}$$

Значение, вычисляемое в знаменателе под каждым коэффициентом, представляет собой отрезок, который отсекает данная прямая линия на оси этого коэффициента. Или, иначе говоря, запись (4.19) дает возможность сразу же определить две точки, лежащие на прямой МНК плоскости коэффициентов:

- первая точка с координатами

$$\left(\frac{\sum_t y_t}{T}; 0\right)$$

принадлежит этой прямой линии и лежит на оси $0a_0$;

- вторая точка имеет координаты

$$\left(0; \frac{\sum_t y_t}{\sum_t x_t}\right)$$

и также принадлежит этой прямой линии, но лежит на оси $0a_1$.

Точно так же и второе уравнение системы МНК (4.18) можно рассматривать как уравнение прямой линии, расположенной на плоскости коэффициентов модели. И ее удобнее всего изучать, представив в форме «уравнения в отрезках»:

$$\frac{a_0}{\frac{\sum_t y_t x_t}{\sum_t x_t}} + \frac{a_1}{\frac{\sum_t y_t x_t}{\sum_t x_t^2}} = 1. \tag{4.20}$$

Эта линия, как можно увидеть из отрезков (4.20), также пересекает оси координат, причем точки пересечения легко определить из этой записи:

- ось $0a_0$ модель пересекает в точке

$$\left(\frac{\sum_t y_t x_t}{\sum_t x_t};0\right);$$

- ось $0a_1$ — в точке

$$\left(0;\frac{\sum_t y_t x_t}{\sum_t x_t^2}\right).$$

Нанесем эти прямые линии на плоскость коэффициентов (рис. 4.10). Их пересечение, как следует из логики решения систем уравнений, и будет представлять собой на плоскости точку, координаты которой соответствуют оценкам МНК.

По поводу этой точки можно сказать не только то, что она является пересечением двух прямых системы нормальных уравнений МНК (4.18), но и то, что через нее, как и через любую другую точку, лежащую на плоскости, можно провести бесконечное число различных прямых. Из этого следует простой вывод: точку на плоскости коэффициентов модели, соответствующую условию минимума дисперсии ошибки аппроксимации, можно получить не только с помощью системы (4.18), но и другими способами.

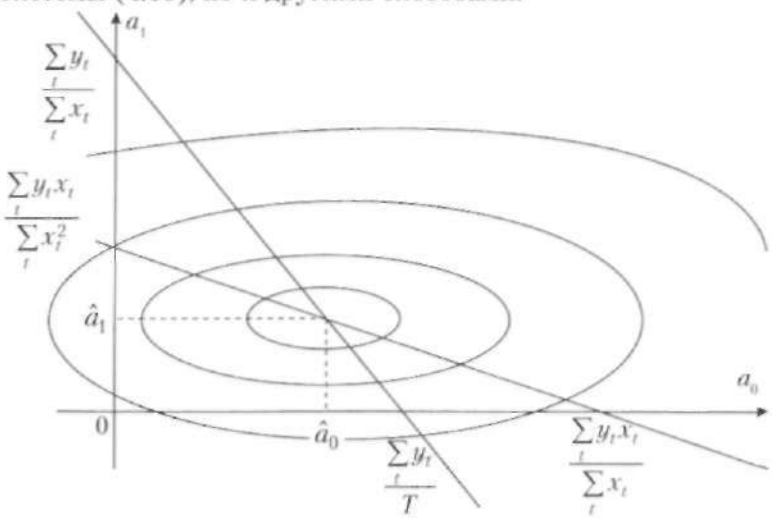

Рис. 4.10. **Уравнения системы МНК на плоскости коэффициентов модели и оценки МНК этих коэффициентов**

Действительно, ранее мы уже пользовались простым приемом для нахождения коэффициента регрессии — центрированием исходных переменных относительно их средней арифметической. Если провести такое центрирование и подставить преобразованные значения переменных во второе уравнение системы (4.18), получим

$$\begin{cases} \sum_t y_t = a_0 T + a_1 \sum_t x_t, \\ \sum_t \left(y_t - \overline{y}\right)\left(x_t - \overline{x}\right) = a_0 0 + a_1 \sum_t \left(x_t - \overline{x}\right)^2. \end{cases} \tag{4.21}$$

Естественно, что решая эту систему, мы вновь получим оценки коэффициентов, соответствующие критерию МНК. Графически это означает, что две прямые на плоскости коэффициентов пересекаются в той же точке, что и прямые линии системы (4.18), но второе равенство системы (4.21) представляет собой уравнение прямой, которая лежит на плоскости не так, как прежде — она перпендикулярна оси $0a_1$ и параллельна оси $0a_0$.

Пример

Пусть в распоряжении прогнозиста имеются следующие данные об изменении показателя y в зависимости от изменений показателя x:

t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
y_t	3,5	4,4	3,9	5,1	6,9	7,3	8,9	9,5	9,9	10,7	12	12,4	13,9	15,2	17
x_t	0,5	1,1	1,6	1,9	2,5	3,2	3,7	3,9	4,6	5,1	6,1	6,4	6,5	7	8,1

Построим по этим данным систему нормальных уравнений (4.18), а также систему уравнений в отрезках (4.19)–(4.20).

Система нормальных уравнений для данного случая будет иметь вид:

$$\begin{cases} 140{,}6 = a_0 15 + a_1 62{,}2, \\ 720{,}23 = a_0 62{,}2 + a_1 335{,}22. \end{cases}$$

Решая эту систему, получим следующие оценки МНК модели рассматриваемой зависимости: $y_t = 1{,}775 x_t + 2{,}013$.

Теперь приведем это уравнение к виду «уравнения в отрезках»:

$$\begin{cases} 1 = \dfrac{a_0}{9{,}3733} + \dfrac{a_1}{2{,}2605}, \\ 1 = \dfrac{a_0}{11{,}5793} + \dfrac{a_1}{2{,}1485}. \end{cases}$$

Анализ последней формы системы МНК показывает, что прямые линии, составляющие эту систему, на плоскости коэффициентов пересекаются под острым углом — отрезки на оси a_0 отличаются друг от друга на 2,2059 единицы (19%), а на оси a_1 — на 0,1119 единиц (4,9%). Этот угол легко вычислить. Если первое уравнение записать как

$$Ax + By + D = 0, \tag{4.22}$$

а второе — как

$$A'x + B'y + D' = 0, \tag{4.23}$$

то косинус угла между ними можно вычислить по формуле[1]:

$$\cos\gamma = \frac{AA' + BB'}{\sqrt{A^2 + B^2}\sqrt{A'^2 + B'^2}}. \tag{4.24}$$

Подставляя в эту формулу значения коэффициентов для каждого из уравнений системы нормальных уравнений МНК, получим

$$\cos\gamma = \frac{15 \cdot 62{,}2 + 62{,}2 \cdot 335{,}22}{\sqrt{15^2 + 62{,}2^2}\sqrt{62{,}2^2 + 335{,}22^2}} = 0{,}998586495.$$

Это означает, что прямые пересекаются под углом в 3,04 градуса. Незначительные ошибки округления будут приводить к изменению точки пересечения прямых условий МНК, и решение оказывается неустойчивым.

Осуществим такую же процедуру применительно к центрированным относительно средних арифметических переменным. Тогда система МНК (4.21) будет иметь вид

$$\begin{cases} 140{,}6 = a_0 15 + a_1 62{,}2, \\ 137{,}2 = a_0 0 + a_1 77{,}3. \end{cases}$$

Представим эту же систему в виде «уравнений в отрезках»:

$$\begin{cases} 1 = \dfrac{a_0}{9{,}3733} + \dfrac{a_1}{2{,}2605}, \\ 1 = \dfrac{a_0}{\infty} + \dfrac{a_1}{1{,}775}. \end{cases}$$

Знак бесконечности в знаменателе второго уравнения означает, что эта прямая линия не пересекает соответствующую ось координат. Легко заметить, что второй отрезок этой прямой линии отсекает

[1] *Корн Г., Корн Т.* Справочник по математике для научных работников и инженеров. М. : Наука, 1984. С. 85.

на оси $0a_1$ отрезок, численно равный оценке МНК коэффициента a_1. Прямые линии данной системы пересекаются под другим углом. Вычислим косинус этого угла, воспользовавшись формулой (4.24):

$$\cos\gamma = \frac{15 \cdot 62{,}2 + 62{,}2 \cdot 335{,}22}{\sqrt{15^2 + 62{,}2^2}\sqrt{62{,}2^2 + 335{,}22^2}} = 0{,}972131357.$$

Этот косинус соответствует углу, равному 13,5 градусов. Прямые, как видно, действительно пересекаются под более тупым углом, чем в системе нормальных уравнений. Это значит, что данные оценки МНК более устойчивы к ошибкам округления, чем оценки МНК, полученные без центрирования.

4.3. Прогнозирование стационарных неоднородных процессов: многофакторные модели

При моделировании и прогнозировании социально-экономической динамики объективно приходится иметь дело с многофакторными зависимостями, когда значение показателя или группы показателей определяется поведением не одного, а многих факторов одновременно. Действительно, если взять любой показатель социально-экономического развития и посмотреть, какие факторы оказывают на него влияние, то, пожалуй, невозможно встретить хотя бы один из них, который не формировался бы под воздействием множества различных причин и условий. Например, на величину цены товара оказывают влияние объемы предложения товара и спроса на него, наличие конкурентов, товаров-заменителей и цены на них и множество других факторов. Численность занятых на производстве основных рабочих, на первый взгляд, определяется только величиной производственного задания. На самом же деле она зависит и от того, на каком оборудовании работает рабочий, какова организация труда на предприятии, насколько развит менеджмент, в конце концов, и от условий оплаты труда. Поэтому применительно к задаче прогнозирования социально-экономических явлений, пожалуй, труднее найти однофакторную зависимость, чем многофакторную.

Очевидно, что при построении однофакторной модели приходится упрощать представление о действительности — вместо моделирования воздействия на прогнозируемый показатель многих факторов приходится из этого множества отбирать только один фактор, который прогнозист посчитает наиболее существенным. Любая однофакторная модель в этой ситуации настолько условна и груба, что ее применение в прогнозировании может давать лишь очень приближенные ориентиры.

Если экономист собирается выполнить точный прогноз с помощью экономико-математической модели, то он должен подобрать такую модель, которая бы лучше всего отражала суть происходящих процессов и описывала их. А так как практически все социально-экономические показатели формируются под воздействием множества факторов, то и модель, прогнозирующая их, также должна учитывать это, т.е. быть многофакторной. Следовательно, от многофакторной прогнозной модели можно ожидать большей точности, чем от однофакторной, поскольку она вскрывает особенности и моделирует экономическую реальность более подробно.

Многофакторную модель в общем виде можно представить так:

$$M\left(y_t \mid x_{j,t}\right) = f\left(x_{j,t}\right), \tag{4.25}$$

где $M(y_t|x_{j,t})$ — условное математическое ожидание значений прогнозируемого процесса; x_j — значения факторов внешней среды, оказывающих влияние на прогнозируемый процесс; j — номер фактора внешней среды ($j = 1, 2, 3, ..., k$).

Многофакторные модели могут быть как линейными, так и нелинейными. Исходя из общенаучного принципа «от простого — к сложному», рассмотрим вначале задачу построения линейной многофакторной модели:

$$\hat{y}_t = a_0 + a_1 x_{1,t} + a_2 x_{2,t} + ... + a_k x_{k,t}. \tag{4.26}$$

Чтобы построить такую модель, вначале необходимо отобрать факторы, которые оказывают наибольшее влияние на результирующий показатель, ведь общее количество факторов, влияющих на социально-экономические показатели, очень велико. Как было показано ранее, формальные процедуры отбора факторов следует использовать только как дополнительный инструмент, подтверждающий или отвергающий выдвинутую прогнозистом гипотезу. В начале этого процесса должен быть проведен экономический анализ задачи, а после того, как определены основные факторы, влияющие на поведение прогнозируемого объекта, осуществляется отбор наиболее важных факторов, в том числе и с помощью формальных процедур.

Покажем, насколько непростой является эта задача.

Пример

Допустим, необходимо сформировать гипотезу о том, какие факторы оказывают влияние на цену одного килограмма картофеля на конкретном рынке у конкретного продавца.

Для этого просто перечислим первую десятку факторов, которые с позиций экономической сути поставленной задачи оказывают влияние на цену:

1) цена, по которой был куплен картофель у поставщика;

2) стоимость аренды одного торгового места на рынке и различных сборов;

3) разновидность картофеля;

4) качество клубней и их товарный вид;

5) цена, по которой продают картофель конкуренты;

6) общее количество конкурентов и объемы их товара;

7) сезон (осенью много других овощей, поэтому есть заменители картофеля);

8) объем спроса на картофель со стороны посетителей рынка;

9) маркетинговое сопровождение продаж (оформление павильона, вид ценника, вид и характеристики упаковочного материала, рекламные лозунги и т.п.);

10) день недели (в выходные дни посетителей больше).

Легко заметить, что хотя все эти факторы оказывают влияние на цену товара, приведенный перечень факторов далеко не исчерпывающий. Но если даже остановиться на этом перечне, то выбрать из этих 10 факторов несколько самых важных и включить их в прогнозную модель будет непросто. Например, третий и четвертый факторы измеряются в классифицирующей шкале (вид картофеля и его сорт), а бóльшая часть других — в метрической, как и сам прогнозируемый показатель. Как с помощью формальных процедур корреляционного анализа оценить степень влияния сортности на цену картофеля? Для этого придется такой показатель, как цена картофеля, переводить в классифицирующую шкалу — тогда можно рассчитать некоторые коэффициенты, характеризующие степень взаимосвязи между этими показателями или использовать коэффициенты ранговой корреляции. Еще сложнее обстоит дело с девятым фактором — маркетинговое сопровождение товара. Действительно: как измерить этот фактор? В каких единицах? Кроме того, данных по некоторым из выбранных факторов по различным причинам может и не оказаться (либо данные могут быть неполными или искаженными). Это создает дополнительные трудности.

Поэтому волей-неволей прогнозисту придется из множества тех факторов, которые оказывают влияние на прогнозируемый показатель, отобрать не те из них, которые являются самыми важными, а те, с которыми можно работать и с помощью которых можно построить прогнозную модель.

Чаще всего после аналитического этапа отбора факторов, оказывающих влияние на прогнозируемый показатель, их количество все еще довольно велико. Поэтому возникает необходимость отбора из этого множества наиболее значимых факторов. Уже на этом этапе можно формализовать процедуру, прибегнув к помощи корреляционного анализа.

Основной коэффициент корреляционного анализа — коэффициент парной корреляции — характеризует степень приближения зависимости между двумя случайными факторами к линейной форме. Нелинейные зависимости он не оценивает. С его помощью заполняют матрицу коэффициентов парной корреляции между факторами, а также между факторами и моделируемым показателем так, как это показано в табл. 4.1. Такая матрица называется корреляционной.

Таблица 4.1

Корреляционная матрица

Переменная	y	x_1	x_2	...	x_j	...	x_k
y	1						
x_1	r_{yx1}	1					
x_2	r_{yx2}	r_{x2x1}	1				
...	...	...	...	...			
x_j	r_{yxj}	r_{xjx1}	r_{xjx2}	...	1		
...	...	...	...	...	...	...	
x_k	r_{yxk}	r_{xkx1}	r_{xkx2}	...	r_{xjxk}	...	1

Поскольку коэффициент парной корреляции симметричен, т.е. $r_{yx} = r_{xy}$, то и матрица в табл. 4.1 симметрична относительно диагонали, значения которой, очевидно, равны единице. Поэтому заполняют либо верхнюю правую часть матрицы, либо ее нижнюю левую часть, как это сделано в табл. 4.1. Анализ этой матрицы дает много дополнительной информации прогнозисту — как о степени линейного влияния каждого фактора на результирующий признак, так и о степени линейной взаимосвязи между факторами. Последнее особенно важно, поскольку наличие линейной взаимосвязи между факторами означает построение модели в условиях мультиколлинеарности, а это имеет очень неприятные последствия для вычислимости модели.

Для того чтобы упростить процесс построения регрессионной модели и определить, каким образом включить те или иные факторы в итоговую модель, имеет смысл обратиться к точечной диаграмме, которую бывает удобно представлять в форме матрицы.

Рассмотрим пример определения формы многофакторной регрессионной модели на условном примере по данным табл. 4.2.

Таблица 4.2

Данные условного примера

Себестоимость продукции, product_price	Затраты на рекламу, advertising	Стоимость основных производственных фондов, capital	Заработная плата, labor	Стоимость материалов material_price
35,55	25	1002,48	63	22,25
44,53	25	1034,84	64	21,92
44,44	24	1027,09	65	20,32
38,57	24	994,00	65	23,79
43,82	24	994,78	66	22,56
42,72	23	992,12	67	24,23
44,84	24	1016,39	67	23,74
36,17	24	1047,74	68	22,84
39,5	23	1029,78	68	23,17
41,2	23	1065,39	69	24,1
43,71	23	1054,37	70	26,19
43,79	24	1059,57	70	27,47
43,16	24	1077,31	71	28,43
45,86	23	1083,06	71	29,85
41,85	24	1063,27	72	31,36
37,79	23	1050,69	72	29,44
41,37	24	1179,48	73	29,98
45,58	25	1276,52	74	31,32
48,64	24	1298,20	74	30,47
48,55	24	1268,24	75	30,73

Окончание табл. 4.2

Себестоимость продукции, product_price	Затраты на рекламу, advertising	Стоимость основных производственных фондов, capital	Заработная плата, labor	Стоимость материалов material_price
47,38	24	1302,19	76	30,74
47,29	24	1302,12	76	31,23
49,74	24	1290,49	77	32,79
50,87	24	1313,14	77	34,24
46,59	24	1287,13	78	36
46,75	24	1274,33	78	37,75
49,21	24	1321,27	79	36,8
49,03	24	1323,17	80	37,88
46,86	24	1272,80	80	38,2
42,61	23	1280,03	81	37,69
48,97	23	1336,00	82	36,63
49,41	23	1346,69	82	37,37
50,03	22	1341,26	83	37,02
43,72	23	1275,43	83	39,03
50,23	23	1334,42	84	40,43
51,98	23	1459,33	84	40,29

На основе этих данных была составлена корреляционная матрица, для каждого коэффициента корреляции была рассчитана его статистическая значимость (проверка гипотезы о том, что коэффициент равен нулю). В табл. 4.3 приведена полученная корреляционная матрица; коэффициенты, оказавшиеся статистически незначимыми на 5%, отмечены символом «*».

Таблица 4.3

Корреляционная матрица по данным условного примера табл. 4.2

	Product_price	Advertising	Capital	Labor	Material_price
Product_price	1				
Advertising	−0,166*	1			
Capital	0,802	−0,192*	1		
Labor	0,708	−0,433	0,925	1	
Material_price	0,675	−0,336	0,891	0,972	1

Данные табл. 4.3 показывают, что на себестоимость продукции оказывают среднее линейное влияние такие переменные, как стоимость материалов и заработная плата. Связь между стоимостью основных производственных фондов и себестоимостью ближе к линейной, чем между указанными выше показателями. Затраты на рекламу практически не связаны с себестоимостью линейно. Возможно, эта связь носит нелинейный характер. Кроме того, стоит отметить связь, близкую к линейной, между капиталом и трудом ($r = 0{,}925$), капиталом и стоимостью материалов ($r = 0{,}891$), а также трудом и стоимостью материалов ($r = 0{,}972$). Последние два коэффициента корреляции, возможно, оказались такими высокими чисто из технических соображений (например, из-за инфляции росла стоимость материалов, а заработная плата росла из-за индексирования на величину инфляции) — вряд ли между этими переменными может существовать какая-то реальная взаимосвязь.

Рассмотрим связи между нашими переменными на матрице точечных диаграмм (рис. 4.11).

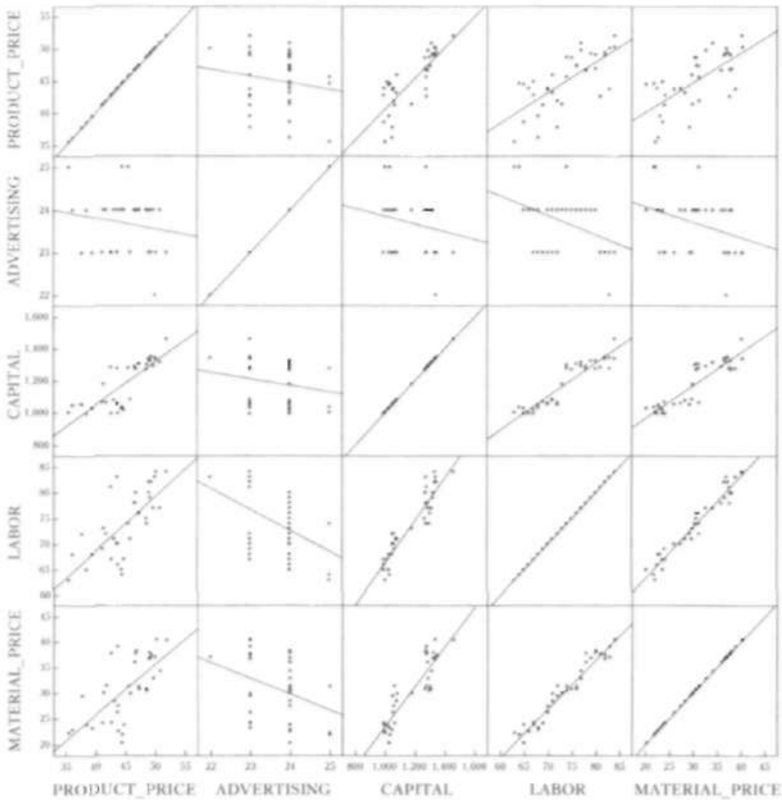

Рис. 4.11. **Матрица точечных диаграмм для условного примера**

По рис. 4.11 можно судить в целом о зависимости между переменными и о том, имеются ли какие-нибудь особенности в этих зависимостях. По диагонали данной матрицы изображены точки, укладывающиеся в прямую линию, так как в этом случае мы рассматриваем зависимость фактора от самого себя. Прямыми линиями представлены однофакторные регрессионные модели линейного вида. Они добавлены на график для наглядности — так проще оценить вид зависимости.

В нашем условном примере нас больше всего интересует, как затраты на рекламу (advertising — x_1), стоимость основных производственных фондов (capital — x_2), среднемесячная заработная плата (labor — x_3) и стоимость материалов (material_price — x_4) влияют на себестоимость продукции (product_price — y). Для этого нам надо проанализировать зависимости на графиках в первой строке.

Как видим, зависимость между затратами на рекламу и себестоимостью не носит явного характера: на рекламу выделялись суммы в 22, 23, 24 и 25 тыс. руб., при этом четкого роста либо снижения себестоимости не наблюдалось (есть лишь незначительное снижение), хотя экономическая теория говорит о том, что себестоимость должна была бы увеличится. По-видимому, в данном случае зависимость носит сложный нелинейный характер, но так как мы не знаем, как ее выявить, опишем ее линейной моделью:

$$\hat{y}_t = a_0 + a_1 x_{1,t}. \qquad (4.27)$$

Стоимость основных производственных фондов, судя по всему, влияет на себестоимость продукции нелинейно — виден небольшой перегиб в тенденции. Кроме того, если сравнивать эту зависимость с прямой линией, проведенной на графике рис. 4.11, мы увидим, что ряд точек систематически лежит под прямой (при малых значениях x), а ряд — над ней (при больших значениях x). Такой вид зависимости может быть описан, например, логарифмической функцией. Добавим этот элемент в модель (4.27): $\hat{y}_t = a_0 + a_1 x_{1,t} + a_2 \ln x_{2,t}$.

Влияние заработной платы и стоимости материалов носит характер, близкий к линейному, с достаточно большим разбросом, поэтому оба эти фактора можно включить в модель как есть. Хотя стоимость материалов, возможно, может быть лучше описана при использовании функции логарифма. Итоговая модель будет иметь вид $\hat{y}_t = a_0 + a_1 x_{1,t} + a_2 \ln x_{2,t} + a_3 x_{3,t} + a_4 x_{4,t}$.

Очевидно, что вид модели определяется исследователем, и каждый исследователь может предложить свои формы моделей, которые бы лучше описали ту или иную зависимость. Однозначного вывода о правильности спецификации модели на этом этапе сделать нельзя (к тому же, при построении многофакторных моделей зависимость между y и x_1 может искажаться под влиянием других факторов, учтенных в модели), можно лишь приблизиться к наилучшему варианту, изучив объект исследования и обсудив с экспертами, какие из зависимостей действительно имеют место в реальности.

Обратим внимание, что как по корреляционной матрице, так и по матрице точечных диаграмм достоверно можно судить лишь об однофакторных зависимостях. В случае построения многофакторных моделей в описание зависимости между любыми двумя факторами вносят искажения другие включенные факторы, поэтому даже включение факторов в линейном виде будет приводить к нелинейному влиянию на результат из-за одновременного влияния всех факторов в модели и связи их друг с другом. Идеальную картину, в которой каждый из факторов влияет на результат по отдельности, можно наблюдать лишь в случае полного отсутствия мультиколлинеарности (о которой мы расскажем позже в этом же параграфе). На рис. 4.12 приведен пример точечной диаграммы между затратами на рекламу и себестоимостью продукции.

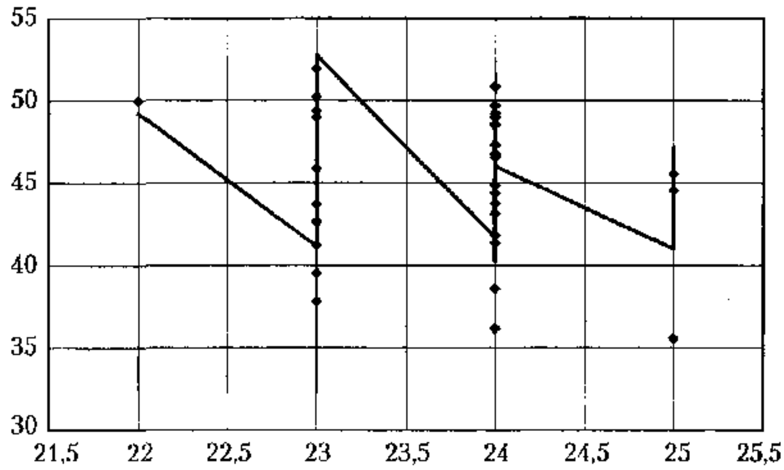

Рис. 4.12. **Точечная диаграмма для условного примера табл. 4.2 и регрессионная линия на плоскости «затраты на рекламу» — «себестоимость продукции»**

Точками здесь показаны пары фактических значений по рекламе и себестоимости, а линией — связь между рекламой и расчетными значениями по себестоимости. Фактически рисунок представляет собой проекцию сложной многомерной зависимости на плоскости. Как видим, эта зависимость носит сложный нелинейный характер, и в линию не укладывается. Этот пример демонстрирует, что построение корреляционной матрицы и матрицы точечных диаграмм может выступать лишь в роли ориентира при выборе формы модели (показывающей связь между переменными в случае, если остальные переменные не меняют своих значений), но не позволяет четко определить ее. Выбор формы модели, в конечном счете, должен осуществляться самим прогнозистом на основе знаний об объекте исследования.

По значениям табл. 4.1 сложно судить о том, насколько все приведенные факторы взаимосвязаны, или, иначе говоря, какова множественная корреляция между факторами. В таблице приведена парная корреляция, но что можно сказать о многофакторной зависимости в целом?

Для того чтобы получить ответ на этот вопрос, вычисляют коэффициент множественной корреляции. Есть несколько способов его вычисления, отличающихся друг от друга трудоемкостью процесса. Приведем самый простой из них[1].

Вначале, воспользовавшись данными табл. 4.1, найдем определитель матрицы:

$$\Delta^* = \begin{vmatrix} 1 & r_{y,x_1} & r_{y,x_2} & \dots & r_{y,x_k} \\ r_{x_1,y} & 1 & r_{x_1,x_2} & \dots & r_{x_1,x_k} \\ r_{x_2,y} & r_{x_2,x_1} & 1 & \dots & r_{x_2,x_k} \\ \dots & \dots & \dots & 1 & \dots \\ r_{x_k,y} & r_{x_k,x_1} & r_{x_k,x_2} & \dots & 1 \end{vmatrix}. \tag{4.28}$$

Затем, убирая из матрицы все коэффициенты парной корреляции с переменной y_t (в нашем случае это первый столбец и первая строка), вычислим определитель другой матрицы:

$$\Delta = \begin{vmatrix} 1 & r_{x_1,x_2} & \dots & r_{x_k,x_1} \\ r_{x_1,x_2} & 1 & \dots & r_{x_k,x2} \\ \dots & \dots & \dots & \dots \\ r_{x_1,xk} & r_{x_2,x_k} & \dots & 1 \end{vmatrix}. \tag{4.29}$$

[1] *Елисеева И. И., Юзбашев М. М.* Общая теория статистики : учебник / под ред. И. И. Елисеевой. М. : Финансы и статистика, 2004. С. 372.

Коэффициент множественной корреляции будет равен:

$$R = \pm\sqrt{\frac{\Delta^*}{\Delta}}. \qquad (4.30)$$

Для линейных многофакторных моделей квадрат коэффициента множественной корреляции будет равен коэффициенту детерминации.

Из вышеприведенного следует и логика подбора факторов модели: включая или исключая тот или иной фактор из матриц (4.28) и (4.29) и вычисляя коэффициент множественной корреляции (4.30), останавливаются на том наборе факторов, при котором коэффициент множественной корреляции принимает наибольшие значения.

Многофакторную регрессионную модель удобнее представлять в матричном виде:

$$\hat{Y} = XA. \qquad (4.31)$$

Здесь исходные факторные переменные представлены в форме матрицы:

$$X = \begin{pmatrix} 1 & x_{1,1} & x_{2,1} & \dots & x_{k,1} \\ 1 & x_{1,2} & x_{2,2} & \dots & x_{k,2} \\ 1 & x_{1,3} & x_{2,3} & \dots & x_{k,3} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{1,T} & x_{2,T} & \dots & x_{k,T} \end{pmatrix},$$

а зависимая переменная и коэффициенты — в виде векторов-столбцов:

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_T \end{pmatrix}, \; A = \begin{pmatrix} a_0 \\ a_1 \\ \dots \\ a_k \end{pmatrix}.$$

Добавление столбца с единицами в матрице X необходимо для учета константы a_0, которая, как видно из формулы (4.31), будет умножаться именно на этот столбец.

Эта модель описывает реальные значения с некоторой ошибкой аппроксимации, что в матричной форме может быть выражено так:

$$Y = \hat{Y} + \varepsilon = XA + \varepsilon. \quad (4.32)$$

Здесь $\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_T \end{pmatrix}$ — вектор ошибок модели, относительно которых обычно делается стандартное предположение о том, что они распределены нормально с нулевым математическим ожиданием и некоторой дисперсией.

Сумма квадратов отклонений фактических значений от расчетных будет представлена как

$$Q = \sum_t \varepsilon_t^2 = \varepsilon'\varepsilon = (Y - XA)'(Y - XA) = Y'Y - 2A'X'Y + A'X'XA, \quad (4.33)$$

где X' — обозначения транспонированной матрицы X.

Дифференцируя по переменным задачи (т.е. по коэффициентам модели) и приравнивая к нулю полученные значения, можно в итоге получить выражение для определения вектора оценок коэффициентов многофакторной модели:

$$X'Y = X'XA, \quad (4.34)$$

откуда получается, что:

$$A = \left(X'X\right)^{-1} X'Y. \quad (4.35)$$

После построения множественных регрессий также можно проверить ряд статистических гипотез. Однако возможно это лишь при условии, что ошибки в модели оказались нормально распределенными. В случае несоблюдения этого условия тесты не имеют смысла.

Одна из самых популярных и простых гипотез — это гипотеза о равенстве коэффициента регрессии какому-нибудь значению. Для этого используется t-тест (см. параграф 3.4). Так, гипотеза о том, что коэффициент модели a_j равен некоторому b, будет формализована следующим образом:

$$H_0: a_j = b;$$

$$H_1: a_j \neq b;$$

$$t = \frac{a_j - b}{s.e.\left(a_j - b\right)}.$$

где $s.e.$ — стандартная ошибка. В случае если b — это некоторая константа, то стандартная ошибка в знаменателе будет равна стандартной ошибке коэффициента a_j.

Число степеней свободы в t-тестах в регрессионном анализе равно $T - k$, где T — число наблюдений, а k — число коэффициентов модели.

С помощью этого же теста можно проверять более сложные гипотезы, касающиеся линейных ограничений в модели, например, гипотезу о том, что несколько коэффициентов в сумме дают значение третьего коэффициента:

H_0: $a_1 + a_2 + a_3 = a_4$;

H_1: $a_1 + a_2 + a_3 \neq a_4$.

Для того чтобы проверить такую гипотезу, ее удобно представить в виде равенства относительно нуля:

H_0: $a_1 + a_2 + a_3 - a_4 = 0$;

H_1: $a_1 + a_2 + a_3 - a_4 \neq 0$.

Расчетное значение t — статистики в таком случае находится по формуле:

$$t = \frac{a_1 + a_2 + a_3 - a_4}{s.e.(a_1 + a_2 + a_3 - a_4)}.$$

Как видим, в этих простых тестах все упирается в стандартные ошибки модели. А в случае со сложными линейными ограничениями исследователю могут понадобиться и ковариации между коэффициентами модели. Например, для расчета стандартной ошибки $a_1 + a_2$ нам будут нужны дисперсии этих коэффициентов и ковариация между ними:

$$s.e.(a_1 + a_2) = \sqrt{D(a_1 + a_2)} = \sqrt{D(a_1) + D(a_2) + 2\operatorname{cov}(a_1, a_2)}.$$

Эти значения обычно получаются по формуле:

$$V_a = \frac{RSS}{T - m}(X'X)^{-1}, \tag{4.36}$$

где RSS — сумма квадратов отклонений остатков модели.

В результате расчетов по формуле (4.36) будет получена матрица, имеющая следующую структуру:

$$D_a = \begin{pmatrix} D(a_0) & \operatorname{cov}(a_1, a_0) & \dots & \operatorname{cov}(a_k, a_0) \\ \operatorname{cov}(a_0, a_1) & D(a_1) & \dots & \operatorname{cov}(a_k, a_1) \\ \dots & \dots & \dots & \dots \\ \operatorname{cov}(a_0, a_k) & \operatorname{cov}(a_1, a_k) & \dots & D(a_k) \end{pmatrix}. \quad (4.37)$$

Матрица (4.37) называется ковариационно-вариационной матрицей коэффициентов. Как видим, она симметрична, по ее диагонали откладываются значения дисперсий соответствующих коэффициентов. Стандартные ошибки коэффициентов будут равны квадратному корню из соответствующих дисперсий.

Для проверки более сложных гипотез используются распределения Фишера и Пирсона. Например, для того чтобы проверить гипотезу о том, что все коэффициенты в модели равны нулю, используется F-тест:

H_0: $a_1 = a_2 = a_3 = \dots = a_k = 0$;

H_1: $a_1 \neq 0 \vee a_2 \neq 0 \vee a_3 \neq 0 \vee \dots \vee a_k \neq 0$.

Обратим внимание, что альтернативная гипотеза в этом тесте заключается в том, что хотя бы один коэффициент модели статистически значим. Отклонение нулевой гипотезы в принципе не указывает на хорошее качество построенной модели, а лишь говорит о том, что хотя бы один из факторов, выбранных исследователем, действительно оказывает существенное влияние на результат.

Расчетное значение для этого теста вычисляется по формуле

$$F = \frac{ESS / (k-1)}{RSS / (T-k)} = \frac{(T-k) ESS}{(k-1) RSS},$$

где ESS — сумма квадратов отклонений модели.

Полученное значение будет распределено по Фишеру с $(k - 1, T - k)$ степенями свободы.

При построении многофакторной модели, прогнозист неминуемо столкнется с рядом проблем, которые хорошо изучены в курсе эконометрики. Мы приведем их краткую характеристику, укажем, какие последствия они могут иметь для прогнозиста и что можно с ними сделать.

4.3.1. Пропущенные переменные

В случае с построением многофакторной модели прогнозист, скорее всего, столкнется с проблемой, в эконометрике получившей название «проблема пропущенных переменных». Главная идея в данном случае заключается в том, что, если исследователь не учтет какие-то существенные переменные, то оценки коэффициентов его регрессионной модели будут **смещенными** и **несостоятельными**. С точки зрения экономического анализа это означает, что коэффициенты полученной модели нельзя будет корректно интерпретировать, а с точки зрения прогнозирования, — при расчете зависимой переменной $\hat{y}$ с использованием набора факторов X будут получаться значения, существенно отличающиеся от фактических y.

Проблема «пропущенных переменных», к сожалению, абсолютно естественна в общественных науках по обозначенным выше (в примере о цене картофеля) причинам. Исследователь сможет приблизиться к «идеальной модели» при наличии очень хорошей статистики, т.е. большого числа наблюдений по максимально возможному числу факторов без грубых и систематических ошибок. Но даже в таком случае нет гарантий, что он сможет правильно выявить влияние всех собранных факторов на результат и четко определить спецификацию модели, которая соответствовала бы «идеальной модели». Например, в ряде случаев на цену картофеля может влиять не только цена, по которой был куплен картофель на этой неделе, но и цена картофеля, купленного на прошлой неделе — в модели могут быть так называемые «лаговые переменные». Также возможно, что какие-то факторы должны нелинейно влиять на цену на картофель, но исследователь, не учтя это, построит линейную модель. Все это будет приводить к искажению коэффициентов модели по сравнению с «идеалом».

Однако на практике «идеальных» ситуаций с полным набором данных все равно не бывает, поэтому проблема пропущенных переменных оказывается актуальной, хотя исследователь может об этом и не задумываться.

Чтобы показать, что происходит с оценками регрессии в случае пропуска существенных переменных, рассмотрим «идеальную» и «реальную» модели:

- идеальная: $y = a_0 + a_1x_1 + a_2x_2 + ... + a_kx_k + a_{k+1}x_{k+1} + ... + a_mx_m + \varepsilon$;
- реальная: $y = b_0 + b_1x_1 + b_2x_2 + ... + b_kx_k + e$.

В «реальной» модели не учтено $m - k$ последних факторов, из-за чего коэффициенты и ошибки ее будут отличаться от коэффициентов и ошибок «идеальной» модели.

Обратим внимание на то, что ошибки в «идеальной» модели представляют собой некий «белый шум», который вбирает в себя незначительное влияние различных мелких факторов на результат.

В матричном виде эти две модели могут быть представлены следующим образом:

$$Y = XA + \tilde{X}\tilde{A} + \varepsilon \tag{4.38}$$

и

$$Y = XB + e, \tag{4.39}$$

где X — матрица первых k учтенных регрессоров; A и B — векторы первых k коэффициентов моделей; $\tilde{X}$ — матрица оставшихся неучтенных $m - k$ регрессоров; $\tilde{A}$ — соответствующие этим регрессорам $m - k$ коэффициентов.

Сравним коэффициенты A и B. Из модели (4.38) нужно убрать невключенные в модель (4.39) регрессоры $\tilde{X}$. Для этого вычтем из левой и правой частей (4.38) $\tilde{X}\tilde{A}$. Получим

$$Y - \tilde{X}\tilde{A} = XA + \varepsilon. \tag{4.40}$$

Теперь проведем следующую замену в (4.40):

$$\tilde{Y} = Y - \tilde{X}\tilde{A}. \tag{4.41}$$

В результате этого уравнение (4.38) примет вид

$$\tilde{Y} = XA + \varepsilon. \tag{4.42}$$

Вектор коэффициентов в (4.42), как мы уже знаем, легко находится с помощью МНК по формуле (4.35):

$$A = \left(X'X\right)^{-1} X'\tilde{Y}. \tag{4.43}$$

Коэффициенты в модели (4.39) находятся аналогично:

$$B = \left(X'X\right)^{-1} X'Y. \tag{4.44}$$

Теперь сравним смещение коэффициентов (4.44) и идеальные коэффициенты (4.43). Для этого рассчитаем разность (4.44) и (4.43):

$$B - A = \left(X'X\right)^{-1} X'Y - \left(X'X\right)^{-1} X'\tilde{Y} = \left(X'X\right)^{-1}\left(X'Y - X'\tilde{Y}\right). \tag{4.45}$$

Подставим в (4.45) значение из (4.41):

$$B - A = \left(X'X\right)^{-1}\left(X'Y - X'\left(Y - \tilde{X}\tilde{A}\right)\right),$$

раскроем скобки

$$B-A=(X'X)^{-1}(X'Y-X'Y+X'\tilde{X}\tilde{A})$$

и сократим слагаемые:

$$B-A=(X'X)^{-1}X'\tilde{X}\tilde{A}. \tag{4.46}$$

Как видим, разница в коэффициентах «идеальной» и «реальной» моделей составляет величину, выражаемую формулой (4.45). В случае если значения коэффициентов $\tilde{A}$ невелики (т.е. влияние пропущенных переменных близко к нулю), это смещение будет несущественным. Оно также будет несущественным, если пропущенные переменные не будут коррелировать с учтенными переменными (что на практике выполняется достаточно редко).

Рассмотрим, как будут связаны между собой ошибки в «идеальной» и «реальной» моделях. Для этого выразим вектор коэффициентов A в (4.46): $A=B-(X'X)^{-1}X'\tilde{X}\tilde{A}$.

Теперь подставим полученный вектор в модель (4.38):
$$Y=X\left(B-(X'X)^{-1}X'\tilde{X}\tilde{A}\right)+\tilde{X}\tilde{A}+\varepsilon=XB-X(X'X)^{-1}X'\tilde{X}\tilde{A}+\tilde{X}\tilde{A}+\varepsilon.$$

Вынося за скобки одинаковые слагаемые, получим

$$Y=XB+\left(I_T-X(X'X)^{-1}X'\right)\tilde{X}\tilde{A}+\varepsilon, \tag{4.47}$$

где I_T — единичная матрица с T числом столбцов и строк.

Приравнивая (4.39) и (4.47), получим

$$XB+e=XB+\left(I_T-X(X'X)^{-1}X'\right)\tilde{X}\tilde{A}+\varepsilon,$$

откуда очевидно, что:

$$e=\left(I_T-X(X'X)^{-1}X'\right)\tilde{X}\tilde{A}+\varepsilon. \tag{4.48}$$

Как видим, ошибки в «реальной» модели будут соответствовать ошибкам в «идеальной» модели либо в случае, если влияние неучтенных факторов ничтожно (т.е. $\tilde{A}$ близки к нулю), *либо в случае выполнения условия* $X(X'X)^{-1}X'=I_T$, которое само по себе смысла не имеет и на практике не достижимо.

Итак, в результате пропуска существенных переменных эти переменные в «реальной» модели исказят не только коэффициенты, но и ошибки модели. В итоге, в случае, если пропущенная переменная коррелирует с переменными, включенными в модель, ее дисперсия «распределится» по модели: коррелирующая часть «уйдет» в имеющиеся факторы и исказит значения коэффициентов, а не коррелирующая — в ошибки e. И если ошибки «идеальной» модели еще могут быть распределены нормально, то ожидать этого от ошибок «реальной» модели не стоит, так как на их распределение уже оказывает непосредственное влияние распределение не только учтенных, но и неучтенных факторов. Это, в свою очередь, (как мы видели в параграфе 3.4) приводит к невозможности проверять параметрические статистические гипотезы.

Пример

Пусть прогнозируемый социально-экономический процесс в точности подчиняется следующей функциональной зависимости:

$$\hat{y}_t = 2 - 0{,}5x_{1t} + 3x_{2t} + x_{3t}. \tag{4.49}$$

Если зависимость однозначная и функциональная, то можно ожидать полного и однозначного выявления зависимости, ведь влияние случайного фактора полностью исключено. Проверим, так ли это.

В табл. 4.4 для разных значений переменных по данной формуле вычислены значения моделируемой переменной.

Таблица 4.4

Данные условного примера

t	x_{1t}	x_{2t}	x_{3t}	Расчетное значение y_t, полученное по (4.49)
1	1	7	2	24,5
2	2	8	4	29
3	3	9	3	30,5
4	4	10	5	35
5	5	11	7	39,5
6	6	12	4	39

Для приближения задачи к реальной ситуации будем считать, что прогнозисту не известно о влиянии второго фактора на моделируемый показатель или же, что он заметил, что между первым и вторым фактором имеется линейная функциональная зависимость (коэффициент парной корреляции равен единице) и исключил второй фактор из рассмотрения. Тогда он захочет построить следующую двухфакторную линейную модель: $\hat{y}_t = a_0 + a_1x_{1t} = a_3x_{3t}$.

Решая систему нормальных уравнений, можно определить оценки МНК и получить модель вида $\hat{y}_t = 20 + 2{,}5x_{1t} + x_{3t}$.

Как легко убедиться, коэффициент перед первым фактором здесь положителен, хотя влияние этого фактора, как следует из исходной модели (4.49), — отрицательно — увеличение фактора x_{1t} в анализируемом процессе приводит к уменьшению результирующего признака, а в полученной с помощью МНК модели — моделируется противоположный результат. Очевидно, что построенная без учета второго фактора модель ошибочна.

Для решения проблемы пропущенных переменных эконометристы рекомендуют обращаться к методу инструментальных переменных[1], суть которого сводится к тому, чтобы заменить часть имеющихся регрессоров на их расчетные значения. Чтобы получить эти значения, исследователь должен подобрать такие переменные Z (называемые «инструментальными»), которые не коррелировали бы с ошибкой e в модели (4.39), но при этом коррелировали бы с факторами, которые нужно заменить. Определение заменяемых факторов обычно осуществляется экспертным путем: предполагается, что какие-то факторы могут коррелировать с ошибками. Проблема, когда регрессоры коррелируют с ошибками, в эконометрике называется **проблемой эндогенности** и математически может быть записана так: $M(\varepsilon | X) = M(\varepsilon X) \neq 0$.

Однако здесь кроется серьезный изъян, создающий сложности в понимании самого механизма инструментальных переменных. Если рассматривать идеальную модель (4.38), то в ней ошибки не могут коррелировать с регрессорами, так как подразумевается, что в них находятся мелкие неучтенные факторы, не оказывающие существенного влияния на результат, а значит и не коррелирующие с существенными факторами, так как сама модель описывает объект исследования наилучшим образом. Если же рассматривать реальную модель (4.39), то при использовании МНК (к которому обычно и прибегают при построении регрессий) ошибки в модели просто не могут коррелировать с факторами (подробней этот вопрос будет рассмотрен нами во втором томе учебника в параграфе 10.2). Обычно обоснование такой корреляции сводится к предположению о том, что регрессоры носят стохастический характер, т.е. говорится о том, что они формируются случайно, а значит, — теоретически могут коррелировать с ошибками (что объясняется в форме асимптотической корреляции). Такое предположение могло бы иметь место в пространственных

---

[1] *Эббес Питер.* Инструментальные переменные и эндогенность: нетехнический обзор // Квантиль. 2007. № 2. С. 3—20.

данных (в которых выборка формируется случайным образом), но во временных рядах (с которыми в основном и приходится работать прогнозисту) оно бессмысленно.

Получается, что для решения проблемы пропущенных переменных вводится искусственное, не имеющее ничего общего с реальностью предположение, с которым дальше предлагается бороться.

На самом же деле задача, которую призван решать метод инструментальных переменных, заключается в подборе таких переменных, которые коррелировали бы с нуждающимися в замене факторами, включенными в модель, но при этом не коррелировали бы с пропущенными переменными. Только в таком случае связь между используемыми и пропущенными переменными можно разорвать, в результате чего оценки коэффициентов реальной модели (4.39) могут стать ближе к реальности.

Однако здесь кроется другой изъян: измерить корреляцию между пропущенными и инструментальными переменными невозможно. В результате этого инструменты включаются исключительно на основе экспертного мнения экономиста. Естественно, что применение метода инструментальных переменных во многих случаях приводит к значительным потерям в качестве модели (так как фактор заменяется на его расчетное значение, и таким образом убирается часть дисперсии фактора, нужная для объяснения зависимой переменной), что в итоге сказывается на точности прогнозов.

Достаточно часто благодаря инструментальным переменным значимость некоторых регрессоров может изменяться (например, незначимые становятся значимыми), чем экономисты успешно пользуются. Так, для получения модели со значимыми регрессорами, которые в других условиях не прошли бы эконометрические тесты, достаточно использовать инструментальные переменные для замены этих переменных. Например, в случае, если на переменную y_t некоторый фактор x_t влияет слабо (или практически не влияет), достаточно заменить x_t расчетным значением с помощью инструментальной переменной z_t, коррелирующей с y_t. Очевидно, что часть дисперсии z_t, объясняющая y_t, точно так же будет объяснять его в новой модели: $y_t = a_0 + a_1\hat{x}_t + e_t$, где $\hat{x}_t = \hat{b}_0 + b_1 z_t$. Прогнозисту, нацеленному на получение практического результата, прибегать к таким ухищрениям опасно, так как внесение искажений может привести к неточным прогнозам. Однако это не говорит о том, что метод в принципе неприменим, — в некоторых случаях он действительно может помочь получить более точные оценки коэффициентов.

Для решения проблемы пропущенных переменных также иногда прибегают к замещающим переменным. Идея заключается в поиске такой переменной z_t, которая влияла бы на пропущенный $\tilde{x}_t$. Найденную переменную (которая в английском называется «proxy») далее используют непосредственно в модели регрессии. В случае, если все пропущенные переменные учтены и вместо них использованы замещающие переменные, оценки коэффициентов перед имеющимися в распоряжении переменными X будут несмещенными. Оценки перед замещающими переменными, естественно, не имеют смысла. Однако надо иметь в виду, что все это возможно только в том случае, если исследователь смог выявить все недостающие переменные и заменить их «proxy» переменными.

Сделаем одну ремарку. В прогнозировании выделяют *два типа регрессионных моделей*: «описательная» и «причинная»[1].

Первый вид представляет собой модель, которая отражает некоторые сложившиеся закономерности в отношениях между исследуемыми признаками. Примером такой модели может быть зависимость цены автомобиля от его пробега. Так, высокий пробег не является причиной низкой цены. Скорее всего, на самом деле на цену влияет износ автомобиля, объем двигателя, тип салона и прочие характеристики. Конечно же, интерпретировать коэффициенты такой модели бессмысленно. Однако для прогнозирования стоимости автомобиля с определенным пробегом такой моделью вполне можно воспользоваться.

Второй тип модели требует, чтобы в качестве входных переменных использовались те, которые являются именно причинами изменения зависимой переменной. В данном случае задача значительно усложняется, так как выявить истинные причины экономических явлений часто оказывается крайне сложно, а включение переменных в такой модели осуществляется исключительно на основе экспертного суждения. Примером такой модели может выступать зависимость объема выпуска тапочек от затраченного труда на их производство. Конечно же, в случае, если в причин-

---

[1] *Makridakis Spyros G., Wheelwright Steven C., Hyndman Rob J.* Forecasting: Methods and Applications. Wiley, 1998. P. 61.

ной модели учтены все факторы, вызывающие изменение изучаемого показателя, в той форме, в какой это происходит на самом деле, интерпретация коэффициентов имеет смысл, а в прогнозировании такая модель оказывается крайне полезной. Однако такую ситуацию скорее можно отнести к классу идеальных, редко достижимых на практике.

Очевидно, что и однофакторные, и многофакторные модели, являясь только некоторым более или менее удачным приближением к истине, при прогнозировании социально-экономической динамики не могут дать 100%-но достоверные прогнозы. Невозможность учета всех факторов, влияющих на результат, приводит к тому, что модели становятся очень и очень условными — они моделируют только некоторую часть свойств, не охватывая всю сложную совокупность характеристик прогнозируемых процессов. Поэтому для прогнозирования социально-экономической динамики характерно наличие высокой неопределенности как объективного свойства экономической и социальной жизни.

В меньшей мере это касается процессов, изучаемых по пространственным данным. В этом случае довольно часто удается выявить систему причинно-следственной связи и промоделировать ее, поскольку часть факторов, которая оказывает влияние на прогнозируемый показатель, неизменна из-за отсутствия влияния фактора времени и вариация прогнозируемого показателя определяется в этот момент конкретно-исторического развития вариацией определенного фиксированного набора факторов.

В связи с этим необходимо сказать несколько слов и о попытке экономической интерпретации коэффициентов эконометрических моделей, которая все еще встречается в работах отдельных ученых. Построив с помощью методов эконометрики некоторую эконометрическую модель, исследователи пытаются дать каждому из коэффициентов некоторую глубокую интерпретацию. Это было бы уместно, если бы моделировалась некая замкнутая совокупность, входы и выходы которой известны, так же, как известны состав и характер элементов и взаимосвязь между ними. Экономика — открытая, непрерывно эволюционирующая система, и выявить и описать всю сложную совокупность причинно-следственных связей как внутри этой системы, так и при ее взаимодействии с внешней средой, невозможно. В таком случае любая многофакторная модель будет являться лишь некоторым в той или иной степени удачным приближением к описанию истинных процессов.

Поэтому, получив отрицательное значение коэффициента при каком-то факторе, нельзя утверждать, что этот фактор отрицательно влияет на результирующий показатель. Точно так же, нельзя утверждать и о положительном влиянии фактора на результат, если коэффициент при нем получился положительным. Тем более странно, получив, например, коэффициент пропорциональности перед фактором труда, равный 1,23, утверждать, что каждая единица труда увеличивает результат в 1,23 раза.

В приведенном выше примере истинный процесс описывался функцией $\hat{y}_t = 2 - 0{,}5x_{1t} + 3x_{2t} + x_{3t}$, а регрессионная модель этого процесса имела вид $\hat{y}_t = 20 + 2{,}5x_{1t} + x_{3t}$.

Разве вклад первой переменной в изменение результата определяется коэффициентом 2,5? Конечно, нет, ведь реальное влияние этой переменной на результат равно –0,5 (т.е. рост первого фактора на единицу будет приводить к уменьшению результата на 0,5 единиц).

Основная мысль, на которую нам хотелось бы обратить внимание, заключается в том, что в любой регрессионной модели обязательно будут пропущенные переменные. В каких-то случаях, с точки зрения исследователя, это будут несущественные переменные, но выявить их «существенность» в принципе невозможно. При построении регрессий нельзя забывать о том, что модель — это всего лишь упрощенное представление об объекте исследования, а значит и делать по ней какие бы то ни было умозаключения стоит с осторожностью. При этом, в ряде случаев для прогнозирования достаточно иметь и простую модель, учитывающую не все факторы, а лишь основные имеющиеся в распоряжении. Конечно же, делать по такой модели заключение о том, что она есть истина, вскрывает и отражает какие-то внутренние процессы, неразумно.

В социально-экономической динамике практически всегда приходится сталкиваться с ситуацией, в которой неизвестный характер изменения объекта описывается слабо соответствующей ему моделью, коэффициенты которой только отражают статистическое наполнение модели, но вовсе не отражают суть происходящих процессов.

4.3.2. Проблема гетероскедастичности

Одна из насущных эконометрических проблем — это гетероскедастичность — ситуация, при которой с изменением значения фактора дисперсия зависимой переменной каким-то образом также изменяется. Математически данная проблема записывается следующим образом: $D(\varepsilon|X) \neq \text{const}$.

В результате этого нарушается предположении о постоянстве дисперсии ошибки в модели и, как следствие, оценки коэффициентов становятся не эффективными (дисперсии коэффициентов становятся завышенными). На прогнозирование проблема негативно воздействует следующим образом: при построении модели в условиях гетероскедастичности прогнозные интервалы оказываются либо слишком узкими, либо слишком широкими.

На рис. 4.13 приведен пример гетероскедастичности.

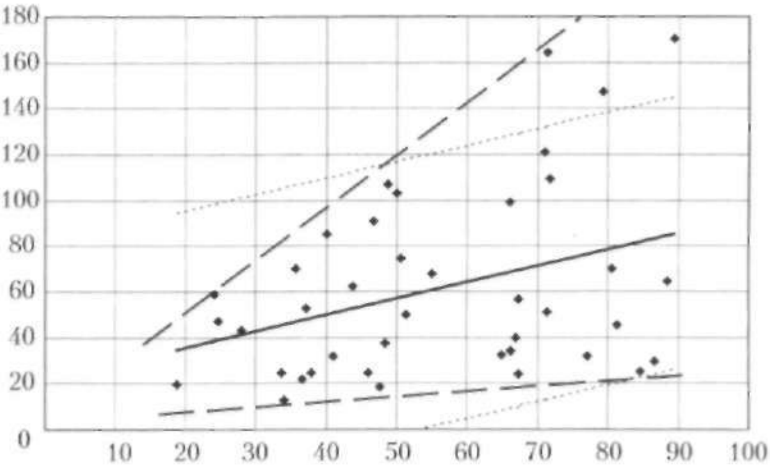

Рис. 4.13. **Графическое представление истинной гетероскедастичности**

С увеличением показателя x (по оси абсцисс) растет и дисперсия показателя y (ось ординат). Линяя регрессии (крупная пунктирная линия на рис. 4.13) пройдет сквозь облако точек так, чтобы квадраты расстояний были минимальными и в начале ряда, и в конце. Однако расчет дисперсии приведет к тому, что отклонения от этой линии будут усреднены на всех наблюдениях. В итоге доверительные интервалы будут совершенно неадекватными (тонкий пунктир на рис. 4.13): например, в начале ряда интервалы будут слишком широкими, а к концу ряда — излишне зауженными. Конечно же, это будет приводить к неточным прогнозам.

Обычно выделяют два типа гетероскедастичности: истинную и ложную. Подразумевается, что истинная гетероскедастичность вызвана какими-то объективными причинами, а ложная — неправильно выбранной спецификацией. Однако грань здесь достаточно тонка. Так, например, при оценке объема выпуска продукции в зависимости от ВВП по пространственным данным по нескольким странам можно столкнуться с тем, что разброс объема выпуска у стран с бо́льшим ВВП будет выше, чем у стран с меньшим ВВП (примерное графическое изображение этой ситуации отражено на рис. 4.13). Это вроде бы сигнализирует об истинной гетероскедастичности. Однако при построении модели с учетом численности населения (например, если рассчитывать все величины не как валовые, а как приходящиеся на душу населения) гетероскедастичность исчезает. Это уже указывает на то, что в данном случае мы столкнулись с ложной гетероскедастичностью — нужно было изначально выбрать корректную форму зависимости в модели. Полагаем, что гетероскедастичность можно назвать проблемой, вызванной неправильной идентификацией процесса со стороны исследователя.

На тему идентификации и борьбы с гетероскедастичностью написано множество статей. Для идентификации гетероскедастичности были предложены многочисленные статистические тесты (например, тест Уайта, тест Бройша-Пагана, тест Глейзера). Для каждого конкретного случая гетероскедастичности были разработаны свои методы избавления или хотя бы получения эффективных оценок коэффициентов регрессии. Все они подробно рассматриваются в учебниках по эконометрике, поэтому здесь мы не будем подробно освещать решения этой проблемы. В большинстве своем методы направлены на то, чтобы либо по-другому рассчитать стандартные ошибки коэффициентов модели, либо иначе оценить коэффициенты модели.

Во временных рядах гетероскедастичность принимает свои виды. Так, например, ряды данных с явной сезонностью могут содержать гетероскедастичность — с ростом показателя будет расти и амплитуда сезонных колебаний. Однако у такого поведения исследуемого показателя есть свое объяснение и свои объективные причины, учет которых позволяет избавиться от гетероскедастичности.

В случаях же, когда выявить причины гетероскедастичности не получается, можно прибегнуть к преобразованию исходного ряда данных. Для этого исходный ряд данных логарифмируют, что позволяет уменьшить условную дисперсию показателя. Правда, в таком случае любая построенная регрессионная модель будет эквивалентна степенной модели, а ошибки в такой модели будут уже смещенными (см. параграф 3.5).

4.3.3. Проблема автокорреляции остатков

Еще одна проблема, приводящая к неэффективным оценкам в регрессии, — это автокорреляция. Она заключается в том, что ошибки в модели могут коррелировать со значениями ошибок на предыдущих наблюдениях. Математически эта проблема записывается как $M(\varepsilon_t | \varepsilon_{t-\tau}) = M(\varepsilon_t \varepsilon_{t-\tau}) \neq 0$ при любых $\tau \neq 0$.

Так же, как и в случае с гетероскедастичностью, автокорреляция приводит к неэффективным оценкам и может быть истинной (вызванной особенностями процесса) и ложной (вызванной неправильной спецификацией). Обычно автокорреляция остатков сигнализирует о том, что какие-то факторы в модели не учтены (например, в модели нужно учесть лаговые переменные), что, в свою очередь, указывает на проблему пропущенных переменных. Также автокорреляция остатков может говорить о том, что исследователь сталкивается с нестационарными временными рядами, которые по всем канонам эконометрики должны быть предварительно приведены к стационарному виду путем взятия разностей (мы снова приходим к тому, что все эти проблемы опять же сводятся к ошибкам исследователя).

Формально говоря, в самом простом случае автокорреляция остатков выражается в том, что за положительными значениями остатков также следуют положительные значения, что указывает на систематические отклонения модели на некоторых временных промежутках. Более сложные виды автокорреляции могут приводить к более сложным зависимостям, общая черта которых заключается в том, что в остатках образуется некоторая зависимость, что сигнализирует о том, что в ряде данных подобная зависимость не была учтена.

Для идентификации автокорреляции остатков используются тест Дарбина—Уотсона (который, правда, проверяет лишь наличие автокорреляции первого порядка) и тест Бройша—Годфри. В ряде случаев более удобным является графическое изучение коррелограмм по остаткам модели *(см. параграф 5.1 второго тома учебника)*.

Для того чтобы избавиться от автокорреляции ошибок, в модель обычно включают либо лаговые переменные, либо значения y на предыдущих наблюдениях. Определение того, что же именно включить, осуществляется с помощью все тех же коррелограмм.

В прогнозировании автокорреляция остатков выражается в том, что неучтенные или неправильно учтенные факторы приводят к неточным точечным прогнозам. Кроме того, в случае с автокорреляцией ошибок адекватный интервальный прогноз на несколько наблюдений вперед дать не представляется возможным, так как нарушается важное условие независимости распределения ошибок.

Ранее, в параграфе 3.6, мы уже обсуждали смысл коэффициента корреляции и теперь, обратившись к проблеме автокорреляции остатков, понимаем, что отсутствие автокорреляции не указывает на то, что зависимости в остатках модели нет, а значит и сама модель хорошая, с эффективными и несмещенными оценками. Все, о чем говорит отсутствие автокорреляции, — это отсутствие линейной связи между значениями остатков на разных наблюдениях. Возможно, в остатках наблюдается сложная нелинейная зависимость, которая не может быть выявлена стандартными методами, что сигнализирует о все той же проблеме пропущенных переменных. В итоге исследователь будет интерпретировать коэффициенты этой модели, даже не подозревая о том, что его модель далека от истины.

С другой стороны, мы также уже обсуждали такой эффект как «ложная корреляция». Автокорреляция остатков может быть сродни корреляции между температурой на улице и длиной очереди в метро, т.е. сама проблема автокорреляции может быть переоценена исследователем. Возможно, именно поэтому в ряде случаев модели с формально диагностируемой автокорреляцией на практике дают более точные прогнозы, чем модели без нее[1].

---

[1] *Makridakis S.* The art and science of forecasting: An assessment and future directions // International Journal of Forecasting. 1986. № 2. P. 15—39.

4.3.4. Проблема мультиколлинеарности

При построении многофакторных моделей социально-экономической динамики уже на первом этапе — построения корреляционной матрицы — практически всегда сталкиваются с тем, что многие, а весьма часто и все коэффициенты парной корреляции этой матрицы оказываются близки по модулю к единице. Это сигнализирует о наличии в имеющихся данных такого явления, как «мультиколлинеарность». Кроме того, некоторые входящие в модель регрессоры могут статистически определяться через другие регрессоры (получается своеобразная модель множественной регрессии в модели множественной регрессии). В этом случае коэффициент множественной корреляции между факторами может быть по модулю близким к единице. Это также будет сигнализировать о наличии мультиколлинеарности.

Мультиколлинеарность, как следует из самого названия, возникает тогда, когда факторы модели имеют одинаковые, монотонные относительно друг друга, тенденции в динамике.

Оценка коэффициентов прогнозной многофакторной модели в условиях мультиколлинеарности затруднена и зачастую при этом получаются такие коэффициенты, которые делают прогнозную модель непригодной к практическому использованию.

Попробуем разобраться в последствиях мультиколлинеарности для оценивания коэффициентов многофакторных моделей на примере из реальной практики.

Пример

По статистическим данным табл. 4.5, в которой отражены показатели экономической динамики одного из регионов России[1], приведенные к относительным величинам, построим с помощью МНК многофакторную линейную модель. Она имеет вид

$$\hat{y} = -1{,}9883 + 0{,}0808x_1 - 0{,}1077x_2 + 2{,}4942x_3 + 0{,}5178x_4. \quad (4.50)$$

Здесь y_t — производство продукции; $x_{1,t}$ — потребление электроэнергии; $x_{2,t}$ — основные производственные фонды; $x_{3,t}$ — численность занятых; $x_{4,t}$ — фонд оплаты труда.

---

[1] *Светуньков С. Г., Благов А. А., Инешин К. А., Козлов М. А.* Развитие Ульяновской области: проблемы моделирования : Ученые записки экономического факультета // Внутривузовский сборник филиала МГУ в Ульяновске. 1991. С. 3.

Если попытаться дать экономическое толкование полученным результатам, то мы вынуждены будем утверждать, что увеличение выпуска промышленной продукции было обусловлено положительным влиянием всех факторов, за исключением стоимости основных производственных фондов, которая отрицательно влияла на рост производства. Отсюда следует, например, рекомендация для руководства региона — для роста объема регионального продукта необходимо уменьшать величину основных фондов предприятий. Очевидно, что полученный вывод, как и результат, не имеет смысла, хотя данная модель хорошо описывает имеющийся ряд наблюдений.

Таблица 4.5

Относительные значения показателей развития региона России

Производство продукции, y_t	Потребление электроэнергии, $x_{1,t}$	Основные производственные фонды, $x_{2,t}$	Численность занятых, $x_{3,t}$	Фонд оплаты труда, $x_{4,t}$
1,0000	1,0000	1,0000	1,0000	1,0000
1,0410	0,9480	1,0994	1,0145	1,0374
1,0910	0,9800	1,2108	1,0323	1,0878
1,1390	1,1338	1,3346	1,0403	1,1267
1,1902	1,1380	1,4840	1,0565	1,1778
1,2438	1,1335	1,6759	1,0780	1,2276
1,3408	1,4320	1,8039	1,0968	1,3105
1,4293	1,5980	1,9623	1,1127	1,3978
1,4936	1,7681	2,0949	1,1109	1,5018
1,5220	2,1061	2,2011	1,1029	1,6133

Если теперь исходные значения показателей табл. 4.5 округлить не до четвертого знака после запятой, а до второго, и вновь оценить параметры многофакторной модели по округленным значениям, модель примет иной вид

$$\hat{y} = -2{,}1715 + 0{,}0772x_1 - 0{,}1949x_2 + 2{,}6025x_3 + 0{,}6863x_4 . \qquad (4.51)$$

Если сравнить полученные коэффициенты с коэффициентами модели (4.50), то можно увидеть, что коэффициенты различаются.

Для подтверждения того, что модель строилась в условиях мультиколлинеарности, необходимо вычислить значения корреляционной матрицы. Результаты расчета коэффициентов парной корреляции для факторов табл. 4.5 приведены в табл. 4.6.

Таблица 4.6

Матрица коэффициентов парной корреляции для многофакторной модели

Признак	y_t	$x_{1,t}$	$x_{2,t}$	$x_{3,t}$	$x_{4,t}$
y_t	1				
$x_{1,t}$	0,9470	1			
$x_{2,t}$	0,9953	0,9283	1		
$x_{3,t}$	0,9646	0,8348	0,9730	1	
$x_{4,t}$	0,9874	0,9777	0,9810	0,9159	1

Все коэффициенты парной корреляции, приведенные в табл. 4.6, близки к единице, что подтверждает множественную коллинеарность и факторов, и результирующего показателя. Поэтому все те неприятности с оценкой коэффициентов модели, которые были продемонстрированы выше, как раз и вызваны наличием мультиколлинеарности.

Если рассчитать коэффициенты множественной корреляции последовательно между x_1 и остальными факторами, x_2 и другими факторами и т.п., то мы получим следующие значения:

- $R_{x_1} = 0{,}9908$;
- $R_{x_2} = 0{,}9987$;
- $R_{x_3} = 0{,}9933$;
- $R_{x_4} = 0{,}9983$.

Как видим, все они очень близки к 1. Это так же указывает на наличие мультиколлинеарности, которая в данном случае проявляется в сильной форме.

Обычно для оценки силы мультиколлинеарности в эконометрике рассчитывают показатель *VIF* («Variance Inflation Factor» — фактор инфляции дисперсии):

$$VIF = \frac{1}{1 - R_{x_j}^2}.$$

Этот показатель лежит в пределах от 0 до бесконечности, что неудобно интерпретировать. В том случае, если значения *VIF* превышают 10 (хотя в некоторых источниках встречается и критическое значение в 5), обычно говорят о мультиколлинеарности в сильной форме. В целом можно заметить, что значения *VIF* будут больше 10 в том случае, если значение коэффициента множественной корреляции будет по модулю превышать 0,95. Таким образом, вместо того, чтобы рассчитывать сложно интерпретируемый показатель, для оценки мультиколлинеарности можно использовать просто коэффициент множественной корреляции.

Как видно из приведенного выше примера, оценки коэффициентов многофакторной модели, если она строится в условиях мультиколлинеарности, оказываются неустойчивыми как к ошибкам наблюдений, так и к новой информации. Чтобы понять, почему это происходит, рассмотрим *задачу оценивания коэффициентов многофакторной модели с помощью МНК в матричном виде*.

1. Исходя из формальной постановки задачи, никаких проблем возникнуть не может — все логично и легко вычисляется. Но в условиях мультиколлинеарности квадратная матрица $X'X$ в (4.34) является слабо обусловленной, поэтому при вычислении обратной матрицы как раз и возникают проблемы — определитель матрицы практически равен нулю, поэтому матрица является необратимой. В результате этого точно оценить коэффициенты модели не представляется возможным.

2. Если все же вычислить определитель, используя современную технику и вычисляя его значения, например, до 20 знака после запятой, можно исчислить обратную матрицу и вычислить вектор оценок A, но оценки параметров многофакторных моделей будут очень неточными из-за указанной выше причины.

3. Оценки параметров модели оказываются крайне неустойчивыми — малейшее изменение исходных данных или даже ошибки округления приводят к значительным изменениям параметров.

4. Ценность таких моделей крайне низка, так как неустойчивая модель дает очень сильную вариацию своих коэффициентов, а значит и расчетных значений прогнозируемого показателя. В результате этого исследователь будет получать некорректные доверительные интервалы как по коэффициентам модели, так и по прогнозным значениям.

5. Модель, с помощью которой сделана попытка описать сложное многофакторное явление, не описывает это явление. Причинно-следственные связи, выявленные на предыдущем этапе, не описываются этой моделью — в некоторых случаях можно даже получить отрицательные коэффициенты при факторах, оказывающих прямое положительное влияние на результат! Модель ведет себя хуже, чем однофакторная.

6. Так как оценки параметров оказываются неточными, то интерпретация влияния факторов на прогнозируемый показатель будет совершенно не той, которая есть на самом деле. Именно поэтому построение многофакторных моделей в условиях мультиколлинеарности является предметом многолетнего внимания ученых. Ученые единодушны в том, что применение МНК к задаче оценивания коэффициентов многофакторных моделей приводит к необходимости решения системы нормальных уравнений, которая в условиях мультиколлинеарности является почти «вырожденной», а потому и неустойчивой к малейшим ошибкам округления. Поэтому, если исходить из такой постановки проблемы, понятны направления борьбы с проклятием «мультиколлинеарности» — либо устранять саму мультиколлинеарность, либо разрабатывать математические методы, устойчивые в этих условиях.

В первом направлении выделяются два подхода:

- исключают из совокупности факторов одну или несколько линейно связанных факторных переменных, чтобы вновь полученные элементы корреляционной матрицы были меньше порогового значения 0,8. Однако само пороговое значение в разных источниках разниться. В некоторых случаях рекомендуется избавляться от факторов, коэффициент корреляции между которыми оказывается больше 0,95;
- преобразуют факторы в новые переменные, уменьшая тем самым количество переменных (факторный анализ).

Во втором направлении — совершенствования математического аппарата оценивания коэффициентов модели — используют методы регуляризации — раздел, называющийся «робастной статистикой» (например, метод главных компонент, гребневой регрессии, стабилизированных оценок или «сжатых» оценок и т.п.).

Оба подхода обладают серьезными недостатками, которые снижают эффективность использования на практике многофакторных моделей.

В первом случае попытки устранения самого явления мультиколлинеарности рекомендуется исключать из модели один из двух факторов, если значение парного коэффициента корреляции между ними превышает по модулю «пороговое» значение, равное 0,8. Логика такого подхода базируется на том, что если этот фактор линейно связан с другими, то их влияние на результат отражает и влияние этого фактора, поэтому без него можно и обойтись. Но сама процедура исключения из совокупности факторов одной или нескольких факторных переменных весьма сомнительна — так можно прийти и к простой однофакторной линейной модели. И с экономических позиций эта процедура неверна — если специалист отобрал совокупность факторов, объясняющих динамику показателя, то уменьшение их числа только ухудшит свойства модели. А если вернуться к рассмотренному выше примеру с построением многофакторной модели динамики Ульяновской области, то анализ матрицы коэффициентов парной корреляции показывает, что все ее значения превышают пороговое значение в 0,8, поэтому, следуя этой логике, надо отбросить все факторы, оставив в рассмотрении только один из них. Но ведь желание построить многофакторную модель вызвано вполне очевидным стремлением вскрыть и описать сложную причинно-следственную связь между социально-экономическими явлениями и тем самым повысить точность прогноза, а вместо этого мы опять приходим к элементарной однофакторной модели.

Другой способ побороть мультиколлинеарность связан с использованием методов факторного анализа, когда пытаются заменить реальные переменные $x_{j,\,t}$ на другие факторы $z_{k,\,t}$, которые как бы являются скрытой причиной их динамики. Такие переменные неизвестны исследователю и называются латентными. С помощью математических преобразований (наиболее известен метод главных компонент) находят эти новые переменные и строят новую многофакторную модель зависимости переменной y_t от новых факторов $z_{k,\,t}$. Однако при этом модель теряет какой-либо экономический смысл, так как латентные переменные в данной ситуации являются математической абстракцией. И хотя модель становится более устойчивой к ошибкам округления, для выполнения прогноза показателя y_t необходимо вначале вычислить прогнозные значения каждой из этих новых переменных $z_{k,\,t}$ с помощью формул преобразования с учетом исходных переменных $x_{j,\,t}$. А поскольку исходные переменные наблюдаются с ошибками, то эти ошибки способствуют тому, что и новые переменные будут содержать их в себе. Но так как находится регрессионная зависимость между y_t и $z_{k,\,t}$, то и эта зависимость генерирует ошибку регрессии. Все эти пересчеты вносят дополнительную погрешность в прогнозирование результирующего признака, и модель становится менее точной, чем хотелось бы.

Второе направление, связанное с совершенствованием математического аппарата оценивания коэффициентов в условиях мультиколлинеарности, представляет собой некоторую попытку «утяжеления» задачи посредством намеренного введения «засоряющей» величины. Здесь вполне допустима следующая аналогия: если какую-либо конструкция неустойчива, то повысить ее устойчивость можно, загрузив конструкцию балластом. В случае многофакторного моделирования построенная таким образом модель в результате утяжеления станет не оптимальной, но вполне устойчивой, что уже следует признать положительным моментом. Этот путь решения проблемы мультиколлинеарности более приемлем, чем первые, однако он требует от исследователя хорошей математической подготовки, так как аппарат методов регуляризации достаточно сложен. Тем не менее, квалифицированные специалисты предпочитают использовать именно методы робастной статистики для построения многофакторных моделей в условиях мультиколлинеарности.

Как это ни кажется парадоксальным, но для случая нелинейной многофакторной модели эта проблема решается достаточно просто! Если прогнозист использует в качестве прогнозной модели нелинейную многофакторную модель, например, такую: $\hat{y}_t = a_0 + a_1 x_{1,t} x_{2,t} + a_2 x_{1,t} x_{2,t} x_{3,t} + a_3 x_{3,t}^2$, то для нахождения параметров a_0, a_1, a_2 и a_3 осуществляют замену переменных и представляют модель в простой линейной форме: $\hat{y}'_t = a_0 + a_1 x'_{1,t} + a_2 x'_{2,t} + a_3 x'_{3,t}$, где $x'_{1,t} = x_{1,t} x_{2,t}$, $x'_{2,t} = x_{1,t} x_{2,t} x_{3,t}$, $x'_{3,t} = x_{3,t}^2$.

Для вновь полученной модели и новых переменных строят корреляционную матрицу и если преобразованные новые переменные коррелируют друг с другом и коэффициенты парной корреляции выше по абсолютному значению, чем 0,8, то для уменьшения взаимной коррелированности переменных предлагается центрировать исходные переменные $x_{j,t}$.

В этом случае новые переменные существенно изменятся нелинейно друг относительно друга, в результате чего корреляционная матрица новых переменных сильно изменит свои значения в сторону уменьшения коэффициентов парных корреляций между ними. Это доказал еще в 1969 г. Г. Ф. Филаретов[1]. Однако для линейной многофакторной модели центрирование исходных данных (как и любые другие аффинные преобразования переменных) не изменит коэффициентов корреляционной матрицы и мультиколлинеарность не устранится. Поэтому складывается парадоксальная ситуация: сложные многофакторные нелинейные

---

[1] *Филаретов Г. Ф.* К вопросу о построении нелинейной регрессионной модели по данным пассивного эксперимента // Проблемы планирования эксперимента / под ред. Г. К. Круга. М. : Наука, 1969.

модели в условиях мультиколлинеарности можно построить достаточно просто, а элементарные линейные многофакторные модели в этих же условиях построить не удается!

Для того чтобы проследить направление действий при решении сформулированной задачи, рассмотрим ее постановку в пространстве коэффициентов модели. В параграфе 4.2 учебника мы рассматривали задачу оценивания коэффициентов однофакторной модели на плоскости коэффициентов этой модели. Для случая многофакторной модели число коэффициентов чаще всего больше двух, поэтому следует рассматривать не плоскость коэффициентов (две переменные), а пространство коэффициентов.

Для этого вначале запишем систему нормальных уравнений МНК для многофакторной линейной модели (4.26), для простоты записи опуская пределы суммирования, понимая, что оно ведется для всех t, начиная с $t = 1$, и заканчивая $t = T$:

$$\begin{cases} \sum\limits_t y_t = Ta_0 + a_1\sum\limits_t x_{1,t} + \ldots + a_k\sum\limits_t x_{k,t}, \\ \sum\limits_t y_t x_{1,t} = a_0\sum\limits_t x_{1,t} + a_1\sum\limits_t x_{1,t}^2 + \ldots + a_k\sum\limits_t x_{1,t}x_{k,t}, \\ \ldots \\ \sum\limits_t y_t x_{k,t} = a_0\sum\limits_t x_{k,t} + a_1\sum\limits_t x_{1,t}x_{k,t} + \ldots + a_k\sum\limits_t x_{k,t}^2. \end{cases} \quad (4.52)$$

Теперь представим эту систему уравнений в виде уравнений в отрезках наподобие того, как мы это делали в параграфе 4.2. Получим

$$\begin{cases} 1 = \dfrac{a_0}{\dfrac{\sum_t y_t}{T}} + \dfrac{a_1}{\dfrac{\sum_t y_t}{\sum_t x_{1,t}}} + \ldots + \dfrac{a_k}{\dfrac{\sum_t y_t}{\sum_t x_{k,t}}}, \\ 1 = \dfrac{a_0}{\dfrac{\sum_t y_t x_{1,t}}{\sum_t x_{1,t}}} + \dfrac{a_1}{\dfrac{\sum_t y_t x_{1,t}}{\sum_t x_{1,t}^2}} + \ldots + \dfrac{a_k}{\dfrac{\sum_t y_t x_{1,t}}{\sum_t x_{1,t}x_{k,t}}}, \\ \ldots \\ 1 = \dfrac{a_0}{\dfrac{\sum_t y_t x_{k,t}}{\sum_t x_{k,t}}} + \dfrac{a_1}{\dfrac{\sum_t y_t x_{k,t}}{\sum_t x_{1,t}x_{k,t}}} + \ldots + \dfrac{a_k}{\dfrac{\sum_t y_t x_{k,t}}{\sum_t x_{k,t}^2}}. \end{cases} \quad (4.53)$$

Уравнения системы МНК здесь представлены в виде уравнений гиперплоскостей в отрезках в гиперпространстве коэффициентов модели. Если в однофакторном случае оценки МНК представляют собой точку пересечения на плоскости коэффициентов двух прямых условий МНК, поскольку неизвестных параметров всего два (a_0 и a_1) и задачу можно изобразить на плоскости, то уже при числе факторов, равном двум, число коэффициентов модели будет равно трем — a_0, a_1 и a_2. Графически такую задачу оценивания параметров многофакторной модели следует рассмотреть не на плоскости, а в трехмерном пространстве параметров — число неизвестных параметров станет равным трем и их можно будет изобразить в качестве осей трехмерного пространства — $0a_0$, $0a_1$ и $0a_2$. В этом случае условия МНК представляют собой систему из трех уравнений с тремя неизвестными, причем каждое из уравнений является уравнением плоскости в пространстве. Решение системы МНК в данном случае будет представлять собой точку пересечения трех плоскостей в пространстве. Координаты точки пересечения этих плоскостей и дадут искомые с помощью МНК значения коэффициентов модели.

Пример

Воспользуемся уравнениями в отрезках (4.53), чтобы более тщательно изучить причины неустойчивости оценок параметров в условиях мультиколлинеарности на примере показателей экономической динамики Ульяновской области, рассмотренном выше. Используя систему МНК для абсолютных значений показателей развития области, приведем эту систему к форме уравнений в отрезках. Такое преобразование системы покажет нам отрезки, которые отсекают гиперплоскости условий многофакторного МНК на осях гиперпространства параметров:

$$\begin{cases} 1 = \dfrac{a_0}{1{,}2491} + \dfrac{a_1}{0{,}9436} + \dfrac{a_2}{0{,}7872} + \dfrac{a_3}{1{,}1734} + \dfrac{a_4}{1{,}0008}, \\ 1 = \dfrac{a_0}{1{,}2965} + \dfrac{a_1}{0{,}9085} + \dfrac{a_2}{0{,}7665} + \dfrac{a_3}{1{,}2075} + \dfrac{a_k}{0{,}9965}, \\ 1 = \dfrac{a_0}{1{,}2944} + \dfrac{a_1}{0{,}9174} + \dfrac{a_2}{0{,}7662} + \dfrac{a_3}{1{,}2050} + \dfrac{a_k}{0{,}9985}, \\ 1 = \dfrac{a_0}{1{,}2555} + \dfrac{a_1}{0{,}9403} + \dfrac{a_2}{0{,}7841} + \dfrac{a_3}{1{,}1778} + \dfrac{a_4}{1{,}0007}, \\ 1 = \dfrac{a_0}{1{,}2766} + \dfrac{a_1}{0{,}9251} + \dfrac{a_2}{0{,}7745} + \dfrac{a_3}{1{,}1929} + \dfrac{a_4}{0{,}9987}. \end{cases}$$

Из анализа величин полученных отрезков легко убедиться в том, что отрезки, отсекаемые гиперплоскостями условий МНК на каждой из осей гиперпространства коэффициентов, практически совпадают друг с другом, т.е. сами гиперплоскости почти параллельны друг другу. Очевидно, поэтому, что решение системы МНК, которое представляет собой точку пересечения этих практически параллельных друг другу гиперплоскостей в гиперпространстве, является чрезвычайно неустойчивым — малейшая ошибка в округлении может привести к тому, что гиперплоскости могут, перемещаясь незначительно, иметь новую точку их пересечения, значительно удаленную от первоначальной точки. Например, отрезок на оси a_4 первой гиперплоскости величиной 1,0008 может стать равным 1,0007 (например, в результате округления). При этом первая гиперплоскость практически незаметно изменит свое положение в пространстве, но решение системы МНК, как точка пересечения гиперплоскостей, меняется так, что искажается не только абсолютная величина коэффициентов многофакторной модели, но и сам знак этих коэффициентов, что и было получено выше.

Следовательно, последствия мультиколлинеарности для оценивания коэффициентов многофакторных линейных моделей вызваны особенностями существующего алгоритма оценивания параметров многофакторных моделей с помощью МНК. Исследования в этом случае должны быть направлены не на борьбу с объективно существующей реальностью — сильной коллинеарностью практически всех показателей и факторов социально-экономической динамики (что подтверждается явлением ложной корреляции), а на улучшение используемого аппарата оценивания таких моделей. Этот вывод справедлив и при построении простых линейных однофакторных моделей, правда, неустойчивость полученных оценок, в отличие от многофакторного случая, не особенно заметна — в последнем случае ситуация усугубляется еще и вычислительными сложностями, поэтому и становится такой явной.

4.4. Один метод получения устойчивых оценок в частном случае мультиколлинеарности

Если взаимосвязь между факторами близка к линейной, но не является таковой, то так или иначе удается построить многофакторную модель. Коэффициенты корреляции между факторами лежат в пределах до 0,95. Однако иногда на практике встречаются ситуации, когда между ними существует практически линейная функциональная зависимость, и коэффициенты парной корреляции почти равны единице (лежат в пределах от 0,95 до 1). В этих случаях оценки многофакторных моделей найти невозможно.

Пример

Пусть мы наблюдаем некоторый социально-экономический процесс, который описывается функциональной двухфакторной линейной моделью со свободным членом, равным нулю:

$$y_t = 7{,}3x_{1,t} + 2x_{2,t}. \quad (4.54)$$

Факторы $x_{1,t}$ и $x_{2,t}$ в рассматриваемый период меняются линейно и это изменение описывается линейной функциональной зависимостью одного фактора от другого:

$$x_{2,t} = \frac{2}{7{,}3}x_{1,t} \approx 0{,}274x_{1,t}. \quad (4.55)$$

Очевидно, что в таком случае коэффициент парной корреляции между ними будет равен единице, что свидетельствует о крайнем случае мультиколлинеарности — о линейной функциональной зависимости между ними. В табл. 4.7 в соответствии с приведенной формулой (4.54) при линейно изменяющихся переменных рассчитаны величины результирующего показателя y_t. Следует еще раз подчеркнуть, что все данные табл. 4.7 — результаты вычисления строгой функциональной зависимости (4.54) между всеми переменными.

Таблица 4.7

Данные условного примера

t	$x_{1,t}$	$x_{2,t}$	Расчетное значение y_t, полученное по (4.54)
1	2	0,547945205	15,69589
2	4	1,095890411	31,39178
3	6	1,643835616	47,08767
4	8	2,191780822	62,78356
5	10	2,739726027	78,47945
6	12	3,287671233	94,17534
7	14	3,835616438	109,8712
8	16	4,383561644	125,5671
9	18	4,931506849	141,263
10	20	5,479452055	156,9589
11	22	6,02739726	172,6548
12	24	6,575342466	188,3507

Окончание табл. 4.7

t	$x_{1,t}$	$x_{2,t}$	Расчетное значение y_t, полученное по (4.54)
13	26	7,123287671	204,0466
14	28	7,671232877	219,7425
15	30	8,219178082	235,4384
16	32	8,767123288	251,1342
17	34	9,315068493	266,8301
18	36	9,863013699	282,526
19	38	10,4109589	298,2219
20	40	10,95890411	313,9178

Теперь поставим обратную задачу — по данным этой таблицы, не зная коэффициенты зависимости (4.54), построить многофакторную линейную модель, т.е. рассчитать по данным таблицы значения коэффициентов модели со свободным членом, равным нулю: $y_t = a_1 x_{1,t} + a_2 x_{2,t}$.

Поскольку между всеми переменными задачи имеется строго функциональная зависимость, на первый взгляд может показаться, что задача будет легко решена, поскольку в ней нет ни одной случайной ошибки и неопределенности — задача полностью детерминирована. И, конечно же, хотелось бы ожидать особой точности от применения МНК, такой, чтобы он позволил найти коэффициенты модели так, чтобы сумма квадратов отклонений расчетных значений от фактических была бы равна нулю.

Используем этот метод для нахождения оценок исходной модели (4.54) по табличным данным, зная, что свободный член равен нулю. Система нормальных уравнений для такой модели будет иметь вид

$$\begin{cases} \sum_t y_t x_{1,t} = a_1 \sum_t x_{1,t}^2 + a_2 \sum_t x_{2,t} x_{1,t}, \\ \sum_t y_t x_{2,t} = a_1 \sum_t x_{1,t} x_{2,t} + a_2 \sum_t x_{2,t}^2. \end{cases}$$

Уравнения, с помощью которых можно вычислить коэффициенты, рассчитанные по данным табл. 4.7, примут вид такой системы:

$$\begin{cases} 90\,094,411 = a_1 11\,480 + a_2 3145,206, \\ 24\,683,403 = a_1 3145,206 + a_2 861,700. \end{cases}$$

Следующий шаг — решить систему и получить искомые значения коэффициентов. Но прежде, чем сделать это, приведем эту систему уравнений к системе уравнений в отрезках. Получим

$$\begin{cases} 1 = \dfrac{a_1}{7{,}8479452} + \dfrac{a_2}{28{,}645}, \\ 1 = \dfrac{a_1}{7{,}8479452} + \dfrac{a_2}{28{,}645}. \end{cases} \tag{4.56}$$

Как видно, система уравнений вырождена и не имеет решений. Плоскости уравнений МНК полностью совпадают друг с другом, т.е. не могут иметь точку пересечения. Значит, многофакторный МНК в данном случае не позволяет построить многофакторную модель.

В рассмотренной задаче между всеми переменными есть строгая линейная зависимость, которая в трехмерном пространстве исходных переменных представляет собой точки, лежащие на одной прямой линии. Поскольку эти точки лежат на одной прямой, то существует бесконечное количество плоскостей, которым будет принадлежать эта прямая линия. Система (4.56) это и демонстрирует — она дает возможность бесконечного сочетания пар значений коэффициентов модели плоскости в трехмерном пространстве, т.е. позволяет получить бесконечное множество уравнений плоскостей, пересекающихся друг с другом и содержащих в себе прямую линию. Одной из таких плоскостей и является плоскость, описываемая уравнением (4.54). Но выделить эту плоскость из бесконечного числа других, как видно, с помощью МНК не представляется возможным.

Одним из эффективных методов решения поставленной задачи является переход из области действительных переменных в область комплексных переменных[1]. Для этого рассмотрим линейные модели комплексного аргумента.

В общем виде линейная модель комплексного аргумента может быть записана так:

$$\hat{y}_t = (a_1 + ia_2)(x_{1,t} + ix_{2,t}) + (b_1 + ib_2). \tag{4.57}$$

Здесь i — мнимая единица, о которой известно, что ее квадрат равен минус единице ($i = \sqrt{-1}$).

---

[1] *Svetunkov S.* Complex-Valued Modeling in Economics and Finance. New York : Springer Science + Business Media, 2012.

После центрирования исходных переменных относительно их средних арифметических эта модель будет выглядеть проще, поскольку сокращается свободный член:

$$\hat{y}_t = (a_1 + ia_2)(x_{1,t} + ix_{2,t}). \tag{4.58}$$

Выделяя действительную и мнимую части этого равенства, получим систему двух уравнений:

$$\begin{cases} \hat{y}_t = a_1 x_{1,t} - a_2 x_{2,t}, \\ 0 = a_2 x_{1,t} + a_1 x_{2,t}. \end{cases}$$

откуда

$$\begin{cases} \hat{y}_t = a_1 x_{1,t} - a_2 x_{2,t}, \\ x_{1,t} = -\dfrac{a_2}{a_1} x_{2,t}. \end{cases} \tag{4.59}$$

Первое уравнение полученной системы (4.59) говорит о том, что результат линейно зависит от факторных переменных, а второе — свидетельствует о том, что для линейной функции комплексного аргумента свойственна линейная функциональная зависимость между факторными переменными. Иначе говоря, *линейная функция комплексного аргумента имеет право на существование исключительно в условиях линейной взаимосвязи между факторами модели, т.е. в условиях проявления крайнего случая мультиколлинеарности — линейной функциональной зависимости между факторами*! Ученые всеми силами борются за то, чтобы устранить последствия действия мультиколлинеарности, а функция комплексного аргумента в иных случаях использоваться и не должна.

Применим МНК к оценке коэффициентов комплекснозначной модели (4.58)[1].

Комплекснозначная функция, минимум которой соответствует оценкам МНК коэффициентов линейной модели комплексного аргумента, запишется так:

$$f(z) = \sum_t \left[y_t - (a_0 + ia_1)(x_{rt} + ix_{it}) \right]^2. \tag{4.59}$$

[1] *Svetunkov S.* Complex-Valued Modeling in Economics and Finance. New York : Springer Science + Business Media, 2012.

Воспользовавшись формулами Римана-Коши, можно найти частные производные действительной части этой функции по каждому из коэффициентов и приравнять их нулю:

$$\frac{\partial \mathrm{Re}(f(z))}{\partial a_0} = 2(a_0 \sum_t x_{rt}^2 - a_0 \sum_t x_{it}^2 - 2a_1 \sum_t x_{rt} x_{it} - \sum_t x_{rt} y_t) = 0;$$

$$\frac{\partial \mathrm{Re}(f(z))}{\partial q_1} = 2(-a_1 \sum_t x_{rt}^2 + a_1 \sum_t x_{it}^2 - 2a_0 \sum_t x_{rt} x_{it} + \sum_t x_{it} y_t) = 0.$$

Отсюда система МНК будет записана в виде

$$\begin{cases} \sum_t y_t x_{1,t} = a_1 \sum_t \left(x_{1,t}^2 - x_{2,t}^2\right) - 2a_2 \sum_t x_{2,t} x_{1,t}, \\ \sum_t y_t x_{2,t} = a_2 \sum_t \left(x_{1,t}^2 - x_{2,t}^2\right) + 2a_1 \sum_t x_{2,t} x_{1,t}. \end{cases} \qquad (4.60)$$

или в отрезках:

$$\begin{cases} 1 = \dfrac{a_1}{\dfrac{\sum_t y_t x_{1,t}}{\sum_t \left(x_{1,t}^2 - x_{2,t}^2\right)}} - \dfrac{a_2}{\dfrac{\sum_t y_t x_{1,t}}{2\sum_t x_{2,t} x_{1,t}}}, \\ 1 = \dfrac{a_1}{\dfrac{\sum_t y_t x_{2,t}}{2\sum_t x_{2,t} x_{1,t}}} + \dfrac{a_2}{\dfrac{\sum_t y_t x_{2,t}}{\sum_t \left(x_{1,t}^2 - x_{2,t}^2\right)}}. \end{cases} \qquad (4.61)$$

Как видно из последней системы уравнений, отрезки уравнений в общем случае не совпадают, поэтому можно ожидать, что пересечение этих линий на плоскости коэффициентов будет иметь место и в случае крайней мультиколлинеарности, т.е. задача имеет решение.

Пример

Применительно к данным табл. 4.7, по которым не удалось с помощью МНК оценить значения двухфакторной линейной модели, используем линейную функцию комплексного аргумента и с помощью МНК найдем значения коэффициентов этой модели. Рассчитаем все необходимые суммы системы (4.60). Тогда для рассматриваемых данных она примет вид

$$\begin{cases} 90\,094{,}41 = a_1 10\,618{,}29 + a_2 6290{,}41, \\ 24\,683{,}4 = a_1 10\,618{,}29 + a_2 6290{,}41. \end{cases}$$

Приведем ее к виду уравнений в отрезках (4.61):

$$\begin{cases} 1 = \dfrac{a_0}{8,4848} + \dfrac{a_1}{14,3225}, \\ 1 = \dfrac{a_1}{2,3246} + \dfrac{a_0}{3,9239}. \end{cases}$$

Легко убедиться в том, что, решив эту систему нормальных уравнений, мы получим единственное и очень устойчивое решение. Сделаем это. Модель комплексного аргумента, коэффициенты которой найдены с помощью МНК по данным табл. 4.7, имеет вид $\hat{y}_t = (7,3 - 2i)(x_{1,t} + ix_{2,t})$.

Если теперь раскрыть скобки и сгруппировать действительную и мнимую части полученного равенства, то оно преобразуется в следующее: $\hat{y}_t = (7,3x_{1,t} + 2x_{2,t}) + i(-2x_{1,t} + 7,3x_{2,t})$.

Действительная часть полученного равенства полностью соответствует исходной функции (4.54), а мнимая часть легко преобразуется в такое равенство: $0 = i(-2x_{1,t} + 7,3x_{2,t}) \rightarrow x_{2,t} = \dfrac{2}{7,3}x_{1,t} = 0,273972603x_{1,t}$.

Таким образом, мнимой частью линейной модели комплексного аргумента абсолютно точно идентифицируется исходная зависимость между факторами (4.55).

Прежде, чем сделать некоторый вывод об использовании на практике этого частного случая построения двухфакторной линейной модели в условиях сильнейшей мультиколлинеарности, следует указать на принципиальное отличие линейной двухфакторной модели действительных переменных

$$y_t = a_1x_{1,t} + a_2x_{2,t} \tag{4.62}$$

от линейной модели комплексного аргумента

$$\hat{y}_t = (a_1 + ia_2)(x_{1,t} + ix_{2,t}). \tag{4.63}$$

Несмотря на то, что в модели (4.62), как и в модели (4.63) две переменные и два коэффициента, они описывают разные фигуры.

Модель действительных переменных (4.62) представляет собой уравнение плоскости в пространстве исходных переменных. Модель же комплексного аргумента (4.63) — уравнение прямой линии в этом пространстве.

Это значит, что в ситуации, когда между переменными $x_{1,t}$ и $x_{2,t}$ нет линейной зависимости, модель комплексного аргумента будет давать плохие результаты.

Наблюдаемые точки в этой ситуации расположены на некоторой плоскости в пространстве с некоторым отклонением от прямой линии, причем, чем меньше коэффициент парной корреляции между ними, тем больше отклоняются точки на плоскости, описываемой моделью (4.62) от прямой линии. Но чем ближе точки в трехмерном пространстве приближаются к одной прямой линии, чем ближе к единице значение модуля коэффициента парной корреляции между объясняющими факторами, тем сильнее влияние мультиколлинеарности и тем больше оснований для использования линейной модели комплексного аргумента.

Следует сделать еще одно замечание относительно области применения линейной модели комплексного аргумента — она может использоваться в случае двухфакторной модели. Методы использования этого инструмента в моделях с большим числом факторов на данный момент еще не разработаны.

4.5. Синтез однофакторных моделей в многофакторную

Как следует из приведенных выше материалов, в условиях мультиколлинеарности в большей части случаев не удается построить приемлемую с позиций устойчивости оценок коэффициентов модель.

Но устойчивость модели, если ее и удается достичь, еще не означает того, что модель действительно описала вскрытые причинно-следственные связи и пригодна для прогнозирования.

Как мы показали в некоторых примерах, приведенных в предыдущих параграфах, очень часто приходится встречаться с ситуациями, когда коэффициенты перед переменными становятся отрицательными, хотя исследователю очевиден положительный вклад переменной в динамику результирующего признака. Здесь можно использовать способ синтеза однофакторных моделей в многофакторную, суть которого заключается в следующем.

Строится система однофакторных моделей, которые с разной степенью точности аппроксимируют результирующий показатель:

$$\begin{cases} y_t = f_1\left(x_{1,t}\right) + \varepsilon_{1,t}, \\ y_t = f_2\left(x_{2,t}\right) + \varepsilon_{2,t}, \\ y_t = f_3\left(x_{3,t}\right) + \varepsilon_{3,t}, \\ \ldots \qquad\qquad\qquad , \\ y_t = f_k\left(x_{k,t}\right) + \varepsilon_{k,t}. \end{cases} \tag{4.64}$$

Левые части равенств равны друг другу, как равны друг другу правые части равенств этой системы. Сложим левые и правые части равенств, получим

$$ky_t = f_1\left(x_{1,t}\right) + \varepsilon_{1,t} + f_2\left(x_{2,t}\right) + \varepsilon_{2,t} + \ldots + f_k\left(x_{k,t}\right) + \varepsilon_{k,t}. \quad (4.65)$$

Группируя и разделив левую и правую части равенства на число моделей k, получим

$$y_t = \frac{1}{k}\left(f_1\left(x_{1,t}\right) + f_2\left(x_{2,t}\right) + \ldots + f_k\left(x_{k,t}\right)\right) + \frac{1}{k}\left(\varepsilon_{1,t} + \varepsilon_{2,t} + \ldots + \varepsilon_{k,t}\right). \quad (4.66)$$

Таким образом, результирующий показатель описывается с помощью многофакторной модели такого вида:

$$\hat{y}_t = \frac{1}{k}\left(f_1\left(x_{1,t}\right) + f_2\left(x_{2,t}\right) + \ldots + f_k\left(x_{k,t}\right)\right) = \\ = \frac{1}{k}f_1\left(x_{1,t}\right) + \frac{1}{k}f_2\left(x_{2,t}\right) + \ldots + \frac{1}{k}f_k\left(x_{k,t}\right). \quad (4.67)$$

Ошибка ее аппроксимации на каждом наблюдении равна

$$\varepsilon_t = \frac{1}{k}\left(\varepsilon_{1,t} + \varepsilon_{2,t} + \ldots + \varepsilon_{k,t}\right) = \frac{1}{k}\varepsilon_{1,t} + \frac{1}{k}\varepsilon_{2,t} + \ldots + \frac{1}{k}\varepsilon_{k,t}. \quad (4.68)$$

Как видно, в этом случае почти всегда удается избежать ситуации, когда перед фактором, оказывающим положительное влияние на результат, в многофакторной модели оказывается знак «минус». Однако однофакторные модели в таком случае страдают от проблемы пропущенных переменных, поэтому определить истинный знак перед регрессором может быть крайне затруднительно. В ряде случаев исследователю потребуется самостоятельно вносить коррективы в то, с каким знаком должен быть учтен фактор в модели.

Пример

Построим многофакторную линейную модель по данным табл. 4.5, используя метод синтеза однофакторных моделей в многофакторную (4.67).

Первая модель зависимости объема производства продукции от объема потребленной промышленностью электроэнергии, найденная с помощью МНК, имеет вид

$$\hat{y}_t = 0{,}4589x_{1,t} + 0{,}6416. \quad (4.69)$$

Вторая модель, описывающая зависимость объема производства от величины основных производственных фондов региональной экономики, оцененная по данным этой таблицы с помощью МНК, имеет вид

$$\hat{y}_t = 0{,}4419x_{2,t} + 0{,}5479. \qquad (4.70)$$

Третья модель, которая описывает зависимость объема регионального производства от численности занятых в промышленности, будет записана так:

$$\hat{y}_t = 4{,}3869x_{3,t} - 3{,}4208. \qquad (4.71)$$

Последняя однофакторная модель, необходимая для дальнейших построений, — это модель зависимости объема регионального производства от фонда оплаты труда:

$$\hat{y}_t = 0{,}9128x_{4,t} + 0{,}1099. \qquad (4.72)$$

Теперь легко построить общую многофакторную линейную модель:

$$\hat{y}_t = \frac{1}{4}\begin{pmatrix} 0{,}4589x_{1,t} + 0{,}6416 + 0{,}4419x_{2,t} + 0{,}5479 + 4{,}3869x_{3,t} - \\ - 3{,}4208 + 0{,}9128x_{4,t} + 0{,}1099 \end{pmatrix}$$

или, в более удобном для использования виде:

$$\hat{y}_t = -0{,}5303 + 0{,}1147x_{1,t} + 0{,}1105x_{2,t} + 1{,}0967x_{3,t} + 0{,}2282x_{4,t}. \qquad (4.73)$$

Как видно, действительно получена многофакторная линейная модель, у которой все коэффициенты имеют более разумный экономический смысл, нежели у тех многофакторных моделей, которые были получены ранее в параграфе 4.3.

Многофакторная модель (4.67) формируется, исходя из предположения, что вклад каждого фактора, а, значит, и каждой однофакторной модели в общую дисперсию одинаков и равен $1/k$. Но если перед прогнозистом стоит задача построить многофакторную модель на основе синтеза однофакторных так, чтобы она лучше всего описывала прошлые значения, то естественно положить такое правило: модели с меньшей дисперсией ошибок учитываются в большей степени, а модели с большей дисперсией — в меньшей.

Тогда возникает необходимость вычисления весовых коэффициентов v_j, которые следует задавать каждой однофакторной модели для их синтеза в многофакторную модель, но делать это необходимо так, чтобы каждый коэффициент был положителен и сумма этих весовых коэффициентов была равна единице:

$$\sum_{j=1}^{k} \nu_j = 1. \tag{4.74}$$

Это требование является не обязательным, но очень удобным — тем самым определяется доля вклада каждой модели в общую вариацию прогнозируемого показателя.

С учетом весовых коэффициентов, модель (4.67) может быть представлена так:

$$\hat{y}_t = \nu_1 f_1\left(x_{1,t}\right) + \nu_2 f_2\left(x_{2,t}\right) + \ldots + \nu_k f_k\left(x_{k,t}\right). \tag{4.75}$$

Остается только найти эти весовые коэффициенты. Одна из наиболее простых формул для их расчета, при которой вес модели, имеющей максимальную дисперсию ошибок, является минимальным и имеет такой вид:

$$\nu_j = \frac{1}{k-1} \frac{\sigma^2 - \sigma_j^2}{\sigma^2}. \tag{4.76}$$

Здесь σ^2_j — дисперсия ошибок j-й однофакторной модели, с помощью которой вычисляется суммарная дисперсия:

$$\sigma^2 = \sum_{j=1}^{k} \sigma_j^2. \tag{4.77}$$

Вместо дисперсии ошибок с успехом можно использовать сумму квадратов отклонений модели. Тогда формула (4.76) примет вид

$$\nu_j = \frac{1}{k-1} \frac{sRSS - RSS_j}{sRSS},$$

где $sRSS = \sum_{j=1}^{k} RSS_j$.

Пример

Допустим, нам необходимо получить на рассматриваемых данных многофакторную прогнозную линейную модель синтезом однофакторных с учетом дисперсии ошибок каждой модели.

Вычислим дисперсии для каждой из построенных с помощью МНК однофакторных моделей. Для первой модели сумма квадратов отклонений равна 0,00331, для второй модели $RSS_2 = 0,0003$, для третьей – $RSS_3 = 0,0022$, для четвертой – $RSS_4 = 0,0009$.

Соответственно суммарная сумма квадратов: $RSS = 0,0665$.

Теперь легко определяются веса каждой модели:

для первой: $\nu_1 = \frac{1}{3}\frac{0,0665 - 0,0033}{0,0665} = 0,1675$;

для второй: $\nu_2 = 0,3182$;

для третьей: $\nu_2 = 0,2213$;

для четвертой: $\nu_2 = 0,2931$.

Сумма этих весовых коэффициентов равна единице, и, как легко заметить, самый большой вес имеет вторая модель с минимальной дисперсией ошибок. А первая однофакторная модель с максимальной дисперсией ошибок, имеет минимальный вес.

Теперь, зная эти веса, легко построить многофакторную линейную модель:

$$\hat{y}_t = -0,4430 + 0,0768x_{1,t} + 0,1406x_{2,t} + 0,9708x_{3,t} + 0,2675x_{4,t}. \quad (4.78)$$

Модель, как и ожидалось, отличается от той, которая была построена с помощью синтеза однофакторных моделей в многофакторную с помощью одинаковых весов.

Помимо этих двух основных методов построения многофакторных моделей их синтезом из однофакторных, существуют и другие разновидности, отличающиеся способами задания весов однофакторных моделей[1].

К числу очевидных преимуществ метода построения многофакторных моделей их синтезом из однофакторных следует отнести и возможность учета экспертной оценки степени влияния каждого из факторов на результат. Зачастую точность аппроксимации результирующего показателя с помощью какой-либо однофакторной модели фактора j вовсе не означает, что этот фактор является важнейшим для объяснения динамики показателя. Это говорит о том, что прогнозисту удалось подобрать хорошую аппроксимирующую модель и только. Экономический анализ сути происходящих процессов позволяет экономисту утверждать, что влияние одних факторов весомее, чем других и поэтому формировать формальную многофакторную модель с помощью весов, вычисляемых в зависимости от дисперсии (4.76),

---

[1] Рабочая книга по прогнозированию / отв. ред. И. В. Бестужев-Лада. М. : Мысль, 1982.

будет неверно. В этом случае веса каждого фактора задаются экспертным путем с учетом условия (4.74) и однофакторные модели синтезируются в многофакторную с помощью этих весов.

Прогнозист самостоятельно должен выбрать метод построения многофакторной модели, либо используя многофакторный МНК, либо построив однофакторные модели и синтезировав их в многофакторную.

4.6. Учет качественных характеристик при построении регрессий

В ряде случаев для получения более точных прогнозов в регрессионных моделях требуется учесть характеристики, измеренные в номинальной или порядковой шкалах. Например, при прогнозировании спроса на смартфоны, можно учесть такие их уникальные характеристики, как установленная операционная система, тип корпуса, материал, из которого изготовлен телефон и т.д. Все эти характеристики невозможно измерить в интервальной или метрической шкалах. Можно лишь обозначить, что у данного телефона корпус металлический, есть поддержка двух SIM-карт и т.д. Все эти характеристики можно зафиксировать в виде категориальной переменной, которая будет обозначать, к какой категории можно отнести данный продукт. Например, материал корпуса телефона может быть:

1) стальной;
2) золотой;
3) пластиковый;
4) картонный;
5) бумажный.

В случае с ранговыми переменными есть некоторое естественное упорядочение характеристик. Например, средний месячный доход респондента может быть:

1) незначительным;
2) маленьким;
3) средним;
4) большим;
5) огромным.

В этой шкале нельзя дать никакую оценку отдельным категориям, однако они естественным образом упорядочены (от меньшего дохода к большему), поэтому есть соблазн включить ранговую переменную в регрессию как есть. Смысл переменной достаточно прост: чем выше ее значение, тем выше доход респондента. Однако если ее включить в регрессию в таком виде, мы сделаем грубое допущение о том, что при переходе от одной категории к другой зависимая переменная (например, продажи телефонов) изменяется одинаково. Это допущение выполнимо лишь в отдельных случаях. Действительно, продажи дешевых телефонов в группах потребителей с «незначительным», «маленьким» и «средним» среднемесячным доходом, скорее всего, будут различаться не на одну и ту же фиксированную величину. Возможно, в данном примере первая группа вообще не будет покупать телефоны, вторая будет покупать дешевые телефоны (являясь целевой аудиторией данного товара), а третья, — скорее всего, будет переключаться на более дорогие телефоны.

Итак, если в распоряжении исследователя есть данные, измеренные в номинальной либо порядковой (ранговой) шкалах, он может поступить одним из двух способов:

1) построить ряд регрессий по каждой категории;

2) построить одну регрессию, в которой учитывалось бы наличие нескольких категорий.

Второй способ зачастую бывает более целесообразным и легко осуществимым, так как, например, можно ожидать, что продажи телефонов, отличающихся только одной деталью, будут схожими, т.е. ряд факторов, включенных в регрессию, будет одинаково влиять на величину продаж. Однако при этом очевидно, что какие-то различия будут.

Чтобы учесть различные категории на основе одной имеющейся категориальной переменной, состоящей из m категорий, нужно создать m так называемых «фиктивных переменных» (по-английски это называется «dummy») — переменных, которые принимают значение «1», если данное наблюдение обладает указанной характеристикой, и «0» — если не обладает.

Для нашего примера с корпусами будет создано пять фиктивных переменных, которые мы условно назовем по-английски (так, чтобы потом было легко идентифицировать переменную):

1. *Steel* — принимает значение «1», если мы рассматриваем телефон со стальным корпусом и «0» в обратном случае.

2. *Gold* — «1», если корпус из золота и «0» в противоположном случае.

3. *Plastic* — «1», если пластик и «0», если нет.

4. *Carton* — «1», если сделан из картона и «0», если не из картона

5. *Paper* — «1», если корпус выполнен в бумаге и «0», если нет.

Для любого телефона, рассматриваемого нами, какая-то одна из фиктивных переменных будет равна «1», а все остальные — «0».

Покажем теперь, как можно включить созданные нами переменные в модель регрессии. Предположим, что мы выяснили, что продажи мобильных телефонов, скорее всего, зависят от их характеристик, цены и доходов потребителей, что может быть записано в виде модели регрессии:

$$\hat{y}_t = a_0 + a_1 x_{1,t} + a_2 x_{2,t}, \tag{4.64}$$

где y_t — объем продаж в штуках; $x_{1,t}$ — стоимость телефона в рублях; $x_{2,t}$ — средний месячный доход потребителей в рублях.

Если мы считаем, что разные типы продукции будут продаваться по-разному, вне зависимости от каких бы то ни было факторов (например, картонные телефоны будут в среднем продаваться хуже, чем стальные), то мы можем просто напрямую включить наши фиктивные переменные в модель и найти соответствующие коэффициенты МНК:

$$\hat{y}_t = a_0 + a_1 x_{1,t} + a_2 x_{2,t} + b_0 steel + c_0 gold + d_0 plastic + e_0 carton. \tag{4.65}$$

Такое представление позволяет легко определить, какими будут продажи отдельных категорий продукции. Например, продажи золотых мобильных телефонов будут:

$$\hat{y}_t = a_0 + a_1 x_{1,t} + a_2 x_{2,t} + c_0, \tag{4.66}$$

так как в данном случае $gold = 1$, в то время как все остальные фиктивные переменные равны нулю.

Стоит, однако, заметить, что в модель мы включили лишь четыре фиктивные переменные из пяти. Вызвано это чисто техническими соображениями. Так, если включить все пять переменных, то коэффициенты такой модели рассчитать не удастся из-за идеальной мультиколлинеарности. Возникает она из-затого, что в этом случае для любого наблюдения сумма фиктивных переменных будет равна 1. Одновременно с этим константу в модели можно представить как коэффициент, умноженный на 1. В результате этого в модели появляется функциональная зависимость между значениями перед константой и суммой фиктивных переменных. Эта проблема называется «ловушкой фиктивных переменных». Чтобы ее избежать, можно либо включить в модель $m - 1$ фиктивную переменную, либо включить все, но убрать константу. Если воспользоваться первым вариантом, то отброшенная фиктивная переменная будет выступать «эталоном». Действительно, в уравнении (4.65) была отброшена фиктивная переменная «*paper*», в результате чего для мобильных телефонов, сделанных из бумаги, мы будем получать простую регрессию вида (4.64) (т.е. в нашем случае для оценки продаж мы выбрали бумажные телефоны в виде «эталона»). В модели (4.66) при этом коэффициент b_2 будет показывать среднюю разницу в продажах золотых и бумажных мобильных телефонов.

Чтобы лучше понять смысл фиктивных переменных, рассмотрим следующий графический пример (рис. 4.14).

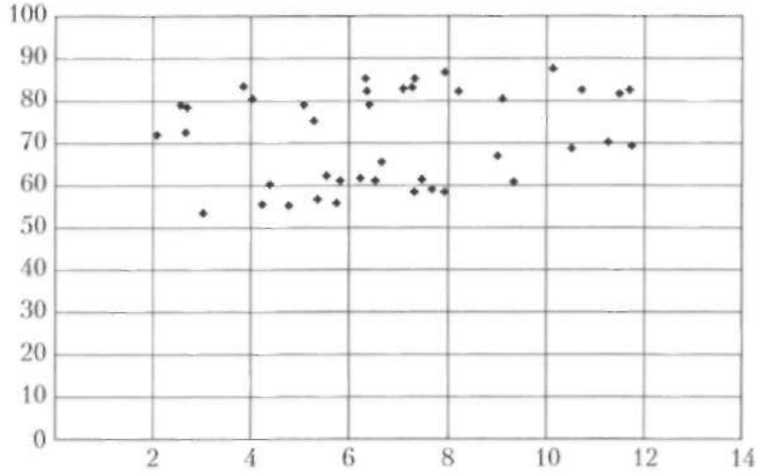

Рис. 4.14. **Точечная диаграмма по условному ряду данных зависимости y от x**

Данную точечную диаграмму можно воспринять по-разному, в зависимости от информированности исследователя относительно объекта исследования и данных, с которыми он работает:

1. Если не учитывать возможность разных категорий в объекте исследования (либо не учитывать, что с течением

времени зависимость между y и x могла измениться), то трактовка диаграммы на рис. 4.14 сводится к тому, что между y и x имеется средняя связь с некоторой постоянной дисперсией. Эта мысль может быть представлена графически, если провести прямую линию через все облако точек. Получим график как на рис. 4.15, в котором не делается никакого деления на категории.

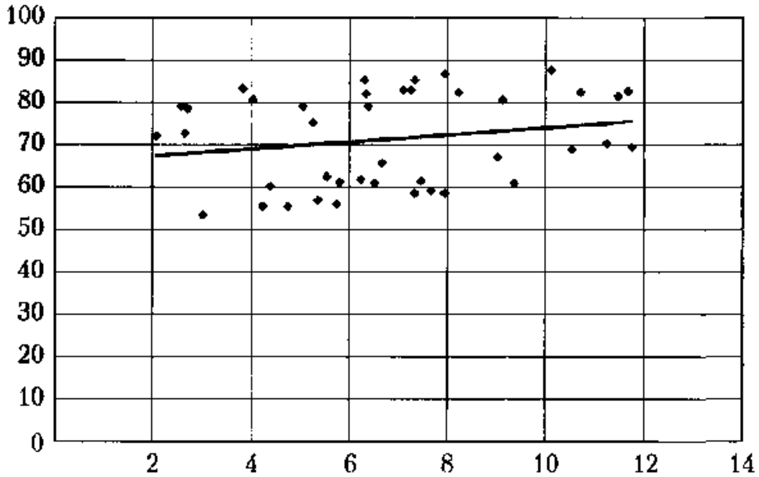

Рис. 4.15. **Точечная диаграмма по условному ряду данных зависимости y от x и регрессионная линия, построенная по всему ряду**

2. Если же, изучив данные, мы придем к тому, что имеем дело с разными категориями, то можно обратить внимание на то, что на рис. 4.14 наблюдается два облака точек: одно чуть выше другого. Линия, построенная на рисунке 4.15, прошла прямо посередине между этими облаками. Воспринимая ситуацию таким образом, можно сказать, что между y и x имеется достаточно сильная зависимость, которая меняется в категориях (либо изменилась каким-то образом во времени). Этому случаю соответствует график на рис. 4.16, на котором явно наблюдается разница в константах для двух ситуаций на 20 единиц (точки пересечения прямых с осью ординат примерно соответствуют значениям 70 и 50).

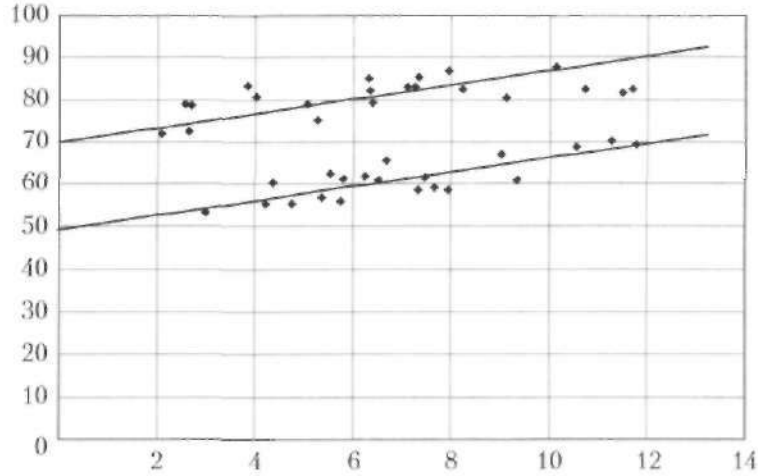

Рис. 4.16. **Точечная диаграмма по условному ряду данных зависимости y от x и регрессионные линии, построенные по разным категориям**

Вторая ситуация в нашем примере более правдоподобна и, как видим, такой подход позволяет более точно описать зависимость между y и x. Введение фиктивных переменных таким образом, как это было сделано в (4.65), позволяет достичь эффекта, показанного на рис. 4.16 за счет изменения значения константы в зависимости от категории.

Однако возможна и другая ситуация, требующая включения фиктивных переменных в модель (рис. 4.17).

На этом рисунке на себя обращает внимание то, что два облака, сформированных из-за наличия различных категорий объекта, отличаются друг от друга, скорее, не уровнем ряда, а углом наклона. Это говорит о том, что в разных категориях влияние x на y различно: в одном случае рост x на 1 приводит к более значительному изменению y, нежели во втором. Чтобы смоделировать такую зависимость, фиктивные переменные нужно добавить уже не в константу, а в угол наклона. Делается это путем перемножения соответствующих фиктивных переменных на значение x.

Обратимся к нашему примеру с мобильными телефонами. Для того чтобы учесть, что изменение цены на телефоны с различными видами корпусов по-разному отразится

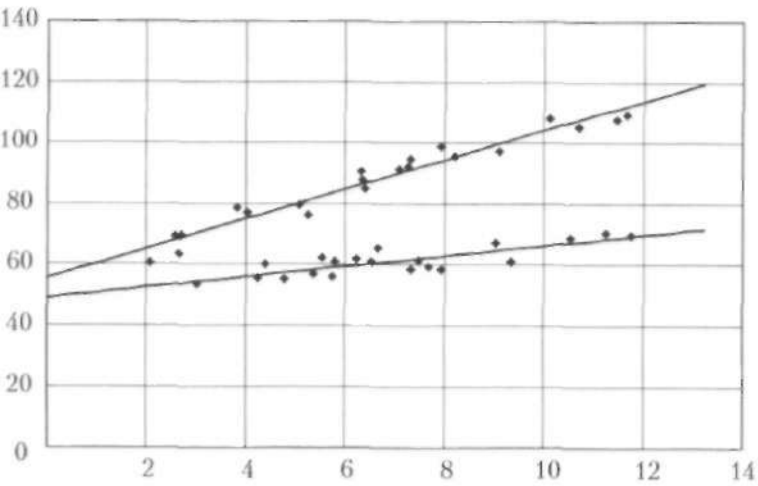

Рис. 4.17. **Точечная диаграмма по условному ряду данных зависимости y от x и регрессионные линии, построенные по разным категориям**

на их продажах, включим фиктивные переменные следующим образом:

$$\hat{y}_t = a_0 + a_1 x_{1,t} + a_2 x_{2,t} + b_1 steel \cdot x_{1,t} + c_1 gold \cdot x_{1,t} + \\ + d_1 plastic \cdot x_{1,t} + e_1 carton \cdot x_{1,t}. \quad (4.67)$$

В полученной регрессии (4.67) так же, как и в регрессии (4.65) мы включили только четыре из пяти фиктивных переменных для того, чтобы избежать «ловушки фиктивных переменных». Эталоном здесь так же, как и в (4.65) выступает телефон с бумажным корпусом — для которого мы все так же будем приходить к регрессии (4.64). Если же, например, рассмотреть, каковы будут продажи золотых телефонов, то нужно переписать уравнение (4.67) с учетом того, что для таких телефонов $gold = 1$, а остальные фиктивные переменные равны нулю:

$$\hat{y}_t = a_0 + a_1 x_{1,t} + a_2 x_{2,t} + c_1 x_{1,t} = a_0 + x_{1,t}(a_1 + c_1) + a_2 x_{2,t}. \quad (4.68)$$

Как видим, для таких телефонов угол наклона составляет уже не a_1, а $(a_1 + c_1)$, т.е. таким введением фиктивных переменных мы добились желаемого эффекта.

Оценка коэффициентов регрессии (4.67) осуществляется простым МНК. Фактически в данном случае мы создаем ряд переменных «$steel \cdot x_{1,t}$», «$gold \cdot x_{1,t}$» и т.п., которые включаются в финальную множественную регрессию.

При желании фиктивные переменные можно ввести, как в константу, так и в углы наклона для каждого из факторов. Тогда в общем случае модель с фиктивными переменными для нашего примера будет выглядеть следующим образом:

$$\hat{y}_t = a_0 + a_1 x_{1,t} + a_2 x_{2,t} + steel\left(b_0 + b_1 x_{1,t} + b_2 x_{2,t}\right) + \\ + gold\left(c_0 + c_1 x_{1,t} + c_2 x_{2,t}\right) + plastic\left(d_0 + d_1 x_{1,t} + d_2 x_{2,t}\right) + \\ + carton\left(e_0 + e_1 x_{1,t} + e_2 x_{2,t}\right). \tag{4.69}$$

В (4.69) мы имеем дело с несколькими видами товара, которые продаются совершенно разным образом. В данном случае возникает вопрос целесообразности рассмотрения общей регрессии с фиктивными переменными. Действительно, если мы приходим к тому, что для каждого типа продукции фактически получается отдельная регрессия, зачем же строить нечто общее и заниматься введением фиктивных переменных и оценкой коэффициентов такой большой модели. Единственное преимущество в таком подходе в данном случае заключается в чисто формальных соображениях, — в том, что модель (4.69) будет иметь большее число степеней свободы.

Например, если в нашем распоряжении имеется выборка из 50 наблюдений, в которых данные собраны по пяти категориям продукции (по 10 наблюдений на каждую категорию), то в случае построения отдельных регрессий вида (4.64) мы будем получать $10 - 3 = 7$ степеней свободы для каждой из них. В случае же построения регрессии вида (4.69) мы будем получать $50 - 3 \cdot 5 = 35$ степеней свободы. В результате этого модель (4.69), скорее всего, будет статистически значимой, а аналогичная ей модель (4.64), построенная по отдельному типу продукции, может оказаться незначимой при той же описательной способности. Точечные прогнозы по продажам при этом будут абсолютно идентичными. Единственное существенное различие в данном случае будет заключаться в дисперсии остатков. Действительно, в первой модели дисперсия будет рассчитываться лишь по остаткам по продажам одного типа продукции, в то время как во второй модели к этим остаткам добавятся остатки по всем остальным категориям товара. Однако однозначно заключить, что лучше, нельзя. Получив усредненную дисперсию остатков для всех наблюдений (не учитывая индивидуальные особенности остатков по категориям), мы исказим дисперсию продаж каждого конкретного типа продукции, что приведет к некорректному прогнозному интервалу. Если же мы рассчитаем дисперсию остатков для продаж отдельного типа продукции, то можем получить недостоверное значение из-за малого числа наблюдений, что также может привести к слишком широкому или же слишком узкому прогнозному интервалу.

Однозначного выбора в пользу простой регрессии по отдельной категории или регрессии с фиктивными переменными нет. Исследователю в каждом конкретном случае нужно основывать свой выбор не на формальных статистических критериях, а исходя из основных принципов прогнозирования (принципы системности, природной специфичности и оптимальности затрат на прогнозирование).

Практикум

Вопросы для самоконтроля

1. Какие типовые прогнозные модели вы знаете? Перечислите их, исходя из общенаучного принципа «от простого — к сложному».

2. Что собой представляют «уравнения в отрезках»?

3. Обычно задачу оценки коэффициентов регрессионной модели рассматривают на плоскости переменных. Какую дополнительную информацию можно получить, если рассматривать задачу на плоскости искомых коэффициентов?

4. Как с помощью уравнений в отрезках оценить устойчивость оценок коэффициентов регрессионной модели?

5. Что нужно сделать для получения более устойчивых оценок коэффициентов прогнозных моделей?

6. Почему от многофакторных прогнозных моделей следует ожидать более точных прогнозов, чем от моделей однофакторных, а тем более — трендов?

7. Как построить многофакторную модель? Какую роль при этом играет корреляционная матрица?

8. Что такое «мультиколлинеарность»?

9. Что представляет собой матрица точечных диаграмм?

10. Что характеризует коэффициент множественной корреляции?

11. Как использовать t-тест при построении многофакторных прогнозных моделей?

12. Чем могут помочь при многофакторном моделировании распределения Фишера и Пирсона?

13. В чем опасность игнорирования проблемы «пропущенных переменных»?

14. Что собой представляет метод инструментальных переменных?

15. Что такое «гомоскедастичность» и «гетероскедастичность»?

16. Как можно определить понятия «ложная» и «истинная» гетероскедастичность?

17. Обычно предполагается, что остатки регрессионной модели некоррелируемые. О чем свидетельствует наличие автокорреляции остатков? Что нужно делать в этом случае?

18. Почему построение многофакторной прогнозной модели в условиях мультиколлинеарности зачастую приводит к получению модели, непригодной для прогнозирования?

19. Как можно уменьшить влияние мультиколлинеарности на неустойчивость регрессионных оценок?

20. Что представляет собой метод синтеза однофакторных моделей в многофакторную? Каковы его преимущества и недостатки?

21. Социально-экономические явления и процессы отражаются множеством показателей, часть из которых измеряется в неметрических шкалах. Действия с переменными, измеренными в разных шкалах, требуют особой осторожности. Как построить прогнозную модель в случае, если факторы этой модели измеряются в разных шкалах?

Задания

Задание 1. Студент продолжает проводить исследование по стоимости обедов. Он не любит гречневую кашу, поэтому, когда в столовой не остается ничего другого в качестве гарнира, он не берет гарнир вообще. В этот раз он решил зафиксировать такие ситуации в качестве фиктивных переменных. Кроме того, он подумал, что на затраты на обед может влиять температура на улице, поэтому зафиксировал и эти данные. Все полученные данные он свел в следующую таблицу:

Затраты на обед в столовой, руб.	150	120	125	130	150	145	135	120	100	150
Количество пар в день	5	2	2	2	4	3	3	2	1	4
Есть только гречневая каша	0	1	1	0	0	0	0	0	1	0
Температура на улице, °C	15	13	14	10	14	13	12	9	11	15

1. Оцените наличие возможной мультиколлинеарности в данных путем построения матрицы корреляций и расчета коэффици-

ентов множественной корреляции. Есть ли мультиколлинеарность в данных? Чем она может быть вызвана? Нужно ли и можно ли ее устранить?

2. Постройте матрицу точечных диаграмм. Попытайтесь определить, каким может быть влияние предложенных переменных на затраты на обед.

3. На основе предыдущих пунктов попробуйте оценить, с какими проблемами вы могли бы столкнуться при построении линейной регрессии. Как можно было бы их решить?

4. Постройте простую линейную модель множественной регрессии, объясняющую влияние указанных переменных на затраты на обед.

5. Рассчитайте значимость коэффициентов полученной модели. Действительно ли наличие гречневой каши существенно влияет на затраты на обед студента?

6. Используя принцип синтеза однофакторных моделей в многофакторную, постройте модель множественной регрессии по своим данным.

Задание 2. Работник магазина, специализирующегося на продаже настольных ламп, решил спрогнозировать продажи ламп на основе их характеристик. Он счел, что цвет и форма лампы могут влиять на ее продаваемость. Чтобы проверить эту гипотезу и спрогнозировать объем продаж новой партии красных ламп на длинных ножках, он отнес все лампы к нескольким категориям и собрал данные за несколько дней (см. таблицу ниже).

Форма лампы	Цвет лампы	Стоимость лампы	Количество проданных ламп
Приплюснутая	Красная	470	456
На длинной ножке	Красная	620	548
Приплюснутая	Черная	470	206
Вытянутая	Красная	350	361
Вытянутая	Черная	330	200
Приплюснутая	Красная	470	423
На длинной ножке	Черная	600	295
Приплюснутая	Красная	470	433
Вытянутая	Черная	330	232
Приплюснутая	Черная	470	257
На длинной ножке	Черная	600	290
Вытянутая	Красная	350	369

Окончание таблицы

Форма лампы	Цвет лампы	Стоимость лампы	Количество проданных ламп
Приплюснутая	Черная	470	226
На длинной ножке	Красная	620	522
На длинной ножке	Красная	620	511
Приплюснутая	Черная	470	197
Вытянутая	Черная	330	220
Вытянутая	Черная	330	250
Вытянутая	Красная	350	310
Приплюснутая	Черная	470	217
На длинной ножке	Красная	620	542
Вытянутая	Красная	350	351
Приплюснутая	Черная	470	262
Приплюснутая	Красная	470	438

1. На основе данных по форме и цвету ламп создайте группы фиктивных переменных. Проанализируйте, какие лампы продаются чаще. Может ли существовать какая-то связь между цветом и формой лампы? Если да, то обоснуйте эту связь.

2. Постройте простую модель парной регрессии зависимости количества проданных ламп от их стоимости. Нанесите фактические и расчетные значения по количеству продаж на точечный график «Стоимость — количество». Проанализируйте полученный график.

3. Постройте модель множественной регрессии с фиктивными переменными, отвечающими за цвет ламп. Постройте точечный график по аналогии с графиком в п. 2. Проанализируйте полученный результат: лучше ли модель аппроксимирует ряд по сравнению с моделью в п. 2? Почему?

4. Постройте две модели (для черных и красных ламп). Сравните эти модели с моделью в п. 3. Какими преимуществами и недостатками, по вашему мнению, они обладают по сравнению с моделью в п. 3?

5. Выполните п. 3, используя фиктивные переменные для формы ламп.

6. Постройте три модели для разных форм ламп. Сравните эти модели с моделью в п. 5.

7. Постройте модель на основе всех фиктивных переменных. Сравните ее точность с точностью предыдущих моделей.

8. Какой из всех построенных моделей вы бы отдали предпочтение и почему?

Наши книги можно приобрести:

Учебным заведениям и библиотекам:
в отделе по работе с вузами
тел.: (495) 744-00-12, e-mail: vuz@urait.ru

Частным лицам:
список магазинов смотрите на сайте urait.ru
в разделе «Частным лицам»

Магазинам и корпоративным клиентам:
в отделе продаж
тел.: (495) 744-00-12, e-mail: sales@urait.ru

Отзывы об издании присылайте в редакцию
e-mail: gred@urait.ru

Новые издания и дополнительные материалы доступны
в электронной библиотеке biblio-online.ru,
а также в мобильном приложении «Юрайт.Библиотека»

Учебное издание

Светуньков Иван Сергеевич,
Светуньков Сергей Геннадьевич

МЕТОДЫ СОЦИАЛЬНО-ЭКОНОМИЧЕСКОГО ПРОГНОЗИРОВАНИЯ

Том 1. Теория и методология

Учебник и практикум для академического бакалавриата

Формат 60×90 1/16.
Гарнитура «Petersburg». Печать цифровая.
Усл. печ. л. 18,43.

ООО «Издательство Юрайт»
111123, г. Москва, ул. Плеханова, д. 4а.
Тел.: (495) 744-00-12. E-mail: izdat@urait.ru, www.urait.ru