Прогнозирование и временные ряды

К. О. Кизбиленов

ПРОГНОЗИРОВАНИЕ И ВРЕМЕННЫЕ РЯДЫ

© Алтайский государственный педагогический университет, 2017
К. О. Кизбикенов
Прогнозирование и временные ряды
Учебное пособие

Барнаул
ФГБОУ ВО «АлтГПУ»
2017
УДК 519.2(075)
ББК 22.17я73+65в631я73
К385

Кизбиенков, К. О.
Прогнозирование и временные ряды [Электронный ресурс] : учебное пособие / К. О. Кизбиенков. – Барнаул : АлтГПУ, 2017. – Систем. требования : Процессор с тактовой частотой 1,5 ГГц и выше; 512 Mb RAM ; Windows XP/Vista/7/8/10 ; Adobe Acrobat Reader ; SVGA монитор с разрешением 1024x768 ; мышь.


Рецензенты:
Родионов Е.Д., доктор физико-математических наук, профессор (Алтайский государственный университет);
Махаева Т.П., доцент, кандидат педагогических наук (Алтайский государственный педагогический университет);
Пономарев И.В., кандидат физико-математических наук, доцент (Алтайский государственный университет)

Учет временной структуры данных о реальных экономических процессах позволяет адекватно отразить их в математических и эконометрических моделях. Осознание этого факта привело как к ревизии многих макроэкономических теорий и построений, так и к бурному развитию специфических методов анализа таких данных. Знание этих методов и способов применения их к анализу конкретных экономических процессов является в настоящее время необходимой составляющей подготовки математиков-исследователей (аналитиков) на магистерском уровне.
Пособие рекомендовано для студентов четвертого курса и магистров, обучающихся по направлению «Прикладная математика», написано в соответствии с программой одноименного курса.

Рекомендовано к изданию редакционно-издательским советом АлтГПУ 23.03.2017 г.

Текстовое (символьное) электронное издание.

Системные требования:
Процессор с тактовой частотой 1,5 ГГц и выше; 512 Mb RAM ; Windows XP/Vista/7/8/10 ; Adobe Acrobat Reader ; SVGA монитор с разрешением 1024x768 ; мышь.

© Алтайский государственный педагогический университет, 2017
Объём издания - 2 800 КБ.
Дата подписания к использованию: 31.05.2017

Федеральное государственное бюджетное образовательное учреждение высшего образования
«Алтайский государственный педагогический университет» (ФГБОУ ВО «АлтГПУ»)
ул. Молодежная, 55, г. Барнаул, 656031
Тел. (385-2) 36-82-71, факс (385-2) 24-18-72
e-mail: rector@altspu.ru, http://www.altspu.ru
## Содержание

**Предисловие** ........................................................................................................................................... 4

1. **Значение и содержание социально-экономического прогнозирования**
   1.1. Понятие прогноза .......................................................................................................................... 5
   1.2. Инструментарий прогнозирования .......................................................................................... 6

2. **Временные ряды и прогнозирование**
   2.1. Общие сведения о временных рядах и задачах их анализа ................................................. 9
   2.2. Стационарные временные ряды и их характеристики. Автокорреляционная функция ..... 10
   2.3. Эргодичность ........................................................................................................................... 13
   2.4. Тренды ........................................................................................................................................ 14
   2.5. Случайные величины, необходимые для тестирования уравнений регрессии ............... 17
   2.6. Сезонные колебания ................................................................................................................ 21
   2.7. Сезонные фиктивные переменные ............................................................................................ 26
   2.8. Сглаживание временного ряда. Метод скользящих средних ............................................. 30
   2.9. Экспоненциальное сглаживание .............................................................................................. 33
   2.10. Критерии, используемые в анализе временных рядов ......................................................... 35
   2.11. Критерии, основанные на автокорреляционной функции ................................................ 36
   2.12. Лаговый оператор .................................................................................................................. 40
   2.13. Модели с распределенным лагом. Лаги Алмон ..................................................................... 42
   2.14. Пример применения метода Алмон ....................................................................................... 46

3. **Модели ARIMA**
   3.1 Разложение Вольда .................................................................................................................... 50
   3.2 Процессы автокоррекции (AR) ............................................................................................... 51
   3.3. Процесс Маркова ..................................................................................................................... 52
   3.4. Процессы скользящего среднего MA(q) .................................................................................. 58
   3.5. Смешанные процессы авторегрессии и скользящего среднего ARMA ............................. 61
   3.6. Модель авторегрессии проинтегрированного скользящего среднего ARIMA ................ 66
   3.7. Проверка на стационарность. Интеграционная статистика Дарбина-Уотсона ................ 68
   3.8. Прогнозирование с помощью ARMA моделей .................................................................... 70
   3.9. ARCH и GARCH ....................................................................................................................... 73

4. **Нестационарные временные ряды**
   4.1. Ряды TS и DS .......................................................................................................................... 77
   4.2. Тест Дикки-Фуллера на единичный корень .......................................................................... 80
   4.3. ADF — тест Дикки-Фуллера .................................................................................................... 83
   4.4. Авторегрессионные модели с распределенными лагами (ADL-модели) .......................... 87
   4.5. Причинность по Грэнджеру ................................................................................................... 89
   4.6. Многомерные процессы .......................................................................................................... 91
   4.7. Коинтеграция ........................................................................................................................... 95
   4.8. Коинтеграционная регрессия и тест Йохансена ................................................................. 101
   4.9. Пример применения коинтеграции ....................................................................................... 104

5. **Задачи** .............................................................................................................................................. 106

**Предметный указатель** ......................................................................................................................... 112

**Библиографический список** ............................................................................................................... 113
Предисловие

Прогнозирование социально-экономического развития является важнейшим разделом экономической науки, призванным обеспечить государственные органы власти, общественность и субъекты экономической деятельности информацией о развитии экономики, а также связанных с ней социальных процессов. В этой связи изучение методов социально-экономического прогнозирования является одной из важных задач в системе экономического образования. Основу всей совокупности названных методов традиционно составляют статистические методы, применяемые для прогнозирования развития социальных и экономических явлений и процессов, построения адекватных моделей временных рядов и выбора наиболее приемлемых вариантов из всех возможных способов прогнозирования. Важное место в системе методов прогнозирования отводится также экспертным методам.

Дипломированный специалист по прикладной математике должен уметь осуществлять прогнозирование и многовариантные аналитические расчеты в области экономической и управленческой деятельности. Поэтому одним из требований при подготовке специалистов в высшей школе является выработка у обучающихся профессиональных навыков в сборе, обработке статистических данных и построении многовариантных прогнозов динамики социально-экономических явлений и процессов.
1. Значение и содержание социально-экономического прогнозирования

1.1. Понятие прогноза

В Российской энциклопедическом словаре приводится следующее определение: «Прогноз (от греческого prognosis – предвидение, предсказание) – конкретное предсказание, суждение о состоянии какого-либо явления в будущем».

Типология прогнозов может строиться по различным критериям в зависимости от целей, задач, объектов, предметов, проблем, характера, периода упреждения, методов, организации прогнозирования и т. д.

1. По проблемно-целевому критерию (для чего разрабатывается прогноз?) различают два типа прогнозов:

поисковый прогноз (исследовательский, изыскательский, трендовый, генетический и т. п.) – определение возможных состояний явления в будущем. Имеется в виду условное продолжение в будущее тенденций развития изучаемого явления в прошлом и настоящем, абстрагируясь от возможных решений, действий на основе которых способны радикально изменить тенденции, вызвать в ряде случаев самоосуществление или саморазрушение прогноза. Такой прогноз отвечает на вопрос: что вероятнее всего произойдет при условии сохранения существующих тенденций?

нормативный прогноз (программный, целевой) – определение путей и сроков достижения возможных состояний явления, принимаемых в качестве цели. Имеется в виду прогнозирование достижения желательных состояний на основе заранее заданных норм, идеалов, стимулов, целей. Такой прогноз отвечает на вопрос: какими путями достичь желаемого?

Поисковый прогноз строится с использованием определенной шкалы (поля, спектра) возможностей, на которой затем устанавливается степень вероятности прогнозируемого явления. При нормативном прогнозировании происходит такое же распределение вероятностей, но уже в обратном порядке: от заданного состояния к наблюдаемым тенденциям.

2. По периоду упреждения – промежутку времени, на который рассчитан прогноз, – различаются:

оперативный, как правило, рассчитан на перспективу, на протяжении которой не ожидается существенных изменений объекта исследования – ни количественных, ни качественных;

краткосрочный – рассчитан на перспективу, на протяжении которой не ожидается существенных количественных изменений объекта исследования;

среднесрочный – охватывает перспективу между кратко- и долгосрочным прогнозами с преобладанием количественных изменений над качественными;

длительный – рассчитан на перспективу, на протяжении которой ожидаются существенные не только количественные, но и существенные качественные изменения объекта исследования;

дальнее срочные (сверхдлинные) – охватывает перспективу, в течение которой ожидаются столь значительные качественные изменения, что можно говорить лишь о самых общих перспективах развития исследуемого явления или процесса.
Временная градация прогнозов является в определенной мере условной и зависит от характера и цели данного прогноза. В социально-экономических прогнозах эмпирически установлен следующий временной масштаб: оперативные прогнозы имеют продолжительность до одного месяца, краткосрочные – до одного года, среднесрочные – рассчитаны на несколько (до пяти) лет, долгосрочные – на период свыше пяти и, примерно, до пятнадцати-двадцати лет, дальнесрочные – за пределами долгосрочных.

3. По объекту исследования различаются естествоведческие, научно-технические и обществоведческие (социальные в широком значении этого термина) прогнозы.

В естествоведческих прогнозах взаимосвязь между предсказанием и предуказанием незначительна, близка или практически равна нулю из-за невозможности управления объектом. Поэтому здесь в принципе применимо только поисковое прогнозирование с ориентацией на возможно более точное безусловное предсказание будущего состояния явления. Яркий пример – метеорологический прогноз.

В обществоведческих прогнозах эта взаимосвязь настолько значительна, что способна давать эффект самосуществования или, напротив, саморазрушения прогнозов действиями людей на основе целей, планов, программ, проектов, вообще решений (включая принятые с учетом сделанных прогнозов). В связи с этим здесь необходимо сочетание поисковых и нормативных разработок, т.е. условных предсказаний с ориентацией на повыше ние эффективности управления. Примером может являться прогноз результатов выборов для той или иной политической партии.

Научно-технические прогнозы занимают промежуточное место между естествоведческими и обществоведческими прогнозами. Они, как правило, основываются на имеющемся опыте (нормативные разработки) и могут быть поисковыми. Примером такого прогноза являются тенденции в технологиях производства кристаллов для компьютеров, отслеживаемые и прогнозируемые заинтересованными фирмами.

1.2. Инструментарий прогнозирования

В основе прогнозирования лежат три взаимодополняющих источника информации о будущем:

- оценка перспектив развития, будущего состояния прогнозируемого явления на основе опыта, чаще всего при помощи аналогий с достаточно хорошо известными сходными явлениями и процессами;

- условное продолжение в будущее (экстраполяция) тенденций, закономерности развития которых в прошлом и настоящем обладают высокой степенью инертности;

- модель будущего состояния того или иного явления, процесса, построенная сообразно ожидаемым или желательным изменениям ряда условий, перспективы развития которых достаточно хорошо известны.

В соответствии с этим существуют три дополняющих друг друга способа разработки прогнозов:

анкетирование (интервью, опрос) – метод изучения мнений населения, специалистов (экспертов) с целью упорядочить, сделать объективными субъективные оценки прогнозного характера. Особенно большое значение имеют экспертные оценки. Опросы населения в практике прогнозирования применяются сравнительно редко;
экстраполирование и интерполирование — построение динамических рядов развития показателей прогнозируемого явления на протяжении периодов основания прогноза в прошлом и упреждения прогноза в будущем (ретроспекции и проспекции прогнозных разработок);

моделирование – построение поисковых и нормативных моделей с учетом вероятного или желательного изменения прогнозируемого явления на период упреждения прогноза по имеющимся прямым или косвенным данным о масштабах и направлениях изменений. Наиболее эффективная прогнозная модель – система уравнений. Однако имеют значение все возможные виды моделей в широком смысле этого термина: сценарии, имитации, графы, матрицы, подборки показателей, графические изображения и т. д.

Приведенное разделение способов прогнозирования условно, потому что на практике эти способы взаимно перекрещиваются и дополняют друг друга. Прогнозная оценка обязательно включает в себя элементы экстраполяции и моделирования. Процесс экстраполяции невозможен без элементов оценки и моделирования. Моделирование подразумевает предварительную оценку и экстраполирование.

Общая логическая последовательность важнейших операций разработки прогноза сводится к следующим основным этапам:

Предпрогнозная ориентация (программа исследования). Уточнение задания на прогноз: характер, масштабы, объект, периоды основания и упреждения и т. д. Формулирование целей и задач, предмета, проблемы и рабочих гипотез, определение методов, структуры и организации исследования.

Построение исходной (базовой) модели прогнозируемого объекта методами системного анализа. Для уточнения модели возможен опрос населения и экспертов.

Сбор данных методами, о которых говорилось выше.

Построение динамических рядов показателей – основы, стержня будущих прогнозных моделей методами экстраполяции; возможно обобщение этого материала в виде прогнозных предмодельных сценариев.

Построение серии гипотетических (предварительных) поисковых моделей прогнозируемого объекта методами поискового анализа профильных и фоновых показателей с конкретизацией минимального, максимального и наиболее вероятного значений.

Построение серии гипотетических нормативных моделей прогнозируемого объекта методами нормативного анализа с конкретизацией значений абсолютного (т. е. не ограниченного рамками прогнозного фона) и относительного (т. е. привязанного к этим рамкам) оптимума по заранее определенным критериям сообразно заданным нормам, идеалам, целям.

Оценка достоверности и точности, а также обоснованности прогноза (его верификация) посредством уточнения гипотетических моделей обычно методами опроса экспертов.

Выработка рекомендаций для вариантных решений в сфере управления на основе сопоставления поисковых и нормативных моделей. Для уточнения рекомендаций возможно проведение опросов населения и экспертов. Иногда при этом строятся серии поствероятностных прогнозных моделей-сценариев с учетом возможных последствий реализации выработанных рекомендаций для их дальнейшего уточнения.

Экспертное обсуждение (экспертiza) прогноза и научных рекомендаций, их доработка с учетом выявленных несовершенств и сдача заказчику.
Вновь предпрогнозная ориентация на основе сопоставления материалов уже разработанного прогноза с новыми данными прогнозного фона и на новый цикл исследования, ибо прогнозирование должно быть таким же непрерывным, как управление, повышению эффективности которого оно призвано служить.

Можно выделить следующие виды социально-экономических объектов прогнозирования:

- с полным обеспечением количественной информацией, для которых имеется в наличии ретроспективная количественная информация в необходимом объеме;

- с неполным обеспечением количественной информацией, для которых имеющаяся в наличии ретроспективная информация допускает использование статистических методов, однако не обеспечивает на заданном времени упреждения заданную точность прогноза;

- с наличием качественной ретроспективной информации, относительно прошлого развития которых имеется только качественная информация и полностью отсутствует либо очень ограничена количественная;

- с полным отсутствием ретроспективной информации — это, как правило, несуществующие, проектируемые объекты.

Статистические методы могут с уверенностью применяться для первого случая, с некоторым уменьшением точности прогноза — для второго случая. Для двух последних случаев более эффективно применение экспертных методов.

На практике при построении прогнозов социально-экономических явлений исследователь чаще всего имеет дело с исходными данными поперечного или продольного срезов. В первом случае он применяет регрессионные модели, во втором — модели временных рядов. Если же имеется недостаток количественной информации, то наиболее распространенными по применению являются экспертные методы.

В этой связи дальнейшее изложение материала сосредоточено на статистических и экспертных методах.

2. Временные ряды и прогнозирование

При рассмотрении классической модели регрессии характер экспериментальных данных, как правило, не имеет принципиального значения. Однако это оказывается не так, если условия классической модели нарушены. Методы исследования моделей, основанных на данных пространственных выборок и временных рядов, вообще говоря, существенно отличаются. Объясняется это тем, что в отличие от пространственных выборок наблюдения во временных рядах, как правило, нельзя считать независимыми. В этой главе мы остановимся на некоторых общих понятиях и вопросах, связанных с временными рядами, использованием регрессионных моделей временных рядов для прогнозирования. При анализе точности этих моделей и определении интервальных ошибок прогноза на их основе будем полагать, что рассматриваемые в главе регрессионные модели временных рядов удовлетворяют условиям классической модели. Модели временных рядов, в которых нарушены эти условия, будут рассмотрены ниже.
2.1. Общие сведения о временных рядах и задачах их анализа

Под временным рядом (динамическим рядом, или рядом динамики) в экономике подразумевается последовательность наблюдений некоторого признака (случайной величины) \( x \) в последовательные моменты времени. Отдельные наблюдения называются уровнями ряда, которые будем обозначать \( x_t \) \( (t = 1, 2, ..., n) \), где \( n \) – число уровней.

В таблице приведены данные, отражающие спрос на некоторый товар за восьмилетний период (усл. ед), т. е. временной ряд спроса \( x_t \).

<table>
<thead>
<tr>
<th>год, ( t )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>спрос, ( x_t )</td>
<td>213</td>
<td>171</td>
<td>291</td>
<td>309</td>
<td>317</td>
<td>362</td>
<td>351</td>
<td>361</td>
</tr>
</tbody>
</table>

В качестве примера на рис.1 временной ряд \( x_t \) изображен графически.

В общем виде при исследовании экономического временного ряда \( x_t \) выделяются несколько составляющих (аддитивная модель):

\[
x_t = T + S + C + E,
\]

где \( T = 181.32 + 25.679 \cdot t \) (1)

или мультипликативная модель:

\[
x_t = T \cdot S \cdot C \cdot E,
\]

где \( T \) – тренд, плавно меняющаяся компонента, описывающая чистое влияние долговременных факторов, т. е. длительную (« вековую ») тенденцию изменения признака (например, рост населения, экономическое развитие, изменение структуры потребления и т. п.).
Сезонная компонента, отражающая повторяемость экономических процессов в течение не очень длительного периода (года, иногда месяца, недели и т.д., например, объем продаж товаров или перевозок пассажиров в различные времена года);

Циклическая компонента, отражающая повторяемость экономических процессов в течение длительных периодов (например, влияние волн экономической активности Кондратьева, демографических «ям», циклов солнечной активности и т.п.);

Случайная компонента, отражающая влияние не поддающихся учету и регистрации случайных факторов.

Следует обратить внимание на то, что в отличие от случайной компоненты первые три составляющие (компоненты) \( T, S, C \) являются закономерными, неслучайными. Важнейшей классической задачей при исследовании экономических временных рядов является выявление и статистическая оценка основной тенденции развития изучаемого процесса и отклонений от нее. Отметим основные этапы анализа временных рядов:

- графическое представление и описание поведения временного ряда;
- выделение и удаление закономерных (неслучайных) составляющих временного ряда (тренда, сезонных и циклических составляющих);
- сглаживание и фильтрация (удаление низко- или высокоочастотных составляющих временного ряда);
- исследование случайной составляющей временного ряда, построение и проверка адекватности математической модели для ее описания;
- прогнозирование развития изучаемого процесса на основе имеющегося временного ряда;
- исследование взаимосвязи между различными временными рядами.

Среди наиболее распространенных методов анализа временных рядов выделим корреляционный и спектральный анализ, модели авторегрессии и скользящей средней. О некоторых из них речь пойдет ниже. Если выборка \( x_1, x_2, ..., x_n \) рассматривается как одна из реализаций случайной величины \( X \), временной ряд \( x_1, x_2, ..., x_n \) рассматривается как одна из реализаций (траекторий) случайного процесса. \( X(t) \). Вместе с тем следует иметь в виду принципиальные отличия временного ряда \( x_t, (t = 1, 2, ..., n) \) от последовательности наблюдений \( x_1, x_2, ..., x_n \) образующих случайную выборку.

Во-первых, в отличие от элементов случайной выборки члены временного ряда, как правило, не являются статистически независимыми.

Во-вторых, члены временного ряда не являются одинаково распределенными.

2.2. Стационарные временные ряды и их характеристики. Автокорреляционная функция

Важное значение в анализе временных рядов имеют стационарные временные ряды, вероятностные свойства которых не изменяются во времени. Стационарные временные ряды применяются, в частности, при описании случайных составляющих анализируемых рядов.

\(^1\) Случайным процессом (или случайной функцией) \( X(t) \) неслучайного аргумента \( t \) называется функция, которая при любом значении \( t \) является случайной величиной.
Временной ряд \( x_t (t = 1, 2, ..., n) \) называется строго стационарным (или стационарным в узком смысле), если совместное распределение вероятностей \( n \) наблюдений \( x_1, x_2, ..., x_n \) такое же, как и \( n \) наблюдений \( x_{1+k}, x_{2+k}, ..., x_{n+k} \) при любых \( n \) и \( k \). Другими словами, свойства строго стационарных рядов \( x_t \) не зависят от момента \( k \), т. е. закон распределения и его числовые характеристики не зависят от \( k \). Следовательно, математическое ожидание \( M(x_t) = \mu \), среднее квадратическое отклонение \( \sigma(x_t) = \sigma \) могут быть оценены по наблюдениям \( x_t (t = 1, 2, ..., n) \) по формулам:

\[
\bar{x}_t = \frac{\sum_{t=1}^{n} x_t}{n}; \\
\sigma_t^2 = \frac{\sum_{t=1}^{n} (x_t - \bar{x}_t)^2}{n}.
\]

Простейшим примером стационарного временного ряда, у которого математическое ожидание равно нулю, а ошибки е, некоррелированны, является «белый шум». Следовательно, можно сказать, что возмущения (ошибки) \( e \), в классической линейной регрессионной модели образуют белый шум, а в случае их нормального распределения − нормальный (гауссовский) белый шум.

Степень тесноты связи между последовательностями наблюдений стационарного временного ряда \( x_1, x_2, ..., x_n \) и \( x_{1+k}, x_{2+k}, ..., x_{n+k} \) (сдвинутых относительно друг друга на \( k \) единиц, или, как говорят, с лагом \( k \)) может быть определена с помощью коэффициента корреляции

\[
\rho(k) = \frac{M[(x_t - \mu)(x_{t+k} - \mu)]]}{\sigma(t)\sigma(t + k)} = \frac{M[(x_t - \mu)(x_{t+k} - \mu)]}{\sigma^2}
\]

ибо \( M(x_t) = M(x_{t+k}) = \mu, \sigma(t) = \sigma(t + k) = \sigma \).

Так как коэффициент \( \rho(k) \) измеряет корреляцию между членами одного и того же ряда, его называют коэффициентом автокорреляции, а зависимость \( \rho(k) \) − автокорреляционной функцией (\( ACF \) − autocorrelation function). Автокорреляционная функция безразмерна, т. е. не зависит от масштаба измерения анализируемого временного ряда. Ее значения могут изменяться в пределах от \(-1\) до \(+1\); при этом \( \rho(0) = 1 \).

Кроме того, используется и автоковариационная функция

\[
\gamma(k) = M((x_t - \bar{x}_t)(x_{t+k} - \bar{x}_t)),
\]

где \( \bar{x}_t = M(x_t) \). Эти функции, очевидно, связаны соотношением \( \rho(k) = \gamma(k)/\gamma(0) \).

В силу стационарности временного ряда \( x_t (t = 1, 2, ..., n) \) автокорреляционная функция \( \rho(k) \) зависит только от лага \( k \), причем \( \rho(-k) = \rho(k) \), т. е. при изучении \( \rho(k) \) можно ограничиться рассмотрением только положительных значений \( k \). Статистической оценкой \( \rho(k) \) является выборочный коэффициент автокорреляции \( r(k) \), определяемый по формуле коэффициента корреляции 5, в которой \( n \) заменяется на \( n - k \):

\[
r(k) = \frac{(n-k) \sum_{t=1}^{n-k} x_t x_{t+k} - \sum_{t=1}^{n-k} x_t \sum_{t=1}^{n-k} x_{t+k}}{\sqrt{(n-k) \sum_{t=1}^{n-k} x_t^2} - (\sum_{t=1}^{n-k} x_t)^2} \sqrt{(n-k) \sum_{t=1+k}^{n-k} x_t^2} - (\sum_{t=1+k}^{n-k} x_t)^2}
\]

2 Наряду со строго стационарными временными рядами (в узком смысле) в эконометрике рассматриваются стационарные ряды (в широком смысле или слабо стационарные ряды), в которых требование неизменности при любых \( n, t \) и \( k \) распространяется лишь на числовые характеристики указанного распределения.
Функцию $r(k)$ называют выборочной автокорреляционной функцией, сокращенно ACF, а ее график – коррелограммой. Для дальнейшего заметим, что если $x_t$ – стационарный временной ряд и $C$ – некоторая постоянная, то временные ряды $x_t$ и $(x_t + c)$ имеют одинаковые коррелограммы.

При расчете $r(k)$ следует помнить, что с увеличением $k$ число $n - k$ пар наблюдений $x_t, x_{t+k}$ уменьшается, поэтому лаг $k$ должен быть таким, чтобы число $n - k$ было достаточным для определения $r(k)$. Обычно ориентируются на соотношение $k < n/4$.

Для стационарного временного ряда с увеличением лага $k$ взаимосвязь членов временного ряда $x_t, x_{t+k}$ ослабевает и автокорреляционная функция $\rho(k)$ должна убывать (по абсолютной величине). В то же время для ее выборочного (эмпирического) аналога $r(k)$, особенно при небольшом числе пар наблюдений $n - k$, свойство монотонного убывания (по абсолютной величине) при возрастании $k$ может нарушаться.

Наряду с автокорреляционной функцией при исследовании стационарных временных рядов рассматривается частная автокорреляционная функция или PACF (PACF – partial autocorrelation function) $\rho_{\text{part}}(k)$, где $\rho_{\text{part}}(k)$ есть частный коэффициент корреляции между членами временного ряда $x_t$ и $x_{t+k}$, т. е. коэффициент корреляции между $x_t$ и $x_{t+k}$ при устранении (элиминировании) влияния промежуточных (между $x_t$ и $x_{t+k}$) членов.

Напомним, что выборочным частным коэффициентом корреляции (или просто частным коэффициентом корреляции) между переменными $x_i$ и $x_j$ при фиксированных значениях остальных ($p - 2$) переменных называется выражение

$$ r_{ij,12i} = \frac{-q_{ij}}{\sqrt{q_{ii}q_{jj}}} \quad (7) $$

где $q_{ii}$, $q_{ij}$ и $q_{jj}$ – алгебраические дополнения элементов $r_{ii}$, $r_{ij}$ и $r_{jj}$ матрицы выборочных коэффициентов корреляции

$$
\begin{pmatrix}
1 & r_{12} & \ldots & r_{1p} \\
 r_{21} & 1 & \ldots & r_{2p} \\
 \ldots & \ldots & \ldots & \ldots \\
 r_{p1} & r_{p2} & \ldots & 1
\end{pmatrix},
$$

а $r_{ij}$ – коэффициент корреляции между $X_i$ и $X_j$.

Статистической оценкой $\rho_{\text{part}}(k)$ является выборочная частная автокорреляционная функция $r_{\text{part}}(k)$, где $r_{\text{part}}(k)$ – выборочный частный коэффициент корреляции, определяемый по формуле (7). Например, выборочный частный коэффициент автокорреляции порядка 2 между членами временного ряда $x_t$ и $x_{t+2}$ при устранении влияния $x_{t+1}$ может быть вычислен по формуле (7) из матрицы

$$
\begin{pmatrix}
1 & r(1) & r(2) \\
 r(1) & 1 & r(1) \\
 r(2) & r(1) & 1
\end{pmatrix},
$$

тогда

$$ r_{\text{part}}(2) = r_{02,1} = \frac{r(2) - r^2(1)}{1 - r^2(1)}, $$

где $r(1)$, $r(2)$ – выборочные коэффициенты автокорреляции между $x_t$ и $x_{t+1}$, $x_t$ и $x_{t+2}$, $t = 1, \ldots, n$. 
Пример. По данным таблицы 1 для временного ряда \( x_t \) найти среднее значение, среднее квадратическое отклонение, коэффициенты автокорреляции (для лагов \( k = 1; 2 \)) и частный коэффициент автокорреляции 1-го порядка.

Решение. Среднее значение временного ряда находим по формуле (3):
\[
\bar{x}_t = \frac{213 + 171 + \cdots + 361}{8} = 296.88 \text{ (ед.)}
\]

Дисперсию и среднее квадратическое отклонение можно вычислить по формуле (4), но в данном случае проще использовать соотношение
\[
s^2 = \frac{\sum_{t=1}^{n} x_t^2}{n} = \frac{213^2 + 171^2 + i + 361^2}{8} = 92478.38 \]

Найдём коэффициент автокорреляции \( r_k \) временного ряда (для лага \( k = 1 \)), т. е. коэффициент корреляции между последовательностями семи пар наблюдений \( x_t \) и \( x_{t+1} \) (\( t = 1, 2, \ldots, 7 \)).

<table>
<thead>
<tr>
<th>Данные для примера</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_t )</td>
</tr>
<tr>
<td>( x_{t+1} )</td>
</tr>
</tbody>
</table>

Таблица 1

Проведя необходимые вычисления, найдем, что \( r(1) = 0.725 \).

Коэффициент автокорреляции \( r(2) \) для лага \( k = 2 \) между членами ряда \( x_t \) и \( x_{t+2} \) (\( t = 1, 2, \ldots, 6 \)) по шести парам наблюдений вычисляем аналогично: \( r(2) = 0.842 \).

Для определения частного коэффициента корреляции 1-го порядка \( r_{par}(2) = r_{0.2.1} \) между членами ряда \( x_t \) и \( x_{t+2} \) при исключении влияния \( x_{t+1} \) вначале найдем (по аналогии с предыдущим) коэффициент автокорреляции \( r(1, 2) \) между членами ряда \( x_{t+1} \) и \( x_{t+2} \): \( r(1, 2) = 0.825 \), а затем вычислим \( r_{par}(2) \) по формуле (7): \( r_{par}(2) = 0.627 \).

Знание автокорреляционных функций \( r(k) \) и \( r_{par}(k) \) может оказать существенную помощь при подборе и идентификации модели анализируемого временного ряда и статистической оценке его параметров (см. об этом дальше).

2.3. Эргодичность

Мы говорили, что стохастический процесс – это функция как бы двух величин: времени и случайности. Когда случайность фиксирована, мы рассматриваем одну реализацию. Однако все наши рассуждения касаются характеристик случайных величин, таких как математическое ожидание \( M(x_t) \) и другие. Это подразумевает, что в каждый заданный момент времени мы имеем всю генеральную совокупность случайной величины, соответствующую этому моменту времени. Всякий раз усреднение идет по повторяющейся выборке. Однако в нашем распоряжении одна единственная реализация. Поэтому любая статистика, с которой мы будем иметь дело, может использовать только реализацию, а не повторяющуюся выборку. Единственной, по сути, возможность остается усреднение по реализации, по времени. Поэтому нам нужны основания, чтобы считать, что усреднение по времени в каком-то смысле эквивалентно усреднению по всевозможным значениям генеральной совокупности.
Процессы, которые обладают таким свойством, называются эргодическими (ergodic). Недостаточно для процесса быть стационарным, вообще говоря, нужно еще, чтобы процесс был эргодическим. Иногда это свойство называют свойством хорошего перемешивания. Разбросанные в разные моменты времени значения временного ряда должны составить такую же качественно выборку, как повторная выборка в один и тот же момент времени.

Эргодичность – это свойство, позволяющее для оценки математических ожиданий использовать усреднение по времени (по реализации). Например, мы хотим оценить математическое ожидание. Мы должны взять все возможные значения в один и тот же момент времени $t$. У нас таких нет. Но у нас есть значения в другие моменты времени. Свойство хорошего перемешивания означает, что если у нас достаточно длинная реализация, то можно заменить усреднение по ансамблю, по множеству, усреднением по времени. Для того, чтобы стационарный процесс был эргодичным, достаточно выполнения следующего условия

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} y(k) = 0.$$

Мы видим, что стационарные процессы ARMA(p,q) обладают свойством эргодичности. Нестационарный процесс не может быть эргодическим. Но не всякий стационарный процесс эргодичен, хотя для практических целей наличие стационарности неявно подразумевается эргодичность. Важно понимать, что одной стационарности при статистической обработке реализаций недостаточно.

2.4. Тренды

Существует три основных типа трендов.

**Первый** и самим очевидным типом тренда представляется тренд среднего, когда временной ряд выглядит как колебания около медленно возрастающей или убывающей величины.

**Второй** тип трендов – это тренд дисперсии. В этом случае во времени меняется амплитуда колебаний переменной. Иными словами, процесс гетероскедастичен. Часто экономические процессы с возрастающим средним имеют и возрастающую дисперсию.

**Третий** и более тонкий тип тренда, визуально не всегда наблюдаемый, – изменение величины корреляции между текущим и предшествующим значениями ряда, т. е. тренд автоковариации и автокорреляции.

Проведя разложение ряда на компоненты, мы, как правило, подразумеваем под трендом изменение среднего уровня переменной, то есть тренд среднего. В рамках анализа тренда среднего выделяют следующие основные способы аппроксимации временных рядов и соответствующие основные виды трендов среднего.

**Полиномиальный тренд**:

$$\bar{x}_t = a_0 + a_1 \cdot t + \cdots + a_p \cdot t^p.$$  \hspace{1cm} (8)

Для $p = 1$ имеем линейный тренд.

**Экспоненциальный тренд**:

$$\bar{x}_t = e^{a_0} \cdot e^{a_1 \cdot t} \cdot \cdots \cdot e^{a_p \cdot t^p}.$$  \hspace{1cm} (9)

**Гармонический тренд**:

$$\bar{x}_t = R \cdot \cos(\omega t + \phi),$$  \hspace{1cm} (10)

где $R$ – амплитуда колебаний, $\omega$ – угловая частота, $\phi$ – фаза.
Тренд, выражаемый логистической функцией:

\[
\bar{x}_t = \frac{k}{1 + b \cdot e^{-at}}. 
\]  (11)

Оценивание параметров полиномиального и экспоненциального трендов (после введения обозначения \(z_i = t^i, i = 1, ..., p\) – в первом случае и логарифмирования функции во втором случае) производится с помощью обычного МНК. Гармонический тренд оправдан, когда в составе временного ряда отчетливо прослеживаются периодические колебания. При этом если частота \(\omega\) известна (или ее можно оценить), то функцию (10) несложно представить в виде линейной комбинации синуса и косинуса: \(\bar{x}_t = \alpha \cos(\omega t) + \beta \sin(\omega t)\) и, рассчитав векторы \(\cos(\omega t)\) и \(\sin(\omega t)\), также воспользоваться МНК для оценивания параметров \(\alpha\) и \(\beta\). Логистическая кривая нуждается в особом рассмотрении.

Оценка логистической функции

Проанализируем логистическую функцию: (11) \(\bar{x}_t = \frac{k}{1 + b \cdot e^{-at}}\), где \(a, b, k\) – параметры, подлежащие оцениванию. Функция ограничена и имеет горизонтальную асимптоту \(\lim_{t \to \infty} \bar{x}_t = k\) (рис. 2).

В этом преимущество логистической функции перед полиномиальной или экспоненциальной функциями, которые по мере роста \(t\) стремятся в бесконечность и, следовательно, не всегда годятся для прогнозирования. Логистическая кривая наиболее часто используется при изучении социальных, в частности демографических процессов. Особенностью логистической кривой является нелинейность по оцениваемым параметрам \(a, b, k\), поэтому система уравнений, получаемая с помощью МНК, нелинейна относительно неизвестных параметров и для ее решения могут применяться только интерактивные численные методы. Гарольд Готтелинг (H. Hotteling) предложил интересный метод для оценки этих параметров, основанный на использовании дифференциального уравнения логистической функции. Дифференцирование функции \(\bar{x}_t\) по времени \(t\) дает первую производную:

\[
\frac{d\bar{x}_t}{dt} = \frac{kabe^{-at}}{(1 + be^{-at})^2}. 
\]
Поскольку
\[ \bar{x}_t^2 = \frac{k}{(1 + be^{-at})^2} \text{ и } be^{-at} = \frac{k}{\bar{x}_t} - 1, \]
то, подставляя эти выражения в формулу первой производной, получаем дифференциальное уравнение, выражающее зависимость темпа прироста исследуемой переменной от абсолютного уровня показателя в момент времени \( t \):

\[ \frac{d\bar{x}_t}{\bar{x}_t} = a - \frac{a}{k} \frac{1}{\bar{x}_t}. \] (12)

Исходя из этого соотношения, можно предположить, что в реальности абсолютный прирост показателя \( \Delta x_i \) связан с фактическим его уровнем \( x_t \) следующей статистической зависимостью:

\[ \Delta x_i = ax_t + \left( -\frac{a}{k} \right) x_t^2 + e_t, \] где \( e_t \) – белый шум. К этому уравнению теперь можно применить непосредственно метод наименьших квадратов, получить оценки параметров \( a \) и \( -a/k \) и, следовательно, найти \( k \). Оценка параметра \( b \) методом моментов впервые предложена Родсом. Так как \( be^{-at} = k/x_t - 1 \), то \( \ln b = at + \ln(k/x_t - 1) \) и с помощью метода моментов получаем:

\[ \ln b = \frac{1}{n} \left( a \cdot \frac{n(n + 1)}{2} + \sum_{t=1}^{n} \ln \left( \frac{k}{x_t} - 1 \right) \right). \]

Описанный выше метод Готтенинга имеет ограниченную сферу применения, его использование оправдано лишь в том случае, если наблюдения в исходном временному ряду представлены через равные промежутки времени (например, ежегодные или ежедневные данные).

Пример. По данным таблицы 1 выделить линейный тренд.

Решение. Нам нужно методом МНК (метод наименьших квадратов) оценить коэффициенты уравнения \( \bar{x}_t = a + b \cdot t \). Для этого рассмотрим функцию \( F(a, b) \), равную сумме квадратов разности \((x_t - a - b \cdot t)^2\):

\[ F(a, b) = \sum_{t=1}^{n} \left( \bar{x}_t - (a + bt) \right)^2. \]

Это сумма квадратов отклонений от тренда, то есть ошибки. Мы хотим добиться, чтобы эта сумма квадратов ошибок была минимальной. Для этого продифференцируем \( F \) по \( a \) и \( b \) и приравняем производные к нулю

\[ \frac{dF}{da} = -2 \sum_{t=1}^{n} (x_t - a - bt) = 0, \]

\[ \frac{dF}{db} = -2 \sum_{t=1}^{n} t(x_t - a - bt) = 0. \]
Или

\[ an + b \sum_{t=1}^{n} t = \sum_{t=1}^{n} x_t, \]
\[ a \sum_{t=1}^{n} t + b \sum_{t=1}^{n} t^2 = \sum_{t=1}^{n} x_t. \]

Учитывая, что значения переменной \( t = 1, 2, \ldots, n \) образуют натуральный ряд чисел от 1 до \( n \), суммы \( \sum_{t=1}^{n} t, \sum_{t=1}^{n} t^2 \) можно выразить через число членов ряда \( n \) по известным в математике формулам:

\[ \sum_{t=1}^{n} t = \frac{n(n + 1)}{2}; \quad \sum_{t=1}^{n} t^2 = \frac{n(n + 1)(2n + 1)}{6}. \]

Система уравнений (13) примет вид:

\[ \begin{cases} 8a + 36b = 2.375, \\ 36a + 204b = 11.766, \end{cases} \]

откуда \( a = 181.32; \ b = 25.679 \) и уравнение тренда \( x_t = 181.32 + 25.679 \cdot t \) (Рис 1), т. е. спрос ежегодно увеличивается в среднем на 25.7 ед.

Регрессионный анализ позволяет определить оценки коэффициентов регрессии. Но, являясь лишь оценками, они не позволяют сделать вывод, насколько точно эмпирическое уравнение регрессии соответствует уравнению для всей генеральной совокупности, насколько близки оценки \( a \) и \( b \) коэффициентов своим теоретическим прототипам \( \alpha \) и \( \beta \), насколько надежны найденные оценки. Для ответа на эти вопросы необходимы определенные дополнительные исследования.

2.5. Случайные величины, необходимые для тестирования уравнений регрессии

Определение 1. Если случайные величины \( Z_1, Z_2, \ldots, Z_n \) независимы и все имеют стандартное нормальное распределение \( N(1, 0) \) (единичную дисперсию и нулевую среднюю), то тогда говорят, что случайная величина \( X \), являющаяся суммой квадратов стандартных нормальных величин в количестве \( n \) штук, имеет распределение хи-квадрат с \( n \) степенями свободы (\( \chi_n^2 \)):

\[ x^2 = \sum_{t=1}^{n} Z_t^2. \]

Определение 2. Если случайная величина \( Z \) имеет стандартное нормальное распределение \( (N(0, 1)) \), случайная величина \( X \) имеет распределение хи-квадрат с \( n \) степенями свободы (\( \chi_n^2 \)) и \( Z \) и \( X \) независимы (их корреляция равна нулю), то случайная величина

\[ t = \frac{Z}{\sqrt{\frac{1}{n} X}}, \text{ где } X = \sum_{t=1}^{n} X_t^2, \]

называется распределением Стьюдента с \( n \) степенями свободы (\( t_n \)).
Определение 3. Если случайная величина $X_1$ имеет распределение хи-квадрат с $n$ степенями свободы, а случайная величина $X_2$ имеет распределение хи-квадрат с $m$ степенями свободы, то случайная величина

$$ F = \frac{X_1}{n} / \frac{X_2}{m} $$

имеет распределение Фишера-Снедекора с $n$ и $m$ степенями свободы:

$$ F = \frac{X_1^2/n}{X_2^2/m} = F_{n,m}. $$

Проверка значимости линейной регрессии.

Построение уравнения регрессии является начальным этапом эконометрического анализа. Первое же построенное по выборке уравнение регрессии очень редко является удовлетворительным по тем или иным характеристикам. Поэтому следующей важнейшей оценкой является проверка качества уравнения регрессии. В эконометрике принята устоявшаяся схема такой проверки, которая проводится по следующим направлениям:

- проверка статистической значимости коэффициентов уравнения регрессии;
- проверка общего качества уравнения регрессии;
- проверка свойств данных, выполнимость которых предполагалась при оценивании уравнения (проверка выполнимости предпосылок МНК).

После того как найдено уравнение линейной регрессии, проводится оценка значимости как уравнения в целом, так и его параметров. Оценка значимости уравнения регрессии в целом дается с помощью $F$-критерия Фишера. При этом выдвигается нулевая гипотеза, что коэффициент регрессии равен нулю, т. е. $b = 0$, и, следовательно, фактор $x$ не оказывает влияния на результат $y$. Непосредственному расчету $F$-критерия предшествует анализ дисперсии. Центральное место в нем занимает разложение общей суммы квадратов отклонений переменной $y$ от среднего значения на две части – «объясненную» и «необъясненную»:

$$ \sum (y - \bar{y})^2 = \sum (\hat{y}_x - \bar{y})^2 + \sum (y - \hat{y}_x)^2 $$

(15)

Общая сумма квадратов отклонений
Сумма квадратов отклонений, объясненная регрессией
Остаточная сумма отклонений

Общая сумма квадратов отклонений индивидуальных значений результативного признака $y$ от среднего значения $y$ вызвана влиянием множества причин. Условно разделим всю совокупность причин на две группы: изучаемый фактор $x$ и прочие факторы. Если фактор $x$ не оказывает влияния на результат, то линия регрессии на графике параллельна оси ох. Тогда все дисперсия результативного признака обусловлена воздействием прочих факторов и общая сумма квадратов отклонений совпадает с остаточной. Если же прочие факторы не влияют на результат, то $y$ связан с $x$ функционально и остаточная сумма квадратов равна нулю. В этом случае сумма квадратов отклонений, объясненная регрессией, совпадает с общей суммой квадратов.
Поскольку не все точки графика лежат на линии регрессии, то всегда имеет место их разброс как обусловленный влиянием фактора \( x \), т. е. регрессией \( y \) по \( x \), так и вызванный действием прочих причин (необъясненная вариация). Пригодность линии регрессии для прогноза зависит от того, какая часть общей вариации признака \( y \) приходится на объясненную вариацию. Очевидно, что если сумма квадратов отклонений, обусловленная регрессией, будет больше остаточной суммы квадратов, то уравнение регрессии статистически значимо и фактор \( x \) оказывает существенное воздействие на результат \( y \). Это равносильно тому, что коэффициент детерминации \( R^2 \) будет приближаться к единице.

Любая сумма квадратов отклонений связана с числом степеней свободы (\( df – degrees of freedom \)), т. е. с числом свободы независимого варьирования признака. Число степеней свободы связано с числом единиц совокупности \( n \) и с числом определяемых по ней констант. Применимо к исследуемой проблеме число степеней свободы должно показать, сколько независимых отклонений из \( n \) возможных

\[
(y_1 - \bar{y}), (y_2 - \bar{y}), \ldots, (y_n - \bar{y})
\]

требуется для образования данной суммы квадратов. Так, для общей суммы квадратов \( \sum(y - \bar{y})^2 \) требуется \((n - 1)\) независимых отклонений, ибо по совокупности из \( n \) единиц после расчета среднего уровня свободно варьируют лишь \((n - 1)\) число отклонений. Например, имеем ряд значений \( y \): 1, 2, 3, 4, 5. Среднее из них равно 3, и тогда \( n \) отклонений от среднего составят: \(-2; -1; 0; 1; 2\). Так как \( \sum(y - \bar{y}) = 0 \), то свободно варьируют лишь четыре отклонения, а пятое отклонение может быть определено, если предыдущие четыре известны.

При расчете объясненной или факторной суммы квадратов \( \sum(\hat{y} - \bar{y})^2 \) используются теоретические (расчетные) значения результативного признака \( \hat{y}_x \), найденные по уравнению регрессии: \( \hat{y}_x = a + b \cdot x \).

Поскольку при заданном объеме наблюдений по \( x \) и \( y \) факторная сумма квадратов при линейной регрессии зависит только от одной константы коэффициента регрессии \( b \), то данная сумма квадратов имеет одну степень свободы. К этому же выводу придем, если рассмотрим содержательную сторону расчетного значения признака \( y \), т. е. \( \hat{y}_x \). Величина \( \hat{y}_x \) определяется по уравнению линейной регрессии: \( \hat{y}_x = a + b \cdot x \). Параметр \( a \) можно определить, как \( a = \bar{y} - b \cdot \bar{x} \). Подставив выражение параметра \( a \) в линейную модель, получим:

\[
\hat{y}_x = \bar{y} - b \cdot \bar{x} + b \cdot x = \bar{y} - b \cdot (x - \bar{x}).
\]

Отсюда видно, что при заданном наборе переменных \( y \) и \( x \) расчетное значение \( \hat{y}_x \) является в линейной регрессии функцией только одного параметра – коэффициента регрессии. Соответственно и факторная сумма квадратов отклонений имеет число степеней свободы, равное 1.

Существует равенство между числом степеней свободы общей, факторной и остаточной суммами квадратов. Число степеней свободы остаточной суммы квадратов при линейной регрессии составляет \( n - 2 \). Число степеней свободы для общей суммы квадратов определяется числом единиц, и поскольку мы используем среднюю вычисленную по данным выборки, то теряем одну степень свободы, т. е. \( df_{общ} = n - 1 \). Итак, имеем два равенства:

\[
\sum (y - \bar{y})^2 = \sum (\hat{y}_x - \bar{y})^2 + \sum (y - \hat{y}_x)^2,
\]

\[
(n - 1) = 1 + (n - 2).
\]
Разделив каждую сумму квадратов на соответствующее ей число степеней свободы, получим средний квадрат отклонений, или, что тоже самое, дисперсию на одну степень свободы \( D \).

\[
D_{\text{общ}} = \frac{(y - \bar{y})^2}{n - 1};
\]

(18)

\[
D_{\text{факт}} = \frac{(\bar{y}_x - \bar{y})^2}{1};
\]

(19)

\[
D_{\text{общ}} = \frac{(y - \bar{y}_x)^2}{n - 2}.
\]

(20)

Определение дисперсии на одну степень свободы приводит дисперсии к сравнимому виду. Разделив факторную на остаточную дисперсии в расчете на одну степень свободы, получим величину \( F \)-отношения (\( F \)-критерий):

\[
F = \frac{D_{\text{факт}}}{D_{\text{ост}}},
\]

где \( F \) – критерий для проверки нулевой гипотезы \( H_0 \): \( D_{\text{факт}} = D_{\text{ост}} \). Если нулевая гипотеза справедлива, то факторная и остаточная дисперсии не отличаются друг от друга. Для \( H_0 \) необходимо опровержение, чтобы факторная дисперсия превышала остаточную в несколько раз. Английским статистиком Джорджем Снедекором разработаны таблицы критических значений \( F \)-отношений при разных уровнях существенности нулевой гипотезы и различном числе степеней свободы. Табличное значение \( F \)-критерия – это максимальная величина отношения дисперсий, которая может иметь при случайном их расхождении для данного уровня вероятности наличие нулевой гипотезы. Вычисленное значение \( F \)-отношения признается достоверным (отличным от единицы), если оно больше табличного. В этом случае нулевая гипотеза об отсутствии связи признаков отклоняется и делается вывод о существенности этой связи: \( F_{\text{факт}} > F_{\text{табл}} \). \( H_0 \) отклоняется.

Если же величина окажется меньше табличной \( F_{\text{факт}} < F_{\text{табл}} \), то вероятность нулевой гипотезы выше заданного уровня (например, 0,05) и она не может быть отклонена без серьезного риска сделать неправильный вывод о наличии связи. В этом случае уравнение регрессии считается статистически незначимым. \( F_{\text{табл}} = 6.61 \), то уравнение тренда значимо.

В линейной регрессии

\[
\hat{y} = c + b_1x_1 + b_2x_2 + \ldots + b_mx_m
\]

обычно оценивается значимость не только уравнения в целом, но и отдельных его параметров. С этой целью по каждому из параметров определяется его стандартная ошибка \( b_i \). Стандартная ошибка коэффициента регрессии определяется по формуле

\[
b_i = \sqrt{\frac{\sum (y - \hat{y}_x)^2}{(n - m - 1)} \cdot \frac{1}{\sum (x - \bar{x})^2}} = \sqrt{\frac{S^2}{\sum (x - \bar{x})^2}}.
\]

(21)

где \( S^2 \) – остаточная дисперсия на одну степень свободы. В нашем примере \( b_i = 6.501570 \). Фактическое значение \( t_{0.05,5} = 2.0150 \). Так как наблюдаемое значение \( t \) статистики 6.501570 > 2.0150, то коэффициент \( b \) в регрессии с вероятностью 0.95 значим.
2.6. Сезонные колебания

Известно несколько подходов к анализу структуры временных рядов, содержащих сезонные или циклические колебания. Простейший подход — расчет значений сезонной компоненты методом скользящей средней и построение аддитивной или мультипликативной модели временного ряда. Напомним, что общий вид аддитивной модели следующий:

\[ Y = T + S + E. \] (22)

Эта модель предполагает, что каждый уровень временного ряда может быть представлен как сумма трендовой \( T \), сезонной \( S \) и случайной \( E \) компонент. Общий вид мультипликативной модели выглядит так:

\[ Y = TSE. \] (22)

Данная модель предполагает, что каждый уровень временного ряда может быть представлен как произведение трендовой \( T \), сезонной \( S \) и случайной \( E \) компонент. Выбор одной из двух моделей проводится на основе анализа структуры сезонных колебаний. Если амплитуда колебаний приблизительно постоянна, строят аддитивную модель временного ряда, в которой значения сезонной компоненты предполагаются постоянными для различных циклов. Если амплитуда сезонных колебаний возрастает или уменьшается, строят мультипликативную модель временного ряда, которая ставит уровни ряда в зависимость от значений сезонной компоненты. Построение аддитивной и мультипликативной моделей сводится к расчету значений \( T, S \) и \( E \) для каждого уровня ряда. Процесс построения модели включает в себя следующие шаги.

Шаг 1. Выравнивание исходного ряда методом скользящей средней.

Шаг 2. Расчет значений сезонной компоненты \( S \).

Шаг 3. Устранение сезонной компоненты из исходных уровней ряда и получение выровненных данных \((T + E)\) в аддитивной или \((T \cdot E)\) в мультипликативной модели.

Шаг 4. Аналитическое выравнивание уровней \((T + E)\) или \((T \cdot E)\) и расчет значений \( T \) с использованием полученного уравнения тренда.

Шаг 5. Расчет полученных по модели значений \((T + S)\) или \((T \cdot E)\).

Шаг 6. Расчет абсолютных и/или относительных ошибок. Если полученные значения ошибок не содержат автокорреляции, ими можно заменить исходные уровни ряда и в дальнейшем использовать временной ряд ошибок \( E \) для анализа взаимосвязи исходного ряда и других временных рядов.

Рассмотрим методику построения аддитивной модели временного ряда на примере.

\[ \text{Моделирование циклических колебаний в целом осуществляется аналогично моделированию сезонных колебаний, поэтому мы рассмотрим только методы моделирования последних.} \]
Пример. Обратимся к данным об объеме потребления электроэнергии жителями региона за последние четыре года в Млн. Кв (см. табл. 2).

Вычислим значения автокорреляционной функции ряда \( x_t \). Получим

\[
\begin{array}{cccccccccccc}
  k & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\
  ACF & 0.142 & 0.478 & 0.068 & 0.684 & 0.016 & -0.448 & -0.078 & 0.355 & -0.067 & -0.348 & -0.134 & 0.096 \\
\end{array}
\]

Изобразим график этой функции

Рис. 3. График ряда и коррелограмма

Так как \( r(1) = 0.142 \), то это говорит о слабой зависимости текущих значений уровней ряда от непосредственно предшествующих им уровней. Коррелограмма (рис. 3) показывает, что максимум ACF достигается при лаге 4, что свидетельствует о сезонных колебаниях с лагом 4. Об этом свидетельствует и график самого ряда. Объемы потребления электроэнергии в осенне-зимний период времени (I и IV кварталы) выше, чем весной и летом (II и III кварталы). По графику этого ряда (см. рис. 3) можно установить наличие приблизительно равной амплитуды колебаний. Это свидетельствует о соответствии этого ряда аддитивной модели. Рассчитаем ее компоненты.

Шаг 1. Проведем выравнивание исходных уровней ряда методом скользящей средней. Для этого:
- просуммируем уровни ряда последовательно за каждые четыре квартала со сдвигом на один момент времени и определим условные годовые объемы потребления электроэнергии (гр. 3 табл. 2);
- разделив полученные суммы на 4, найдем скользящие средние (гр. 4 табл. 2). Отметим, что полученные таким образом выраженные значения уже не содержат сезонной компоненты;
- приведем эти значения в соответствие с фактическими моментами времени, для чего найдем средние значения из двух последовательных скользящих средних – центрированные скользящие средние (гр. 5 табл. 2). Заметим, что оценки сезонной компоненты расположены ровно по середине относительно исходных данных (второго столбца) таблицы 2.
Шаг 2. Найдем оценки сезонной компоненты как разность между фактическими уровнями ряда и центрированными скользящими средними (гр. 6 табл. 2). Используем эти оценки для расчета значений сезонной компоненты $S$ (табл. 3). Для этого найдем среднее за каждый квартал (по всем годам) оценки сезонной компоненты $S_i$. В моделях с сезонной компонентой обычно предполагается, что сезонные воздействия за период взаимопогашаются. В аддитивной модели это выражается в том, что сумма значений сезонной компоненты по всем кварталам должна быть равна нулю. Для данной модели имеем: $0.6 - 1.938 – 1.275 + 2.708 = 0.075$.

### Расчет оценок сезонной компоненты

<table>
<thead>
<tr>
<th>Номер квartaла, $t$</th>
<th>Потребление электроэнергии, $x_t$</th>
<th>Скользящая средняя за четыре квартала</th>
<th>Центрированная скользящая средняя</th>
<th>Оценка сезонной компоненты</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>4.4</td>
<td>6.1</td>
<td>6.4</td>
<td>−1.25</td>
</tr>
<tr>
<td>3.</td>
<td>5</td>
<td>6.5</td>
<td>6.25</td>
<td>2.55</td>
</tr>
<tr>
<td>4.</td>
<td>9</td>
<td>6.75</td>
<td>6.45</td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>7.2</td>
<td>7</td>
<td>6.625</td>
<td>0.575</td>
</tr>
<tr>
<td>6.</td>
<td>4.8</td>
<td>7.2</td>
<td>6.875</td>
<td>−2.075</td>
</tr>
<tr>
<td>7.</td>
<td>6</td>
<td>7.4</td>
<td>7.1</td>
<td>−1.1</td>
</tr>
<tr>
<td>8.</td>
<td>10</td>
<td>7.5</td>
<td>7.3</td>
<td>2.7</td>
</tr>
<tr>
<td>9.</td>
<td>8</td>
<td>7.75</td>
<td>7.45</td>
<td>0.55</td>
</tr>
<tr>
<td>10.</td>
<td>5.6</td>
<td>8</td>
<td>7.625</td>
<td>−2.025</td>
</tr>
<tr>
<td>11.</td>
<td>6.4</td>
<td>8.25</td>
<td>7.875</td>
<td>−1.475</td>
</tr>
<tr>
<td>12.</td>
<td>11</td>
<td>8.4</td>
<td>8.125</td>
<td>2.875</td>
</tr>
<tr>
<td>13.</td>
<td>9</td>
<td>8.5</td>
<td>8.325</td>
<td>0.675</td>
</tr>
<tr>
<td>14.</td>
<td>6.6</td>
<td>8.13333</td>
<td>8.375</td>
<td>−1.775</td>
</tr>
<tr>
<td>15.</td>
<td>7</td>
<td>8.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16.</td>
<td>10.8</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Расчет значений сезонной компоненты

<table>
<thead>
<tr>
<th>Показатель</th>
<th>Год</th>
<th>Номер квартала, $i$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>I</td>
</tr>
<tr>
<td></td>
<td></td>
<td>II</td>
</tr>
<tr>
<td></td>
<td></td>
<td>III</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IV</td>
</tr>
<tr>
<td>Значения сезонной компоненты</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Итого за $i$-й квартал (за все годы)</td>
<td>X</td>
<td>1,800</td>
</tr>
<tr>
<td>Средняя оценка сезонной компоненты для $i$-го квартала, $S_i$</td>
<td>X</td>
<td>0,600</td>
</tr>
<tr>
<td>Скорректированная сезонная компонента, $S_i$</td>
<td>X</td>
<td>0,581</td>
</tr>
</tbody>
</table>

Определим корректирующий коэффициент:

$$k = 0.075/4 = 0.01875.$$
Рассчитаем скорректированные значения сезонной компоненты как разность между ее средней оценкой и корректирующим коэффициентом \( k \):
\[
S_i = S - k, \text{ где } i = 1 \ldots 4.
\]

Проверим условие равенства нулю суммы значений сезонной компоненты:
\[
0.581 - 1.977 - 1.294 + 2.690 = 0.
\]

Таким образом получены следующие значения сезонной компоненты:
I квартал: \( S_1 = 0.581 \);
II квартал: \( S_2 = -1.979 \);
III квартал: \( S_3 = -1.294 \);
IV квартал: \( S_4 = 2.690 \).

Занесем полученные значения в табл. 3 для соответствующих кварталов каждого года.

Шаг 3. Элиминируем влияние сезонной компоненты, вычитая ее значение из каждого уровня исходного временного ряда. Получим: \( T + E = x_t - S \) (гр. 4 табл. 4). Эти значения рассчитываются для каждого момента времени и содержат только тенденцию и случайную компоненту.

<table>
<thead>
<tr>
<th>( t )</th>
<th>( x_t )</th>
<th>( S_i )</th>
<th>( T + E = x_t - S_i )</th>
<th>( T )</th>
<th>( T+S )</th>
<th>( E = x_t - (T + S) )</th>
<th>( E^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>1</td>
<td>6.0</td>
<td>0.581</td>
<td>5.419</td>
<td>5.902</td>
<td>6.483</td>
<td>-0.483</td>
<td>0.2333</td>
</tr>
<tr>
<td>2</td>
<td>4.4</td>
<td>-1.977</td>
<td>6.337</td>
<td>6.088</td>
<td>4.111</td>
<td>0.289</td>
<td>0.0835</td>
</tr>
<tr>
<td>3</td>
<td>5.0</td>
<td>-1.294</td>
<td>6.294</td>
<td>6.275</td>
<td>4.981</td>
<td>0.019</td>
<td>0.0004</td>
</tr>
<tr>
<td>4</td>
<td>9.0</td>
<td>2.690</td>
<td>6.310</td>
<td>6.461</td>
<td>9.151</td>
<td>-0.151</td>
<td>0.0228</td>
</tr>
<tr>
<td>5</td>
<td>7.2</td>
<td>0.581</td>
<td>6.619</td>
<td>6.648</td>
<td>7.229</td>
<td>-0.029</td>
<td>0.0008</td>
</tr>
<tr>
<td>6</td>
<td>4.8</td>
<td>-1.977</td>
<td>6.777</td>
<td>6.834</td>
<td>4.857</td>
<td>-0.057</td>
<td>0.0032</td>
</tr>
<tr>
<td>7</td>
<td>6.0</td>
<td>-1.294</td>
<td>7.294</td>
<td>7.020</td>
<td>5.727</td>
<td>0.273</td>
<td>0.0745</td>
</tr>
<tr>
<td>8</td>
<td>10.0</td>
<td>2.690</td>
<td>7.310</td>
<td>7.207</td>
<td>9.896</td>
<td>0.104</td>
<td>0.0108</td>
</tr>
<tr>
<td>9</td>
<td>8.0</td>
<td>0.581</td>
<td>7.419</td>
<td>7.393</td>
<td>7.974</td>
<td>0.026</td>
<td>0.0007</td>
</tr>
<tr>
<td>10</td>
<td>5.6</td>
<td>-1.977</td>
<td>7.577</td>
<td>7.580</td>
<td>5.603</td>
<td>-0.030</td>
<td>0.0009</td>
</tr>
<tr>
<td>11</td>
<td>6.4</td>
<td>-1.294</td>
<td>7.694</td>
<td>7.766</td>
<td>6.472</td>
<td>-0.072</td>
<td>0.0052</td>
</tr>
<tr>
<td>12</td>
<td>11.0</td>
<td>2.690</td>
<td>8.310</td>
<td>7.952</td>
<td>10.642</td>
<td>0.358</td>
<td>0.1282</td>
</tr>
<tr>
<td>13</td>
<td>9.0</td>
<td>0.581</td>
<td>8.419</td>
<td>8.139</td>
<td>8.720</td>
<td>0.280</td>
<td>0.0784</td>
</tr>
<tr>
<td>14</td>
<td>6.6</td>
<td>-1.977</td>
<td>8.577</td>
<td>8.325</td>
<td>6.348</td>
<td>0.252</td>
<td>0.0635</td>
</tr>
<tr>
<td>15</td>
<td>7.0</td>
<td>-1.294</td>
<td>8.294</td>
<td>8.519</td>
<td>7.218</td>
<td>-0.218</td>
<td>0.0475</td>
</tr>
<tr>
<td>16</td>
<td>10.8</td>
<td>2.690</td>
<td>8.110</td>
<td>8.698</td>
<td>11.388</td>
<td>-0.588</td>
<td>0.3457</td>
</tr>
</tbody>
</table>

Шаг 4. Определим компоненту \( T \) данной модели. Для этого проведем аналитическое выравнивание ряда \( (T+E) \) с помощью линейного тренда. Результаты аналитического выравнивания следующие (табл. 5).
Таблица 5

<table>
<thead>
<tr>
<th>Название показателя</th>
<th>Значение</th>
</tr>
</thead>
<tbody>
<tr>
<td>Константа</td>
<td>5.715416</td>
</tr>
<tr>
<td>Коэффициент регрессии</td>
<td>0.186421</td>
</tr>
<tr>
<td>Стандартная ошибка коэффициента регрессии, $R$-квадрат</td>
<td>0.015188, 0.914971</td>
</tr>
<tr>
<td>Число наблюдений</td>
<td>16</td>
</tr>
<tr>
<td>Число степеней свободы</td>
<td>14</td>
</tr>
</tbody>
</table>

Таким образом имеем линейный тренд: $T = 5.715 + 0.186 \cdot t$. Подставив в это уравнение значения $t = 1, \ldots, 16$, найдем уровни $T$ для каждого момента времени (гр. 5 табл. 4). График уравнения тренда приведен на рис. 4.

Шаг 5. Найдем значения уровня ряда, полученные по аддитивной модели. Для этого прибавим к уровням $T$ значения сезонной компоненты для соответствующих кварталов. Графически значения $(T + S)$ представлены на рис. 4.

Шаг 6. В соответствии с методикой построения аддитивной модели расчет ошибки проводится по формуле:

$$E = Y - (T + S).$$

Это абсолютная ошибка. Численные значения абсолютных ошибок приведены в гр. 7 табл. 4. По аналогии с моделью регрессии для оценки качества построения модели, а также для выбора наилучшей модели можно использовать сумму квадратов абсолютных ошибок. Для данной аддитивной модели сумма квадратов абсолютных ошибок равна 1.10. По отношению к общей сумме квадратов отклонений уровней ряда от его среднего уровня, равной 71.59, эта величина составляет чуть более 1.5 %:

$$(1 - 1.10/71.59) - 100 = 1.536.$$
Следовательно, можно сказать, что аддитивная модель объясняет 98.5 % общей вариации уровней временного ряда потребления электроэнергии за последние 16 кварталов.

2.7. Сезонные фиктивные переменные

Выявление и устранение сезонного эффекта (в некоторых источниках употребляется термин «десезонализация уровней ряда») используются в двух направлениях. Во-первых, воздействие сезонных колебаний следует устранять на этапе предварительной обработки исходных данных при изучении взаимосвязи нескольких временных рядов. Поэтому в российских и международных статистических сборниках часто публикуются данные, в которых устранено влияние сезонной компоненты (если это помесячная или поквартальная статистика), например, показатели объемов производства в отдельных отраслях промышленности, уровня безработицы и т. д. Во-вторых, выявление сезонного эффекта производится в анализе структуры одномерных временных рядов с целью прогнозирования уровней ряда в будущие моменты времени.

Использование сезонных фиктивных переменных является одним из методов моделирования сезонных составляющих временного ряда. При этом подходе строится регрессионная модель, которая, помимо фактора времени, включает сезонные фиктивные переменные. Фиктивной переменной (dummy variable) называется атрибутивный (или качественный) фактор, представленный с помощью определенного цифрового кода. Регрессионная модель, включающая в качестве фактора (факторов) фиктивную переменную, называется регрессионной моделью с переменной структурой. Рассмотрим временной ряд $X_{ij}$, где $i$ — это номер сезона (периода времени внутри года, например, месяца или квартала); $L$ — число сезонов в году; $j$ — номер года; $m$ — общее количество лет. Количество уровней исходного ряда равно $L \times m = n$. Число сезонных фиктивных переменных в регрессионной модели всегда должно быть на единицу меньше сезонов внутри года, т. е. должно быть равно величине $L - 1$. При моделировании ежемесячных данных регрессионная модель, помимо фактора времени, должна содержать одиннадцать фиктивных компонент ($12 - 1$). При моделировании поквартальных данных регрессионная модель должна содержать три фиктивных компоненты ($4 - 1$) и т. д. Каждому из сезонов соответствует определенное сочетание фиктивных переменных. Сезон, для которого значения всех фиктивных переменных равны нулю, принимается за базу сравнения. Для остальных сезонов одна из фиктивных переменных принимает значение, равное единице.

Если имеются поквартальные данные, то значения фиктивных переменных $d_2, d_3, d_4$ будут принимать следующие значения для каждого из кварталов (табл. 6).

<table>
<thead>
<tr>
<th>Квартал</th>
<th>$d_2$</th>
<th>$d_3$</th>
<th>$d_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Общий вид регрессионной модели с переменной структурой в данном случае будет иметь вид:

$$ y_t = \beta_0 + \beta_1 \cdot t + c_1 \cdot d_2 + c_2 \cdot d_3 + c_3 \cdot d_4 + \epsilon_t. $$
Построенная модель регрессии является разновидностью аддитивной модели временного ряда. Базисным уравнением исследуемой регрессионной зависимости будет являться уравнение тренда для первого квартала:

$$y_t = \beta_0 + \beta_1 \cdot t + \epsilon_t.$$

Частными случаями регрессионной зависимости будут являться уравнения трендов для остальных кварталов:

$$y_t = \beta_0 + \beta_1 \cdot t + c_2 + \epsilon_t,$$

$$y_t = \beta_0 + \beta_1 \cdot t + c_3 + \epsilon_t,$$

$$y_t = \beta_0 + \beta_1 \cdot t + c_4 + \epsilon_t.$$

Частные регрессионные уравнения отличаются друг от друга только на величину свободного члена уравнения регрессии $c_i$. Коэффициент $\beta_1$ в данной регрессионной зависимости характеризует среднее абсолютное изменение уровней динамического ряда под влиянием основной тенденции.

Оценку сезонной компоненты для каждого сезона можно рассчитать, как разность между средним значением свободных членов всех частных регрессионных уравнений и значением постоянного члена одного из уравнений.

Среднее значение свободных членов всех регрессионных уравнений рассчитывается по формуле:

$$\bar{\beta}_0 = \frac{\beta_0 + (\beta_0 + c_2) + (\beta_0 + c_3) + (\beta_0 + c_4)}{4}.$$  

(27)

Оценки сезонных отклонений в случае поквартальных данных могут быть рассчитаны следующим образом (табл. 7):

<table>
<thead>
<tr>
<th>№</th>
<th>Оценка компоненты</th>
<th>Квартал</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>$\beta_0 - \bar{\beta}_0$</td>
<td>для первого квартала</td>
</tr>
<tr>
<td>2.</td>
<td>$\beta_0 + c_2 - \bar{\beta}_0$</td>
<td>для второго квартала</td>
</tr>
<tr>
<td>3.</td>
<td>$\beta_0 + c_3 - \bar{\beta}_0$</td>
<td>для третьего квартала</td>
</tr>
<tr>
<td>4.</td>
<td>$\beta_0 + c_4 - \bar{\beta}_0$</td>
<td>для четвертого квартала</td>
</tr>
</tbody>
</table>

При пример. Рассмотрим временной ряд оценка среднего сложившегося уровня торговой наценки организации в текущем квартале (в % к стоимости проданных товаров) (процент), таблица 8.
Таблица 8

Оценка среднего сложившегося уровня торговой наценки

<table>
<thead>
<tr>
<th>Квартал</th>
<th>Уровень торговой наценки</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 квартал 2011</td>
<td>18.66</td>
</tr>
<tr>
<td>2 квартал 2011</td>
<td>19.6</td>
</tr>
<tr>
<td>3 квартал 2011</td>
<td>18.99</td>
</tr>
<tr>
<td>4 квартал 2011</td>
<td>18.54</td>
</tr>
<tr>
<td>1 квартал 2012</td>
<td>21.97</td>
</tr>
<tr>
<td>2 квартал 2012</td>
<td>21.61</td>
</tr>
<tr>
<td>2 квартал 2013</td>
<td>22.4</td>
</tr>
<tr>
<td>3 квартал 2013</td>
<td>21.8</td>
</tr>
<tr>
<td>4 квартал 2013</td>
<td>21.6</td>
</tr>
<tr>
<td>1 квартал 2014</td>
<td>23.9</td>
</tr>
<tr>
<td>2 квартал 2014</td>
<td>23</td>
</tr>
<tr>
<td>3 квартал 2014</td>
<td>23.3</td>
</tr>
<tr>
<td>4 квартал 2014</td>
<td>23</td>
</tr>
<tr>
<td>1 квартал 2015</td>
<td>24.7</td>
</tr>
<tr>
<td>2 квартал 2015</td>
<td>23.5</td>
</tr>
<tr>
<td>3 квартал 2015</td>
<td>22.9</td>
</tr>
<tr>
<td>4 квартал 2015</td>
<td>22.1</td>
</tr>
</tbody>
</table>

График этого ряда (рис. 5) показывает наличие сезонных колебаний.

Рис. 5. Исходные данные

Применим к нему метод фиктивных переменных. Так как данные являются поквартальными, то необходимо ввести три фиктивных переменных. (табл. 9)
Таблица 9

Фиктивные данные

<table>
<thead>
<tr>
<th>Тренд</th>
<th>$d_2$</th>
<th>$d_3$</th>
<th>$d_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

После оценки уравнения $x_t = \beta_0 + \beta_1 \cdot t + c_2 d_2 + c_3 d_3 + c_4 d_4 + \epsilon_t$, получим (табл. 10).

Таблица 10

Результаты оценки

<table>
<thead>
<tr>
<th>Перемен.</th>
<th>Коэффициент</th>
<th>Ошибки</th>
<th>$t$-Stat.</th>
<th>Вероятность</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>19.52960</td>
<td>0.737264</td>
<td>26.48928</td>
<td>0.0000</td>
</tr>
<tr>
<td>T</td>
<td>0.294262</td>
<td>0.062026</td>
<td>4.744175</td>
<td>0.0005</td>
</tr>
<tr>
<td>S2</td>
<td>-0.475094</td>
<td>0.781592</td>
<td>-0.607854</td>
<td>0.5546</td>
</tr>
<tr>
<td>S3</td>
<td>-0.620367</td>
<td>0.863074</td>
<td>-0.718787</td>
<td>0.4860</td>
</tr>
<tr>
<td>S4</td>
<td>-0.564629</td>
<td>0.860204</td>
<td>-0.656390</td>
<td>0.5240</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.665702</td>
<td>Средняя ряда $x_t$</td>
<td>21.85706</td>
<td></td>
</tr>
<tr>
<td>Скорр. $R^2$</td>
<td>0.554269</td>
<td>S.D. $x_t$</td>
<td>1.867032</td>
<td></td>
</tr>
<tr>
<td>S.E. регрессии</td>
<td>1.246488</td>
<td>Крит. Акайка</td>
<td>3.518465</td>
<td></td>
</tr>
<tr>
<td>Сумма квадратов $e_t$</td>
<td>18.64478</td>
<td>Крит. Шварца</td>
<td>3.763528</td>
<td></td>
</tr>
<tr>
<td>Ln функции правд.</td>
<td>-24.90696</td>
<td>Крит. Hannan-Quinn</td>
<td>3.542825</td>
<td></td>
</tr>
<tr>
<td>$F$-стат</td>
<td>5.974032</td>
<td>DW критерий</td>
<td>1.206635</td>
<td></td>
</tr>
</tbody>
</table>

Мы видим, что $R^2 = 0.6657$, отсутствует автокорреляция и остатки образуют белый шум.
Уравнение примет вид (28):

\[ x_t = 19.52960 + 0.294262469 \cdot t - 0.4750937 \cdot d_2 - 0.6203667 \cdot d_3 - 0.56462917 \cdot d_4. \]  (28)

Найдем теперь среднее значение свободных членов по формуле (27)

\[ \bar{\beta}_0 = \frac{(19.5296 + (19.5296 - 0.47509) + (19.5296 - 0.62036) + (19.5296 - 0.564629))}{4} = 19.1146. \]  (29)

По формуле (7) найдем сезонные компоненты \( d_1 = 0.215968, d_2 = 0.0122223, d_3 = -0.148206, d_4 = -0.0799837. \)

Рис. 6. Исходный ряд и ряд без сезонных компонент

Мы видим (рис. 6), что сезонные компоненты сглажены.

### 2.8. Сглаживание временного ряда. Метод скользящих средних

Одним из альтернативных по отношению к функциональному описанию тренда вариантов сглаживания временного ряда \( x_t \) является метод скользящих или, как еще говорят, подвижных средних. Метод скользящих средних

Суть метода заключается в замене исходного временного ряда последовательностью средних, вычисляемых на отрезке, который перемещается вдоль временного ряда, как бы скользит по нему. Задается длина отрезка скольжения \( 2m + 1 \) по временной оси, т. е. берется нечетное число наблюдений. Подбирается полином \( p \)-го порядка

\[ P_t = \sum_{k=1}^{p} a_k t^k \]  (30)

к группе первых \((2m + 1)\) членов ряда, и этот полином используется для определения значения тренда в средней \((m + 1)\)-й точке группы. Затем производится сдвиг на один уровень ряда вперед и подбирается полином того же порядка к группе точек, состоящей из 2-го, 3-го, ..., \((2m+3)\)-го наблюдения. Находится значение тренда в \((m + 2)\)-й точке и т. д. тем же способом вдоль всего ряда до последней группы из \((2m + 1)\) наблюдения. В действительности нет необходимости строить полином для каждого отрезка. Как будет показано, эта процедура эквивалентна нахождению линейной комбинации уровней временного ряда с коэффициентами, которые могут быть определены раз и навсегда и зависят только от длины отрезка скольжения и степени полинома.
Для определения коэффициентов \( a_0, a_1, \ldots, a_p \) полинома (30) с помощью МНК по первым \((2m + 1)\) точкам минимизируется функционал:

\[
f = \sum_{t=-m}^{m} (x_t - a_0 - a_1 t - \cdots - a_p t^p)^2.
\]  

(31)

Заметим, что \( t \) принимает условные значения от \(-m\) до \( m\). Это весьма удобный прием, существенно упрощающий расчеты. Дифференцирование функционала по \( a_0, a_1, \ldots, a_p \) дает систему из \( p + 1 \) уравнения типа:

\[
a_0 \sum_{t=-m}^{m} t^j + a_1 \sum_{t=-m}^{m} t^{j+1} + a_2 \sum_{t=-m}^{m} t^{j+2} + \cdots + a_p \sum_{t=-m}^{m} t^{j+p} = a_0 \sum_{t=-m}^{m} x_t t^j, \quad j = 0, 1, \ldots, p.
\]  

(32)

Решение этой системы уравнений относительно неизвестных параметров \( a_0, a_1, \ldots, a_p \) \((j = 0, 1, \ldots, p)\) облегчается тем, что все суммы \( \sum_{t=-m}^{m} t^j \) при нечетных \( j \) равны нулю. Кроме того, т. к. полином, подобранный по \( 2m + 1 \) точкам, используется для определения значения тренда в средней точке, а в этой точке \( t = 0 \), то, положив в уравнении (30) \( t = 0 \), получаем значение тренда, равное \( a_0 \). Стало быть, задача сглаживания временного ряда сводится к поиску \( a_0 \).

Система нормальных уравнений (32), которую нужно разрешить относительно \( a_0 \), разбивается на две подсистемы: одну – содержащую коэффициенты с четными индексами \( a_0, a_2, a_4, \ldots \), другую – включающую коэффициенты с нечетными индексами \( a_1, a_3, a_5, \ldots \). Решение системы относительно \( a_0 \) зависит от численных значений \( \sum_{t=-m}^{m} t^j \) и линейных функций от \( x_t \) типа \( \sum_{t=-m}^{m} x_t t^j \).

В итоге значением тренда в центральной точке отрезка будет средняя арифметическая, взвешенная из значений временного ряда от \( -m \) до \( m \) с весовыми коэффициентами \( b_t \), которые зависят от значений \( m \) и \( p \):

\[
a_0 = \sum_{t=-m}^{m} b_t x_t.
\]

Указанная формула применяется для всех последующих отрезков скольжения, с вычислением значений тренда в их средних точках.

Пример. Продемонстрируем рассматриваемый метод на примере полинома второй степени \((p = 2)\) и длины отрезка скольжения, равной пяти точкам \((m = 2)\). Здесь надо свести к минимуму сумму:

\[
\sum_{t=-2}^{2} (x_t - a_0 - a_1 t - a_2 t^2)^2.
\]

Получается система уравнений:

\[
\begin{align*}
\sum_{t=-2}^{2} a_0 + a_1 \sum_{t=-2}^{2} t + a_2 \sum_{t=-2}^{2} t^2 &= \sum_{t=-2}^{2} x_t, \\
a_0 \sum_{t=-2}^{2} t + a_1 \sum_{t=-2}^{2} t^2 + a_2 \sum_{t=-2}^{2} t^4 &= \sum_{t=-2}^{2} x_t t, \\
a_0 \sum_{t=-2}^{2} t^2 + a_1 \sum_{t=-2}^{2} t^3 + a_2 \sum_{t=-2}^{2} t^4 &= \sum_{t=-2}^{2} x_t t^2.
\end{align*}
\]
Вычисляя суммы при \( a_i \), получим систему:

\[
\begin{aligned}
5a_0 + 10a_2 &= \sum_{t=-2}^{2} x_t, \\
10a_1 &= \sum_{t=-2}^{2} x_t t, \\
10a_0 + 34a_2 &= \sum_{t=-2}^{2} x_t t^2.
\end{aligned}
\]

Решение этой системы относительно \( a_0 \) дает следующий результат:

\[
a_0 = \frac{1}{35} \left( 17 \sum_{t=-2}^{2} x_t - 5 \sum_{t=-2}^{2} x_t t^2 \right) = \frac{1}{35} (-3x_{-2} + 12x_{-1} + 17x_0 + 12x_1 - 3x_2).
\]

Решая эту систему относительно \( a_1 \) и \( a_2 \), получим:

\[
\begin{aligned}
a_1 &= -0.2x_{-2} - 0.1x_{-1} + 0.1x_1 - 0.2x_2, \\
a_2 &= \frac{1}{7} x_{-2} - \frac{1}{14} x_{-1} - \frac{1}{7} x_0 - \frac{1}{14} x_1 + \frac{1}{7} x_2.
\end{aligned}
\]

**Свойства скользящих средних**

1. Сумма весов \( b_t \) в формуле \( a_0 = \sum_{t=-m}^{m} b_t x_t \) равна единице.

2. Веса симметричны относительно нулевого значения \( t \), т. е. \( b_t = b_{-t} \).

3. Для полиномов четного порядка \( p = 2k \) формулы расчета \( a_0 \) будут теми же самыми, что и для полиномов нечетного порядка \( p = 2k + 1 \).

4. Оценки параметров \( a_1, \ldots, a_p \) тоже выражены в виде линейной комбинации уровней временного ряда, входящих в отрезок, но весовые коэффициенты в этих формулах в сумме равны нулю и не симметричны.

Естественным образом возникает вопрос, какой степени полином следует выбирать и какой должна быть длина отрезка скольжения. Закономерность такова: чем выше степень полинома и короче отрезок скольжения, тем ближе расчетные значения к первоначальным данным. При этом помимо тенденции могут воспроизводиться и случайные колебания, нарушающие ее смысл. И наоборот, чем ниже степень полинома и чем длиннее отрезок скольжения, тем более гладкой является сглаживающая кривая, тем в большей мере она отвечает свойствам тенденции, хотя ошибка аппроксимации будет при этом выше.

В принципе, если ставится задача выявления тренда, то, с учетом особенностей покомпонентного разложения временного ряда, следует ориентироваться не на минимальную остаточную дисперсию, а на стационарность остатков, получающихся после исключения тренда.

**Пример.** Рассмотрим некоторый временной ряд

\[ x_t = -4, -1, -5, -3, -2, -4, 0, -6, -2, 3, -1, 0, 8, 3, 7, 10, 12, 16, 20, 19. \]

Выполним сглаживание его с помощью скользящей средней с параметрами \( m = 2 \) и \( p = 2 \). Возьмем первые пять элементов ряда и вычислим взвешенную среднюю по формуле \( \frac{1}{35} (-3x_{-2} + +12x_{-1} + 17x_0 + 12x_1 - 3x_2) \). Получим

\[
y_3 = \frac{1}{35} (-3(-4) + 12(-1) + 17(-5) + 12(-3) - 3(-2)) = -3.2857.
\]

32
Полученное число –3.2857 будет 3-м элементом \( y_3 \) сглаженного ряда.

Теперь рассмотрим элементы со 2-го по 6-й ряды \( x_t \). Пользуясь теми же формулами вычислим взвешенную среднюю \( y_t = -3.4286 \) и т. д. Получим сглаженный ряд

\[
y_t = -3.2857, -3.4286, -2.9429, -1.8571, -3.0857, -3.5143, -1.9143, 0.9429, 0.0286, 1.8857, 4.40, 5.7429, 6.1429, 9.7429, 12.4286, 16.2571.
\]

Его длина укоротилась на \( 2m \) членов по сравнению с исходным за счет первых \( m \) и последних \( m \).

Чтобы вычислить первые 2 значения сглаженного ряда, вычислим коэффициенты \( a_0, a_1, a_2 \) полинома по первым пяти значениям ряда \( x_t \). Получим \( a_0 = -3.2857, a_1 = -0.2(-4) - 0.1(-1) + +0.1(-3) + 0.2(-2) = 0.2, a_2 = \frac{1}{7}(-4) - \frac{1}{14}(-1) - \frac{1}{7}(-5) - \frac{1}{14}(-3) + \frac{1}{7}(-2) = 0.1429. \)

Получим многочлен \( f(t) = -3.2857 + 0.2t + 0.1429t^2 \), который аппроксимирует первые пять значений ряда. Чтобы получить первые элементы сглаженного ряда необходимо подставить вместо \( T \) значения \(-2 \) и \(-1 \). Имеем \( x_1 = f(-2) = -3.2857 + 0.2 \cdot (-2) + 0.1429 \cdot (-2)^2 = = -3.1141, x_2 = f(-1) = -3.2857 + 0.2 \cdot (-1) + 0.1429 \cdot (-2)^2 = -3.3428. \) Проводя аналогичные вычисления, получим \( x_9 = 17.3599, x_{20} = 21.39929. \) На рисунке 7 изображены исходный ряд \( X \) и сглаженный ряд \( Y \).

![Рис. 7. Исходный и сглаженный ряды](image)

На рисунке можно увидеть наличие полиномиального тренда.

### 2.9. Экспоненциальное сглаживание

Кроме метода скользящей средней как способа фильтрации временного ряда известностью пользуется экспоненциальное сглаживание, в основе которого лежит расчет экспоненциальных средних. Экспоненциальная средняя рассчитывается по рекуррентной формуле:

\[
u_t = ax_t + bu_{t-1}, \tag{33}
\]

где \( u_t \) – значение экспоненциальной средней в момент \( t \),

\( a \) – параметр сглаживания (вес последнего наблюдения), \( 0 < a < 1 \),

\( b = 1 - a \).
Экспоненциальную среднюю, используя рекуррентность формулы (33), можно выразить через значения временного ряда:

\[ u_t = a x_t + b(ax_{t-1} + bu_{t-2}) = ax_t + abx_{t-1} + b^2u_{t-2} = \ldots = \]

\[ = ax_t + abx_{t-1} + ab^2x_{t-2} + \cdots + ab^{t-1}x_1 + b^t u_0 = a \sum_{j=0}^{t-1} b^j x_{t-j} + b^t u_0, \tag{34} \]

где \( t \) – количество уровней ряда, \( u_0 \) – некоторая величина, характеризующая начальные условия для первого применения формулы (33) при \( t = 1 \). В качестве \( u_0 \) можно использовать первое значение временного ряда, т. э. \( x_1 \).

Так как \( b < 1 \), то при \( t \to \infty \) величина \( b^t \to 0 \), а сумма коэффициентов \( a \sum_{j=0}^{t-1} b^j \to 1 \).

Действительно,

\[ a \sum_{j=0}^{t-1} b^j = a \frac{1}{1 - b} = (1 - b) \frac{1}{1 - b} = 1. \]

Тогда последним слагаемым в формуле (34) можно пренебречь и

\[ u_t = a \sum_{j=0}^{\infty} b^j x_{t-j} = a \sum_{j=0}^{\infty} (1 - a)^j x_{t-j}. \]

Таким образом, величина \( u_t \) оказывается взвешенной суммой всех уровней ряда, причем веса уменьшаются экспоненциально, по мере углубления в историю процесса, отсюда название – экспоненциальная средняя.

Несложно показать, что экспоненциальная средняя имеет то же математическое ожидание, что и исходный временной ряд, но меньшую дисперсию (покажите, что дисперсия слаженного ряда \( \sigma_u^2 \) выражается через дисперсию \( \sigma^2 \) ряда \( x_t \) по формуле \( \sigma_u^2 = a\sigma^2/(2 - a) \)). Что касается параметра сглаживания \( a \), то, чем ближе \( a \) к единице, тем менее ощутимо расхождение между слаженным рядом и исходным. И наоборот, чем меньше \( a \), тем в большей степени подавляются случайные колебания ряда и отчетливее вырисовывается его тенденция.

Экспоненциальное сглаживание можно представить в виде фильтра, на вход которого поступают значения исходного временного ряда, а на выходе формируется экспоненциальная средняя.

Пример. Рассмотрим

\[ x_t = -9, -11, -9, -10, -8, -8, -8, -5, -7, -5, -5, -2, -3, -3, -1, -2, 3, 2, 6. \]

Применим экспоненциальное сглаживание к ряду \( x_t \). Для этого воспользуемся равенством (33), где в качестве \( a \) возьмем числа 0.4 и 0.6. А в качестве \( u_1 \) первый член ряда \( x_t \). Начнем

\[ u_2 = 0.4 \cdot (-11) + 0.6 \cdot (-9) = -9.8, \quad u_3 = 0.4 \cdot (-9) + 0.6 \cdot (-9.8) = -9.48 \quad \text{т. д.} \]

В результате получим слаженный ряд


34
Сгладим ряд $a_t$ при $a = 0.6$. Получим

$u'_t = -9, -10.2, -9.48, -9.792, -8.7168, -8.286720, -8.114688, -6.245875, -6.698350,$

$-5.679340, -5.271736, -5.108694, -3.243478, -3.097391, -3.038956, -1.815583, -1.926233,$

$1.029507, 1.611803, 4.244721.$

Полученные ряды изображены на рисунке 8.

Рис. 8. Экспоненциальное сглаживание

Видно, чем ближе $a$ к единице, тем менее ощутимо расхождение между сглаженным рядом и исходным.

2.10. Критерии, используемые в анализе временных рядов

В анализе временных рядов наиболее разработанными критериями являются критерии случайности, которые призваны определить, является ли ряд чисто случайным либо в его поведении проявляются определенные закономерности, которые позволяют делать предсказания.

«Чисто случайный ряд» — это в данном случае неформальный термин, подчеркивающий отсутствие закономерностей. Здесь может, например, подразумеваться ряд, состоящий из независимых и одинаково распределенных наблюдений (что соответствует понятию выборки в обычной статистике), либо белый шум, в том смысле, который указан ранее. Среди экономических временных рядов редко встречаются такие, которые подходят под это описание. Типичный экономический ряд характеризуется сильной положительной корреляцией. Очень часто экономические ряды содержат тенденцию, сезонность и т. д. В связи с этим применение критериев случайности по прямому назначению не имеет особого смысла. Тем не менее критерии случайности играют очень важную роль в анализе временных рядов, и существуют различные способы их использования:

1. Критерий может быть чувствительным к определенным отклонениям от «случайности». Тогда большое значение соответствующей статистики может указывать на наличие именно такого отклонения. Таким образом, статистика критерия может использоваться просто как описательная статистика. При этом формальная проверка гипотезы не производится. Так, например, автокорреляционная функция, о которой речь пойдет ниже, очень чувствительна к наличию периодичностей и трендов. Кроме того, по автокорреляционной функции можно определить, насколько быстро затухает временная зависимость в рядах.
2. Критерий можно применять к остаткам от модели, а не к самому исходному ряду. Пусть, например, была оценена модель вида «тренд плюс шум». После вычитания из ряда выявленного тренда получаются остатки, которые можно рассматривать как оценки случайной компоненты. Наличие в остатках каких-либо закономерностей свидетельствует о том, что модель неполна, либо в принципе некорректна. Поэтому критерии случайности можно использовать в качестве диагностических критериев при моделировании.

Следует помнить, однако, что распределение статистики, рассчитанной по остаткам, и распределение статистики, рассчитанной по исходному случайному шуму, вообще говоря, не совпадают. В некоторых случаях при большом количестве наблюдений это различие несущественно, но часто в результате критерий становится несостоятельным и критические значения в исходном виде применять нельзя. Существует большое количество различных критериев случайности. По-видимому, наиболее популярными являются критерии, основанные на автокорреляционной функции.

2.11. Критерии, основанные на автокорреляционной функции

Для того чтобы сконструировать критерий, следует рассмотреть, какими статистическими свойствами характеризуется автокорреляционная функция стационарного процесса. Известно, что выборочные автокорреляции имеют нормальное асимптотическое распределение. При большом количестве наблюдений математическое ожидание $r(k)$ приближенно равно $\rho(k)$. Дисперсия автокорреляции приближенно равна

$$D(r(k)) \approx \frac{1}{n} \sum_{i=-\infty}^{+\infty} [\rho^2(i) + \rho(i + k) - 4\rho(k)\rho(i)\rho(i + k) + 2\rho^2(k)\rho^2(i)].$$

(35)

Для ковариации двух коэффициентов автокорреляции верно приближение

$$cov(r(k), r(l)) \approx \frac{1}{n} \sum_{i=-\infty}^{+\infty} [\rho(i + k)\rho(i + l) + \rho(i - k)\rho(i + l) - 2\rho(k)\rho(i)\rho(i + l) - 2\rho(l)\rho(i)\rho(i + k) + 2\rho(k)\rho(l)\rho^2(i)].$$

(36)

Эти аппроксимации были выведены Бартлеттом.

В частности, для белого шума (учитывая, что $\rho(k) = 0$ при $k \neq 0$) получаем согласно формуле (35)

$$D(r(k)) = \frac{1}{n}$$

(37)

Это только грубое приближение для дисперсии. Для гауссовского белого шума известна точная формула для дисперсии коэффициента автокорреляции:

$$D(r(k)) = \frac{n - k}{n(n + 2)}.$$  

(38)

4 Так, Q-статистика, о которой идет речь ниже, в случае остатков модели ARMA(p, q) будет распределена не как $\chi^2_n$, а как $\chi^2_{n-p-q}$. Применение распределения $\chi^2_n$ приводит к тому, что нулевая гипотеза о "случайности" принимается слишком часто.
Кроме того, из приближенной формулы (36) следует, что автокорреляции \( r(k) \) и \( r(l) \), соответствующие разным порядкам \( k \neq l \), некоррелированы. Эти формулы позволяют проверять гипотезы относительно автокорреляционных коэффициентов. Так, в предположении, что ряд представляет собой белый шум, можно использовать следующий доверительный интервал для отдельного коэффициента автокорреляции:

\[
\left[ r(k) - \frac{\sqrt{n-k}}{\sqrt{n(n+2)}} \varepsilon_{1-\alpha}, r(k) + \frac{\sqrt{n-k}}{\sqrt{n(n+2)}} \varepsilon_{1-\alpha} \right],
\]

(39) где \( \varepsilon_{1-\alpha} \) – квантиль нормального распределения. При больших \( n \) и малых \( k \) оправдано использование более простой формулы

\[
\left[ r(k) - \frac{\varepsilon_{1-\alpha}}{\sqrt{n}}, r(k) + \frac{\varepsilon_{1-\alpha}}{\sqrt{n}} \right].
\]

(40)

Вместо того чтобы проверять отсутствие автокорреляции для каждого отдельного коэффициента, имеет смысл использовать критерий случайности, основанный на нескольких ближних автокорреляциях. Рассмотрим \( m \) первых автокорреляций: \( r(1), \ldots, r(m) \). В предположении, что ряд является белым шумом, при большом количестве наблюдений их совместное распределение приближенно равно \((0, 1/n)\). На основе этого приближения Дж. Бокс и Г. Пирс предложили следующую статистику, называемую Q-статистикой Бокса-Пирса

\[
Q(r) = n \sum_{k=1}^{m} r^2(k).
\]

(41)

Она имеет асимптотическое распределение \( \chi^2_m \).

Работа Бокса и Пирса была опубликована в 1970 г. В том же году Дарбин опубликовал в другом журнале свою работу, в которой показал, что статистика Дарбина-Уотсона не применима для проверки автокорреляции остатков при наличии в качестве регрессора объясняемой переменной с лагом. Одновременно Дарбин предложил так называемую \( h \)-статистику и альтернативную процедуру Дарбина для проверки наличия автокорреляции в таких моделях. Эти работы появились в разных журналах примерно в одно и то же время, и каждая породила последователей. Не сразу было замечено, что в моделях с лаговой объясняемой переменной в качестве регрессора статистика Бокса-Пирса не применима по тем же причинам, что и статистика Дарбина-Уотсона. Поскольку в моделях временных рядов проверка наличия автокорреляции критически важна, статистика Бокса-Пирса нашла широкое применение и до сих пор входит в состав большинства специализированных эконометрических пакетов. Поскольку обнаружилось, что статистика Бокса-Пирса имеет малую мощность, Бокс и Льонг в 1978 г. предложили использовать для тех же целей улучшенную \( Q \)-статистику Бокса-Льонга, которую, как и статистику Бокса-Пирса, часто называют портманто-статистикой (portmanteau-statistics). По сравнению со статистикой Бокса-Пирса различным слагаемым приданы разные веса.

\[
Q(r) = n(n+2) \sum_{k=1}^{m} \frac{r^2(k)}{n-k}.
\]

(42)

Авторы показали, что эта статистика имеет то же асимптотическое распределение \( \chi^2 \), но лучше им аппроксимируется при конечном числе наблюдений. К сожалению, и статистика Бокса-Льонга теоретически не применима для тестирования автокорреляции остатков в моделях ARMA по тем же причинам, что и статистики Дарбина-Уотсона и Бокса-Пирса. Тем не менее статистика Бокса-Льонга входит во все специализированные пакеты, множество исследователей ею пользуются, хотя она теоретически не состоятельна.
Нулевая гипотеза в Q-критерии заключается в том, что ряд представляет собой белый шум, то есть является чисто случайным процессом. Используется стандартная процедура проверки: если расчетное значение Q-статистики больше заданного квантиля распределения $\chi^2_m$, то нулевая гипотеза отвергается и признается наличие автокорреляции до $m$-го порядка в исследуемом ряду.

Для проверки наличия автокорреляции в моделях ARMA лучше воспользоваться более мощным и универсальным способом, а именно методом множителей Лагранжа (Lagrange multiplier – LM), применительно к проверке автокорреляции остатков его еще называют тестом Бреуша-Годфри (Breusch-Godfrey). Он входит в «триаду» классических асимптотических тестов: отношения правдоподобий, Вальда, множителей Лагранжа и применим для широкого класса задач проверки ограничений на коэффициенты модели. Сформулируем его применительно к анализу автокорреляции остатков. Пусть рассматривается модель множественной регрессии

$$x_t = b_0 + b_1 x_1 + \ldots + b_k x_k + e_t,$$

где $x_1, \ldots, x_k$ – разные регрессоры, в том числе, возможно, и лаговые значения как объясняющих, так и объясняемой переменных. Проверим предположение, что $e_t$ подчиняется авторегрессионной схеме порядка $p$, т. е. задается уравнением

$$e_t = \gamma_1 e_{t-1} + \ldots + \gamma_p e_{t-p} + u_t,$$

где $u_t$ – белый шум. В терминах коэффициентов модели основная и альтернативная гипотезы принимают вид:

$$H_0: \gamma_1 = \ldots = \gamma_p = 0,$$

$$H_1: \gamma_1^2 + \ldots + -\gamma_p^2 > 0.$$

Метод множителей Лагранжа для проверки этой гипотезы заключается в следующем. Методом МНК строится обычная регрессия вида

$$x_t = b_0 + b_1 x_1 + \ldots + b_k x_k + e_r.$$

Обозначим ее остатки через $\varepsilon_t$

1. Строится регрессия либо той же объясняемой переменной $x_o$, либо остатков $e_t$ на старые регрессоры и остатки с лагом до $p$ включительно (то есть в качестве дополнительных объясняющих переменных используем $\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots, \varepsilon_{t-p}$).
2. Проверяется гипотеза о том, что группа дополнительных переменных является излишней. Если в качестве объясняемой переменной используются остатки, то статистика $nR^2$, где $n$ – число наблюдений, а $R^2$ – коэффициент множественной детерминации регрессии из пункта 1, имеет асимптотическое (при увеличении числа наблюдений) распределение $c\chi^2$ с $p$ степенями свободы.

Для проверки гипотезы о равенстве нулю группы переменных мы привыкли использовать $F$-статистику, проверяющую, фактически, статистическую значимость уменьшения остаточной суммы квадратов от включения дополнительных переменных. Разумеется, $F$-статистика применима и в этом случае. Но, напротив, она применима только при нормальном распределении случайного члена. Применение же теста множителей Лагранжа не требует нормальности распределения, но (работает) только асимптотически. Современные эмпирические пакеты, в частности Eviews, сообщают пользователю при применении LM-теста значения и критические вероятности ($p$-values) для обеих статистик, так что вы можете выбирать: использовать ли $F$-отношение, верное для конечных выборок, но в предположении нормальности, либо не требовать нормальности и использовать статистику $nR^2$, верную лишь асимптотически.

Для проверки автокорреляции в остатках широко используется критерий Дарбина-Уотсона (DW). Он основывается на расчете величины

$$d = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}. \quad (43)$$

38
Таким образом, \( d \) – это отношение суммы квадратов разностей последовательных значений остатков к остаточной сумме квадратов по модели регрессии. Практически во всех статистических пакетах прикладных программ значение критерия Дарбина-Уотсона устанавливается наряду с коэффициентом детерминации, значениями \( t \)- и \( F \)-критериев. Известно, что \( d \approx 2(1 - \rho(1)), \) где \( \rho(1) \) коэффициент корреляции между рядами \( e_i \) и \( e_{i-1}. \) Если в остатках есть полная отрицательная автокорреляция, то \( \rho(1) = -1 \) и, следовательно, \( d \approx 4. \) Если \( \rho(1) = 1, \) то \( d = 0, \) если же \( \rho(1) = 0, \) то \( d = 2. \) Значит, \( 0 \leq d \leq 4. \)

Алгоритм выявления автокорреляции остатков на основе критерия Дарбина-Уотсона следующий. Выдвигается гипотеза: \( H_0 \) об отсутствии автокорреляции остатков. Альтернативные гипотезы \( H_1 \) и \( H_1^* \) состоят соответственно в наличии положительной или отрицательной автокорреляции в остатках. Далее по специальным таблицам (табл.11) определяются критические значения критерия Дарбина-Уотсона \( d_L \) и \( d_U \) для заданного числа наблюдений \( n, \) числа независимых переменных модели \( k \) и уровня значимости \( \alpha. \) По этим значениям числовой промежуток \([0; 4]\) разбивают на пять отрезков. Принятие или отклонение каждой из гипотез с вероятностью \( (1 - \alpha) \) представлено в таблице 12. Если фактическое значение критерия Дарбина-Уотсона попадает в зону неопределенности, то на практике предполагают существование автокорреляции остатков и отклоняют гипотезу \( H_0. \)

Важное замечание

В 1970 году Дарбин опубликовал свою работу, в которой показал, что статистика Дарбина-Уотсона не применима для проверки автокорреляции остатков при наличии в качестве регрессора объясняемой переменной с лагом.
Для проверки нормальности существует множество тестов. Мы рассмотрим тест Жарге-Бера (Jarque-Bera). Этот тест выглядит следующим образом. Он вычисляет выборочные значения для коэффициентов асимметрии

\[ as = \frac{1}{n\sigma^3} \sum (e_t - \overline{e})^3 \]

и эксцесса

\[ k = \frac{1}{n\sigma^4} \sum (e_t - \overline{e})^4, \]

где \( \overline{e} \) – выборочное среднее, а \( \sigma \) – выборочное среднеквадратичное остатков модели. При условии нормальности остатков статистика Жарге-Бера

\[ \frac{n - p - q - 1}{6} \left[ as^2 + \frac{1}{4} (K - 3)^2 \right] \]

имеет \( \chi^2 \) распределение с двумя степенями свободы.

Кроме критериев случайности можно строить и другие критерии на основе автокорреляций. Пусть, например, \( \rho(i) = 0 \) при \( i \leq k \), т. е. процесс автокоррелирован, но автокорреляция пропадает после порядка \( k^5 \). Тогда по формуле (35) получаем

\[ D(r(k)) \approx 1 + 2 \sum_{i=1}^{k} \rho^2(i). \]

Если в этой формуле заменить теоретические автокорреляции выборочными, то получим следующее приближение:

\[ D(r(k)) \approx 1 + 2 \sum_{i=1}^{k} r^2(i). \] (44)

На основе этого приближения (приближения Бартлетта) с учетом асимптотической нормальности можно стандартным образом построить доверительный интервал для \( r(k) \):

\[ \left[ r(k) - \varepsilon_{1-\alpha} \sqrt{D(r(k))}, r(k) + \varepsilon_{1-\alpha} \sqrt{D(r(k))} \right]. \] (45)

2.12. Лаговый оператор

Одним из основных понятий, употребляемых при моделировании временных рядов, является понятие лага. В буквальном смысле в переводе с английского лаг – запаздывание. Под лагом некоторой переменной понимают ее значение в предыдущие периоды времени. Например, для переменной \( x_t \) лагом в \( k \) периодов будет \( x_{t-k} \).

При работе с временными рядами удобно использовать лаговый оператор \( L \), т. е. оператор запаздывания, сдвигая назад во времени. Хотя часто использование этого оператора сопряжено с некоторой потерей математической строгости, однако это окупаются значительным упрощением вычислений. Если к переменной применить лаговый оператор, то в результате получится лаг этой переменной:

\[ Lx_t = x_{t-1}. \]

5 Это предположение выполнено для процессов скользящего среднего MA(q) при \( q < k \) (см. ниже).
Использование лагового оператора \( L \) обеспечивает сжатую запись разностных уравнений и помогает изучать свойства целого ряда процессов. Удобство использования лагового оператора состоит в том, что с ним можно обращаться как с обычной переменной, т. е. операторы можно преобразовывать сами по себе, без учета тех временных рядов, к которым они применяются. Основное отличие лагового оператора от обычной переменной состоит в том, что оператор должен стоять перед тем рядом, к которому применяется, т. е. нельзя переставлять местами лаговый оператор и временной ряд.

Как и для обычных переменных, существуют функции от лагового оператора, они, в свою очередь, тоже являются операторами. Простейшая функция – степенная.

По определению, для целых \( m \)
\[
L^m x_t = x_{t-m},
\]
т. е. \( L^m \), действующий на \( x_t \), означает запаздывание этой переменной на \( m \) периодов. Продолжая эту же логику, можно определить многочлен от лагового оператора, или лаговый многочлен:
\[
\alpha(L) = \sum_{i=0}^{m} \alpha_i L^i = \alpha_0 + \alpha_1 L + \cdots + \alpha_m L^m.
\]
Если применить лаговый многочлен к переменной \( x_t \), то получается
\[
\alpha(L)x_t = (\alpha_0 + \alpha_1 L + \cdots + \alpha_m L^m)x_t = \alpha_0 + \alpha_1 x_{t-1} + \cdots + \alpha_m x_{t-m}.
\]

Нетрудно проверить, что лаговые многочлены можно перемножать как обычные многочлены. Например,
\[
(\alpha_0 + \alpha_1 L)(\beta_0 + \beta_1 L) = \alpha_0 \beta_0 + (\alpha_1 \beta_0 + \alpha_0 \beta_1)L + \alpha_1 \beta_1 L^2.
\]
При \( m \to \infty \) получается бесконечный степенной ряд от лагового оператора:
\[
\sum_{i=0}^{\infty} \alpha_i L^i = (\alpha_0 + \alpha_1 L + \alpha_2 L^2 + \cdots)x_t = \alpha_0 x_t + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \cdots = \sum_{i=0}^{\infty} \alpha_i x_{t-i}.
\]
Полезно помнить следующие свойства лаговых операторов:
1. Лаг константы есть константа: \( LC = C \).
2. Дистрибутивность: \( (L^i + L^j)x_t = L^i x_t + L^j x_t = x_{t-i} + x_{t-j} \).
3. Ассоциативность: \( L^i L^j x_t = L^j (L^i x_t) = L^i x_{t-j} = x_{t-i-j} \). Заметим, что: \( L^0 x_t = x_t \), т. е. \( L^0 = 1 \).
4. L, возведенный в отрицательную степень, – опережающий оператор:
\[
L^{-i}x_t = x_{t+i}.
\]
5. При \(|\alpha| < 1\) бесконечная сумма
\[
(1 + \alpha L + \alpha^2 L^2 + \alpha^3 L^3 + \cdots)x_t = (1 - L)^{-1} x_t
\]
(бесконечная геометрическая прогрессия).

Кроме лагового оператора в теории временных рядов широко используют разностный оператор \( \Delta \), который определяется следующим образом:
\[
\Delta = 1 - L,
\]
так что \( \Delta x_t = (1-L)x_t = x_t - x_{t-1} \).
Разностный оператор превращает исходный ряд в ряд первых разностей. Ряд d-го порядка (разностей d-го порядка) получается как степень разностного оператора, то есть применением разностного оператора d раз. При d = 2 получается \( \Delta^2 = (1- L)^2 = 1 - 2L + L^2 \), поэтому \( \Delta^2 x_t = (1 - 2L + L^2)x_t = x_t - 2x_{t-1} + x_{t-2} \). Для произвольного порядка d следует использовать формулу бинома Ньютонана:

\[
\Delta^d = (1-L)^d = \sum_{k=0}^{d} (-1)^k C_d^k L^k, \text{где } C_d^k = \frac{d!}{k! (d-k)!}
\]

(46)

так что

\[
\Delta^d x_t = (1-L)^d x_t = \sum_{k=0}^{d} (-1)^k C_d^k x_{t-k}.
\]

2.13. Модели с распределенным лагом. Лаги Алмон

При исследовании экономических процессов нередко приходится моделировать ситуации, когда значение результирующего признака в текущий момент времени T формируется под воздействием ряда факторов, действовавших в прошлые моменты времени t – 1, t – 2, ..., t – l. Например, на выручку от реализации или прибыль компании текущего периода могут оказывать влияние расходы на рекламу или проведение маркетинговых исследований, сделанные компанией в предшествующие моменты времени. Величину l, характеризующую запаздывание в воздействии фактора на результат, называют в эконометрике лагом, а временные ряды таких факторных переменных, сдвинутые на один или более моментов времени, лаговыми переменными. Разработка экономической политики как на макро, так и на микроуровне требует решения обратного типа задач, т. е. задач, определяющих, какое воздействие окажут значения управляемых переменных текущего периода на будущие значения экономических показателей. Например, как повлияет инвестиции в промышленность на валовую добавленную стоимость этой отрасли экономики будущих периодов или как может измениться объем ВВП, произведенного в периоде (t + l), под воздействием увеличения денежной массы в периоде t? Эконометрическое моделирование охарактеризованных выше процессов осуществляется с применением моделей, содержащих не только текущие, но и лаговые значения факторных переменных. Эти модели называются моделями с распределенным лагом. Модель вида

\[ y_t = a + b_0 \cdot t + b_1 \cdot x_{t-1} + b_2 \cdot x_{t-2} + e_t \]

(47)

является примером модели с распределенным лагом.

Рассмотрим модель с распределенным лагом в ее общем виде в предположении, что максимальная величина лага конечна:

\[ y_t = a + b_0 \cdot t + b_1 \cdot x_{t-1} + \ldots + b_l \cdot x_{t-l} + e_t. \]

(48)

Эта модель говорит о том, что если в некоторый момент времени t происходит изменение независимой переменной x, то это изменение будет влиять на значения переменной y в течение p следующих моментов времени. Коэффициент регрессии b0 при переменной xт характеризует среднее абсолютное изменение y, при изменении xт на 1 ед. своего измерения в некоторый фиксированный момент времени t, без учета воздействия лаговых значений фактора x. Этот коэффициент называют краткосрочным мультипликатором.
В момент \((t + 1)\) совокупное воздействие факторной переменной \(x_t\) на результат \(y_t\) составит \((b_0 + b_1)\) усл. ед., в момент \((t + 2)\) это воздействие можно охарактеризовать суммой \((b_0 + b_1 + b_2)\) и т. д. Полученные таким образом суммы называют промежуточными мультипликаторами.

С учетом конечной величины лага можно сказать, что изменение переменной \(x_t\) в момент \(t\) на 1 усл. ед. приведет к общему изменению результата через \(l\) моментов времени на \((b_0 + b_1 + + \cdots + b_l)\) абсолютных единиц.

Введем следующее обозначение:

\[ b_0 + b_1 + \cdots + b_l = b. \]

Величину \(b\) называют долгосрочным мультипликатором. Он показывает абсолютное изменение в долгосрочном периоде \(t + l\) изменение результата \(y\) под влиянием изменения на \(l\) ед. фактора \(x\). Предположим

\[ \beta_j = \frac{b_j}{b}, \quad y = 0 \ldots l. \]

Назовем полученные величины относительными коэффициентами относительные коэффициенты модели с распределенным лагом. Если все коэффициенты \(b_i\) имеют одинаковые знаки, то для любого \(j\)

\[ 0 < \beta_j < 1, \text{ и } \sum_{j=0}^{i} \beta_j = 1. \]

В этом случае относительные коэффициенты \(\beta_j\) являются весами для соответствующих коэффициентов \(b_j\). Каждый из них измеряет долю общего изменения результативного признака в момент времени \((t + j)\).

Зная величины \(\beta_j\), с помощью стандартных формул можно определить еще две важные характеристики модели множественной регрессии: величину среднего лага и медианного лага. Средний лаг определяется по формуле средней арифметической взвешенной:

\[ \bar{l} = \sum_{j=0}^{l} j \cdot \beta_j \]

и представляет собой средний период, в течение которого будет происходить изменение результата под воздействием изменения фактора в момент времени \(t\). Небольшая величина среднего лага свидетельствует об относительно быстром реагировании результата на изменение фактора, тогда как высокое его значение говорит о том, что воздействие фактора на результат будет сказываться в течение длительного периода времени.

Медианный лаг – это величина лага, для которого

\[ \sum_{j=0}^{M_t} \beta_j \approx 0.5. \]

Это тот период времени \(M_t\), в течение которого с момента времени \(t\) будет реализована половина общего воздействия фактора на результат. Рассмотрим пример.

Пример. Интерпретация параметров модели с распределенным лагом.

По результатам изучения зависимости объемов продаж компании в среднем за месяц от расходов на рекламу была получена следующая модель с распределенным лагом (млн руб.):

\[ y_t = -0.67 + 4.5x_t + 3.0x_{t-1} + 1.5x_{t-2} + 0.5x_{t-3}. \]
В этой модели краткосрочный мультипликатор равен 4.5. Это означает, что увеличение расходов на рекламу на 1 млн руб. ведет в среднем к росту объема продаж компании на 4.5 млн руб. в том же периоде. Под влиянием увеличения расходов на рекламу объем продаж компании возрастет в момент времени \( t \) на 1 млн руб., \((t + 1)\) – на 4.5 + 3.0 = 7.5 млн руб., \((t + 2)\) – на 7.5 + 1.5 = 9.0 млн руб. Наконец, долгосрочный мультипликатор для данной модели составит \( b = 4.5 + 3.0 + 1.5 + 0.5 = 9.5. \) В долгосрочной перспективе (например, через 3 месяца) увеличение расходов на рекламу на 1 млн руб. в настоящий момент времени приведет к общему росту объема продаж на 9.5 млн. руб. Относительные коэффициенты регрессии в этой модели равны:

\[
\beta_1 = \frac{4.5}{9.5} = 0.474; \quad \beta_2 = \frac{3.0}{9.5} = 0.316; \quad \beta_3 = \frac{1.5}{9.5} = 0.158; \quad \beta_4 = \frac{0.5}{9.5} = 0.053.
\]

Следовательно, 47.4 % общего увеличения объема продаж, вызванного ростом затрат на рекламу, происходит в текущем моменте времени; 31.6 % – в момент \((t + 1)\); 15.8 % – в момент \((t + 2)\); и только 5.3 % этого увеличения приходится на момент времени \((t + 3)\).

Средний лаг в этой модели определяется как

\[ T = 0 \cdot 0.474 + 1 \cdot 0.316 + 2 \cdot 0.158 + 3 \cdot 0.053 = 0.791 \text{ мес.} \]

Небольшая величина лага (менее 1 месяца) еще раз подтверждает, что большая часть эффекта роста затрат на рекламу проявляется сразу же.

Медианный лаг в данном примере составляет 1.08228, т. е. чуть более 1 месяца. Это значит, что половина воздействия рекламы на рост продаж реализуется в первый месяц.

Изложенные выше приемы анализа параметров модели с распределенным лагом действительны только в предположении, что все коэффициенты при текущем и лаговых значениях исследуемого фактора имеют одинаковые знаки. Это предположение вполне оправдано с экономической точки зрения: воздействие одного и того же фактора на результат должно быть однонаправленным, независимо от того, с каким временным лагом измеряется сила или теснота связи между этими признаками. Однако на практике получить статистически значимую модель, параметры которой имели бы одинаковые знаки, особенно при большой величине лага \( l \), чрезвычайно сложно. Применение обычного МНК к таким моделям в большинстве случаев затруднительно по следующим причинам.

Во-первых, текущие и лаговые значения независимой переменной, как правило, тесно связаны друг с другом. Тем самым оценка параметров модели проводится в условиях высокой многоколлинеарности факторов.

Во-вторых, при большой величине лага снижается число наблюдений, по которому строится модель, и увеличивается число ее факторных признаков. Это ведет к потере числа степеней свободы в модели.

В-третьих, в моделях с распределенным лагом часто возникает проблема автокорреляции остатков. Вышеуказанные обстоятельства приводят к значительной неопределенности относительно оценок параметров модели, снижению их точности и получению неэффективных оценок. Чистое влияние факторов на результат в таких условиях выявить невозможно. Поэтому на практике параметры моделей с распределенным лагом проводят в предположении определенных ограничений на коэффициенты регрессии и в условиях выбранной структуры лага.

Текущее и лаговые значения факторной переменной оценивают различное по силе воздействие на результативную переменную модели. Количественно сила связи между результатом и значениями факторной переменной, относящимися к различным моментам времени, измеряется с помощью коэффициентов регрессии при факторных переменных. Если построить график зависимости этих коэффициентов от величины лага, можно получить графическое изображение структуры лага, или распределения во времени воздействия факторной переменной на результат.
Рассмотрим общую модель с распределенным лагом, имеющую конечную максимальную величину лага $l$, которая описывается соотношением (48). Предположим, было установлено, что в исследуемой модели имеет место полиномиальная структура лага, т.е. зависимость коэффициентов регрессии $b_i$ от величины лага описывается полиномом $k$ степени. Лаги, структуру которых можно описать с помощью полиномов, называют также лагами Алмон, по имени III. Алмон, впервые обратившей внимание на такое представление лагов.

Формально модель зависимости коэффициентов $b_j$ от величины лага $j$ в форме полинома можно записать в следующем виде:
dля полинома 1-й степени: $b_j = c_0 + c_1$;
dля полинома 2-й степени: $b_j = c_0 + c_1 j + c_2 j^2$;
dля полинома 3-й степени: $b_j = c_0 + c_1 j + c_2 j^2 + c_3 j^3$ и т.д. В наиболее общем виде для полинома $k$-й степени имеем:

$$b_j = c_0 + c_1 j + c_2 j^2 + \cdots + c_k j^k.$$  \hspace{1cm} (49)

Замечание 1. В данной модели предполагается, что степень полинома $k$ меньше максимальной величины лага $l$, $k < l$.

Тогда каждый из коэффициентов $b_j$ модели (48) можно выразить следующим образом:

$$b_0 = c_0,$$
$$b_1 = c_0 + c_1 + \cdots + c_k,$$
$$b_2 = c_0 + 2c_1 + 4c_2 + \cdots + 2^k c_k,$$
$$b_3 = c_0 + 3c_1 + 9c_2 + \cdots + 3^k c_k$$  \hspace{1cm} (50)

и т.д.

$$b_l = c_0 + lc_1 + l^2 c_2 + \cdots + l^k c_k.$$  

Подставим в (48) найденные соотношения для $b_j$, получим:

$$y_t = a + (c_0 + c_1 + \cdots + c_k)x_{t-1} + (c_0 + 2c_1 + 4c_2 + \cdots + 2^k c_k)x_{t-2} +$$

$$+ (c_0 + 3c_1 + 9c_2 + \cdots + 3^k c_k)x_{t-3} + \cdots + (c_0 + lc_1 + l^2 c_2 + \cdots + l^k c_k)x_{t-l} + e_t.$$  

Перегруппируем слагаемые:

$$y_t = a + c_0(x_t + x_{t-1} + x_{t-2} + \cdots + x_{t-l}) + c_1(x_{t-1} + 2x_{t-2} + 3x_{t-3} + \cdots + lx_{t-l}) +$$

$$+ c_2(x_{t-1} + 4x_{t-2} + 9x_{t-3} + \cdots + l^2 x_{t-l}) + \cdots + c_k(x_{t-1} + 2^k x_{t-2} + 3^k x_{t-3} + \cdots + l^k x_{t-l}) + e_t.$$  

Обозначим слагаемые в скобках при $c_j$ как новые переменные:

$$z_0 = x_t + x_{t-1} + x_{t-2} + \cdots + x_{t-l},$$
$$z_1 = x_{t-1} + 2x_{t-2} + 3x_{t-3} + \cdots + lx_{t-l},$$
$$z_2 = x_{t-1} + 4x_{t-2} + 9x_{t-3} + \cdots + l^2 x_{t-l},$$

...  \hspace{1cm} (51)

$$z_k = x_{t-1} + 2^k x_{t-2} + 3^k x_{t-3} + \cdots + l^k x_{t-l}.$$  

Перепишем модель (48) с учетом соотношений (51):

$$y_t = a + c_0 z_0 + c_1 z_1 + c_2 z_2 + \cdots + c_k z_k + e_t.$$  \hspace{1cm} (52)
Процедура применения метода Алмон для расчета параметров модели с распределенным лагом выглядит следующим образом.

1. Определяется максимальная величина лага \( l \).
2. Определяется степень полинома \( k \), описывающего структуру лага.
3. По соотношениям (51) рассчитываются значения переменных \( z_0, \ldots, z_k \).
4. Определяются параметры уравнения линейной регрессии (52).
5. С помощью соотношений (50) рассчитываются параметры исходной модели с распределенным лагом.

Применение метода Алмон сопряжено с рядом проблем. Во-первых, величина лага \( l \) должна быть известна заранее. При ее определении лучше исходить из максимально возможного лага, чем ограничиваться лагами небольшой длины. Выбор меньшего лага, чем его реальное значение, приведет к тому, что в модели регрессии не будет учтен фактор, оказывающий значительное влияние на результат, т. е. к неверной спецификации модели. Влияние этого фактора в такой модели будет выражено в остатках. Тем самым в модели не будут соблюдаться предпосылки МНК о случайности остатков, а полученные оценки ее параметров окажутся неэффективными и смещенными. Выбор более большой величины лага по сравнению с ее реальным значением будет означать включение в модель статистически не значимого фактора и снижение эффективности полученных оценок, однако эти оценки все же будут несмещенными.

Метод Алмон имеет два неоспоримых преимущества.

1. Он достаточно универсален и может быть применен для моделирования процессов, которые характеризуются разнообразными структурами лагов.
2. При относительно небольшом количестве переменных в (51) (обычно выбирают \( k = 2 \) или \( k = 3 \)), которое не приводит к потере значительного числа степеней свободы, с помощью метода Алмон можно построить модели с распределенным лагом любой длины.

2.14. Пример применения метода Алмон

Таблица 13

<table>
<thead>
<tr>
<th>Год</th>
<th>1 кв.</th>
<th>2 кв.</th>
<th>3 кв.</th>
<th>4 кв.</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td>104.9</td>
<td>104.8</td>
<td>105.4</td>
<td>105.4</td>
</tr>
<tr>
<td>2006</td>
<td>104.1</td>
<td>106.6</td>
<td>107.2</td>
<td>107.1</td>
</tr>
<tr>
<td>2007</td>
<td>107.8</td>
<td>107.1</td>
<td>106.8</td>
<td>105.6</td>
</tr>
<tr>
<td>2008</td>
<td>106</td>
<td>104.3</td>
<td>101.9</td>
<td>90.9</td>
</tr>
<tr>
<td>2009</td>
<td>83.3</td>
<td>85.4</td>
<td>88.1</td>
<td>101</td>
</tr>
<tr>
<td>2010</td>
<td>108.4</td>
<td>107.6</td>
<td>105.9</td>
<td>107.2</td>
</tr>
<tr>
<td>2011</td>
<td>105</td>
<td>106.1</td>
<td>105.3</td>
<td>103.8</td>
</tr>
<tr>
<td>2012</td>
<td>104.5</td>
<td>102.4</td>
<td>103.4</td>
<td>103.2</td>
</tr>
<tr>
<td>2013</td>
<td>98.8</td>
<td>100.8</td>
<td>100.6</td>
<td>101.4</td>
</tr>
<tr>
<td>2014</td>
<td>101.1</td>
<td>101.8</td>
<td>101.5</td>
<td>102.1</td>
</tr>
<tr>
<td>2015</td>
<td>99.6</td>
<td>95.1</td>
<td>95.8</td>
<td>96.1</td>
</tr>
</tbody>
</table>
Таблица 1

Индекс промышленного производства,
в % к соответствующему периоду предыдущего года

<table>
<thead>
<tr>
<th>Год</th>
<th>1 кв.</th>
<th>2 кв.</th>
<th>3 кв.</th>
<th>4 кв.</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td>94.8</td>
<td>100.1</td>
<td>103.5</td>
<td>107.3</td>
</tr>
<tr>
<td>2006</td>
<td>93.6</td>
<td>102.5</td>
<td>104.1</td>
<td>107.3</td>
</tr>
<tr>
<td>2007</td>
<td>94.2</td>
<td>101.8</td>
<td>103.8</td>
<td>106.1</td>
</tr>
<tr>
<td>2008</td>
<td>94.6</td>
<td>100.2</td>
<td>101.4</td>
<td>94.6</td>
</tr>
<tr>
<td>2009</td>
<td>86.7</td>
<td>102.7</td>
<td>104.6</td>
<td>108.5</td>
</tr>
<tr>
<td>2010</td>
<td>93.1</td>
<td>101.9</td>
<td>102.9</td>
<td>109.9</td>
</tr>
<tr>
<td>2011</td>
<td>91.1</td>
<td>102.9</td>
<td>102.2</td>
<td>108.2</td>
</tr>
<tr>
<td>2012</td>
<td>91.7</td>
<td>100.8</td>
<td>103.3</td>
<td>108</td>
</tr>
<tr>
<td>2013</td>
<td>87.8</td>
<td>102.8</td>
<td>103.1</td>
<td>108.8</td>
</tr>
<tr>
<td>2014</td>
<td>87.6</td>
<td>103.6</td>
<td>102.7</td>
<td>109.6</td>
</tr>
<tr>
<td>2015</td>
<td>85.4</td>
<td>98.9</td>
<td>103.5</td>
<td>110</td>
</tr>
</tbody>
</table>

Графики этих рядов показывают наличие сезонности. Поэтому предварительно необходимо удалить сезонные компоненты у этих рядов.

Для этого воспользуемся известным пакетом эконометрических исследований Eviews компании IHS Global Inc. После стандартной загрузки исходных данных \( y, x \) в Eviews с помощью команды \( \text{seas}(a)y \) в командном окне получим сезонные компоненты аддитивной модели ряда \( y \), а командой \( \text{seas}(a)x \) – сезонные компоненты аддитивной модели ряда \( x \). Кроме того, мы получим два ряда \( v \) и \( p \) – это исходные ряды без сезонных компонент. В этом пакете процедура применения лагов Алмон автоматизирована. В командном окне набираем \( \text{ls } v \text{ c pdl}(p,3,2) \). Здесь \( \text{ls} \) означает метод наименьших квадратов, \( c \) – константу в уравнении. Число 3 равно \( l \) (максимальному лагу), а число 2 показывает степень полинома \( k \) (напомним, что \( k < l \)).

В результате мы получим уравнение (рис. 11)

\[
v = -298.6077 + 0.9860 \cdot p_t + 0.97516 \cdot p_{t-1} + 0.9893 \cdot p_{t-2} + 1.02855 \cdot p_{t-3} + e_t.
\]
При этом $R^2 = 0.990663$. Однако $Q$-статистика на первом лаге $Q = 12.630$, что говорит о наличии автокорреляции. Поэтому увеличим длину лага на 1. Введем в командное окно `ls v c pdl(p,4,2)`, в результате получим уравнение (рис. 12)

$$v = -298.5810 + 0.82687 \cdot p_t + 1.0983 \cdot p_{t-1} + 1.0827 \cdot p_{t-2} + 0.7801 \cdot p_{t-3} + 0.19048 \cdot p_{t-4} + et.$$ 

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>-298.6077</td>
<td>5.565547</td>
<td>-45.3429</td>
<td>0.0000</td>
</tr>
<tr>
<td>pdl.03</td>
<td>0.001571</td>
<td>0.024861</td>
<td>0.067220</td>
<td>0.9480</td>
</tr>
<tr>
<td>pdl.02</td>
<td>0.012511</td>
<td>0.021156</td>
<td>0.591379</td>
<td>0.5579</td>
</tr>
<tr>
<td>pdl.01</td>
<td>0.975167</td>
<td>0.024274</td>
<td>40.1726</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R-squared</td>
<td>0.990606</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjusted R-squared</td>
<td>0.989066</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S.E. of regression</td>
<td>0.003926</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sum squared resid</td>
<td>13.58442</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LogLikelihood</td>
<td>-35.51312</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F-statistic</td>
<td>1308.541</td>
<td></td>
<td></td>
<td>0.820501</td>
</tr>
<tr>
<td>Prob(F-statistic)</td>
<td>0.000000</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Рис. 12. Модель `pdl(p,4,2)`

$R^2 = 0.978406$, значение $Q$-статистики для лага 2 равно 0.0609, что говорит об отсутствии автокорреляции. Статистика Jarge-Bera равна 1.79, значит, остатки образуют белый шум.

Анализ этой модели показывает, что рост промышленного производства на 1 % в текущем периоде приведет через 4 года к росту ВВП России в среднем на $(0.82687+1.0983+1.0827+0.7801+0.1904+0.390448) = 3.97837 \%$, т. е. почти на 4 %.

Определим относительные коэффициенты регрессии:

$$\beta_0 = \frac{0.82687}{3.97837} = 0.207841; \quad \beta_1 = \frac{1.0983}{3.97837} = 0.276068; \quad \beta_2 = \frac{1.0827}{3.97837} = 0.272147;$$

$$\beta_3 = \frac{0.7801}{3.97837} = 0.196085; \quad \beta_4 = \frac{0.1904}{3.97837} = 0.0478588.$$ 

Почти половина воздействия фактора на результат реализуется с лагом в 1 год (так как $\beta_0 + \beta_1 = 0.483909 \approx 0.5$), причем 20 % этого воздействия реализуется сразу же, в текущем периоде, т. е. медианный лаг примерно равен 1 г.

Средний лаг в данной модели составит:

$$t = 0.207835 + 0.276062 \cdot 1 + 0.272145 \cdot 2 + 0.196084 \cdot 3 + 0.0478776 \cdot 4 = 1.80795.$$ 

В среднем увеличение промышленного производства в России приведет к увеличению ВВП через 1.8 года.

Для сравнения приведем результаты применения обычного МНК для расчета параметров этой модели:

$$v = 59.6686 + 1.1002 \cdot p - 0.1353 \cdot p_{t-1} - 0.26189 \cdot p_{t-2} + 0.1445 \cdot p_{t-3} - 0.4232 \cdot p_{t-4} + et.$$ 

$R^2 = 0.252$, $F$-статистика $= 2.288979$, уравнение не значимое, как и все коэффициенты этой модели.
3. Модели ARIMA

3.1. Разложение Вольда

Стационарный стохастический процесс $x_t$ с нулевым математическим ожиданием иногда полезно представлять в виде линейной комбинации последовательности возмущений $e_t, e_{t-1}, e_{t-2}, \ldots$, т. е.

$$x_t = e_t + \varphi_1 e_{t-1} + \varphi_2 e_{t-2} + \ldots = \sum_{i=0}^{\infty} \varphi_i e_i,$$  \hspace{1cm} (53)

или с использованием лагового оператора:

$$x_t = (1 + \varphi_1 L + \varphi_2 L^2 + \ldots)x_t,$$

где $\varphi_0 = 1$ и выполняется:

$$\sum_{i=0}^{\infty} |\varphi_i| < \infty,$$  \hspace{1cm} (54)

t. е. ряд абсолютных значений коэффициентов сходится.

Предполагается, что последовательность $e_t$ представляет собой чисто случайный процесс или, другими словами, белый шум (см. выше). Автокорреляционная функция белого шума имеют очень простую форму:

$$\rho(k) = \begin{cases} 
1, & k = 0, \\
0, & k \neq 0.
\end{cases}$$

Таким образом, белый шум легко идентифицируется с помощью графика автокорреляционной функции. Часто предполагается, что последовательность $e_t$ состоит из независимых одинаково распределенных величин. Упростить анализ помогает дополнительное предположение о том, что $e_t$ имеет нормальное распределение, т. е. представляет собой гауссовский белый шум.

Данная модель не является произвольной. Фактически, согласно теореме Вольда, любой слабо стационарный ряд допускает представление в виде модели (53), а именно: разложение Вольда ряда $x_t$.

Следует помнить, однако, что разложение Вольда единствено, в то время как представление (53), вообще говоря, неоднозначно. Таким образом, разложение Вольда представляет процесс в виде модели линейного фильтра, в то время как модель линейного фильтра не обязательно задает разложение Вольда. Кompактная запись линейного фильтра выглядит следующим образом:

$$x_t = \varphi(L)e_t.$$

Как мы увидим в дальнейшем, модель линейного фильтра (53) применима не только к стационарным процессам, таким, что выполняется (54), – с соответствующими оговорками она упрощает анализ и многих нестационарных процессов.

Если процесс $x_t$ подчинен модели (53), то при выполнении условия (54) он имеет математическое ожидание, равное нулю:

$$M(x_t) = \sum_{i=0}^{\infty} \varphi_i M(e_i) = 0.$$

6 В разложении Вольда произвольного стационарного процесса может присутствовать также полностью предсказуемая (линейно детерминированная) компонента. Однако такая компонента, если ее свойства известны, не создает больших дополнительных сложностей для анализа.

7 См. ниже в этой главе анализ обратимости процесса скользящего среднего.
Если требуется, чтобы математическое ожидание \( x_t \) не было равно нулю, то уравнение модели линейного фильтра должно включать константу:

\[
x_t = \mu + \sum_{i=0}^{\infty} \varphi_i e_{t-i}.
\]

Выведем формулы для **автоковариаций** рассматриваемой модели:

\[
\gamma(k) = M(x_t x_{t+k}) = M\left( \sum_{i=0}^{\infty} \varphi_i e_{t-i} \sum_{j=0}^{\infty} \varphi_j e_{t+k-j} \right) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \varphi_i \varphi_j M(e_{t-i} e_{t+k-j}) = \sigma_e^2 \sum_{i=0}^{\infty} \varphi_i \varphi_{i+k}. \tag{55}
\]

Здесь учитывается, что для белого шума

\[
M(e_{t-i} e_{t+k-j}) = \begin{cases} 
\sigma_e^2, & j = i + k, \\
0, & j \neq i + k.
\end{cases}
\]

Заметим, что из (54) следует сходимость возникающих здесь рядов. Это говорит о том, что данное условие подразумевает стационарность.

Ясно, что модель линейного фильтра (53) в общем виде представляет в основном теоретический интерес, поскольку содержит бесконечное число параметров. Для прикладного моделирования желательно использовать уравнения с конечным числом параметров. В основе таких моделей может лежать так называемая рациональная аппроксимация для \( \varphi(L) \), т.е. приближение в виде частного двух лаговых многочленов

\[
\varphi(L) \approx \frac{\theta(L)}{\phi(L)},
\]

где лаговые многочлены \( \theta(L) \) и \( \phi(L) \) имеют уже конечное число параметров. Как показывает практика, многие ряды можно достаточно хорошо аппроксимировать этим методом. Частными случаями применения рациональной аппроксимации являются модели авторегрессии \( AR(p) \) и скользящего среднего \( MA(q) \). В общем случае получаем смешанные процессы авторегрессии – скользящего среднего \( ARMA(p,q) \).

### 3.2. Процессы авторегрессии (AR)

В модели авторегрессии текущее значение процесса \( x_t \) представляется в виде линейной комбинации конечного числа предыдущих значений процесса и белого шума \( e_t \):

\[
x_t = \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \cdots + \varphi_p x_{t-p} + e_t, \tag{56}
\]

при этом предполагается, что текущее значение \( e_t \) не коррелировано с лагами \( x_t \). Такая модель называется **авторегрессией** \( p \)-го порядка и обозначается \( AR(p) \) (от английского *autoregression*).

Используя лаговый оператор \( L \), представим уравнение авторегрессии в виде:

\[
(1 - \varphi_1 L - \varphi_2 L^2 - \cdots - \varphi_p L^p)x_t = e_t,
\]

или кратко через лаговый многочлен \( \varphi(L) = 1 - \varphi_1 L - \varphi_2 L^2 - \cdots - \varphi_p L^p \):

\[
\varphi(L)x_t = e_t.
\]

Нетрудно показать, что модель авторегрессии является частным случаем модели линейного фильтра:

\[
x_t = \psi(L)e_t,
\]

где \( \psi(L) = \varphi^{-1}(L) \), т.е. \( \psi(L) \) – оператор, обратный оператору \( \varphi(L) \).
Удобным и полезным инструментом для изучения процессов авторегрессии является характеристический многочлен (характеристический полином)

\[ z^p - \varphi_1 z^{p-1} - \cdots - \varphi_p, \]

(57)

и связанное с ним характеристическое уравнение

\[ z^p - \varphi_1 z^{p-1} - \cdots - \varphi_p = 0. \]

(58)

Как мы увидим в дальнейшем, от того, какие корни имеет характеристическое уравнение, зависят свойства процесса авторегрессии, в частности является ли процесс стационарным или нет. Рассмотрим наиболее часто использующиеся частные случаи авторегрессионных процессов.

### 3.3. Процесс Маркова

Процессом Маркова (марковским процессом) называется авторегрессионный процесс первого порядка, AR(1):

\[ x_t = \varphi x_{t-1} + e_t, \]

(59)

где \( e_t \) представляет собой белый шум, который не коррелирует с \( x_{t-1} \). Здесь мы упростили обозначения, обозначив \( \varphi = \varphi_1 \).

Найдем необходимые условия стационарности марковского процесса. Предположим, что процесс \( x_t \) слабо стационарен (стационарен в широком смысле). Тогда его первые и вторые моменты неизменны. Находя дисперсии от обеих частей (59), получим, учитывая, что \( \text{cov}(x_{t-1}, e_t) = 0 \):

\[ D(x_t) = \varphi^2 D(x_{t-1}) + D(e_t) \]

или

\[ \sigma_x^2 = \varphi^2 \sigma_x^2 + \sigma_e^2. \]

Ясно, что при \( |\varphi| \geq 1 \), с учетом \( \sigma_e^2 > 0 \), правая часть этого равенства должна быть больше левой, что невозможно. Получаем, что у стационарного марковского процесса \( |\varphi| < 1 \).

Пусть, с другой стороны, \( |\varphi| < 1 \). Представим \( x_t \) через белый шум \( e_t \). Это можно осуществить с помощью последовательных подстановок по формуле (59):

\[ x_t = \varphi x_{t-1} + e_t = \varphi(\varphi x_{t-2} + e_{t-1}) + e_t = \varphi^2 x_{t-2} + \varphi e_{t-1} + e_t, \]

после чего

\[ x_t = \varphi^2 (\varphi x_{t-3} + e_{t-2}) + \varphi e_{t-1} + e_t = \varphi^3 x_{t-3} + \varphi^2 e_{t-2} + \varphi e_{t-1} + e_t \]

и т. д. В пределе, поскольку множитель при лаге \( x_t \) стремится к нулю, получим следующее представление \( x_t \) в виде модели линейного фильтра:

\[ x_t = e_t + \varphi e_{t-1} + \varphi^2 e_{t-2} + \cdots. \]

Это же представление можно получить с использованием оператора лага. Уравнение (59) запишется в виде

\[ (1 - \varphi L)x_t = e_t. \]
Применив к обеим частям уравнения \((1 - \varphi L)^{-1}\), получим

\[
x_t = (1 - \varphi L)^{-1} e_t = (1 + \varphi L + \varphi^2 L^2 + \cdots) e_t = e_t + \varphi e_{t-1} + \varphi^2 e_{t-2} + \cdots.
\]

В терминах модели линейного фильтра (53) для марковского процесса \(\varphi_i = \varphi^i\). Поэтому

\[
\sum_{i=0}^{\infty} |\varphi_i| = \sum_{i=0}^{\infty} |\varphi|^i = \frac{1}{1 - |\varphi|} < \infty,
\]

т. е. условие стационарности модели линейного фильтра (53) выполняется при \(|\varphi| < 1\).

Можно сделать вывод, что условие стационарности процесса Маркова имеет следующий вид:

\(|\varphi| < 1\).

Свойства стационарного процесса AR(1):

1. Если процесс \(x_t\) слабо стационарен, то его математическое ожидание неизменно, поэтому, беря математическое ожидание от обеих частей (59), получим \(M(x_t) = \varphi M(x_t)\), откуда

\(M(x_t) = 0\).

Если добавить в уравнение (59) константу:

\(x_t = \mu + \varphi x_{t-1} + e_t\),

tо \(M(x_t) = \mu + \varphi M(x_t)\) и

\[M(x_t) = \frac{\mu}{1 - \varphi}.
\]

2. Найдем дисперсию процесса Маркова, используя полученное выше уравнение \(\sigma^2_x = \varphi^2 \sigma^2_x + \sigma^2_e\):

\[D(x_t) = \sigma^2_x = \frac{\sigma^2_e}{1 - \varphi}.\]  

(61)

Можно также применить общую формулу для автоковариации в модели линейного фильтра (53) с \(k = 0\):

\[\sigma^2_x = \sigma^2_e \sum_{i=0}^{\infty} \varphi^2_i = \sigma^2_e (1 + \varphi^2 + \varphi^4 + \cdots) = \frac{\sigma^2_e}{1 - \varphi^2}.
\]

При \(|\varphi| \geq 1\) дисперсия процесса \(x_t\), вычисляемая по этой формуле, неограниченно растет. В частности, при \(\varphi = 1\) получим:

\[\sigma^2_x = \sigma^2_e (1 + 1 + 1 + \cdots) = n \cdot \sigma^2_e \to \infty,
\]

при \(n \to \infty\).

3. Коэффициент автоковариации \(k\)-го порядка по формуле (55) равен

\[
y(k) = \sigma^2_e \sum_{i=0}^{\infty} \varphi^i \varphi_{t+k} = \sigma^2_e \sum_{i=0}^{\infty} \varphi^i \varphi_{i+k} = \sigma^2_e \sum_{i=0}^{\infty} \varphi^{2i+k} = \sigma^2_e \varphi^k \sum_{i=0}^{\infty} \varphi^{2i} = \frac{\sigma^2_e}{1 - \varphi^2} \varphi^k.
\]

(63)

4. Коэффициент автокорреляции (автокорреляционная функция), исходя из (61), равен:

\[
\rho(k) = \frac{y(k)}{D(x_t)} = \varphi^k.
\]

(64)
При $0 < \varphi < 1$ автокорреляционная функция имеет форму затухающей экспоненты (рис. 13), при $-1 < \varphi < 0$ – форму затухающей знакопеременной экспоненты (рис. 13).

Рис. 13. Графики автокорреляционных функций AR(1)

Если $\varphi > 1$, процесс Маркова превращается во «взрывной» процесс.

В случае $\varphi = 1$ имеет место так называемый процесс случайного блуждания, который относится к разряду нестационарных, так как дисперсия случайного блуждания не постоянна, см. (62).

Модель AR(p)

Рассмотрим авторегрессию второго порядка AR(2):

$$x_t = \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + e_t,$$

или через лаговый оператор:

$$(1 - \varphi_1 L - \varphi_2 L^2)x_t = e_t.$$ 

Для стационарности процесса авторегрессии AR(2) необходимо, чтобы корни $\lambda_1$, $\lambda_2$ характеристического уравнения

$$z^2 - \varphi_1 z - \varphi_2 = 0,$$

которые, вообще говоря, могут быть комплексными, находились внутри единичного круга на комплексной плоскости, т. е. $|\lambda_1| < 1$, $|\lambda_2| < 1$.

Обоснуем условия стационарности AR(2). Уравнение 65 запишем в виде:

$$x_t - \varphi_1 x_{t-1} - \varphi_2 x_{t-2} = e_t.$$

Левая часть уравнения представляет собой линейное рекуррентное уравнение второго порядка с постоянными коэффициентами. Так как правая часть уравнения $e_t$ представляет собой стационарный временной ряд, который колеблется около 0, то приближенно можно считать, что правая часть равна нулю. Все решения этого однородного уравнения в случае, когда $\lambda_1 \neq \lambda_2$ имеют вид:

$$C_1 \lambda_1^t + C_2 \lambda_2^t,$$

где $C_1, C_2$ – произвольные параметры, которые определяются исходя из начальных условий. Из вида решения видно, что мы получим ограниченную функцию от $t$, только если $|\lambda_1| < 1$ и $|\lambda_2| < 1$. 

54
Автокорреляционная и автоковариационная функция AR(p)

Для стационарного процесса авторегрессии:

\[ x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + e_t \]  \hspace{1cm} (66)

можно вывести формулу автокорреляционной функции. Умножив обе части уравнения на \( x_{t-k} \):

\[ x_{t-k} x_t = \phi_1 x_{t-k} x_{t-1} + \phi_2 x_{t-k} x_{t-2} + \cdots + \phi_p x_{t-k} x_{t-p} + x_{t-k} e_t, \]

и перейдя к математическим ожиданиям, получим уравнение, связывающее коэффициенты автоковариации различного порядка:

\[ \gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \cdots + \phi_p \gamma_{k-p}, k > 0. \]  \hspace{1cm} (67)

Это выражение является следствием того, что соответствующие ковариации между процессом и ошибкой равны нулю: \( M(x_{t-k} e_t) = 0 \) при \( k > 0 \), т. к. \( x_{t-k} \) может включать лишь ошибки \( e_j \) для \( j \leq t - k \).

Делением уравнения (67) на \( D(x_t) \) получаем важное рекуррентное соотношение для автокорреляционной функции:

\[ \rho(k) = \phi_1 \rho(k - 1) + \phi_2 \rho(k - 2) + \cdots + \phi_p \rho(k - p), k > 0. \]  \hspace{1cm} (68)

Подставляя в выражение (68) \( k = 1, \ldots, p \), получаем, с учетом симметричности автокорреляционной функции, так называемые уравнения Юла-Уокера (Yule-Walker) для AR(p):

\[
\begin{align*}
\rho(1) &= \phi_1 + \phi_2 \rho(1) + \cdots + \phi_p \rho(p - 1), \\
\rho(2) &= \phi_1 \rho(1) + \phi_2 + \cdots + \phi_p \rho(p - 2), \\
&\vdots \\
\rho(p) &= \phi_1 \rho(p - 1) + \phi_2 \rho(p - 2) + \cdots + \phi_p.
\end{align*}
\]  \hspace{1cm} (69)

Мы имеем здесь \( p \) линейных уравнений, связывающих \( p \) автокорреляций, \( \rho(1), \ldots, \rho(p) \). Из этой системы при данных параметрах можно найти автокорреляции. С другой стороны, при данных автокорреляциях из уравнений Юла-Уокера можно найти параметры \( \phi_1, \ldots, \phi_p \). Замена теоретических автокорреляций выборочными дает метод оценивания параметров процесса AR(p).

В частности, для AR(2) (процесса Юлай) получим из (69)

\[ \rho_1 = \frac{\phi_1}{1 - \varphi_2}, \rho_2 = \frac{\phi_1^2}{1 - \varphi_2^2} + \varphi_2. \]

Зная \( \rho(1), \ldots, \rho(p) \), все последующие автокорреляции \( \rho(k), (k > p) \) можем найти по рекуррентной формуле (69).

Для нахождения автоковариационной функции требуется знать \( D(x_t) \), дисперсию процесса \( x_t \). Если умножить обе части (66) на \( e_t \) и взять математические ожидания, получим, что \( M(e_t x_t) = E(e_t^2) = \sigma_e^2 \). Далее, умножая обе части (66) на \( x_t \) и беря математические ожидания, получим

\[ \gamma(0) = \phi_1 \gamma(1) + \phi_2 \gamma(2) + \cdots + \phi_p \gamma(p) + \sigma_e^2. \]

Значит, если известны автокорреляции, то дисперсию \( x_t \) можно вычислять по формуле:

\[ \gamma(0) = \sigma_x^2 = \frac{\sigma_e^2}{1 - \phi_1 \rho(1) - \phi_2 \rho(2) - \cdots - \phi_p \rho(p)}. \]  \hspace{1cm} (70)
Автоковариации затем можно вычислить как $\gamma(j) = \rho(j)\sigma^2_e$. С учетом полученного дополнительного уравнения можно записать вариант уравнений Юла-Уокера для автоковариаций:

$$
\begin{align*}
\gamma(0) &= \varphi_1\gamma(1) + \varphi_2\gamma(2) + \cdots + \varphi_p\gamma(p) + \sigma^2_e, \\
\gamma(1) &= \varphi_1\gamma(0) + \varphi_2\gamma(1) + \cdots + \varphi_p\gamma(p-1), \\
\gamma(2) &= \varphi_1\gamma(1) + \varphi_2\gamma(0) + \cdots + \varphi_p\gamma(p-2), \\
&\vdots \\
\gamma(p) &= \varphi_1\gamma(p-1) + \varphi_2\gamma(p-2) + \cdots + \varphi_p\gamma(0).
\end{align*}$$

В этой системе имеется $p + 1$ уравнений, связывающих $p + 1$ автоковариацию, что позволяет непосредственно вычислять автоковариации при данных параметрах. Заметим, что (67) и (68) имеют вид линейных однородных конечноразностных уравнений, а для подобных уравнений существует общий метод нахождения решения. Решив уравнение (68), можно получить общий вид автокорреляционной функции процесса авторегрессии. Проведем это рассуждение более подробно.

Характеристическое уравнение для разностного уравнения (68) имеет вид:

$$z^p - \varphi_1z^{p-1} - \cdots - \varphi_p = 0.$$ 

Пусть $\lambda_1, \ldots, \lambda_p$ – корни характеристического уравнения (с учетом кратности). Здесь возможны два случая.

1. Все корни различные. Тогда общее решение уравнения 68 имеет вид

$$\rho(k) = c_1\lambda_1^k + c_2\lambda_2^k + \cdots + c_p\lambda_p^k$$

где $c_i$ – некоторые константы, в общем случае комплексные.

2. Пусть, например (не ограничивая общности), корень $\lambda_1$ имеет кратность $l$, а остальные корни $\lambda_{i+1}, \ldots, \lambda_p$ различные. Тогда общее решение разностного уравнения 68 имеет вид

$$\rho(k) = (c_1 + c_2k + \cdots + c_lk^{l-1})\lambda_1^k + c_{l+1}\lambda_{l+1}^k + \cdots + c_p\lambda_p^k.$$ 

Если все корни характеристического уравнения удовлетворяют условию $|\lambda_i| < 1, i = 1, \ldots, p$, то все слагаемые в правой части (72), (73) затухают с ростом $k$. Если же для какого-то корня $|\lambda_i| > 1$, то (при условии $c_i \neq 0$) соответствующее слагаемое «уходит на бесконечность». Если $|\lambda_i| = 1$, то соответствующее слагаемое не затухает. Из этих рассуждений следует условие стационарности AR(p) – все корни соответствующего характеристического уравнения по модулю должны быть меньше единицы.

Таким образом, из соотношений (72), (73) следует, что в общем случае автокорреляционная функция стационарного процесса авторегрессии является комбинацией затухающих экспонент и затухающих синусоид.

Итак, мы вывели общий вид автокорреляционной функции стационарного процесса авторегрессии. Теоретически выборочная автокорреляционная функция может служить инструментом для распознавания авторегрессионного процесса. На практике же для коротких рядов отличительная сила автокорреляционной функции не очень высока. Однако часто изучение автокорреляционной функции является хорошим заделом исследования системы.
Оценивание авторегрессий

Термин «авторегрессия» для обозначения модели (56) используется потому, что она фактически представляет собой модель регрессии, в которой регрессорами служат лаги изучаемого ряда $x_t$. По определению авторегрессии ошибки $e_t$ являются белым шумом и некоррелированы с лагами $x_t$. Таким образом, выполнены все основные предположения регрессионного анализа: ошибки имеют нулевое математическое ожидание, некоррелированы с регрессорами, не автокоррелированы и гомоскедастичны. Следовательно, модель (56) можно оценивать с помощью обычного метода наименьших квадратов.

Частная автокорреляционная функция

Как мы видели, автокорреляционная функция процесса авторегрессии состоит из экспоненциально затухающих компонент. Такая характеристика не очень наглядна, поскольку соседние автокорреляции сильно связаны друг с другом, и, кроме того, для полного описания свойств ряда используется бесконечная последовательность автокорреляций. Более наглядными характеристиками авторегрессии являются частные автокорреляции.

Частная автокорреляция измеряет «чистую» корреляцию между уровнями временного ряда $x_t$ и $x_{t-k}$ при исключении опосредованного влияния промежуточных уровней ряда. Такой показатель корреляции между элементами ряда более информативен.

Пусть $x_t$ — произвольный стационарный ряд (не обязательно авторегрессия) и $\rho(j)$ — его автокорреляция. Применим к нему уравнения Юла-Уокера (69), как если бы процесс представлял собой авторегрессию $k$-го порядка, и найдем по автокорреляциям коэффициенты. Если обозначить $j$-й коэффициент уравнения авторегрессии порядка $k$ через $\varphi_{kj}$, то уравнения Юла-Уокера (69) принимают вид:

$$
\begin{bmatrix}
1 & \rho(1) & \rho(2) & \ldots & \rho(k-1) \\
\rho(1) & 1 & \rho(1) & \ldots & \rho(k-2) \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\rho(k-1) & \rho(k-2) & \rho(k-3) & \ldots & 1
\end{bmatrix}
\begin{bmatrix}
\varphi_{k1} \\
\varphi_{k2} \\
\vdots \\
\varphi_{kk}
\end{bmatrix}
= 
\begin{bmatrix}
\rho(1) \\
\rho(2) \\
\vdots \\
\rho(k)
\end{bmatrix}.
$$

Частная автокорреляция $k$-го порядка определяется как величина $\varphi_{kk}$, полученная из этих уравнений.

Решение этих уравнений соответственно для $k = 1, 2, 3$ дает следующие результаты (здесь используется правило Крамера):

$$
\varphi_{11} = \rho(1), \varphi_{22} = \rho(1) - \frac{\rho(2)}{1 - \rho^2(1)}, \varphi_{33} = \frac{1}{1 - \rho^2(1)}.
$$

Частная автокорреляционная функция рассматривается как функция частной автокорреляции от лага $k$, где $k = 1, 2, 3, \ldots$. Для процесса авторегрессии порядка $p$ частная автокорреляционная функция $\rho_{\text{part}}(k)$ будет ненулевой для $k \leq p$ и равна нулю для $k > p$, то есть обрывается на лаге $p$. 

57
Значение выборочного частного коэффициента автокорреляции $\rho_{part}(k) = \varphi_{kk}$ вычисляется как МНК-оценка последнего коэффициента в уравнении авторегрессии AR($k$). Частная автокорреляционная функция может оказаться полезной в решении задачи идентификации модели временного ряда: если она быстро затухает, то это авторегрессия, причем ее порядок следует выбрать по последнему большому значению частной автокорреляционной функции.

3.4. Процессы скользящего среднего MA($q$)

Другой частный случай модели линейного фильтра, широко распространенный в анализе временных рядов, — модель скользящего среднего, когда $x_t$ линейно зависит от конечного числа $q$ предыдущих значений $e$:

$$x_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}. \quad (74)$$

Модель скользящего среднего $q$-го порядка обозначают MA($q$) (от английского moving average). Данную модель можно записать и более сжато:

$$x_t = \theta(L)e_t,$$

через оператор скользящего среднего:

$$\theta(L) = 1 - \theta_1 L - \theta_2 L^2 - \cdots - \theta_q L^q. \quad (75)$$

Легко видеть, что процесс MA($q$) является стационарным без каких-либо ограничений на параметры $\theta_j$. Действительно, математическое ожидание процесса $M(x_t) = 0$, а дисперсия

$$D(x_t) = (1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2)\sigma_e^2,$$

t. е. равна дисперсии белого шума, умноженной на конечную величину $1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2$.

Остальные моменты второго порядка ($\gamma(k), \rho(k)$) также от времени не зависят.

Автоковариационная функция процесса MA($q$)

Автоковариационная функция MA($q$)

$$\gamma(k) = \begin{cases} \left( -\theta_k + \theta_1 \theta_{k+1} + \cdots + \theta_{q-k} \theta_q \right) \sigma_e^2, & k = 1, 2, \ldots, q, \\ 0, & k > q. \end{cases} \quad (76)$$

В частном случае для MA(1) имеем:

$$\gamma(0) = (1 + \theta_1^2)\sigma_e^2,$$

$$\gamma(1) = -\theta_1 \sigma_e^2,$$

$$\gamma(k) = 0, k > 1,$$

и автоковариационная матрица, соответствующая последовательности $x_1, x_2, \ldots, x_n$, будет иметь следующий трехдиагональный вид:

$$\Gamma = \begin{bmatrix}
1 + \theta_1^2 & -\theta_1 & 0 & \cdots & 0 \\
-\theta_1 & 1 + \theta_1^2 & -\theta_1 & \cdots & 0 \\
0 & -\theta_1 & 1 + \theta_1^2 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & 1 + \theta_1^2
\end{bmatrix}.$$
В общем случае автоковариационная матрица процесса скользящего среднего порядка \( q \) имеет \( q \) ненулевых поддиагоналей и \( q \) ненулевых наддиагоналей, все же остальные элементы матрицы равны нулю. Автокорреляционная функция имеет вид:

\[
\rho(k) = \begin{cases} 
-\theta_k + \theta_1 \theta_{k+1} + \ldots + \theta_{q-k} \theta_q, & k = 1, 2, \ldots, q, \\
1 + \theta_1^2 + \ldots + \theta_q^2, & k > q.
\end{cases}
\] (77)

Таким образом, автокорреляционная функция процесса MA(\( q \)) обрывается на лаге \( q \), и в этом отличительная особенность процессов скользящего среднего.

С другой стороны, частная автокорреляционная функция, в отличие от авторегрессий, не обрывается и затухает экспоненциаль но. Например, для MA(1) частная автокорреляционная функция имеет вид:

\[-\theta^p \frac{1 - \theta_1^2}{1 - \theta_1^{2p+2}}.\]

**Обратимость процесса MA(\( q \))**

Авторегрессию, как мы видели выше, можно представить как \( MA(\infty) \). С другой стороны, процесс скользящего среднего можно представить в виде \( AR(\infty) \).

Рассмотрим, например, MA(1) (будем для упрощения писать \( \theta \) вместо \( \theta_1 \)):

\[x_t = e_t - \theta e_{t-1}, (78)\]

Сдвигом на один период назад получим \( e_{t-1} = x_{t-1} + \theta e_{t-2} \) и подставим в (78):

\[x_t = e_t - \theta x_{t-1} - \theta^2 e_{t-2}.\]

Далее, \( e_{t-2} = x_{t-2} + \theta e_{t-3} \), поэтому

\[x_t = e_t - \theta x_{t-1} - \theta^2 x_{t-2} - \theta^3 e_{t-3}.\]

Продолжая, получим на \( k \)-м шаге

\[x_t = e_t - \theta x_{t-1} - \theta^2 x_{t-2} - \ldots - \theta^k x_{t-k} - \theta^{k+1} e_{t-k-1}.\]

Если \( |\theta| < 1 \), то последнее слагаемое стремится к нулю при \( k \to \infty \). Переходя к пределу, получаем представление \( AR(\infty) \) для MA(1):

\[x_t = -\sum_{j=1}^{\infty} \theta^j x_{t-j} + e_t.\] (79)

С помощью лагового оператора можем записать это как

\[\eta(L)x_t = e_t,\]

где

\[\eta(L) = (1 - \theta L)^{-1} = \sum_{j=1}^{\infty} \theta^j L^j.\]
В то время как процесс (78) стационарен при любом \( \theta \), процесс (79) стационарен только при \( |\theta| < 1 \). При \( |\theta| \geq 1 \) веса \( \theta^j \) в разложении (79) растут (при \( |\theta| = 1 \) не меняются) по абсолютной величине по мере увеличения \( j \). Тем самым нарушается разумная связь текущих событий с событиями в прошлом. Говорят, что при \( |\theta| < 1 \) процесс MA(1) является обратимым, а при \( |\theta| \geq 1 \) — необратимым.

В общем случае уравнение процесса MA(q) в обращенной форме можно записать как

\[
e_t = \theta^{-1}(L)x_t = \eta(L)x_t = \sum_{j=1}^{\infty} \eta_j x_{t-j}.
\]

Процесс MA(q) называется обратимым, если абсолютные значения весов \( \eta_j \) в обращенном разложении образуют сходящийся ряд.

Стационарным процесс MA(q) является всегда, но для того, чтобы он обладал свойством обратимости, параметры процесса должны удовлетворять определенным ограничениям.

Выведем условия, которым должны удовлетворять параметры \( \theta_1, \theta_2, \ldots, \theta_q \) процесса MA(q), чтобы этот процесс был обратимым. Пусть \( h_i, i = 1, \ldots, q \) — корни характеристического уравнения

\[
z^q - \theta_1 z^{q-1} - \cdots - \theta_q = 0
\]

(будем предполагать, что они различны). Тогда справедливо разложение этого многочлена

\[
z^q - \theta_1 z^{q-1} - \cdots - \theta_q = (z - h_1)(z - h_2) \cdots (z - h_p).
\]

В этом равенстве сделаем замену переменной \( y = 1/z \). Получим

\[
1 - \theta_1 y - \theta_2 y^2 - \cdots - \theta_q y^q = (1 - h_1 y)(1 - h_2 y) \cdots (1 - h_q y)
\]

или в операторном виде

\[
\theta(L) = \prod_{i=1}^{q} (1 - h_i L).
\]

Тогда обратный к \( \theta(L) \) оператор \( \eta(L) \) можно представить в следующем виде:

\[
\eta(L) = \theta^{-1}(L) = \frac{1}{\prod_{i=1}^{q} (1 - h_i L)} = \sum_{i=1}^{q} E_i \frac{1}{1 - h_i L}.
\]

Здесь каждое слагаемое можно, по аналогии с MA(1), представить в виде бесконечного ряда:

\[
\frac{E_i}{1 - h_i L} = E_i \sum_{j=0}^{\infty} h_i^j L^j, \quad i = 1, \ldots, q,
\]

который сходится, если \( |h_i| < 1 \).

Тогда процесс MA(q) в обращенном представлении выглядит как

\[
e_t = \sum_{i=1}^{q} E_i \sum_{j=0}^{\infty} h_i^j L^j x_{t-j},
\]

и он стационарен, если корни характеристического уравнения лежат внутри единичного круга. Иными словами, MA(q) обладает свойством обратимости, если для всех корней выполнено \( |h_i| < 1, \forall i \). Если же для одного из корней \( |h_i| \geq 1 \), то ряд не будет сходиться и процесс MA(q) будет необратимым.
3.5. Смешанные процессы авторегрессии и скользящего среднего ARMA

На практике иногда бывает целесообразно ввести в модель как элементы авторегрессии, так и элементы скользящего среднего. Это делается для того, чтобы с использованием как можно меньшего числа параметров уловить характеристики исследуемого эмпирического ряда. Такой процесс называется смешанным процессом авторегрессии-скользящего среднего и обозначается \( \text{ARMA}(p, q) \):

\[
x_t = \varphi_1 x_{t-1} + \cdots + \varphi_p x_{t-p} + e_t - \theta_1 e_{t-1} - \cdots - \theta_q e_{t-q},
\]

или, с использованием оператора лага,

\[
(1 - \varphi_1 L - \varphi_2 L^2 - \cdots - \varphi_p L^p)x_t = (1 - \theta_1 L - \theta_2 L^2 - \cdots - \theta_q L^q)e_t.
\]

В операторной форме смешанная модель выглядит так:

\[
\varphi(L)x_t = \theta(L)e_t,
\]

где \( \varphi(L) \) – оператор авторегрессии, \( \theta(L) \) – оператор скользящего среднего.

Модель (80) получила название модели Бокса-Дженкинса, поскольку была популяризирована Дж. Боксом и Г. Дженкинсом в их известной книге «Анализ временных рядов». Методология моделирования с помощью (80) получила название методологии Бокса-Дженкинса.

Автокорреляционная функция процесса ARMA(p, q)

Рассмотрим, как можно получить автоковариационную и автокорреляционную функции стационарного процесса ARMA(p, q), зная параметры этого процесса. Для этого умножим обе часы уравнения (80) на \( x_{t-k} \), где \( k \geq 0 \), и перейдем к математическим ожиданиям:

\[
M(x_{t-k}x_t) = \varphi_1 M(x_{t-k}x_{t-1}) + \varphi_2 M(x_{t-k}x_{t-2}) + \cdots + \varphi_p M(x_{t-k}x_{t-p}) + \]

\[
+ M(x_{t-k}e_t) - \theta_1 M(x_{t-k}e_{t-1}) - \theta_2 M(x_{t-k}e_{t-2}) - \cdots - \theta_q M(x_{t-k}e_{t-q}).
\]

Обозначим через \( \delta_s \) кросс-ковариацию изучаемого ряда \( x_t \) и ошибки \( e_t \) с лагом \( s \), т. е.

\[
\delta_s = M(x_t e_{t-s}).
\]

Поскольку процесс стационарен, то эта кросс-ковариационная функция не зависит от момента времени \( t \). В этих обозначениях

\[
M(x_{t-k}e_{t-j}) = \delta_{j-k}.
\]

Получаем выражение для автокорреляционной функции:

\[
\gamma(k) = \varphi_1 \gamma(k-1) + \cdots + \varphi_p \gamma(k-p) + \delta_{-k} - \theta_1 \delta_{1-k} - \cdots - \theta_q \delta_{q-k}.
\]

(81)

Так как \( x_{t-k} \) зависит только от импульсов, которые произошли до момента \( t - k \), то

\[
\delta_{j-k} = M(x_{t-k}e_{t-j}) = 0 \text{ при } j < k.
\]

Для того чтобы найти остальные нужные нам кросс-ковариации, \( \delta_0, \ldots, \delta_q \), необходимо поочередно умножить все члены выражения (81) на \( e_t, e_{t-1}, \ldots, e_{t-q} \) и перейти к математическим ожиданиям. В итоге получится следующая система уравнений:

\[
\delta_0 = \sigma_e^2,
\]

\[
\delta_1 = \varphi_1 \delta_0 - \theta_1 \sigma_e^2,
\]

\[
\delta_2 = \varphi_1 \delta_1 + \varphi_2 \delta_0 - \theta_2 \sigma_e^2,
\]

\[
\ldots
\]

61
Общая формула для всех \( 1 \leq s \leq p \) имеет вид:
\[
\delta_s = \varphi_1 \delta_{s-1} + \cdots + \varphi_s \delta_0 - \theta_s \sigma_e^2.
\]
При \( s > p \) (такой случай может встретиться, если \( p < q \))
\[
\delta_s = \varphi_1 \delta_{s-1} + \cdots + \varphi_p \delta_{s-p} - \theta_s \sigma_e^2.
\]
Отсюда рекуррентно, предполагая \( \sigma_e^2 \) и параметры \( \varphi \) и \( \theta \) известными, найдем \( \delta_s \).

Далее, зная \( \delta_s \), по аналогии с уравнениями Юла-Уокера (69) по формуле (81) при \( k = 0, \ldots, p \) с учетом того, что \( \gamma(-k) = \gamma(k) \) найдем автоковариации \( \gamma_0, \ldots, \gamma_p \). Остальные автоковариации вычисляются рекуррентно по формуле (81).

Автокорреляции рассчитываются как \( \rho(k) = \gamma(k)/\gamma(0) \). Заметим, что если требуется найти только автокорреляции, то без потери общности можно взять ошибку \( e_t \) с единичной дисперсией: \( \sigma_e^2 = 1 \).

Если в уравнении (81) \( k > q \), то все кросс-корреляции равны нулю, поэтому
\[
\gamma(k) = \varphi_1 \gamma(k-1) + \cdots + \varphi_p \gamma(k-p), k > q.
\]
Поделив это выражение на \( \gamma(0) \), выводим уравнение автокорреляционной функции для \( k > q \):
\[
\rho(k) = \varphi_1 \rho_{k-1} + \cdots + \varphi_p \rho_{k-p},
\]
(82)
или
\[
\varphi(L) \rho_k = 0, \quad k > q.
\]

Таким образом, начиная с некоторой величины задержки, а точнее, когда \( q < p \), поведение автокорреляционной функции стационарного процесса ARMA(p, q) определяется, как и в случае чистой авторегрессии AR(p), однородным конечно-разностным уравнением (82). В свою очередь, решение этого конечно-разностного уравнения определяется корнями характеристического уравнения. То есть при \( q < p \) автокорреляционная функция будет состоять из комбинации затухающих экспонент и экспоненциально затухающих синусоид\(^8\).

По аналогии с AR(p) условия стационарности ARMA(p, q) определяются корнями характеристического уравнения: если эти корни лежат внутри единичного круга, то процесс стационарен.

Пример. Рассмотрим процесс авторегрессии AR(2) \( x_t = 4.375 + 0.25 x_{t-1} - 0.125 x_{t-2} + e_t \).

Характеристическое уравнение принимает в этом случае вид \( z^2 - 0.25z + 0.125 = 0 \), и имеет корни \( z_{1,2} = 0.125 \pm 0.3307 \cdot i \). Оба корня по абсолютной величине \( |z| = \sqrt{0.125^2 + 0.3307^2} = 0.3536 \) меньше единицы, так что процесс стационарный. Математическое ожидание этого процесса равно \( M(x_t) = \mu/(1 - \varphi_1 - \varphi_2) = 4.375/(1 - 0.25 + 0.125) = 5 \), так что траектории этого процесса колеблются вокруг уровня 5.

Для построения коррелограммы воспользуемся уравнениями Юла-Уокера. У нас \( p = 2 \), так что
\[
\rho(k) = 0.25 \rho(k - 1) - 0.125 \rho(k - 2), k > 0.
\]
По определению \( \rho(0) = 1 \). Для \( \rho(1) \) имеем
\[
\rho(1) = 0.25 \rho(0) - 0.125 \rho(-1) = 0.25 - 0.125 \rho(1),
\]
откуда находим:
\[
\rho(1) = 0.25 / (1 + 0.125) = \frac{2}{9} = 0.222.
\]

\(^8\) Заметим, что граничные условия у AR(p) другие, поэтому автокорреляционные функции не будут совпадать.
Далее последовательно находим:

\[ \rho(2) = 0.25\rho(1) - 0.125\rho(0) = 0.25 \cdot 0.222 - 0.125 = -0.069, \rho(3) = -0.045, \]
\[ \rho(4) = -0.003, \rho(5) = 0.005 \]

и т. д.

**Подбор стационарной модели ARMA для ряда наблюдений**

Если мы предполагаем, что некоторый наблюдаемый временной ряд \( x_1, x_2, ..., x_n \) порождается моделью ARMA, то при этом возникает проблема подбора конкретной модели из этого класса, решение которой предусматривает три этапа:

1) идентификация модели;
2) оценивание модели;
3) диагностика.

На этапе идентификации производится выбор некоторой частной модели из всего класса ARMA, т. е. выбор значений \( p \) и \( q \). Используемые при этом процедуры являются не вполне точными, что может при последующем анализе привести к выводу о непригодности идентифицированной модели и необходимости замены ее альтернативной моделью. На этом же этапе делаются предварительные грубые оценки коэффициентов \( \varphi_1, \varphi_2, ..., \varphi_p, \theta_1, \theta_2, ..., \theta_q \) идентифицированной модели.

На втором этапе производится уточнение оценок коэффициентов модели с использованием эффективных статистических методов. Для оцененных коэффициентов вычисляются приближенные стандартные ошибки, дающие возможность, при дополнительных предположениях о распределениях случайных величин \( X_1, X_2, ..., \), строить доверительные интервалы для этих коэффициентов и проверять гипотезы об их истинных значениях с целью уточнения спецификации модели.

На третьем этапе применяются различные диагностические процедуры проверки адекватности выбранной модели имеющимся данным (misspecification tests). Неадекватности, обнаруженные в процессе такой проверки, могут указать на необходимую корректировку модели, после чего производится новый цикл подбора, и т. д. до тех пор, пока не будет получена удовлетворительная модель. Разумеется, если мы имеем дело с ситуацией, когда уже имеется достаточно отработанная и разумно интерпретируемая модель эволюции того или иного показателя, можно обойтись и без этапа идентификации.

Если ряд порождается моделью ARMA\((p, q)\), то мы будем в дальнейшем для краткости обозначать это как \( x_t \sim ARMA(p, q) \). Соответственно, если ряд порождается моделью AR\((p)\), то \( x_t \sim AR(p) \), и если ряд порождается моделью MA\((q)\), то \( x_t \sim MA(q) \).

**Идентификация стационарной модели ARMA**

Основной отправной точкой для идентификации стационарной модели ARMA является различие поведения автокорреляционных (ACF) и частных автокорреляционных (PACF) функций (ACF – autocorrelation function, PACF – partial autocorrelation function) рядов, соответствующих различным моделям ARMA.

О поведении автокорреляционных функций для различных моделей ARMA мы уже говорили. Однако по поведению только автокорреляционной функции трудно идентифицировать даже порядок чистого (без MA составляющей) процесса авторегрессии. Решению этого вопроса помогает рассмотрение поведения частной автокорреляционной функции (PACF) стационарного процесса \( x_t \). Ее значение \( \rho_{part}(k) \) на лаге \( k \) определяется как значение коэффициента корреляции между случайными величинами \( x_t \) и \( x_{t+k} \), очищенными от влияния случайных величин \( x_{t+1}, ..., x_{t+k-1} \).
Замечательным является тот факт, что если \( x_t \) -- процесс типа AR(p), то тогда

\[
\rho_{part}(p) \neq 0, \rho_{part}(k) = 0 \text{ для } k > p.
\]

Это позволяет по графику PACF определять порядок процесса авторегрессии и отличать процесс авторегрессии от процессов скользящего среднего и ARMA(p, q) с \( q > 0 \).

Напомним, что зануление ACF после лага \( q \) соответствует процессу MA(q). Теперь же мы видим, что зануление PACF после лага \( p \) соответствует процессу AR(p). Поэтому идентификация этих моделей по ACFF и PACF более определена по сравнению с идентификацией моделей ARMA(p, q) с \( p \neq 0, q \neq 0 \).

В то же время вместо не известных нам истинных последовательностей автокорреляций \( \rho(k) \) и частных автокорреляций \( \rho_{part}(k) \) мы можем довольствоваться только их состоятельными оценками -- выборочной ACF, образованной выборочными автокорреляциями \( r(k) \), и выборочной PACF, образованной выборочными частными автокорреляциями \( r_{part}(k) \).

Получить последние можно, заменяя входящие в выражения для \( \rho_{part}(k) \) автокорреляции \( \rho(i) \) их оценками \( r(i) \). Однако проще поступить иначе, исходя из того, что \( \rho_{part}(k) \) является коэффициентом при \( x_{t-k} \) в линейной комбинации случайных величин \( x_{t-1}, \ldots, x_{t-k} \), наилучшим образом приближающей случайную величину \( x_t \). Можно просто оценить методом наименьших квадратов коэффициенты в модели

\[
x_t = \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \cdots + \varphi_k x_{t-k} + e_t.
\]

Полученная в результате оценка коэффициента \( \varphi_k \) и есть \( r_{part}(k) \).

Если \( x_t \) является стационарным процессом типа ARMA(p, q) и \( M(x_t^4) < \infty \), то указанные оценки \( r(k) \) и \( r_{part}(k) \) являются состоятельными оценками для \( \rho(k) \) и \( \rho_{part}(k) \), соответственно. Поскольку \( r(k) \) и \( r_{part}(k) \) всего лишь оценки для \( \rho(k) \) и \( \rho_{part}(k) \), то их наблюдаемые значения могут значительно отличаться от \( \rho(k) \) и \( \rho_{part}(k) \). Более того, характер изменения теоретической автокорреляционной функции вовсе не обязательно будет воспроизводиться в ее выборочном аналоге -- выборочной автокорреляционной функции.

Тем не менее во многих случаях поведение теоретических ACF и PACF в какой-то мере отражается на поведении их выборочных аналогов. Поэтому представление о поведении теоретических ACF и PACF может помочь в решении задачи идентификации соответствующих моделей в рамках общего класса моделей ARMA.

В распечатках анализа временных рядов вместе с графиками выборочных ACF и PACF обычно печатаются значения Q-статистики, относящиеся к критерию проверки гипотезы о том, что наблюдаемые данные являются реализацией процесса белого шума (т. е. отсутствует автокорреляция).

В пакете EVIEWS (Econometric Views) значения Q-статистики Люнга-Бокса распечатываются вместе с приближенными P-значениями, соответствующими распределениям \( \chi^2 \).

Границы для значения ACF вычисляются по формуле (39), причем принято по нулевой гипотезе, что \( r(k) = 0 \), т. е.

\[
\left[ -\frac{n-k}{\sqrt{n(n+2)}} \varepsilon_{1-\alpha}, \frac{n-k}{\sqrt{n(n+2)}} \varepsilon_{1-\alpha} \right],
\]

где \( \varepsilon_{1-\alpha} \) -- квантиль нормального распределения.
В процедурах анализа временных рядов обычно предусмотрена распечатка графиков выборочных ACF и PACF, на которые нанесены границы полосы ±2/√n. В этих границах с вероятностью, близкой к 0.95, должно заключаться значение r(k), если x_t — белый шум, и значение r_{part}(k), если x_t ~ AR(p). Здесь следует сделать одно важное предупреждение. Оба построенных критерия имеют уровень значимости, близкий к 0.05, только когда мы проверяем гипотезу H_0 при некотором фиксированном k.

Пример. Рассмотрим модель процесса AR(2) \( x_t = 0.5x_{t-1} - 0.7x_{t-2} + e_t \) (ряд был сгенерирован с помощью Ewiev). Характеристическое уравнение принимает в этом случае вид \( z^2 - 0.5z + 0.7 = 0 \) и имеет корни \( z_{1,2} = 0.35 \pm 0.798 \cdot i \), по модулю меньше 1, так что процесс, порождаемый такой моделью стационарен. Оценим модель. Для этого вычислим значения ACF (\( r(k), k = 1, \ldots, 12 \))

\[
r = [0.325, -0.567, -0.571, 0.111, 0.530, 0.221, -0.289, -0.339, 0.028, 0.282, 0.129, -0.148],
\]

и PACF

\[
r_{part} = [0.325, -0.752, -0.012, 0.033, 0.072, -0.025, 0.019, 0.009, -0.010, -0.000, -0.033, 0.013].
\]

Судя по значениям \( r(k) \) последнее максимально отличное от нуля значение \( r(k) \) при \( k = 5 \). Поэтому можно предположить, что это процесс MA(5). Оценим его

\[
x_t = 0 + 0.545e_{t-1} - 0.415e_{t-2} - 0.552e_{t-3} + 0.259e_{t-5} + e_t.
\]

Все коэффициенты значимы. \( R^2 = 0.555 \), сумма квадратов ошибок ESS = 188.516. Вероятность того, что значения \( r(k), k = 5, \ldots, 12 \) попадут в интервал (см. 83) равны \( Prob = 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000 \), где \( Prob \) сокращенно от английского probability (вероятность). Нуевые вероятности означают, что модель неудачная, так как остатки коррелированы между собой (присутствует автокорреляция). Попробуем увеличить q. Получим модель AR(10) \( x_t = 0.562e_{t-1} - 0.433e_{t-2} - 0.601e_{t-3} + 0.4998e_{t-5} + 0.340e_{t-6} - 0.224e_{t-7} - 0.297e_{t-8} + 0.182e_{t-10} + e_t \). Все коэффициенты значимы. \( R^2 = 0.613 \), сумма квадратов ошибок ESS = 163.8904. Вероятность того, что значения \( r(k), k = 9, \ldots, 16 \) попадут в интервал (см. 83), равны \( Prob = 0.139, 0.196, 0.347, 0.247, 0.192, 0.217, 0.299, 0.358 \).

Все вероятности больше 0.05, значит, нет автокорреляции и остатки образуют «белый шум».

Теперь будем использовать значения PACF. Видно, что начиная с q = 3 значения PACF близки к нулю. Поэтому это скорее процесс AR(2). Оценим его

\[
x_t = 0.576x_{t-1} - 0.761x_{t-1} + e_t.
\]

Все коэффициенты значимы. \( R^2 = 0.613 \), сумма квадратов ошибок ESS = 163.7040. Вероятность того, что значения \( r(k), k = 9, \ldots, 16 \) попадут в интервал (см. 83), равны \( Prob = 0.488, 0.718, 0.779, 0.889, 0.940, 0.971, 0.987, 0.992 \). Все вероятности больше 0.05, значит, нет автокорреляции и остатки образуют «белый шум». Здесь вероятности более высокие, чем в предыдущем случае.

Оценим модель ARMA(5, 3). Имеем

\[
x_t = -0.283x_{t-2} + 0.385x_{t-5} + 0.644e_{t-1} - 0.352e_{t-3} + e_t.
\]

Все коэффициенты значимы. \( R^2 = 0.60955 \), сумма квадратов ошибок ESS = 163.8445. Вероятность того, что значения \( r(k), k = 5, \ldots, 12 \) попадут в интервал (см. 83), равны \( Prob = 0.010, 0.015, 0.039, 0.071, 0.115, 0.167, 0.192, 0.239 \). Возможно есть небольшая автокорреляция.

В результате получили три модели. Какая из них лучшая? Более определенные выводы при выборе модели на первом этапе можно получить, применяя информационные критерии отбора моделей.
Использование информационных критериев

Если заранее ограничиваться рассмотрением только AR моделей, т. е. полагать, что процесс \( x_t \) следует модели AR\((k)\), с неизвестным истинным порядком \( k \), то для определения \( k \) в таких ситуациях долгое время использовался информационный критерий Акаике. Согласно этому критерию, среди альтернативных значений \( k \) выбирается значение, которое минимизирует величину

\[
AIC(k) = \ln\sigma_e^2 + \frac{2k}{n}, \tag{84}
\]

где \( n \) – количество наблюдений.

Впоследствии было выяснено, что оценка Акаи́ке несостоятельна и асимптотически переоценивает (завышает) истинное значение \( k_0 \) с ненулевой вероятностью.

В связи с этим были предложены состоятельные критерии. Одним из таких критериев является часто используемый в настоящее время информационный критерий Шварца – SIC (BIC),

\[
SIC = \ln\sigma_e^2 + k\frac{\ln n}{n}. \tag{85}
\]

В случае, когда имеем процесс ARIMA, эти критерии примут вид

\[
AIC = (1 + \ln 2\pi) + \ln\sigma_e^2 + 2\frac{p + q}{n}, \tag{86}
\]

\[
SIC = (1 + \ln 2\pi) + \ln\sigma_e^2 + 2\frac{p + q}{n}\ln n. \tag{87}
\]

Вернемся к примеру и выберем лучшую модель, руководствуясь информационными критериями. Вычислим все критерии для каждой модели и все данные сведем в таблицу.

<table>
<thead>
<tr>
<th>Номер модели</th>
<th>( R^2 )</th>
<th>AIC, SIC(BIC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.613075</td>
<td>2.718758 2.850690</td>
</tr>
<tr>
<td>2</td>
<td>0.613497</td>
<td>2.667872 2.701087</td>
</tr>
<tr>
<td>3</td>
<td>0.609553</td>
<td>2.704821 2.771960</td>
</tr>
</tbody>
</table>

По информационным критериям лучшей является 2 модель

\[
x_t = 0.576x_{t-1} - 0.761x_{t-1} + e_t.
\]

3.6. Модель авторегрессии проинтегрированного скользящего среднего ARIMA

Рассмотрим ряд вида \( x_t = a + bt + ct^2 + e_t \). Его полностью детерминированная часть, то, что мы назвали трендом, является параболической функцией времени. Очевидно, что этот ряд нестационарный. Математическое ожидание этого процесса зависит от времени. Приводится ли такой ряд взятием последовательных разностей к стационарному? Если мы возьмем первую последовательную разность, то получим:

\[
\Delta x_t = a + bt + ct^2 + e_t - a - b(t - 1) - c(t - 1)^2 - e_{t-1} = b + 2ct - c + (e_t - e_{t-1}).
\]
Степень полинома, описывающего тренд, понизилась на единицу. Если провести взятие второй разности, то останется $\Delta^2 x_t = \Delta x_t - \Delta x_{t-1} = 2c + (e_t - 2e_{t-1} + e_{t-2})$, то есть получим стационарный процесс. Правда, мы видим, как в уравнение начинает «проникать» скользящее среднее. Полученный двукратным взятием разностей стационарный процесс является процессом MA(2). Но, по крайней мере, взятием последовательных разностей исходный ряд с квадратичным трендом приводится к стационарному виду.

Характерной особенностью стационарных процессов типа ARMA(p, q) является то, что корни находятся внутри единичного круга. Если один или несколько корней лежат на единичной окружности или вне ее, то процесс нестационарен.

Теоретически можно предложить много различных типов нестационарных моделей ARMA(p, q), однако, как показывает практика, наиболее распространенным типом нестационарных стохастических процессов являются интегрированные процессы или, как их еще называют, процессы с единичным корнем. Единичным называют корень характеристического уравнения, равный действительной единице: $\lambda_i = 1$.

Рассмотрим в качестве примера следующий процесс ARMA(4, 2):

$$x_t = 2.8x_{t-1} - 3.1x_{t-2} + 1.7x_{t-3} - 0.4x_{t-4} + e_t + 0.5e_{t-1} - 0.4e_{t-2}.$$ 

Характеристическое уравнение этого процесса имеет следующие корни: $\lambda_1 = 0.8$, $\lambda_2 = 0.5 + 0.5 \cdot i$, $\lambda_3 = 0.5 - 0.5 \cdot i$, $\lambda_4 = 1$. Все корни лежат внутри единичного круга, кроме последнего, который является единичным. Эти корни изображены на рисунке 14.

Рис. 14. Единичный корень

Оператор авторегрессии этого процесса можно представить в следующем виде:

$$1 - 2.8L + 3.1L^2 - 1.7L^3 + 0.4L^4 = (1 - 1.8L + 1.3L^2 - 0.4L^3)(1 - L) = (1 - 1.8L + 1.3L^2 - 0.4L^3)\Delta,$$

где $\Delta = 1 - L$ - оператор первой разности. Введем обозначение $w_t = \Delta x_t = x_t - x_{t-1}$. Полученный процесс $w_t$ является стационарным процессом ARMA(3, 2), задаваемым уравнением: $w_t = 4.25w_{t-1} - 12w_{t-2} + 14.75w_{t-3} + e_t + 0.5e_{t-1} - 0.4e_{t-2}$.

В общем случае, если характеристическое уравнение процесса ARMA(p, d, q) содержит $d$ единичных корней, а все остальные корни по модулю меньше единицы, то $d$-я разность этого временного ряда

$$w_t = \Delta^d x_t = \varphi(L)(1 - L)^d x_t$$

может быть представлена как стационарный процесс ARMA(p, q):

$$\varphi(L)\Delta^d x_t = \Theta(L)e_t \text{ или } \varphi(L)w_t = \Theta(L)e_t.$$

67
Таким образом, если \( x_t \) — нестационарный процесс со стационарными разностями \( d \)-го порядка, т. е. \( w_t = \Delta^d x_t \) — стационарный процесс, а \( \Delta^{d-1} x_t \) — нестационарный. Это означает, что \( x_t \) интегрируем \( d \)-го порядка. Если \( w_t \) — процесс ARMA(\( p, q \)), т. е.

\[
w_t = \varphi_0 + \varphi_1 w_{t-1} + \cdots + \varphi_p w_{t-p} - \theta_1 e_{t-1} - \cdots - \theta_q e_{t-q} + e_t,
\]

тогда \( x_t \) называется процессом ARIMA(\( p, d, q \)). Часто свободный член \( \varphi_0 \) опускается.

Нестационарный процесс, первые разности которого стационарны, называют интегрированным первого порядка и обозначают I(1). Стационарный процесс обозначают I(0). Если \( d \)-е разности случайного процесса стационарны, то его называют интегрированным \( d \)-го порядка и обозначают I(\( d \)). Таким образом, I(\( d \)) — такой процесс, \( d \)-е разности которого являются I(0).

Большинство эмпирических временных рядов можно считать реализациями процессов ARIMA. Другими словами, для большинства временных рядов может быть найден процесс ARIMA, или модель ARIMA, и именно этот процесс можно считать сгенерированным данным конкретный временной ряд. Основная задача в анализе временных рядов — специфицировать порядок модели ARIMA(\( p, d, q \)) в соответствии со свойствами временного ряда и оценить посредством статистических методов параметры уравнения модели и дисперсию остатков. Как уже отмечалось, проблема состоит в том, что обычно мы имеем только одну реализацию изучаемого процесса.

**Идентификация модели и оценивание параметров**

Задачу построения модели типа ARIMA по реализации случайного процесса Бокс и Дженкинс предложили разбить на несколько этапов.

I этап

1. Установить порядок интеграции \( d \), то есть добиться стационарности ряда, взяв достаточное количество последовательных разностей. Другими словами, «остационаризовать» ряд.

2. После этого мы получаем временной ряд \( y_t \), к которому нужно подобрать уже ARMA(\( p, q \)). Исходя из поведения автокорреляционной (ACF) и частной автокорреляционной функции (PACF), установить параметры \( p \) и \( q \). I этап принято называть **идентификацией модели ARIMA(\( p, d, q \)). Это всего лишь определение величин \( p, d, q \), но именно в такой последовательности: сначала \( d \), а потом \( p \) и \( q \).

II этап

Оценивание коэффициентов \( \varphi_1, \varphi_2, \ldots, \varphi_p, \theta_1, \theta_2, \ldots, \theta_q \) при условии, что мы уже знаем \( p \) и \( q \).

III этап

Стандартная для эконометрического подхода процедура. По остаткам осуществляется тестирование или диагностика построенной модели.

IV этап

Использование модели в основном для прогнозирования будущих значений временного ряда.

**3.7. Проверка на стационарность. Интеграционная статистика Дарбина-Уотсона**

Одним из важнейших условий получения корректных оценок в регрессионных моделях является требование стационарности переменных. В экономике довольно часто встречаются стационарные ряды, например, уровень безработицы. Однако, как правило, экономические процессы описываются нестационарными рядами: объем производства, уровень цен и т. д.
Наиболее простой способ проверки на стационарность временного ряда — применение интеграционной статистики Дарбина-Уотсона (IDW-статистики) для авторегрессии первого порядка вида

\[ x_t = \varphi x_{t-1} + e_t. \]  

(88)

Разработанная на основе статистики Дарбина-Уотсона для анализа автокорреляции остатков IDW-статистика имеет следующий вид:

\[ IDW = \frac{\sum (x_t - x_{t-1})^2}{\sum (x_t - \bar{x}_t)^2}, \]

gде \( x_t \) — временной ряд, являющийся реализацией процесса \( X_t \); \( \bar{x}_t \) — выборочное среднее \( x_t \).

Если временной ряд \( x_t \) — нестационарный, т. е. в уравнении (88) \( \varphi = 1 \), тогда имеем выражение в числителе \( \sum (x_t - x_{t-1})^2 = \sum e_t^2 \). Ясно, что для нестационарного ряда это отношение будет близко к 0. Можно сказать, что процесс \( x_t \) — не стационарный, если значение \( IDW \approx 0 \), и достаточно уверенно утверждать, что \( x_t \) — стационарный, если значение \( IDW \approx 2 \).

Утверждение о стационарности процесса не требует подтверждения результатами других тестов, однако нестационарность ставит задачу определения порядка интегрируемости либо заключения о том, что процесс неинтегрируем вообще.

Как правило, исследователю не известно заранее, какие компоненты содержит временной ряд, включает ли он свободный член или тренд. Поэтому использование интеграционной статистики Дарбина-Уотсона на этапе оценки интегрируемости временных рядов без применения дополнительных тестов может привести к ошибочным выводам и повлечь за собой неправильную спецификацию регрессионных уравнений.

Для оценки стационарности или порядка интегрируемости рассматриваемых временных рядов необходимо сопоставить расчетные значения IDW-статистики с критическими. Поскольку распределение IDW-статистики не соответствует ни одному из известных теоретических распределений, критические значения будут представлены не единичными значениями, а интервалами прямой в окрестности точки 2. Для выявления нестационарных временных рядов таблица критических значений составляется из отрезков прямой в окрестности точки 0.

Критические значения из табл. 15 применяются для проверки гипотезы \( H_0: IDW = 2 \) (рассматриваемый процесс стационарный) и альтернативной ей гипотезы \( H_1: IDW \neq 2 \) (рассматриваемый процесс не является стационарным).

<table>
<thead>
<tr>
<th>Число наблюдений</th>
<th>Уровень значимости</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>( IDW_L )</td>
</tr>
<tr>
<td>24</td>
<td>1.0177</td>
</tr>
<tr>
<td>89</td>
<td>1.4340</td>
</tr>
<tr>
<td>119</td>
<td>1.5082</td>
</tr>
<tr>
<td>151</td>
<td>1.5548</td>
</tr>
</tbody>
</table>
В табл. 16 представлены критические значения для проверки гипотезы \( H_0^* : IDW = 0 \) (процесс нестационарный) и альтернативной гипотезы \( H_1^* : IDW \neq 0 \) (процесс не является нестационарным). Механизм проверки гипотезы о стационарности и нестационарности временного ряда представлен в табл. 17.

Таблица 16

**Критические значения интеграционной статистики Дарбина-Уотсона для оценки нестационарных временных рядов**

<table>
<thead>
<tr>
<th>Число наблюдений</th>
<th>Уровень значимости</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>IDW(_L)</td>
</tr>
<tr>
<td>24</td>
<td>0.9409</td>
</tr>
<tr>
<td>89</td>
<td>0.4309</td>
</tr>
<tr>
<td>119</td>
<td>0.3255</td>
</tr>
<tr>
<td>151</td>
<td>0.2590</td>
</tr>
</tbody>
</table>

Таким образом, для применения интеграционного критерия Дарбина-Уотсона расчетное значение IDW-статистики необходимо сравнить с нижним критическим значением из табл. 16, и, если выполняется соотношение \( \text{IDW}_{\text{расч}} < \text{IDW}_{L} \), тогда на соответствующем уровне значимости гипотеза о нестационарности временного ряда не может быть отклонена. Если между расчетным и верхним критическим значением из табл. 16 выполняется соотношение \( \text{IDW}_{\text{расч}} > \text{IDW}_{U} \), тогда на соответствующем уровне значимости нет оснований отклонить гипотезу о стационарности временного ряда.

Таблица 17

**Механизм проверки гипотезы о стационарности временных рядов**

<table>
<thead>
<tr>
<th>Тестируемый временной ряд нестационарен. Нет оснований отклонить гипотезу ( H_0^* )</th>
<th>Зона неопределенности</th>
<th>Тестируемый временной ряд стационарен. Нет оснований отклонить гипотезу ( H_0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 0 )</td>
<td>( IDW_{L} )</td>
<td>( IDW_{U} )</td>
</tr>
</tbody>
</table>

3.8. **Прогнозирование с помощью ARMA моделей**

Прогнозирование будущих значений экономической величины является одним из основных способов применения моделей временных рядов. Использование моделей типа ARMA обладает некоторыми особенностями по сравнению с прогнозированием по модели множественной регрессии. При прогнозировании по правильно специфицированной модели, вообще говоря, существуют 2 источника ошибок прогноза:

- неопределенность будущих значений случайной величины \( e \);
- отсутствие точных значений коэффициентов модели (у нас есть только их оценки, оцененные по имеющейся выборке).
При прогнозировании по модели множественной регрессии мы оцениваем значение зависимой переменной при заданных значениях независимых переменных – регрессоров. Это приводит к тому, что характеристики прогноза, как случайной величины, являются, по сути, условными характеристиками при условии имеющейся выборки независимых переменных. Иная ситуация в моделях типа ARIMA. Значение переменной прогнозируется для некоторого будущего момента времени, при этом лаговые значения этой переменной, служащие регрессорами модели, можно рассматривать фиксированными на выборочных значениях, или случайными. Первая возможность приводит к условному прогнозу, как и для модели множественной регрессии, а вторая – к безусловному прогнозу. Таким образом, при прогнозировании по модели типа ARIMA можно рассматривать как условный, так и безусловный прогнозы. Из курса теории вероятностей известно, что условная дисперсия случайной величины не превышает ее безусловную дисперсию, поэтому точность условного прогноза всегда выше.

При прогнозировании по модели ARIMA от имеющейся выборки зависят как оценки коэффициентов модели, так и значения регрессоров, поэтому сложно аналитически выразить условную дисперсию ошибки прогноза через имеющиеся значения временного ряда. Общепринято ограничиваться не очень реалистичным предположением о том, что коэффициенты модели известны точно. Разумеется, это предположение уменьшает дисперсию ошибки прогноза, чем увеличивает кажущуюся точность как условного, так и безусловного прогнозов.

Показано, что если мы хотим достичь минимума среднеквадратической ошибки (MSE), т. е. не требовать несмещенности, то надо взять условное математическое ожидание: $M(x_{t+h}|x_1, ..., x_t)$. Такое условное математическое ожидание будет гарантировать получение MSE, или иногда ее обозначают MMSE, т. е. Minimum mean square error.

Процесс MA(q)

Начнем с модели MA(q): $x_t = \theta_0 + e_t + \theta_1 e_{t-1} + \cdots + \theta_q e_{t-q}$. Мы полагаем, что коэффициенты модели точно известны и что имеются значения $x_t$ для $t \in [1, n]$. Очевидно, что безусловным точечным прогнозом для любого момента времени будет математическое ожидание процесса, т. е. $\theta_0$. Условным прогнозом для момента времени $n + 1$ будет условное математическое ожидание

$$\bar{x}_{n+1} = M(\theta_0 + \theta_{n+1} + \theta_1 e_n + \cdots + \theta_q e_{n-q+1}|x_1, ..., x_n).$$

Среди случайных величин $e$, которые стоят в левой части, есть такие, которые связаны с имеющимися наблюдениями. Ведь наблюдение складывается из «модельного» значения и ошибки, поэтому условные математические ожидания всех слагаемых, кроме $e_{n+1}$, не равны нулю. Рассмотрим сначала $M(e_n|x_1, ..., x_n)$, потому что схема одна и та же. Это математическое ожидание – остаток между наблюдением и расчетом, прогнозом по модели, т. е. $e_n = x_n - \bar{x}_n$. Поэтому условные математические ожидания от всех предыдущих значений случайной составляющей надо заменить соответствующими остатками. Точно так же строится прогноз не на 1, а на 2 и вообще на $h$ шагов вперед. Все последующие $e$ заменяются нулями, а предыдущие – заменяются реально наблюдаемыми остатками. Таким образом, для модели MA(q) прогноз зависит от того, какие ошибки были на предыдущих шагах. Начиная с шага $(q + 1)$, условный прогноз представляет собой просто математическое ожидание $\theta_0$, т. е. условный прогноз совпадает с безусловным. Рассмотрим условную дисперсию ошибки прогноза на 1 шаг:

$$D(x_{n+1} - \bar{x}_{n+1}|x_1, ..., x_n) = M\left((\theta_0 + e_{n+1} + \theta_1 e_n + \cdots + \theta_q e_{n-q+1} + \cdots - \theta_0 - \theta_1 \epsilon_1 - \cdots - \theta_q \epsilon_{n-q+1})|x_1, ..., x_n\right)^2 = \sigma_e^2.$$
Аналогично дисперсия прогноза на 2 шага равна \(D(x_{n+2} - \tilde{x}_{n+2}|x_1, \ldots, x_n) = (1 + \theta_1^2)\sigma^2_e\), а дисперсия прогноза на \(h\) шагов составляет \((1 + \theta_1^2 + \cdots + \theta_h^2)\sigma^2_e\) при \(h < q\). При \(h \geq q\) дисперсия ошибки условного прогноза становится равной дисперсии ошибки безусловного прогноза, т. е. просто дисперсии случайного процесса \(x_t\).

Теперь рассмотрим модель стационарного процесса AR(p):

\[x_t = \mu + \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \cdots + \varphi_p x_{t-p} + e_t.\]

Для прогноза на 1 шаг вперед можно записать:

\[
\hat{x}_{n+1} = M(x_{n+1}|x_1, \ldots, x_n) = M(\mu + \varphi_1 x_n + \cdots + \varphi_p x_{n-p} | x_1, \ldots, x_n) = \mu + \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \cdots + \varphi_p x_{t-p}.
\]

Мы просто подставляем в уравнение модели \(p\) предыдущих значений.

Дисперсия ошибки прогноза на 2 шага равна:

\[
\sigma^2_e = \frac{\mu}{1 - \varphi_1 - \varphi_2 - \cdots - \varphi_p}.
\]

Условную дисперсию ошибки прогноза можно рассчитать аналогично случаю модели скользящего среднего, но выкладки становятся весьма громоздкими, даже для моделей малого порядка. Рассмотрим, например, модель AR(2) без свободного члена. Тогда \(\hat{x}_{n+1} = \varphi_1 x_n + \varphi_2 x_{n-1}\), а \(x_{n+1} = \varphi_1 x_t + \varphi_2 x_{t-1} + e_t\). Очевидно, что дисперсия ошибки прогноза на 1 шаг равна \(\sigma^2_e\). Для прогноза на 2 шага соответственно получаем:

\[
\hat{x}_{n+2} = \varphi_1 \hat{x}_{n+1} + \varphi_2 x_n = \varphi_1(\varphi_1 x_n + \varphi_2 x_{n-1}) + \varphi_2 x_n,
\]

\[
x_{n+2} = \varphi_1 x_{n+1} + \varphi_2 x_n + e_{n+2} = \varphi_1(\varphi_1 x_n + \varphi_2 x_{n-1} + e_{n+1}) + \varphi_2 x_n + e_{n+2}.
\]

Дисперсия ошибки прогноза на 2 шага равна: \((1 + \varphi^2_1)\sigma^2_e\). Очевидно, что дисперсия ошибки увеличивается от шага к шагу.

Значительно более простые выражения для дисперсии ошибки прогноза получаются, если перейти от AR(p) представления модели к эквивалентному MA представлению \(x_t = \mu + e_t + \theta_1 e_{t-1} + \cdots + \theta_q e_{t-q} + \cdots\), хотя и с бесконечным числом слагаемых. Тогда дисперсия ошибки прогноза на \(h\) шагов выражается очевидной формулой \(\sigma^2_e \sum_{i=0}^{h-1} \theta_i^2\), где \(\theta_1 = 1\).

Для общей модели ARMA(p, q) нужно просто объединить то, что мы сейчас получили. Если мы хотим получить MMSE прогноз, то мы рассчитываем прогнозные значения по нашей модели, подставляя туда для времени \([1, n]\) – известные значения \(x\) и рассчитанные значения остатков, а для всех последующих моментов времени заменяем остатки нулями, а для значений \(x\) подставляем их прогнозные значения. Для получения дисперсии ошибки прогноза переходим к MA представлению и пользуемся только полученной формулой.

Во всех рассмотренных случаях условный точечный прогноз асимптотически приближался к математическому ожиданию ряда, а дисперсия ошибки прогноза к дисперсии ряда. Это означает, что для стационарного процесса влияние имеющейся информации о реализации на прогноз и его точность асимптотически убывает до нуля. К тому же при увеличении горизонта прогноза дисперсия ошибки не превышает дисперсии временного ряда. Напомним, что этот результат является следствием нереалистичного предположения о том, что коэффициенты модели известны точно.

Так же, как и в моделях множественной линейной регрессии, хорошее качество подгонки модели ARIMA не гарантирует высокой точности прогноза, т. е. высокой прогнозной силы. Для оценки прогнозных свойств модели одним из общепринятых приемов является разбиение имеющейся реализации на 2 части. По первым \(m\) наблюдениям выбирается и оценивается модель, а по последним \((n - m)\) наблюдениям проводится сравнение наблюденных и рассчитанных по модели значений. Такая процедура иногда называется постпрогнозом. Сравнивая прогнозную силу моделей, мы выбираем «наилучшую».

72
Мы вывели формулы для расчета точечного прогноза по модели ARMA и дисперсии этого прогноза. Если дополнительно предположить, что ошибки $e_t$ подчиняются нормальному закону (т. е. представляют собой гауссовский процесс), то можно получить также интервальный прогноз. При этом предположении при известных значениях процесса до момента $n$ распределение будущего значения процесса $x_{n+h}$ (т. е. условное распределение $x_{n+h} \mid \Omega_n$) также будет нормальным со средним значением $x_n(h)$ и дисперсией $\sigma^2_e$:

$$x_{n+h} \mid \Omega_n \sim N(x_n(h), \sigma_p^2),$$

где $\sigma_p^2 = \sigma_e^2 \left[ 1 + \sum_{i=1}^{h-1} \theta_i^2 \right]$ — средний квадрат ошибки прогноза. Учитывая это, получаем доверительный интервал для $x_{n+h}$, т. е. интервальный прогноз:

$$\left[ x_n(h) - \zeta_{1-\alpha} \sigma_p, x_n(h) + \zeta_{1-\alpha} \sigma_p \right]$$

или

$$\left[ x_n(h) - \zeta_{1-\alpha} \sigma_e \sqrt{\sum_{i=0}^{h-1} \theta_i^2}, x_n(h) - \zeta_{1-\alpha} \sigma_e \sqrt{\sum_{i=0}^{h-1} \theta_i^2} \right]. \quad (89)$$

где $\zeta_{1-\alpha}$ — двусторонний $(1 - \alpha)$-квантиль стандартного нормального распределения. Это $(1 - \alpha) \cdot 100$-процентный доверительный интервал.

### 3.9. ARCH и GARCH

Традиционные модели временных рядов, такие как модель ARMA, не могут адекватно учесть все характеристики, которыми обладают финансовые временные ряды, и требуют расширения. Одна из характерных особенностей финансовых рынков состоит в том, что присущая рынку неопределенность изменяется во времени. Как следствие, наблюдается «кластеризация волатильности». Имеется в виду чередование периодов, когда финансовый показатель ведет себя непостоянно и относительно спокойно. На рисунке 15 для иллюстрации этого явления показаны темпы прироста индекса РТС\(^{10}\) за несколько лет.

Рис. 15. Темпы роста РТС с 1.09.1995 г. по 13.09.2007 г., в процентах

---

9 Здесь мы ввели сокращенную запись $\Omega_n = (x_1, ..., x_n)$

На графике период 1 – бурный, период 2 – сравнительно спокойный, период 3 – еще более бурный, 4 – более спокойный. Термин «волатильность» (volatility – англ. изменчивость, непостоянство) используется, как правило, для неформального обозначения степени вариабельности, разброса переменной. Формальной мерой волатильности служит дисперсия (или среднеквадратическое отклонение). Эффект кластеризации волатильности отмечен в таких рядах, как изменение цен акций, валютных курсов, доходности спекулятивных активов.

**Модель ARCH**

Модель ARCH, т. е. модель с авторегрессионной условной гетероскедастичностью (autoregressive conditional heteroskedasticity), предложена Р. Энглом в 1982 г. для моделирования кластеризации волатильности. Процесс ARCH q-го порядка задается следующими соотношениями:

\[
e_t | \Omega_{t-1} \sim N(0, \sigma_t^2),
\]
\[
\sigma_t^2 = \omega + \gamma_1 e_{t-1}^2 + \cdots + \gamma_q e_{t-q}^2.
\]

Здесь \( \Omega_{t-1} = (e_{t-1}, e_{t-2}, \ldots) \) – предыстория процесса \( e_t \), а \( \sigma_t^2 \) – условная по предыстории дисперсия \( e_t \), т. е. \( \sigma_t^2 = D(e_t | \Omega_{t-1}) = M(e_t^2 | \Omega_{t-1}) \). Условную дисперсцию часто называют волатильностью процесса. Для того чтобы условная дисперсия оставалась положительной, требуется выполнение соотношений \( \Omega > 0 \) и \( \gamma_1, \ldots, \gamma_q \geq 0 \). Данный процесс можно записать несколько иначе:

\[
e_t \sim N(0,1),
\]
\[
e_t = \varepsilon_t \sigma_t,
\]
\[
\sigma_t^2 = \omega + \gamma_1 e_{t-1}^2 + \cdots + \gamma_q e_{t-q}^2,
\]

где \( \varepsilon_t \) нормально распределены и независимы. Такая запись удобна тем, что нормированный случайный процесс \( \varepsilon_t \) не зависит от предыстории.

Смысл модели ARCH состоит в том, что если абсолютная величина \( e_t \) оказывается большой, то это приводит к повышению условной дисперсии в последующие периоды. В свою очередь, при высокой условной дисперсии более вероятно появление больших (по абсолютной величине) значений \( e_t \). Наоборот, если значения \( e_t \) в течение нескольких периодов близки к 0, то это приводит к понижению условной дисперсии в последующие периоды практически до уровня \( \omega \). В свою очередь, при низкой условной дисперсии более вероятно появление малых (по абсолютной величине) значений \( e_t \). Таким образом, ARCH-процесс характеризуется инерционностью условной дисперсии (кластеризацией волатильности).

Если обозначить разницу между величиной \( e_t^2 \) и ее условным математическим ожиданием \( \sigma_t^2 \) через \( \eta_t \), то получится следующая эквивалентная запись процесса ARCH:

\[
e_t^2 = \Omega + \gamma_1 e_{t-1}^2 + \cdots + \gamma_q e_{t-q}^2 + \eta_t.
\]

Поскольку условное математическое ожидание \( \eta_t \) равно 0, то безусловное математическое ожидание также равно 0. Кроме того, как можно показать, \( \eta_t \) не автокоррелирован. Следовательно, квадраты процесса ARCH(q) следуют авторегрессионному процессу q-го порядка.

Если все корни характеристического уравнения

\[
z^q - \gamma_1 z^{q-1} - \cdots - \gamma_q = 0
\]

лежат внутри единичного круга, то у процесса ARCH(q) существует безусловная дисперсия, и он является слабо стационарным. Поскольку коэффициенты \( \gamma_j \) неотрицательны, то это условие эквивалентно условию \( \sum_{j=1}^{q} \gamma_j < 1 \).
Действительно, вычислим безусловную дисперсию стационарного ARCH-процесса, которую мы обозначим через $\sigma^2$. Для этого возьмем математическое ожидание от обеих частей уравнения условной дисперсии (90):

$$M(\sigma_t^2) = \Omega + \gamma_1 M(e_{t-1}^2) + \cdots + \gamma_q M(e_{t-q}^2).$$

Заметим, что $M(\sigma_t^2) = M(M(e_t^2|\Omega_{t-1})) = M(e_t^2) = D(e_t^2) = \sigma^2$, т. е. математическое ожидание условной дисперсии равно безусловной дисперсии. Следовательно,

$$\sigma^2 = \omega + \gamma_1 \sigma^2 + \cdots + \gamma_q \sigma^2,$$

или

$$\sigma^2 = \frac{\omega}{1 - \gamma_1 - \cdots - \gamma_q}.$$

Таким образом, для всех $e_t$ безусловная дисперсия одинакова, т. е. имеет место гомоскедастичность. Однако условная дисперсия меняется, поэтому одновременно имеет место условная гетероскедастичность.

Если не все корни приведенного выше характеристического уравнения лежат внутри единичного круга, т. е. если $q \sum_{j=1}^{q} \gamma_j \geq 1$, то безусловная дисперсия не существует, и поэтому ARCH-процесс не будет слабо стационарным.

Еще одно свойство ARCH-процессов состоит в том, что безусловное распределение $e_t$ имеет более высокий куртозис (экссесс) (т. е. более толстые хвосты и острую вершину), чем нормальное распределение. У ARCH(1) экссес равен

$$\frac{M(e_t^4)}{\sigma^4} - 3 = \frac{6\gamma_1^2}{1 - 3\gamma_1^2},$$

причем при $3\gamma_1^2 = 1$ четвертый момент распределения не существует (экссес равен бесконечности). Это свойство ARCH-процессов хорошо соответствует финансовым временным рядам, которые обычно характеризуются толстыми (тяжелыми) хвостами. На рисунке 16 изображен график плотности безусловного распределения ARCH(1). Для сравнения на графике приведена плотность нормального распределения.

---

11 Она называется авторегрессионной, поскольку динамика квадратов ARCH-процесса описывается авторегрессией.

12 При этом у ARCH-процессов есть интересная особенность: они могут быть строго стационарны, не будучи слабо стационарны. Дело в том, что определение слабой стационарности требует существования конечных первых и вторых моментов ряда. Строгая же стационарность этого не требует, поэтому даже если условная дисперсия бесконечна (и, следовательно, ряд не является слабо стационарным), ряд все же может быть строго стационарным.
Получить состоятельные оценки коэффициентов ARCH-процесса можно, используя вышеприведенное представление его квадратов в виде авторегрессии (91). Более эффективные оценки получаются при использовании метода максимального правдоподобия.

При применении ARCH-моделей к реальным данным было замечено, что модель ARCH(1) не дает достаточно длительных кластеров волатильности, а только порождает большое число выбросов (выделяющихся наблюдений). Для корректного описания данных требуется довольно большая длина лага q, что создает трудности при оценивании. В частности, зачастую нарушается условие неотрицательности оценок \( \gamma_j \).

Модель GARCH

Модель GARCH (generalized ARCH – обобщенная модель ARCH), предложенная Т. Боллерслевом, является альтернативной модификацией модели ARCH (91), позволяющей получить более длинные кластеры при малом числе параметров. Модель ARMA зачастую позволяет получить более сжатое описание временных зависимостей для условного математического ожидания, чем модель AR. Подобным же образом модель GARCH дает возможность обойтись меньшим количеством параметров по сравнению с моделью ARCH, если речь идет об условной дисперсии.

Модель GARCH(p, q) имеет вид:

\[
\sigma_t^2 = \omega + \delta_1 \sigma_{t-1}^2 + \cdots + \delta_p \sigma_{t-p}^2 + \gamma_1 e_{t-1}^2 + \cdots + \gamma_q e_{t-q}^2 = \omega + \sum_{j=1}^{p} \delta_j \sigma_{t-j}^2 + \sum_{j=1}^{q} \gamma_j e_{t-j}^2. \tag{92}
\]

При этом предполагается, что \( \omega > 0, \delta_1, \ldots, \delta_p \geq 0 \) и \( \gamma_1, \ldots, \gamma_q \geq 0 \). На практике, как правило, достаточно взять \( p = 1 \) и \( q = 1 \). Изредка используют GARCH(1, 2) или GARCH(2, 1).

Как и в модели ARCH, \( \sigma_t^2 \) служит условной дисперсией процесса:

\[ e_t | \Omega_t \sim N(0, \sigma_t^2). \]

Рассчитаем безусловную дисперсию GARCH-процесса, предполагая, что он стационарен. Для этого возьмем математические ожидания от обеих частей уравнения (92) для условной дисперсии:

\[
M(\sigma_t^2) = \sum_{j=1}^{p} \delta_j M(\sigma_{t-j}^2) + \sum_{j=1}^{q} \gamma_j M(e_{t-j}^2),
\]

откуда

\[
\sigma^2 = \sum_{j=1}^{p} \delta_j \sigma^2 + \sum_{j=1}^{q} \gamma_j \sigma^2,
\]

и

\[
\sigma^2 = \frac{1}{1 - \sum_{j=1}^{p} \delta_j - \sum_{j=1}^{q} \gamma_j}.
\]

Таким образом, с точки зрения безусловной дисперсии GARCH-процесс гомоскедастичен.

Для того чтобы дисперсия была конечной, необходимо выполнение условия \( \sum_{j=1}^{p} \delta_j + \sum_{j=1}^{q} \gamma_j < 1 \). В частности, для модели GARCH(1, 1) требуется \( \delta_1 + \gamma_1 < 1 \).

С точки зрения прогнозирования перспективной является модель, сочетающая ARIMA с GARCH. Модель ARIMA в этом случае используется для моделирования поведения условного математического ожидания ряда, а GARCH – для моделирования условной дисперсии.
В обычных моделях временного ряда с неизменными условными дисперсиями (например, ARMA) неопределенность ошибки прогноза — это возрастающая функция горизонта прогноза, которая не зависит от момента прогноза. Однако в присутствии ARCH-ошибок точность прогноза будет непропорционально зависит от текущей информации и, следовательно, от момента прогноза. Поэтому для корректного построения интервальных прогнозов требуется иметь оценки будущих условных дисперсий ошибки.

Кроме того, в некоторых случаях полезно иметь прогнозы не только (условного) математического ожидания изучаемой переменной, но и ее (условной) дисперсии. Это важно, например, при принятии решений об инвестициях в финансовые активы. В этом случае дисперсия (волнатильность) доходности естественно рассматривать как меру рискованности финансового актива. Таким образом, сами по себе прогнозы условной дисперсии могут иметь практическое применение.

4. Нестационарные временные ряды

4.1. Ряды TS и DS

Сейчас мы переходим к рассмотрению нестационарных временных рядов. Теперь ситуация изменилась. Из теоремы Вольда следует, что модели типа ARMA охватывают все стационарные процессы. А вот с нестационарными временными рядами ситуация иная, фактически мы будем рассматривать только частные виды нестационарных временных рядов. С одним из таких видов мы уже встречались.

По определению модели ARIMA(p, d, q), d — это степень интеграции ряда, т. е. ряд становится стационарным после применения d раз операции взятия последовательной разности. Мы начнем рассмотрение именно с нестационарных рядов, которые могут быть приведены к стационарному виду с помощью взятия последовательных разностей. Такие ряды обозначаются I(d) и называются интегрированными рядами порядка d.

Естественно, мы начнем с простейшего вида ряда I(1). Оказалось, что 2 разных по свойствам типов нестационарных рядов приводятся к стационарному виду с помощью взятия последовательных разностей. Мы их уже рассматривали.

1 тип: процесс с детерминированным полиномиальным трендом.

\[ x_t = P_k(t) + e_t, \text{ где } P_k(t) — полином степени } k \text{ от } t, \text{ а } e_t — стационарный процесс, не обязательно белый шум. Если ограничиться рассмотрением только линейного тренда } x_t = a + bt + e_t, \text{ то можно записать: } \Delta x_t = x_t - x_{t-1} = (1 - L)x_t = b + (e_t - e_{t-1}). \text{ Поскольку } e_t — стационарный процесс, то его первая разность — также стационарный процесс, хотя если } e_t — белый шум, то появляется MA-часть. В случае полиномиального тренда для приведения к стационарному виду нужно взять последовательную разность несколько раз.

2 тип: процесс случайного блуждания c дрейфом, если \( b \neq 0 \): \( x_t = m + x_{t-1} + e_t \). В этом случае \( \Delta x_t = m + e_t \), и процесс \( x_t \) называется случайным блужданием c дрейфом. Мы можем записать решение этого разностного уравнения в следующем виде:

\[ x_t = m + m + x_{t-2} + e_t + e_{t-1} = \ldots = mt + \sum_{j=0}^{t-1} e_{t-j}. \]

Дисперсия процесса случайного блуждания с дрейфом \( \sigma_x^2 = D(e_{t-1}) + \ldots + D(e_1) = t\sigma_e^2 \) растет вместе с ростом \( t \). Поэтому этот процесс не является стационарным. Заметим, что соответствующее характеристическое уравнение имеет единичный корень.
Если применить подход Бокса-Дженкинса и, имея некоторую реализацию, перейти к разностям, оценить модель ARMA(p, q), то, чтобы вернуться назад к процессу $x_t$, нужно выбрать, по какой схеме возвращаться. Как мы уже рассматривали, эти процессы ведут себя по-разному. В чем-то они схожи: у обоих есть линейный тренд, но они отличаются случайной частью. В первом случайная часть — это текущий шок, текущие возмущения, а во втором — это накопленные возмущения от всех предыдущих шоков. Если случайное блуждание можно привести к стационарному виду только взятием первой разности, то ряд первого типа можно привести к стационарному виду также выделением линейного тренда, например, построив линейную регрессию на время и рассмотрев стационарный остаток. Применим ли такой подход к случайному блужданию? Что произойдет, если на самом деле процесс является случайным блужданием, пусть даже для простоты с нулевым математическим ожиданием, т. е. имеет вид $\Delta x_t = e_t$, а мы строим регрессию на время вида $x_t = a + bt + e_t$? Можно интуитивно догадаться, что мы получим значимый по $t$-статистике коэффициент $b$. Метод наименьших квадратов как бы преобразует непостоянную дисперсию в значимый тренд, т. е. в непостоянное математическое ожидание. Таким образом, мы эти 2 типа процесса спутаем. Вопрос о том, насколько опасно путать эти 2 процесса между собой, привлек внимание относительно поздно. Давайте введем следующие названия для этих типов нестационарных процессов.

1 тип: процесс, приводимый к стационарному путем выделения линейного тренда — TSP (trend stationary process). Это процесс вида $x_t = a + bt + e_t$, он приводится к стационарному процессу путем включения в регрессию линейного тренда. Это, в принципе, процесс, у которого есть детерминированный тренд. Иногда такой процесс называют TS.

2 тип: процесс, приводимый к стационарному путем взятия первой разности — DSP (differencing stationary process). Вид этого процесса таков: $x_t = x_{t-1} + e_t$. Иногда такой процесс называют DS. Сведем же рассмотренные нами различия в свойствах этих процессов в табл. 18.

### Различия в свойствах процессов TSP и DSP

<table>
<thead>
<tr>
<th>Процесс TSP</th>
<th>Процесс DSP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Не стационарен из-за непостоянного тренда. Конечная память о шоках. Он забывает об ошибке на предыдущем шаге сразу. Если вместо белого шума будет стоять более общий процесс ARMA(p,q), то, конечно, шоки сказываются некоторое время, но их влияние со временем ослабевает.</td>
<td>Не стационарен из-за непостоянной дисперсии. Так как в явном решении стоит сумма всех предыдущих $e$, то шоки помнятся все время. Это процесс с бесконечной памятью. Экономически это не очень понятно, шоки не должны сказываться постоянно.</td>
</tr>
</tbody>
</table>

В 1984 г. была опубликована работа Нельсона и Канга, в которой они задались вопросом: что произойдет, если процесс является DS, а мы будем выделять линейный тренд, как для процесса типа TS? Их результаты были получены с помощью большого количества испытаний Монте- Карло и заключались в следующем.

1. Линейная регрессия случайного блуждания без дрейфа на время дает коэффициент множественной детерминации $R^2 > 0,44$ независимо от размера выборки, т. е. $R^2$ значим даже для малого количества точек наблюдения.

78
2. Если дрейф присутствует \((m \neq 0)\), то \(R^2 > 0.44\) и зависит от размера выборки. Это интуитивно можно понять, потому что, как только \(m \neq 0\), появляется свой тренд – нестационарное среднее. Также оказалось, что \(R^2 \to 1\) при \(n \to \infty\), т. е. для процесса DS с дрейфом \(R^2 \to 1\) при увеличении объема выборки. Это явление значимого коэффициента тренда, когда в истинном уравнении тренд отсутствует, было названо «ложный тренд» (spurious trends).

3. Оценка дисперсии остатков составляет примерно 14% от истинной дисперсии случайного возмущения, т. е. оценка дисперсии сильно занижена. Процедура оценивания указывает на значимый тренд и малую дисперсию, хотя на самом деле процесс случайного блуждания характеризуется растущей без ограничений дисперсией.

4. Остатки регрессии оказываются коррелированными с коэффициентом корреляции примерно равным \(\rho = 1 - 10/n\).

5. Разумеется, в этом случае \(t\)-статистика не годится для проверки гипотезы о значимости коэффициента при времени, она смещена в сторону принятия гипотезы о наличии линейного тренда.

6. Независимые случайные блуждания демонстрируют высокую корреляционную зависимость.

Этот пункт требует пояснения. Если взять 2 независимых случайных блуждания:

\[ y_t = y_{t-1} + u_t, \quad x_t = x_{t-1} + v_t, \]

где \(u_t\) и \(v_t\) – независимые белые шумы, и построить регрессию \(y_t = a + bx_t + e_t\), то из общих соображений мы ожидаем, что коэффициент \(b\) будет незначим. Но вы должны чувствовать из первых четырех результатов, что этого здесь не произойдет. Два независимых случайных блуждания показывают высокую регрессионную зависимость, коэффициент \(b\) оказывается значимым. Как можно это пояснить? Если оба процесса имеют значимые тренды, то они оба зависят от \(t\). Регрессионный анализ покажет, что процессы зависимости между собой. А это означает, что если каждая из величин, между которыми мы ищем регрессионную зависимость, является DS процессом, то регрессия между ними является «кажущейся». Таким образом, некоторые зависимости являются кажущимися. Нам представляется, что свяана динамика денежной массы и инфляции, но мы не учили, что оба процесса – DS, и сделанный экономический вывод неправомочен. Этот результат Нельсона и Канга показывает опасность прямого применения обычного регрессионного анализа в том случае, когда данные являются типа DS. Значимость этого результата особо подчеркнута более ранней, знаменитой работой Нельсона и Плоссера, которая связана с исследованием исторических макроэкономических рядов в США. Результат этого исследования показал, что практически все ряды макроэкономических показателей США, за исключением ряда уровня безработицы, оказались рядами типа DS, т. е. типа случайного блуждания с дрейфом. Это произвело впечатление разорвавшейся бомбы. Известный экономист Саржент заметил, что все, что сделано до сих пор в области макроэкономической динамики, подлежит пересмотру. Правда, дальнейшие исследования показали, что ситуация не столь драматична. Две эти работы, Нельсона-Канга и Нельсона-Плоссера, поставили вопрос ребром. Оказывается, надо проводить четкое разграничение между рядами двух видов: DS и TS. Если тренд в модели типа TS отсутствует, то вопрос сводится к различию между стационарными и нестационарными рядами. Ранше мы рассматривали для этого поведение выборочной автокорреляционной функции.

Мы нуждаемся в методе, позволяющем формально проводить различие между рядами типа TSP и DSP. Для этого используется тест Дикки-Фуллера, и обозначается как DF-тест.
4.2. Тест Дикки-Фуллера на единичный корень

Рассмотрим два ряда.

\[ x_t = m + x_{t-1} + e_t, \quad x_t = a + bt + v_t. \]

Попробуем рассмотреть модель: \( x_t = a + r x_{t-1} + bt + e_t \). Она вобрала черты обеих моделей, и гипотезы о характере ряда можно записать в виде прямых гипотез о ее параметрах.

- \( H_0 \): ряд является DS \( \Rightarrow r = 1, \ b = 0 \).
- \( H_1 \): ряд является TS \( \Rightarrow r < 1 \). Тогда \( e \) будет не просто белый шум, а некоторый стационарный ряд.

Тест, к рассмотрению которого мы сейчас переходим, появился сначала в более простой версии этой модели: \( x_t = a + r x_{t-1} + e_t \), т. е. без включения линейного тренда. В этом случае гипотезы принимают вид:

- \( H_0 \): \( r = 1 \Rightarrow DS \),
- \( H_1 \): \( r < 1 \Rightarrow TS \).

Теперь можно проверить гипотезу о том, что \( r = 1 \) при помощи \( t \)-статистики. Уравнение можно переписать в другом виде. После вычитания из обеих частей \( x_{t-1} \) получим \( \Delta x_t = a + (r - 1)x_{t-1} + e_t \). Пусть \( r - 1 = g \), тогда проверяемые гипотезы примут вид:

- \( H_0 \): \( g = 0 \),
- \( H_1 \): \( g < 0 \).

В классической линейной регрессии для проверки такой гипотезы применяется односторонняя \( t \)-статистика. Но в случае выполнения нулевой гипотезы, ряд \( x_t \) является случайным блужданием, его дисперсия стремится к бесконечности при увеличении \( t \), и распределение статистики \( g/\sigma_g \) не является распределением Стьюдента. Следовательно, и асимптотическое распределение этой статистики не является нормальным. Причиной является невыполнение условий центральной предельной теоремы в этом случае. Аналитическое выражение для асимптотического распределения статистики \( g/\sigma_g \) можно выразить через стохастические интегралы от винеровского случайного процесса, но мы этого делать не будем.

Критические точки этого распределения приходится рассчитывать численно, используя ассимиляционные процедуры Монте-Карло. Впервые это распределение было выведено и затабуировано в работе Дикки и Фуллера и носит их имя. Тест, использующий для проверки типа нестационарности это распределение, при условии \( g = 0 \), т. е. когда процесс принадлежит типу DS, называется и обозначается как DF-test. При условии, что нулевая гипотеза о том, что \( g = 0 \), выполнена, мы имеем процесс типа случайного блуждания (DSP). Именно для этого случая не работает \( t \)-статистика.

Позже MacKinnon расширил эти таблицы, рассмотрел некоторые другие случаи и предложил аппроксимирующие формулы для быстрого расчета критических точек в компьютерных программах. С помощью моделирования Дикки и Фуллера также рассчитали аналог \( F \)-статистики для случая, когда процесс является случайным блужданием. Это распределение также иногда называют распределением Дикки-Фуллера. Обычно из контекста ясно, о каком из распределений идет речь.

Вернемся к общей модели: \( x_t = a + bt + (r - 1)x_{t-1} + e_t \). В этом случае мы имеем следующие гипотезы:

- \( H_0 \): \( r = 1; b = 0 \),
- \( H_1 \): \( r < 1 \).
Сравним, как ведет себя DF-статистика и F-статистика. Нас интересует критическое значение при заданном числе наблюдений и для заданного уровня значимости. Если \( n \) – количество наблюдений, то в зависимости от него получим следующую табл. 19 для уровня значимости 0.05, в которой \( F \)-статистика – это фактически \( F_{2,n-3} \).

<table>
<thead>
<tr>
<th>( n )</th>
<th>( DF )</th>
<th>( F )-статистика</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>7.24</td>
<td>3.42</td>
</tr>
<tr>
<td>50</td>
<td>6.73</td>
<td>3.20</td>
</tr>
<tr>
<td>100</td>
<td>6.49</td>
<td>3.10</td>
</tr>
<tr>
<td>( \infty )</td>
<td>6.25</td>
<td>3.00</td>
</tr>
</tbody>
</table>

Таблица 19

На рис. 17 схематически показаны графики плотностей распределения Дикки-Фулера и Фишера. Мы видим, что DF-распределение значительно правее, чем F-распределение. Это означает, что в большом количестве случаев применение стандартной \( F \)-статистики ведет к тому, что мы считаем, что ряд относится к типу TS, в то время как он типа DS, а именно, если \( F \)-статистика попадает в область \( (a;b) \).

Рис. 17. Графики плотностей распределения Дикки-Фулера и Фишера

Кроме того, оказалось, что аналитическое выражение, а следовательно, и форма DF-распределения зависят от того, включены ли в модель регрессоры \( a, bt \) или нет ни того, ни другого. Поэтому для этих случаев нужны, по сути, разные таблицы.

Прежде чем двигаться дальше, давайте рассмотрим пример оценки конкретного ряда: логарифма индекса производства по данным Федеральной резервной системы США. Рассматриваются данные с I квартала 1950 г. по IV квартал 1977 г. Специфицируем модель в следующем виде:

\[
X_t = b_0 + b_1 t + a_1 X_{t-1} + a_2 (x_{t-1} - x_{t-2}) + e_t.
\]

Естественно, мы предполагаем, что \( e_t \) – белый шум.

Эта модель была оценена методом наименьших квадратов, но перед оцениванием из обеих частей вычли \( x_{t-1} \). Это стандартный прием, тогда можно сравнивать коэффициент с нулем, а не с единицей. А в левой части уравнения оказывается объясняемая переменная \( \Delta x_t \). Поэтому были оценены 2 следующие модели, соответствующие нулевой и альтернативной гипотезе (в скобках – стандартные ошибки коэффициентов).

Первая модель: \( x_t - x_{t-1} = \Delta x_t = 0.52 + 0.0012 t - 0.119 x_{t-1} + 0.498 (x_{t-1} - x_{t-2}) \)

\[
RSS = 0.056448.
\]

81
Вторая модель: Пусть \( b_1 = 0, \ a_1 = 1 \). МНК дает следующую модель:

\[
x_t - x_{t-1} = 0.0054 + 0.447(x_{t-1} - x_{t-2}).
\]

Соответственно, \( RSS = 0.063211 \). Считаем стандартное \( F \)-отношение:

\[
F = \frac{(0.063211 - 0.056448)/2}{0.056448/106} = 6.34.
\]

А теперь посмотрите на приведенную выше табл. 19. Ряд содержит немного более 100 точек, поэтому воспользуемся строкой для 100 наблюдений. Если бы мы сравнивали \( F \)-отношение с квантилями распределения Фишера, то \( F \)-критическое для 5-процентного уровня значимости было бы равно 3.1. Соответственно, мы бы пришли к выводу, что исследуемый ряд – типа TS. Но мы должны сравнивать статистику с квантилями распределения Дикки-Фуллера, т. е. с 6.49. В результате на уровне значимости 0.05 гипотеза о наличии единичного корня не может быть отвергнута, а для нас это означает, что ряд относится к типу DS. Поскольку проверка типа ряда – TS или DS – сводится к тестированию наличия или отсутствия единичного корня, то эта процедура носит стандартное название unit root test. Речь идет о проверке наличия единичных корней в характеристическом уравнении. Но наличие или отсутствие единичного корня – это то, что в подходе Бокса-Дженкинса называется порядком интеграции – единица или нуль. Т. е. unit root test, который появился как возможность различения между TS и DS рядами, дает статистическую процедуру для определения порядка интеграции ряда в подходе Бокса-Дженкинса.

Существуют распределения Дикки-Фуллера двух типов: типа \( F \)-отношения и типа \( t \)-отношения \( g/\sigma_g \). Второе из отношений рассматривалось для различных моделей. Один раз – для модели без линейного тренда, а во второй раз мы использовали таблицу распределения для случая, когда в модель включен линейный тренд. Тест Дикки-Фуллера предназначен для того, чтобы различить временные ряды типа TS и DS.

В соответствии с нулевой гипотезой \( H_0 \) исследуемый ряд принадлежит к типу DS. По альтернативной гипотезе он может быть типа TS, но одновременно быть нестационарным – иметь детерминированный тренд, или не иметь тренда – быть стационарным. Как уже упоминалось, спецификация модели в прямой и альтернативной гипотезе, т. е. включение или не включение в модели свободного члена и/или детерминированного тренда, влияет на распределение \( t \)-отношения. Рассмотрим это подробнее.

Дикки и Фуллер начали с исследования уравнения: \( \Delta x_t = g x_{t-1} + e_t \). В этом случае при условии, что выполнена гипотеза \( H_0 \), т. е. что \( g = 1 \), распределение \( t \)-отношения (отношения оценки коэффициента \( g \), полученной МНК, к оценке его среднеквадратического отклонения) называется DF-распределением. Включив в модель свободный член, получим модель \( \Delta x_t = a + +g x_{t-1} + e_t \). А дополнительное включение линейного тренда дает модель \( \Delta x_t = a + bt + +g x_{t-1} + e_t \). Оказалось, что во всех трех случаях распределение \( t \)-отношения \( g/\sigma_g \) выражается через интегралы от винеровского процесса, но по-разному. Все три распределения принято связывать с именами Дикки и Фуллера. Однако эти распределения разные и зависят от того, какие добавочные регрессоры входят в уравнение. Обозначим критические величины для первого распределения \( t_0 \), для второго распределения – \( t_1 \), для третьего распределения – \( t_2 \). Взаимное расположение этих критических значений для одного и того же уровня значимости и одного и того же числа степеней свободы схематически показано на рис. 18. Также на рисунке показано критическое значение распределения Стьюдента для того же уровня значимости и того же числа степеней свободы. Все эти значения отрицательные. Впрочем, в литературе их иногда приводят со знаком плюс, что обычно не приводит к недоразумениям.
Что будет, если мы ошибочно стационарный процесс примем за не стационарный и возьмем разностную производную? Взятие излишней разности приводит к увеличению дисперсии, в том числе и остаточной дисперсии после применения метода наименьших квадратов. Иногда этот эффект наряду с поведением выборочной автокорреляционной функции может служить некоторым неформальным критерием для ответа на вопрос: нужно брать последовательную разность или нет? Если при переходе к разности дисперсия возрастает, то, скорее всего, этого делать не надо.

4.3. ADF-тест Дикки-Фуллера

Специфицируем процесс типа TS в виде: \( x_t = a + bt + e_t \). Попробуем выписать такую модель, в которой различие между типами TS и DS будет выражаться в значениях коэффициентов. Для этого заменим в модели белый шум \( e_t \) на процесс \( u_t = ru_{t-1} + e_t \). Тогда нулевая гипотеза, заключающаяся в том, что ряд принадлежит типу DS, эквивалентна тому, что \( r = 1 \), а альтернативная – тому, что \( r < 1 \).

Чтобы убедиться в этом, запишем: \( x_{t-1} = a + b(t - 1) + u_{t-1} \). Теперь умножим левую и правую части на \( r \), что дает следующее выражение:

\[
rx_{t-1} = ra + rb(t - 1) + ru_{t-1}.
\]

Вычитая, получаем: \( x_t = [a(1 - r) + rb] + b(1 - r)t + rx_{t-1} + e_t \). Это соотношение включает свободный член, линейный тренд, авторегрессионный член и белый шум. При выполнении нулевой гипотезы \( r = 1 \) процесс принимает вид:

\[
x_t = b + x_{t-1} + e_t.
\]

Мы получили случайное блуждание с дрейфом, если \( b \neq 0 \). Если же \( r < 1 \), то в модели присутствует тренд и авторегрессионный член, причем авторегрессионная часть стационарна. Можно сделать два вывода. Во-первых, нам действительно удалось показать, что рассматриваемая модель охватывает оба наших случая. Во-вторых, что можно относить к классу TS такие процессы, где вместо белого шума \( e_t \) стоит стационарный процесс типа \( ARMA \).

В соответствии с полученным соотношением принято рассматривать следующую модель для проверки наличия единичного корня: \( x_t = m + bt + rx_{t-1} + e_t \). Если оценить эту модель методом наименьших квадратов, то нужно проверить сформулированную нулевую гипотезу: \( H_0: r = 1 \). Чтобы еще ближе свести процедуру к привычной, вычтем из обеих частей \( x_{t-1} \). Тогда получается уравнение \( \Delta x_t = m + bt + (r - 1)x_{t-1} + e_t \). Получаем, что надо проверить следующую нулевую гипотезу против альтернативной (обозначим \( g = r - 1 \)):

\[
H_0: DS \iff g = 0, \quad H_1: TS \iff g < 0.
\]
Поскольку \( r > 1 \) соответствует нестационарному процессу взрывного типа, который исключен из рассмотрения, используется односторонний тест. При обычной регрессии для проверки этих гипотез применяется \( t \)-отношение, т. е. 
\[
    t = \frac{g}{\sigma_g},
\]
которое сравнивается с критическим значением распределения Стьюдента. Однако мы уже видели, что при выполнении гипотезы \( H_0 \) распределение \( t \)-отношения, не подчиняется распределению Стьюдента, а подчиняется распределению, которое сейчас принято называть распределением Дикки-Фуллера. Поэтому схема проверки гипотез такая же, только мы должны сравнивать статистику с квантили не \( t \)-распределения, а \( DF \)-распределения. И решающее правило тоже чрезвычайно простое. Если 
\[
    t < \tau_{кр},
\]
выборочная статистика расположена левее табличного значения (мы «работаем» на левом хвосте распределения), то мы отвергаем нулевую гипотезу и считаем, что ряд относится к типу TS, а если выборочное значение лежит правее табличного значения, то ряд типа DS. Мы уже знаем, что табличные значения распределения Стьюдента и нормального распределения, подчиняется распределению, которое сейчас принято называть распределением Дикки-Фуллера.

Поэтому схема проверки гипотез такая же, только мы должны сравнивать статистику с квантилями \( DF \)-распределения. И решающее правило тоже чрезвычайно простое. Если 
\[
    t < \tau_{кр},
\]
выборочная статистика расположена левее табличного значения (мы «работаем» на левом хвосте распределения), то мы отвергаем нулевую гипотезу и считаем, что ряд относится к типу TS, а если выборочное значение лежит правее табличного значения, то ряд типа DS. Мы уже знаем, что табличные значения распределения Стьюдента и нормального распределения, подчиняется распределению, которое сейчас принято называть распределением Дикки-Фуллера.

Оказалось, что вид распределения, а, следовательно, его квантили, критические значения существенно зависят от того, включены ли в оцениваемую модель свободный член и/или тренд. Мы начали с уравнения 
\[
    x_t = a + b t + u_t,
\]
где \( u_t \) подчиняется марковской схеме. То есть мы сразу включили в модель свободный член и тренд. Далеко не всегда необходимо вводить тренд в уравнение. Если начать с уравнения 
\[
    y_t = a + u_t,
\]
в котором нет тренда, то в оцениваемом уравнении не только тренд пропадет, но и в свободном члене пропадет одно слагаемое. Поэтому в предположении гипотезы \( H_0 \) у процесса нет дрейфа. Действительно, если выполнена гипотеза \( H_0 \), это означает, что \( r = 1 \) и в уравнении пропадет свободный член. У нас будет случайное блуждание без дрейфа, совсем простое. Но тогда оказывается, что у \( t \)-отношения другое распределение. Поэтому нужно рассчитать уже другое критическое значение. Аналогично, если мы исключим еще и \( a \) из исходной модели, что тоже может быть, то получим еще одно, несколько иное распределение. Поэтому в зависимости от того, какую модель типа DS мы специфицируем: модель с линейным трендом и дрейфом, модель только с дрейфом или модель без дрейфа, у нас получаются три разных случая, три разных таблиц распределения. Существует несколько систем обозначений, мы будем пользоваться такими: \( \tau_0, \tau_1, \tau_2 \).

При практическом применении теста Дикки-Фуллера возникает вопрос, каким образом выбрать вид модели, или, что то же самое, как правильно специфицировать прямую и альтернативную гипотезы. Во всех случаях мы проводим различие между рядами типа TS и DS соответственно, но внутри каждого из типов мы должны выбрать подходящий к исследуемым данным конкретный вид модели. Если выборочная статистика попадает между критическими значениями, то возникает неясность, может быть, мы неправильно специфицировали модель и поэтому пытаемся сделать вывод по неправильному критическому значению. Так как у нас существует разнообразие основной и альтернативных гипотез, т. е. точных спецификаций DS и TS процесса, то как быть в этом случае? Понятно, что можно сделать глобальный перебор всех возможных моделей, но ведь вывод по разным моделям может быть противоречивым. И еще один аспект.

Мы рассмотрели ситуацию, когда \( u_t \) подчиняется схеме AR(1), но случайная составляющая может описываться автокорреляционной схемой более высокого порядка. Начнем с распространения DF-теста на процессы, когда \( u_t \) подчиняется ARMA более высокого порядка.

Рассмотрим процесс 
\[
    x_t = u_t,
\]
опустив без потери общности и тренд, и свободный член для уменьшения громоздкости выкладок. Пусть \( u_t \) подчиняется авторегрессионной схемой второго порядка:
\[
    u_t = r_1 u_{t-1} + r_2 u_{t-2} + e_t.
\]
То есть 
\[
    x_t = r_1 x_{t-1} + r_2 x_{t-2} + e_t.
\]
Прежде всего, нужно выразить в терминах коэффициентов этого уравнения $r_1$ и $r_2$ нулевую гипотезу о том, что ряд относится к типу DS, т. е. имеет единичный корень. Перепишем уравнение в операторном виде:

$$(1 - r_1L - r_2L^2)x_t = e_t.$$ 

Что значит, что уравнение имеет единичный корень? Если подставить $L = 1$, получается $1 - r_1 - r_2 = 0$. Следовательно, нулевая гипотеза принимает вид:

$$r_1 + r_2 = 1.$$ 

Далее вычтем из обеих частей $x_{t-1}$, получаем:

$$\Delta x_t = (r_1 - 1)x_{t-1} + r_2x_{t-2} + e_t.$$ 

Добавим и вычтем $r_2x_{t-1}$ в правой части. Получаем:

$$\Delta x_t = (r_1 + r_2 - 1)x_{t-1} - r_2\Delta x_{t-1} + e_t.$$ 

Теперь нулевая гипотеза о том, что ряд принадлежит типу DS или что у него есть единичный корень, сводится к проверке равенства нулю коэффициента $(r_1 + r_2 - 1) = 0$. Это соотношение позволяет переформулировать нулевую гипотезу $H_0$: DS $\iff r_1 + r_2 - 1 = 0$. То есть если построить регрессию $\Delta x_t$ на $x_{t-1}$ и на $\Delta x_{t-1}$, то можно стандартным образом проверить нулевую гипотезу. Для этого нужно рассчитать $t$-отношение и сравнить его величину с табличным значением нужного распределения. Что изменилось от того, что процесс подчиняется схеме AR(2)? В уравнении прибавился регрессор, являющийся конечной разностью – приращением $\Delta x_t$. А если бы была модель AR(3)? Вновь при наличии единичного корня сумма коэффициентов должна быть равна единице. Мы бы сделали аналогичное преобразование, и в преобразованном уравнении добавилось бы новое приращение $\Delta x_{t-2}$. Наличие у исследуемого процесса MA-части не вносит ничего принципиально нового. Следовательно, в наиболее общем случае, когда случайное возмущение относится к типу ARMA($p$, $q$), нужно исследовать следующее уравнение:

$$\Delta x_t = g x_{t-1} + \sum_{i=1}^{p} w_i \Delta x_{t-i} + \sum_{i=1}^{q} \delta_i e_{t-i}.$$ 

И если бы у процесса в исходном уравнении был свободный член и линейный тренд, вид уравнения сохранился бы. То есть общая модель, которую надо оценивать методом наименьших квадратов, принимает вид:

$$\Delta x_t = a + bt + gx_{t-1} + \sum_{i=1}^{p} w_i \Delta x_{t-i} + \sum_{i=1}^{q} \delta_i e_{t-i}.$$ 

Нулевая гипотеза о том, что процесс относится к типу DS, по смыслу эквивалентна тому, что $g = 0$. А альтернативная гипотеза, что ряд относится к типу TS, означает, что $g < 0$. Ключевой вопрос: изменило ли добавление регрессоров распределение $t$-отношения для коэффициента $g$? Мы уже видели, что добавление свободного члена и/или линейного тренда приводило к изменению распределения и его критических значений. Оказалось, что наличие приращений и запаздывающих значений случайного возмущения не меняет распределения, и мы можем пользоваться теми же таблицами Мак-Киннона.
Если в модели \( \Delta x_t = a + bt + gx_{t-1} + \sum_{i=1}^{p} w_i \Delta x_{t-i} + \sum_{i=1}^{q} \delta_i e_{t-i} \) присутствуют и свободный член, и тренд, то нулевую гипотезу надо проверять, используя статистику \( \tau_2 \), если только свободный член \( a \), то статистику \( \tau_1 \), если нет ни того, ни другого, то надо использовать статистику \( \tau_0 \). Этот тест носит название Augmented Dickey-Fuller test из-за того, что в уравнении появились приращения и, как мы только что указали, те же самые статистики \( \tau_2, \tau_1, \tau_0 \) «работают» для этой расширенной модели. Стандартная аббревиатура его названия: ADF-тест. По-русски название этого теста обычно переводят: расширенный тест Дикки-Фуллера. Напомним, что его использование основано на предположении, что процесс \( u_t \) подчиняется \( ARMA(p, q) \), а это значит, имеет постоянную дисперсию. Второе существенное предположение состоит в том, что мы знаем точные значения \( p \) и \( q \).

В ADF-тесте мы проверяем значимость только одного единственного коэффициента. Все остальные коэффициенты нас сейчас не интересуют, они являются вспомогательными. Для применимости ADF-теста важно проверить, что дисперсия случайного возмущения \( e_t \) постоянна, т. е. наличие гомоскедастичности возмущений. Для макроэкономических рядов этот вопрос не столь важен, а для финансовых важен чрезвычайно, потому что в них волатильность часто меняется и дисперсия не всегда постоянна. В случае гетероскедастичности возмущений ADF-тест уже не применим.

И вторая важная проблема: как правило, мы не знаем значений \( p \) и \( q \). Выяснилось, что, к сожалению, ADF-тест весьма чувствителен к правильному выбору параметров \( p \) и \( q \). Моделирование методом Монте-Карло показало, что, если число приращений в ADF-тесте согласовано с длинной реализации ряда, распределение \( t \)-отношения для коэффициента \( g \) является распределением Дикки-Фуллера. Следовательно, количество лагов, которое нужно включать в модель при применении ADF-теста, должно быть увязано с длиной реализации.

Для выбора числа лагов, включаемых при применении ADF-теста, было предложено несколько эвристических критериев, проверенных практикой и моделированием Монте-Карло. Например, было предложено выбирать количество лагов следующим образом: число лагов выбирается равным \( [n^{1/3}] \), где \([x]\) означает целую часть числа \( x \). Для квартальных данных было предложено выбирать число лагов по формуле: \( [4(n/100)^{1/4}] \), и для месячных данных – \( [12(n/100)^{1/4}] \). Наконец, Диколд и Нерлов показали, что на практике хорошо работает приближение \( [n^{1/4}] \).

Получаем простое правило. В макроэкономических рядах, если у вас от 81 до 256 точек, то нужно включать 3 лага. Если у вас меньше 81 точки, то два лага. В российской макроэкономике других случаев у вас не будет. Если вы имеете дело с данными по финансовым рядам, то там может быть другая ситуация. Интересно, что мощность критерия зависит не от числа наблюдений, к чему мы привыкли, но еще от «спэна», т. е. от общей продолжительности ряда. У вас есть возможность выбора. Если есть выбрать: использовать 500 наблюдений на коротком интервале или 500 – на длинном, лучше брать на длинном. Популярный эконометрический пакет Econometric Views включает в себя некоторый алгоритм выбора числа лагов.

В литературе встречается еще один прием по выбору надлежащего числа лагов при применении ADF-теста. Предлагается оставлять такое количество лагов, при котором все оценки МНК-коэффициентов при приращениях будут статистически значимы по обычному \( t \)-распределению Стьюдента.

На практике использование разного числа лагов может привести к разным выводам о типе процесса. Обычно мы стараемся учитывать трудность однозначного определения числа лагов и доверять результату, когда он устойчив к изменению лага, но в некотором разумном диапазоне. То, что иногда трудно различить тип ряда, TS или DS, не должно удивлять, ведь если для процесса AR(1) \( r = 0.998 \), т. е. близко к единице, то на конечных отрезках очень трудно отличить этот стационарный процесс от случайного блуждания.
Пакет Econometric Views ищет ADF-уравнение с тем количеством лагов, которое вы ему указете, или по умолчанию применяет встроенный алгоритм.

4.4. Авторегрессионные модели с распределенными лагами (ADL-модели)

До сих пор мы рассматривали только один случайный процесс $x_t$. Мы исследовали его на стационарность, строили модели процесса, прогнозировали будущие значения процесса. Однако при изучении экономических явлений наибольший интерес представляет взаимозависимость экономических величин. Текущее значение экономической величины будет зависеть не только от ее предыдущих значений, но и от текущего и предыдущих значений других экономических величин. Другими словами, среди регрессоров будут лаговые значения как объясняемой, так и объясняющих величин. Такие модели принято называть авторегрессионными моделями с распределенными лагами, по-английски – ADL (autoregressive distributed lag) models. Другое название таких моделей – ARMAX – напоминает об их сходстве с моделями ARIMA. Регрессор $x_t$ как бы заменяет собой белый шум в модели ARIMA. Для случая только одной объясняющей экономической величины общий вид моделей рассматриваемого типа:

$$y_t = q + a_1 y_{t-1} + a_2 y_{t-2} + \cdots + a_p y_{t-p} + b_0 x_t + b_1 x_{t-1} + b_2 x_{t-2} + \cdots + b_q x_{t-q} + e_t; \quad t = 1, 2, \ldots, T \quad (93)$$

Случайное возмущение по-прежнему полагается белым шумом. Используя операторные полиномы, получим

$$\alpha_p(L)y_t = \theta + \beta_q(L)x_t + e_t.$$  

В отличие от модели ARIMA мы отказываемся от нормирующего условия $\beta_0 = 1$. Мы будем использовать обозначения $ADL(p, q)$, чтобы указать количество лагов независимой и зависимой переменной.

Если процесс $x_t$ является стационарным в широком смысле, то, используя его разложение по теореме Вольда, мы получаем представление процесса $y_t$ в виде модели $ARMA(p, \infty)$, и стационарность процесса $y_t$ полностью определяется авторегрессионной частью исходной модели.

Для нестационарного регрессора $x_t$ процесс $y_t$ является, вообще говоря, также нестационарным. Отдельного рассмотрения заслуживает только случай, когда характеристические уравнения обоих полиномов $\alpha_p(L)$ и $\beta_q(L)$ имеют одинаковые корни, по модулю большие единицы.

Начнем рассмотрение с ситуации, когда регрессор стационарен, и полином $\alpha_p(L)$ не имеет корней вне или на единичной окружности. В этом случае можно переписать модель в виде:

$$y_t = \alpha_p^{-1}(L)\theta + \alpha_p^{-1}(L)\beta_q(L)x_t + \alpha_p^{-1}(L)e_t,$$

t. е. стационарная ADL модель представима в виде модели с распределенными лагами без авторегрессионной части. Это означает, что текущее значение величины $y_t$ зависит от текущего и всех предшествовавших значений величины $x_t$ или что текущее значение $x_t$ влияет на текущее и все последующие значения величины $y_t$. Другими словами, текущее значение величины $y_t$ представляет в виде суммы динамических откликов на изменения величины $x_t$. Коэффициент влияния значения $x_t$ на величину $y_{t+k}$ представляет собой частную производную $\frac{dy_{t+k}}{dx}$, которую можно выразить как значение производной от операторного выражения:

$$1k! \frac{d^k}{dL^k} \left( \frac{\beta_q(L)}{\alpha_p(L)} \right), \text{при } L = 0.$$
При увеличении параметра \( k \) получаем последовательно мгновенный, краткосрочный, среднесрочный и долгосрочный отклики. Для макроэкономических и монетарных моделей, выраженных в уровнях величин, эти отклики принято называть соответствующими мультипликаторами, а для моделей, выраженных в логарифмах величин, – эластичностями. Например, для модели ADL(1, 1) отклики выражаются через коэффициенты модели следующим образом:

\[
\beta_0, \beta_1 + \alpha_1 \beta_0, \alpha_1 \beta_1 + \alpha_1^2 \beta_0 \text{ т. д.}
\]

Долгосрочный отклик определяется как сумма всех промежуточных откликтов. Используя стационарность процессов \( x_t \) и \( y_t \), долгосрочный отклик можно найти, просто взяв математические ожидания от обеих частей соотношения. В результате получаем уравнение статистического или долгосрочного равновесия.

\[
\hat{y}_t = \frac{\theta}{\alpha_p(1)} + \frac{\beta_q(1)}{\alpha_p(1)} \bar{x}_t,
\]

где через \( \hat{y}_t \) и \( \bar{x}_t \) обозначены равновесные значения соответствующих величин.

Для удобства анализа долгосрочного и краткосрочного поведения динамического соотношения, выраженного ADL моделью, удобно провести перепараметризацию модели. Начнем с модели ADL(1, 1): \( y_t = \theta + \alpha_1 y_{t-1} + b_0 x_t + b_1 x_{t-1} + e_t \). Заменив \( y_t \) на \( y_{t-1} + \Delta y_t \) и \( x_t \) на \( x_{t-1} + \Delta x_t \), получаем:

\[
\Delta y_t = \theta + \beta_0 \Delta x_t - (1 - \alpha_1) y_{t-1} - (\beta_0 + \beta_1) x_{t-1} + e_t.
\]

Перегруппировка членов дает:

\[
\Delta y_t = \beta_0 \Delta x_t - (1 - \alpha_1) \left[ y_{t-1} - \frac{\theta}{1 - \alpha_1} - \frac{\beta_0 + \beta_1}{1 - \alpha_1} x_{t-1} \right] + e_t.
\]

Это представление ADL модели называется моделью коррекции ошибками (error correction model, сокращенно – ECM). Смысл модели и названия становится ясен, если обратить внимание, что выражение в квадратных скобках может трактоваться как отклонение от долгосрочного равновесия в момент времени \( t - 1 \). В самом деле, долгосрочное равновесие определяется соотношением

\[
\left[ y_{t-1} - \frac{\theta}{1 - \alpha_1} + \frac{\beta_0 + \beta_1}{1 - \alpha_1} x_{t-1} \right] = -e_t,
\]

получает отклонение от долгосрочного равновесия, выражение в квадратичных скобках положительно, если значение \( y_{t-1} \) превышает равновесное значение, соответствующее \( x_{t-1} \). Таким образом, текущее (краткосрочное) изменение \( y_t \) представлено в виде суммы двух слагаемых. Первое из них – это мгновенный отклик на текущее (краткосрочное) изменение \( x_t \), а второе – поправка на имевшее место в предыдущий момент отклонение от долгосрочного равновесия. При этом, поскольку для стационарности процесса \( y_t \) необходимо выполнение условия \( |\alpha_1| < 1 \), коэффициент при квадратной скобке отрицательный. Это означает, что второе слагаемое «подтягивает» процесс \( y_t \) к долгосрочному соотношению с процессом \( x_t \). Таким образом, модель коррекции ошибками позволяет удобно объединить в рамках одной модели краткосрочную и долгосрочную динамику, а ее коэффициенты имеют содержательную экономическую интерпретацию.

В общем случае модели ADL(\( p, q \)) мы должны заменить значения процессов \( y_t \) и \( x_t \) в различные моменты времени на их значения в момент \( t - 1 \) и отклонения, например, \( y_{t-2} = y_{t-1} - \Delta y_t, y_{t-3} = y_{t-1} - \Delta y_t - \Delta y_{t-1} \). В результате получаем ECM представление:

\[
\Delta y_t = \beta_0 \Delta x_t + \sum_{i=1}^{p-1} \delta_i \Delta y_{t-i} + \sum_{i=1}^{q-1} \gamma_i \Delta x_{t-i} - \alpha_p(1) \left[ y_{t-1} - \frac{\theta}{\alpha_p(1)} + \frac{\beta_q(1)}{\alpha_p(1)} x_{t-1} \right] + e_t.
\]

88
Выражение в квадратных скобках по-прежнему представляет корректирующий член, «подправляющий» лаговую структуру отклонениями от долгосрочного равновесия на предыдущем шаге. Это хорошо видно после замены нулями всех отклонений, что соответствует равновесному состоянию. Важно отметить, что коэффициенты $\delta_i$ и $\gamma_i$, так же, как и остальные коэффициенты ECM представления, линейно выражаются через коэффициенты исходной ADL модели, причем это преобразование невырождено.

Наиболее общая ADL модель получается в случае, когда имеется $k$ различных экономических дисциплин в качестве регрессоров. Будем обозначать ее $ADL(p, q_1, q_2, \ldots, q_k)$, а общее уравнение принимает вид:

$$\alpha_p(L)Y_t = \theta + \beta_{q_1}(L)x_{1t} + \beta_{q_2}(L)x_{2t} + \cdots + \beta_{q_k}(L)x_{kt} + e_t.$$  

Преобразование к модели коррекции ошибками проводится в этом случае точно так же, как и ранее, но выражение в общем случае становится весьма громоздким.

Оценивание и диагностика ADL моделей близки к моделям ARIMA. Поскольку в модели отсутствует лаговая структура случайного возмущения, основным методом оценивания является метод наименьших квадратов. Мы знаем, что при отсутствии корреляции случайных регрессоров и случайного возмущения МНК дает состоятельные оценки. Разумеется, перед оцениванием модели нужно убедиться, что все переменные являются стационарными.

У нас есть две возможности. Первая: оценить ADL модель, а затем пересчитать параметры и их стандартные ошибки для ECM представления. Вторая: оценить параметры ECM модели непосредственно. Поскольку, как уже упоминалось ранее, коэффициенты ADL и ECM моделей связаны невырожденным линейным преобразованием, можно доказать, что оба пути дают идентичные результаты. Поэтому выбор представления модели определяется содержательной задачей исследования, а не особенностями процедуры оценивания.

Определение числа лагов для переменных, входящих в модель, неизбежно сопровождается построением ряда моделей и выбором наилучшей из них. Общепринятой стратегией выбора модели в настоящее время является подход от общего к частному (from general to simple), предложенный и развитый Дэвидом Хендри и его соавторами. В соответствии с этим подходом мы начинаем построение с наиболее общей модели, как по набору объясняющих переменных, так и по количеству включенных лагов. Построенная методом наименьших квадратов модель подвергается различным тестам: на автокорреляцию остатков, нормальность, гетеросkedастичность и т.д. Если модель проходит все диагностические тесты, на следующем шаге исследуется возможность наложения различных ограничений на ее коэффициенты. Такими ограничениями могут быть исключение из модели отдельных переменных или лагов, равенство некоторых коэффициентов между собой и т. п. Обычно проверяемые ограничения порождаются как содержательными теоретическими соображениями, так и численькими значениями оценок коэффициентов.

4.5. Причинность по Грэнджеру

Понятие причинности по Грэнджеру (Granger causality) было введено Грэнджером. Этот термин сегодня расценивается как не очень удачный, прежде всего из-за дословного совпадения с причинно-следственными отношениями в обычном смысле, но он уже прочно укоренился в литературе. Лимер предлагал более удачный термин «предшествование» (precedence), но в практике укоренилась именно причинность по Грэнджеру.
Основной посылкой Грэнджера было то, что будущее не может быть причиной настоящего или прошлого. Поэтому если событие А произошло после события В, то А определенно не может быть причиной В. Но, как знает каждый, «после того не значит вследствие того». При анализе временных рядов часто хотелось бы знать, предшествует ряд \( x_t \) ряду \( y_t \), или \( y_t \) предшествует \( x_t \), или они «одновременны». Например, предшествует сжатие денежной массы падению производства в России 1990-х гг., или между ними нет соотношения предшествования (причинности по Грэнджеру).

Понятие причинности по Грэнджеру имеет более широкое применение, чем то, в котором мы будем его использовать. Это, скорее, понятие информационное. Сначала запишем формальное определение. Рассмотрим некоторый процесс \( z_t \), не важно, многомерный или нет. Представим себе, что мы рассчитываем условное математическое ожидание этого процесса: \( M(z_{t+1}, \sigma_t) \), где \( \sigma_t \) — вся возможная информация, которая существует в мире в момент \( t \). Потом из всевозможной информации, которая известна к моменту \( t \), удаляем информацию о некотором процессе \( x_t \). Если при этом условное математическое ожидание процесса \( z_t \) не изменится, т. е. если \( M(z_{t+1} | \sigma_t) = M(z_{t+1} | \sigma_t - x_t) \), то \( x_t \) не является причиной по Грэнджеру для \( z_t \). Если же они не равны между собой, т. е. если изъятие информации об \( x_t \) меняет условное математическое ожидание, то \( x_t \) является причиной по Грэнджеру для \( z_t \).

Если \( x_t \) — причина по Грэнджеру для \( z_t \), то это не означает, что между этими процессами есть причинно-следственная связь. Единственный вывод состоит в том, что уж если переменная \( x_t \) не является причиной по Грэнджеру для переменной \( z_t \), то она не является ее причиной и в обычном смысле. Грэнджер предложил метод тестирования причинности по Грэнджеру. Он предложил построить регрессию процесса \( z_t \) на его собственные предыдущие значения и на предыдущие значения процесса \( x_t \):

\[
z_t = a_0 + \sum_{i=1}^{k} a_i z_{t-i} + \sum_{i=1}^{k} b_i x_{t-i} + e_t.
\]

А после этого проверить обычную гипотезу о равенстве нулю группы коэффициентов:

\[
H_0 : b_1 = \cdots = b_k = 0.
\]

\[
H_1 : b_1 + \cdots + b_k > 0.
\]

Это обычный \( F \)-тест. Как обычно, строится полное уравнение и укороченное и сравниваются остаточные суммы по \( F \)-статистике. Если нулевая гипотеза отвергается, то \( x_t \) является причиной по Грэнджеру для \( z_t \). Если же нулевая гипотеза не отвергается, то прошлое процесса \( x_t \) не оказывает влияние на процесс \( z_t \) и не является его причиной по Грэнджеру.

Симс предложил другой подход: \( x_t \) не является причиной по Грэнджеру для \( y_t \), если в регрессии \( y_t \) на прошлые, текущие и предыдущие значения \( x_t \) коэффициенты при будущих значениях \( x_t \) совместно равны нулю. Для проведения теста строим регрессию

\[
y_t = \sum_{i=-k}^{m} b_i x_{t-i} + e_t
\]

и проверяем гипотезу \( H_0 : b_{-i} = 0(i = 1, \ldots, k) \) против естественной альтернативы. Если нулевая гипотеза отвергается, то знание будущих значений \( x_t \) не позволяет улучшать прогноз \( y_t \). Хотя с эконометрической точки зрения тесты Грэнджера и Симса не тождественны, они проверяют одно и то же свойство. Оба теста, впрочем, весьма чувствительны к числу лагов, включенных в тестовое уравнение. Идейно тесты должны включать бесконечное число лагов, т. е. всю предысторию \( x_t \). Но нас ограничивает длина реализации, во-первых, в количестве лагов, которое можно применять, во-вторых, падающее число степеней свободы тоже может повлиять на мощность этого теста. Хорошие рекомендации по выбору числа лагов отсутствуют.
4.6. Многомерные процессы

До сих пор мы рассматривали модели, которые состоят только из одного соотношения, связывающего временные ряды. При этом мы выбирали одну из переменных в качестве эндогенной, а остальные переменные являлись экзогенными. Такое разделение не всегда является естественным, часто приходится рассматривать одновременно несколько соотношений, в которые одни и те же переменные входят и как эндогенные, и как экзогенные. Переменная не всегда может рассматриваться как экзогенная, и мы фактически должны рассматривать модель DGP, состоящую из нескольких уравнений. Это означает моделирование нескольких временных рядов одновременно, другими словами – моделирование многомерного случайного процесса.

Начнем с определений. Рассмотрим вектор $X_t = (x^1_t, x^2_t, ..., x^k_t)$, каждая компонента которого является временным рядом. Верхним индексом будем обозначать номер компоненты, а нижним по-прежнему – момент времени. Распределение компонент характеризуется семейством совместных плотностей. Условием стационарности в узком смысле по-прежнему является независимость от сдвига во времени всего семейства совместных плотностей распределения. Только теперь кроме всевозможных комбинаций значений случайного процесса в различные моменты времени аргументами плотностей вероятности также являются всевозможные комбинации различных компонент в различные моменты времени. Например, для двухмерной плотности получаем из условия стационарности: $f(x^1_t, x^2_t) = f(x^1_{t+T}, x^2_{t+T})$ для любого $T$. Совместное распределение компонент для одного и того же момента времени не зависит от времени. Рассмотрим другую функцию распределения, например, трехмерную, в которую входят значения первой компоненты в два разных момента времени и второй компоненты в некоторый третий момент времени. Стационарность означает, что $f(x^1_t, x^2_{t+h}, x^3_{t+s}) = f(x^1_{t+T}, x^2_{t+h+T}, x^3_{t+s+T})$. Можно сказать, что это свойство инвариантности к сдвигу во времени. То есть если к каждому моменту времени прибавить величину $\tau$, то функция плотности не изменится. Понятно, что стационарность многомерного процесса влечет за собой стационарность каждой из его компонент.

Как и в одномерном случае, стационарность в узком смысле влечет за собой ряд свойств характеристик случайных процессов. Прежде всего, начнем с математического ожидания. Математическое ожидание для каждой компоненты не зависит от других компонент. Поэтому если многомерный процесс стационарен, математическое ожидание каждой компоненты не зависит от времени. Вектор математических ожиданий

$$M(X_t) = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_k \end{pmatrix}$$

не зависит от времени.

Теперь рассмотрим моменты второго порядка. Каждая компонента характеризуется дисперсией и автокорреляционной функцией. Если одномерный ряд стационарен, его автокорреляционная и автоковариационная функции зависят только от сдвига $\tau$: $Corr(\tau) = Corr(x^i_t, x^i_{t+\tau}) = \rho(\tau)$. Однако теперь можно рассмотреть второй смешанный момент для различных компонент, а также $Corr(x^i_t, x^j_{t+\tau})$. Такую величину естественно назвать кросс-корреляционной функцией. Если компоненты образуют многомерный стационарный процесс, то кросс-корреляция будет функцией сдвига во времени $\tau$. Обозначим эту функцию $R_{ij}(\tau)$. Довольно очевидно, что $R_{ij}(\tau) = R_{ji}(-\tau)$. При фиксированном значении $\tau$ элементы $R_{ij}(\tau)$ образуют матрицу $R$, зависящую от $\tau$. Значению $\tau$, равному нулю, соответствует корреляционная матрица вектора $X_t$. 

91
Как и в одномерном случае, назовем многомерный процесс стационарным в широком смысле (слабо стационарным), если вектор математических ожиданий не зависит от времени и кросс-корреляции между любыми компонентами зависят только от разности во времени. Можно сказать, что у стационарного многомерного процесса компоненты стационарны и стационарно связаны между собой. Модель многомерного временного ряда обычно будет задаваться в виде уравнения для каждой из компонент, причем в виде объясняющих переменных будут выступать текущие и предыдущие значения, вообще говоря, всех компонент. Другими словами, модель будет представлена системой одновременных уравнений. Одновременность не дает напрямую использовать такой распространенный метод, как метод наименьших квадратов для оценки параметров многомерных моделей. Возвращаясь к временным рядам, мы видим, что одновременность не дает напрямую использовать такой распространенный метод, как метод наименьших квадратов для оценки параметров многомерных моделей. Возвращаясь к временным рядам, мы видим, что при оценивании моделей, состоящих из нескольких уравнений, возникают сложности, связанные с наличием среди регрессоров «одновременных» составляющих. Кроме того, надо попытаться разделить переменные на эндогенные и экзогенные. Иногда это разделение экономически оправдано, иногда не очень. Например, один из подходов к исследованию эффективности рынка – это моделирование временных рядов кросс-курсов различных валют и цен разных финансовых инструментов. Например, рассматривая курсы рубль/доллар, доллар/евро, рубль/евро, трудно сказать, какой из них естественно выбрать эндогенной переменной, а какие – экзогенными.

В 1980 г. Симс предложил подход, который позволяет уйти от разделения переменных на экзогенные и эндогенные и избавиться от сложностей, связанных с одновременностью уравнений. Этот подход является к тому же естественным обобщением подхода Бокса-Дженкинса к моделям ARIMA и носит название VAR, или векторная авторегрессия. В принципе существует и VARIM. Для того, чтобы избежать смещения, связанного с применением МНК непосредственно к каждому уравнению структурной формы, Симс предложил, не производя деления переменных на экзогенные и эндогенные, представить каждую из компонент многомерного случайного процесса как линейную комбинацию от предыдущих значений всех переменных.

Рассмотрим сначала пример двумерного вектора и только одного лага. Пусть, например, \( x_t^1 \) – темп роста денежной массы, а \( x_t^2 \) – дефлятор ВВП. Запишем систему уравнений указанного вида:

\[
\begin{align*}
    x_t^1 &= a_1 + b_{11} x_{t-1}^1 + b_{12} x_{t-1}^2 + e_t^1, \\
    x_t^2 &= a_2 + b_{21} x_{t-1}^1 + b_{22} x_{t-1}^2 + e_t^2.
\end{align*}
\]

Здесь случайные возмущения предполагаются белыми шумами, вообще говоря, коррелированными. Стандартное обозначение этой модели – VAR(1), где 1 – число лагов. Размерность многомерного случайного процесса обычно не указывается.

Модель VAR(p) в матрично-векторных обозначениях имеет вид:

\[
X_t = a + A_1 X_{t-1} + A_2 X_{t-2} + \cdots + A_p X_{t-p} + e_t,
\]

где \( e_t \) – векторный белый шум со следующими свойствами: \( M(e_t) = 0 \) для всех \( t \),

\[
cov(e_t) = M(e_t e_s) = \begin{cases} 
\Omega, & s \neq t, \\
0, & s = t.
\end{cases}
\]

Модель сразу выписывается в приведенной форме. Одновременные компоненты не входят в правую часть. Для двумерного случая получаем:

\[
\begin{pmatrix} x_{t}^1 \\ x_{t}^2 \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} x_{t-1}^1 \\ x_{t-1}^2 \end{pmatrix} + \begin{pmatrix} e_{t}^1 \\ e_{t}^2 \end{pmatrix}.
\]

Модель VAR(p) напоминает модель ARMA, поэтому многие из их свойства схожи. Перепишем модель VAR(p), используя оператор лага:

\[
(M - A_1 L - A_2 L^2 - \cdots - A_p L^p) X_t = e_t.
\]
Для упрощения записи мы предполагаем, что нет свободного члена, это не приводит к потере общности, но немного сокращает запись. Для рассматриваемого примера можем записать:

\[
\begin{pmatrix}
1 - b_{11}L & -b_{12}L \\
-b_{21} & 1 - b_{22}L
\end{pmatrix}
\begin{pmatrix}
x_t^1 \\
x_t^2
\end{pmatrix}
=
\begin{pmatrix}
e_t^1 \\
e_t^2
\end{pmatrix}.
\]

Также, как и в одномерном случае, возникает вопрос, можно ли выразить вектор \( x_t \) через текущие и предыдущие значения вектора случайных возмущений \( e_t \). Очевидно, что если существует обратная матрица, то можно записать:

\[
\begin{pmatrix}
x_t^1 \\
x_t^2
\end{pmatrix}
=
\frac{1}{\Delta}
\begin{pmatrix}
1 - b_{22}L & b_{21}L \\
b_{12}L & 1 - b_{11}L
\end{pmatrix}
\begin{pmatrix}
e_t^1 \\
e_t^2
\end{pmatrix}.
\]

Если обозначить через \( I \) единичную матрицу, то определитель матрицы системы \( П = I - A_1 \) равен:

\[
\Delta = (1 - b_{11}L)(1 - b_{22}L) - b_{12}b_{21}L^2 = 1 - (b_{11} + b_{22})L + (b_{11}b_{22} - b_{12}b_{21})L^2 = (1 - \mu_1L)(1 - \mu_2L),
\]

где \( \mu_1 \) и \( \mu_2 \) — собственные числа матрицы \( П \), возможно комплексные.

Если хотя бы одно из собственных чисел равно 0, то матрица системы уравнений вырождена и \( x_t \) не выражается через текущие и предыдущие значения вектора белого шума. Вспомнив условия стационарности моделей ARMA(\( p, q \)), можно заключить, что при условии \( |\mu_1| < 1 \) и \( |\mu_2| < 1 \) каждая из компонент процесса \( X_t \) выражается в виде бесконечного ряда текущего и предыдущего значений компонент белого шума. Поскольку коэффициенты каждого из разложений будут учитывать в геометрической прогрессии, по теореме Вольда каждая из компонент процесса \( x_t \) будет стационарной в широком смысле.

Для изучения условий стационарности и свойств моделей VAR(1) удобно использовать сведение матриц \( A_1 \) и \( П \) к простейшей форме. Заметим, что матрицы \( П \) и \( A_1 \) очевидно имеют один и те же собственные векторы, а их собственные числа дополняют друг друга до единиц. Обозначим собственные числа матрицы \( A_1 \) через \( \lambda_1 = 1 - \mu_1 \) и \( \lambda_2 = 1 - \mu_2 \). Поэтому условие, что собственные числа матрицы \( П \) лежат в единичном круге, эквивалентно аналогичному условию для собственных чисел матрицы \( A_1 \).

Рассмотрим сначала случай, когда \( \lambda_1 \neq \lambda_2 \). Из линейной алгебры мы знаем, что собственные векторы, соответствующие различным собственным числам, линейно независимы. Введем невырожденную матрицу \( C \), столбцы которой являются собственными векторами, соответствующими собственным числам \( \lambda_1 \) и \( \lambda_2 \) соответственно. Обозначим через \( \Lambda \) диагональную матрицу \( diag \{\lambda_1, \lambda_2\} \). Тогда очевидно, что \( C^{-1}A_1C = \Lambda \) и \( A_1 = CAC^{-1} \). Определим новый вектор переменных соотношением \( Y_t = C^{-1}X_t \), отсюда следует также, что \( X_t = CY_t \). Умножая уравнение \( X_t = A_1X_{t-1} + e_t \) слева на матрицу \( C^{-1} \), получим \( Y_t = \Lambda Y_{t-1} + \eta_t \), где через \( \eta_t \) обозначен «новый» векторный белый шум.

В новых переменных система уравнений распадается на два отдельных уравнения:

\[
\begin{align*}
y_t^1 &= \lambda_1 y_{t-1}^1 + \eta_t^1, \\
y_t^2 &= \lambda_2 y_{t-1}^2 + \eta_t^2,
\end{align*}
\]

свойства которых можно установить, используя уже известные нам приемы. Очевидно, что \( y_t^1 \) и \( y_t^2 \) слабо стационарны при \( |\lambda_1| < 1 \) и \( |\lambda_2| < 1 \) и, по крайней мере, одна из компонент вектора \( Y_t \) не является стационарной при нарушении этого условия. Поскольку \( x_t^1 \) и \( x_t^2 \) являются линейными комбинациями \( y_t^1 \) и \( y_t^2 \), то условием стационарности векторного процесса \( x_t \) является нахождение собственных чисел матрицы \( A_1 \) внутри единичного круга.
Отдельного рассмотрения заслуживает случай, когда \( \lambda_1 = \lambda_2 \). В этом случае двух линейно независимых векторов может и не существовать и матрицу \( A_1 \) нельзя привести к диагональной форме. Однако доказано, что тогда существует матрица \( P \), которая приводит матрицу \( A_1 \) к так называемой жордановой форме. Мы ограничиваемся только кратким рассмотрением этого понятия без доказательств. Итак, доказано, что существует невырожденная матрица \( P \), такая что \( P^{-1} A_1 P = J \) и \( A_1 = PJ P^{-1} \), где \( J \) – жорданова матрица: \( J = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix} \). Определив вектор \( Y_t = P^{-1}X_t = (X_t \cdot \eta_t) \), сведем исходную модель к виду \( Y_t = JY_{t-1} + \eta_t \), где сейчас векторный белый шум \( \eta_t = P^{-1}e_t \). В этом случае система уравнений

\[
\begin{align*}
y^1_t &= \lambda y^1_{t-1} + y^2_{t-1} + \eta^1_t \\
y^2_t &= \lambda y^2_{t-1} + \eta^2_t
\end{align*}
\]

не распадается на отдельные уравнения, но второе из них может быть решено и исследовано отдельно от первого. Очевидно, что при условии \( |\lambda| < 1 \) второе уравнение представляет собой стационарную модель AR(1), а первое – стационарную модель ADL(1,1). При \( |\lambda| \geq 1 \) оба уравнения нестационарны. Таким образом, оба рассмотренных случая дают единое условие стационарности: собственные числа матрицы \( A_1 \) (или \( P \)) должны лежать строго внутри единичного круга.

Этот результат легко обобщить на случай VAR(1) модели произвольной размерности. Если все собственные числа \( \lambda_1, \lambda_2, \ldots, \lambda_k \) матрицы \( A_1 \) различны, то модель распадается на отдельные уравнения \( y^i_t = \lambda y^i_{t-1} + \eta^i_t(i = 1,2, \ldots, k) \). Каждое из них стационарно тогда и только тогда, когда \( |\lambda_i| < 1 \). Если же среди собственных чисел \( \lambda_1, \lambda_2, \ldots, \lambda_k \) есть группа равных по величине, то, обозначив их без потери общности номерами \( 1,2, \ldots, r \), приводим модель к тому же виду: \( Y_t = JY_{t-1} + \eta_t \). Матрица \( J \) содержит так называемую жорданову клетку \( S \) размера \( r \) в левом верхнем углу:

\[
S = \begin{pmatrix}
\lambda & 1 & 0 & 0 \\
0 & \lambda & 0 & 0 \\
0 & 0 & \lambda & 0 \\
0 & 0 & 0 & 0
\end{pmatrix}.
\]

На главной диагонали жордановой клетки стоят собственные числа, первая наддиагональ содержит 1, а остальные элементы равны 0. Отдельные уравнения системы принимают вид:

\[
\begin{align*}
y^1_t &= \lambda y^1_{t-1} + y^2_{t-1} + \eta^1_t, \\
\vdots \\
y^{r-1}_t &= \lambda y^{r-1}_{t-1} + y^r_{t-1} + \eta^1_t, \\
y^r_t &= \lambda y^{r-1}_{t-1} + \eta^r_t.
\end{align*}
\]

Преобразованная система ясно показывает, что условие стационарности имеет тот же вид, что и для случая двумерной модели.

Для модели VAR(p) рассмотрение, подобное проведенному выше, невозможно. Для получения условий стационарности можно толковать модель VAR(p) как систему разностных линейных уравнений с постоянными коэффициентами и правой частью \( e_t \). По аналогии со скалярным случаем будем искать общее решение однородной системы

\[
X_t - A_1X_{t-1} - \cdots - A_pX_{t-p} = 0
\]

в виде \( X_t = C\lambda^t \). Подставив это выражение в уравнение, получим:

\[
(\lambda^p I + \lambda^{p-1}A_1 + \cdots + A_p)C = 0.
\]
Однородная система линейных уравнений имеет ненулевые решения $\lambda$, равные собственным числам матрицы $\lambda^p I + \lambda^{p-1} A_1 + \cdots + A_p$, т. е. корням характеристического уравнения

$$det(\lambda^p I + \lambda^{p-1} A_1 + \cdots + A_p) = 0.$$ 

Соответствующими нетривиальными решениями системы

$$(\lambda^p I + \lambda^{p-1} A_1 + \cdots + A_p) C = 0$$

являются собственные векторы ее матрицы. Тогда слагаемое, соответствующее собственному числу $\lambda_i$, стремится к нулю при неограниченном увеличении $t$. Если среди $p \times k$ собственных чисел есть одинаковые, то нужно заменить в выражении $X_t = C \lambda^t$ постоянный вектор $C$ зависящим от времени вектором вида $C_0 + C_1 t + \cdots + C_{r-1} t^{r-1}$. Во всех случаях условие стационарности выражается компактно: все корни характеристического уравнения должны лежать строго внутри единичного круга.

Если условие стационарности выполнено, и случайное возмущение $e_t$ обладает свойствами белого шума, то уравнения модели VAR(p) можно оценивать обычным методом наименьших квадратов. Причем каждое уравнение можно оценивать отдельно, т. е. строить регрессию компоненты $x^i_t$ на предыдущие значения всех компонент, включая предыдущие значения самой $x^i_t$. Во-первых, нет корреляции регрессоров с возмущениями, и нет никаких проблем с состоятельностью оценок. Кроме того, все переменные совершенно равноправны, и поскольку все регрессоры предшествуют объясняемым переменным, не возникает проблем с экзогенностью. Таким образом, оценивание моделей VAR(p) требует только многократного применения метода наименьших квадратов (МНК). Если же переменные являются нестационарными типа DSP (содержат единичный корень), то, как мы уже видели, возникает опасность кажущихся (spurious) регрессий. Рекомендуемый до сих пор прием состоял в устранении тренда взятием последовательных разностей и дальнейшим построением ADL-модели между преобразованными таким образом переменными.

С содержательной точки зрения полученные модели описывают только краткосрочную взаимосвязь между экономическими переменными. Устранив тренд, мы, по сути, отказываемся анализировать долгосрочное поведение переменной и отрицаем возможность существования долгосрочного равновесия для нестационарных переменных типа DSP. А результаты Нельсона и Плоссера показывают, что большинство макроэкономических переменных относятся именно к типу DSP. Фактически это означало бы отсутствие долгосрочных зависимостей в макроэкономике. Одной из попыток преодоления этого противоречия было расширение Перроном класса TSP-моделей за счет допущения кусочно-линейного тренда, с чем мы уже познакомились. Кардинальным решением проблемы является введенное Грёнжером и Энглом понятие коинтеграции.
По этимологии этот термин означает совместную интеграцию, и до сих пор сосуществуют практики его написания с дефисом и слитно. Начнем с нескольких формальных определений. Напомним, что мы называем процесс $x_t$ интегрированным порядка $d$ и обозначаем это следующим образом: $x_t \sim I(d)$, если ряд его конечных разностей порядка $d$ является стационарным. В этих терминах стационарный случайный процесс является процессом нулевого порядка интеграции, а случайное блуждание (с дрейфом или без) – порядка интеграции 1.

Рассмотрим для начала два временных ряда первого порядка интеграции: $x_t \sim I(1), y_t \sim I(1)$, т. е. оба они имеют единичный корень, являются нестационарными рядами типа DSP. Их линейная комбинация, т. е. сумма с некоторыми коэффициентами, $Z_t = ax_t + by_t \sim I(1)$, вообще говоря, тоже является рядом порядка интеграции 1. Однако может оказаться, что существуют такие коэффициенты $(a, b)$, что линейная комбинация с этими коэффициентами окажется процессом нулевого порядка интеграции, стационарным случайным процессом. И если такая линейная комбинация существует, то ряды $x_t$ и $y_t$ называются коинтегрированными, а вектор с компонентами $(a, b)$ называется коинтегрирующим вектором.

Это определение легко обобщить на общий случай двух рядов порядка интеграции выше первой. Пусть процессы $x_t \sim I(d), y_t \sim I(d)$ имеют одинаковый порядок интеграции $d$. Тогда если существует вектор $(a, b)$, такой что $ax_t + by_t \sim I(d - b)$, где $b > 0$, то процессы $x_t$ и $y_t$ называются коинтегрированными. И это записывается так: $x_t, y_t \sim CI(d, b)$. Принято обозначать через $d$ исходный порядок интеграции, а через $b$ – на сколько порядок понижается. Исходя из таких обозначений, в предыдущем случае мы можем записать $x_t, y_t \sim CI(1,1)$. В этом наиболее распространеннном случае часто опускают параметры в скобках, т. е. пишут просто $x_t, y_t \sim CI$.

Определение легко можно обобщить на многомерный случай. Пусть многомерный случайный процесс $w_t$ имеет порядок интеграции $d$: $w_t \sim I(d)$, т. е. все компоненты вектора являются процессами одного и того же порядка интеграции $d$, и существует некоторая линейная комбинация компонент вектора, которая имеет порядок интегрирования $d - b$ ($b > 0$). Тогда компоненты процесса $w_t$ называются коинтегрированными: $(aw_t) \sim I(d - b)$, а вектор $a$ называется коинтегрирующим вектором. Правда, как мы увидим в дальнейшем, в многомерном случае может существовать несколько линейно независимых коинтегрирующих векторов.

Вернемся к случаю двух процессов первого порядка интеграции. На первый взгляд непонятно, каким образом суммирование случайных процессов может понизить порядок интеграции. Однако идея получения коинтеграции весьма проста. Ситуация несколько напоминает следующую. Представьте себе, что у вас есть две случайных величины. Вообще говоря, их линейная комбинация также есть случайная величина. Однако бывают такие случайные величины, что их некоторая линейная комбинация может оказаться детерминированной величиной. Например, определим случайную величину $Y$ как сумму случайной величины $Z$ и детерминированной величины. Тогда линейная комбинация случайных величин $Y$ и $Z$ с коэффициентами 1 и –1 соответственно, очевидно, является детерминированной величиной. Аналогичная ситуация может быть со случайными процессами.

Может существовать их линейная комбинация, которая является стационарной. Прежде чем анализировать формальные последствия определения коинтеграции и различные примеры, давайте рассмотрим, что означает существование коинтегрирующих процессов содержательно. Если $x_t, y_t \sim CI(1, 1)$, то каждый из процессов является процессом типа случайного блуждания, который имеет стохастический тренд. А коинтеграция означает, что стохастические тренды этих процессов движутся в унисон друг с другом, поэтому еще говорят, что они имеют общий стохастический тренд. Существует некоторое соотношение между двумя экономическими величинами (линейная комбинация $ax_t + by_t$), которое является стационарным процессом.
Соотношение можно переписать в виде \( y_t = g x_t + u_t \), где \( u_t \) – стационарный процесс. Если взять математическое ожидание от обоих частей, то оказывается, что зависящие от времени математические ожидания нестационарных процессов \( x_t \) и \( y_t \) связаны детерминированным соотношением: между ними существует долгосрочная связь. Вот в чем существенная разница. Кointеграция совместима с понятием долгосрочного равновесия. Хотя каждый из нестационарных процессов «блуждает» случайным образом, наличие кointеграции «застывает» их «блуждать» вместе, не уходя далеко друг от друга. Поскольку термин «случайное блуждание» происходит из шуточной задачи описания траектории движения пьяного в чистом поле, можно проиллюстрировать кointеграцию как движение в чистом поле двух пьяных, связанных эластичным жгутом. Каждый движется сам по себе, но далеко разойтись они не могут. Если же любая линейная комбинация двух процессов является нестационарной (процессы не кointегрированы), то окрестность любого соотношения между переменными наблюдается с нулевой вероятностью. Мы отмечали, что одним из свойств нестационарного процесса является бесконечно возрастающая дисперсия, что влечет за собой возвращение к исходной точке с вероятностью нуль и уход от нее сколь угодно далеко. И поэтому говорить о равновесном соотношении бессмысленно с содержательной и статистической точек зрения. Если же существует стационарное соотношение, то экономически это означает, что оно наблюдается очень часто, и его можно рассматривать как долгосрочное равновесие. А если нет кointегрирующего соотношения, то просто бессмысленно называть равновесием то значение, к которому процесс практически никогда не вернется.

Следовательно, кointегрирующее соотношение соответствует тому, что между рассматриваемыми величинами существует долгосрочное равновесие. И тогда общая динамика поведения показателей может быть разложена на две составляющие: долгосрочное поведение и краткосрочное поведение. Причем долгосрочное поведение описывается кointегрирующим соотношением. Посмотрим, как описывается краткосрочное поведение. Пусть кointегрирующее соотношение представляется в нормированном виде следующим образом: \( y_t = g x_t + e_t \), где \( e_t \) – белый шум. Поскольку \( x_t, y_t \) являются процессами \( I(1) \), то процессы их разностей \( \Delta x_t \) и \( \Delta y_t \) являются процессами типа \( I(0) \), т. е. стационарными. Содержательно \( \Delta x_t \) и \( \Delta y_t \) описывают собой краткосрочные изменения исходных переменных. Поэтому уже знакомая нам модель коррекции ошибками (ECM) \( \Delta y_t = a \Delta x_t + b (y_{t-1} - g x_{t-1}) + v_t \) связывает между собой стационарную величину \( \Delta y_t \), стационарную \( \Delta x_t \) и стационарное кointегрирующее соотношение. Поскольку выражение в скобках является стационарным, экономическая интерпретация модели коррекции ошибками сохраняется такой же, как если бы переменные \( y_t \) и \( x_t \) были стационарными. Краткосрочное изменение переменной \( y_t \) зависит от краткосрочного изменения переменной \( x_t \) и от отклонения от долгосрочного равновесия в предыдущий момент времени. То есть вводится коррекция в зависимости от того, на сколько переменные отклонились от своего долгосрочного равновесия. Напомним, что экономически такой подход оказался очень плодотворным. Выведено, что можно в одном уравнении совместить вместе краткосрочное и долгосрочное поведение. Отметим, что ECM имеет один и тот же вид как для стационарных, так и для нестационарных, но кointегрированных переменных.

Теперь убедимся, что кointегрированные процессы существуют. Рассмотрим, следуя Энглу и Грэнжеру, следующий пример. Пусть случайные процессы \( x_t \) и \( y_t \) определены следующим образом: \( x_t + b y_t = u_t, x_t + a y_t = v_t \). Пусть также \( u_t = u_{t-1} + e_1 t \) и \( v_t = r v_{t-1} + e_2 t \), где \( e_1 t \) и \( e_2 t \) – некоррелированные белые шумы. При условии, что \(|r| < 1, v_t \) является стационарным процессом, а \( u_t \) – случайным блужданием без дрейфа. Другими словами: \( u_t \sim I(1), v_t \sim I(0) \). Совместное распределение \( x_t \) и \( y_t \) полностью определяется приведенными уравнениями. Отметим, что имеет смысл рассматривать модель только при \( a \neq b \), так как иначе система противоречива.

97
Определим порядок интеграции процессов \( x_t \) и \( y_t \). Для этого перейдем от структурной к приведенной форме модели, т. е. разрешим систему относительно \( x_t \) и \( y_t \). Получаем:

\[
x_t = \frac{a}{a-b} u_t - \frac{b}{a-b} v_t \sim I(1),
\]
\[
y_t = -\frac{1}{a-b} u_t + \frac{1}{a-b} v_t \sim I(1).
\]

Очевидно, что каждый из процессов \( x_t \) и \( y_t \) является процессом первого порядка интеграции. В то же время, очевидно, что существует их линейная комбинация \( x_t + ay_t \), равная стационарному процессу, поэтому они являются коинтегрированными в смысле ранее данного определения. Поэтому \( x_t, y_t \sim CI(1, 1) \).

Приведенный простой пример подсказывает, что свойство коинтеграции связано с представлением многомерного случайного процесса в виде линейной модели. Особенно удобным оказалось использование VAR-модели. Наличие или отсутствие коинтеграции можно установить по значениям характеристических корней VAR-модели. Начнем с простейшего случая модели VAR(1) для двумерного процесса. Сохраняя обозначения предыдущего параграфа, рассмотрим модель

\[
\begin{pmatrix}
X^1_t \\
X^2_t
\end{pmatrix} = \begin{pmatrix}
a_1 & b_{12} \\
b_{21} & b_{22}
\end{pmatrix} \begin{pmatrix}
X^1_{t-1} \\
X^2_{t-1}
\end{pmatrix} + \begin{pmatrix}
e^1_t \\
e^2_t
\end{pmatrix}.
\]

Обозначим

\[
A_1 = \begin{pmatrix}
b_{11} & b_{12} \\
b_{21} & b_{22}
\end{pmatrix}.
\]

Выше мы уже использовали приведение системы к простейшей форме, применяя диагональный вид матрицы \( A_1 \) для случая, когда собственные числа этой матрицы различны. Оказывается, это представление позволяет исчерпывающе исследовать свойство коинтеграции.

Без потери общности перейдем к центрированным переменным. Пусть собственные числа матрицы \( A_1 \) различны: \( \lambda_1 \neq \lambda_2 \). Как и выше, определим вектор новых переменных соотношением \( \bar{\eta}_t = C^{-1} \bar{x}_t \), где \( C \) – матрица, столбцы которой являются собственными векторами, соответствующими собственным числам \( \lambda_1 \) и \( \lambda_2 \).

В новых переменных система уравнений распадается на два отдельных уравнения:

\[
\begin{align*}
\gamma^1_t &= \lambda_1 \gamma^1_{t-1} + \eta^1_t, \\
\gamma^2_t &= \lambda_2 \gamma^2_{t-1} + \eta^2_t,
\end{align*}
\]
где через \( \eta_t \) обозначен преобразованный векторный белый шум.

Как мы уже установили ранее, если хотя бы одно из собственных чисел превышает по модулю единицу, то хотя бы одна из компонент вектора \( \bar{y}_t \) не приводится к стационарному процессу взятием последовательных разностей. Но тогда этим свойством обладают и компоненты процесса \( \bar{x}_t \), так как его компоненты являются линейными комбинациями компонент вектора \( \bar{y}_t \) (напомним, что \( \bar{x}_t = C \bar{y}_t \)).

Поэтому в дальнейшем мы не будем рассматривать случай «взрывной» нестационарности. Если же оба корня по модулю меньше единицы, то компоненты процесса \( \bar{x}_t \) являются стационарными, и вопрос о коинтеграции не возникает.
Пусть теперь $\lambda_1 = 1$ и $\lambda_2 < 1$. Тогда $y_t^1 \sim I(1)$, $y_t^2 \sim I(0)$, и обе компоненты вектора $\tilde{x}_t$, как линейные комбинации процессов нулевого и первого порядков интеграции, будут первого порядка интеграции. Но нельзя забывать, что справедливо и обратное: компоненты вектора $\tilde{y}_t$ являются линейными комбинациями компонент вектора $\tilde{x}_t$. Следовательно, существует линейная комбинация процессов $x_t^1$ и $x_t^2$, первого порядка интеграции каждый, являющаяся стационарным процессом. Коэффициенты этой линейной комбинации являются элементами второй строки матрицы $C^{-1}$. В соответствии с определением процессы $x_t^1$ и $x_t^2$ коинтегрированы.

Ранее мы ввели в рассмотрение матрицу $\Pi = I - A_1$ и отметили, что собственные векторы матриц $\Pi$ и $A_1$ совпадают, а их собственные числа дополняют друг друга до единицы. Поэтому условие наличия одного единичного собственного числа у матрицы $A_1$ означает наличие одного нулевого собственного числа у матрицы $\Pi$, а это, в свою очередь, эквивалентно равенству единице ранга матрицы $\Pi$. Используя диагональное представление матрицы $\Pi$, получаем:

$$\Pi = C \begin{pmatrix} 0 & 0 \\ 0 & 1 - \lambda_2 \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 1 - \lambda_2 \end{pmatrix} \begin{pmatrix} \tilde{c}_{11} & \tilde{c}_{12} \\ \tilde{c}_{21} & \tilde{c}_{22} \end{pmatrix} = \begin{pmatrix} (1 - \lambda_2)c_{12} \\ (1 - \lambda_2)c_{22} \end{pmatrix} \begin{pmatrix} \tilde{c}_{11} & \tilde{c}_{12} \\ \tilde{c}_{21} & \tilde{c}_{22} \end{pmatrix}.$$

Здесь через $c_{ij}$ обозначены элементы матрицы $C$, а через $\tilde{c}_{ij} - \text{ элементы матрицы } C^{-1}$. Обратим внимание, что матрица $\Pi$ разложена в произведение ( faktorizovana) двух прямоугольных матриц: вектора-столбца на вектор-строку, причем вектор-строка представляет собой коинтегрирующий вектор. Еще раз подчеркнем, что это просто вторая строка матрицы, обратной матрице, составленной из собственных векторов матрицы $\Pi$ или $A_1$.

Пусть теперь $\lambda_1 = \lambda_2 = \lambda$. В этом случае, как мы уже упоминали, существует невырожденная матрица $P$, такая что $P^{-1}A_1P = J$ и $A_1 = PJP^{-1}$, где $J - жорданова матрица$:

$$\begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$$

Определив вектор $\tilde{y}_t = P^{-1}\tilde{x}_t (\tilde{x}_t = P\tilde{y}_t)$, получаем систему соотношений:

$$y_t^1 = \lambda y_{t-1}^1 + y_{t-1}^2 + \eta_t^1,$n
$$y_t^2 = \lambda y_{t-1}^2 + \eta_t^2.$$

При $|\lambda| < 1$ второе уравнение представляет собой стационарную модель AR(1), а первое – стационарную модель ADL(1, 1), и о коинтеграции речь не идет.

Если же $\lambda = 1$, то из второго уравнения следует, что $y_t^2$ является случайным блужданием, т. е. $y_t^2 \sim I(1)$. Поскольку первое уравнение сводится к виду $\Delta y_t^1 = y_{t-1}^2 + \eta_t^1$, то для достижения стационарности нужно еще раз взять последовательную разность. Это означает, что $y_t^1 \sim I(2)$. Поскольку $\tilde{x}_t = P\tilde{y}_t$, его обе компоненты являются процессами порядка интеграции 2. В то же время существует их линейная комбинация ($y_t^2$), чей порядок интеграции равен 1. Следовательно, эти компоненты коинтегрированы: $x_t^1$, $x_t^2 \sim CI(2, 1)$. Оба собственных числа матрицы $\Pi$ равны в этом случае нулю, а ее ранг равен 1. В этом случае матрица

$$\Pi = P \begin{pmatrix} 0 & -1 \\ 0 & 0 \end{pmatrix} P^{-1} = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \tilde{p}_{11} & \tilde{p}_{12} \\ \tilde{p}_{21} & \tilde{p}_{22} \end{pmatrix} = \begin{pmatrix} -p_{11} \\ -p_{21} \end{pmatrix} \begin{pmatrix} \tilde{p}_{11} & \tilde{p}_{12} \\ \tilde{p}_{21} & \tilde{p}_{22} \end{pmatrix}$$

также факторизуется в произведение прямоугольных матриц, и вторая из них даст коинтегрирующий вектор.

Наконец, при равенстве собственных чисел может тем не менее существовать два линейно независимых собственных вектора, соответствующих этому собственному числу. В этом случае матрицы $A_1$ и $\Pi$ приводятся к диагональной форме, и система уравнений для вектора $\tilde{y}_t$ принимает вид:

$$y_t^1 = y_{t-1}^1 + \eta_t^1, \quad y_t^2 = y_{t-1}^2 + \eta_t^2.$$
Обе компоненты относятся к типу \( I(1) \), поэтому обе компоненты вектора \( \tilde{x}_t \) также относятся к типу \( I(1) \), а коинтегрирующей комбинации нет. Заметим, что матрица \( \Pi \) в этом случае состоит из нулей и ее ранг равен 0.

Оказывается, результат рассмотрения удобнее и элегантнее выразить в терминах ранга матрицы \( \Pi \), чем в количестве единичных корней матрицы \( A_1 \). Мы фактически установили, что если двумерный случайный процесс имеет представление в виде модели VAR(1) и среди его характерических корней ни один по модулю не превышает единицу, то количество коинтегрирующих линейно независимых векторов между компонентами процесса в точности равно рангу матрицы \( \Pi = I - A_1 \). При этом вовсе не обязательно пользоваться VAR-моделью, важно лишь, чтобы многомерный процесс допускал VAR(1) представление. Сразу отметим, что ранг матрицы \( \Pi \) равен числу нулевых собственных чисел этой матрицы.

Осталось обобщить этот вывод на случай процесса любой размерности и общую модель VAR(\( p \). Обобщение на многомерный процесс достаточно очевидно из цепочки предыдущих рассуждений. Рассмотрим общий случай модели VAR(\( p \):

\[
\tilde{y}_t = A_1 \tilde{y}_{t-1} + \cdots + A_p \tilde{y}_{t-p+1} + \tilde{e}_t.
\]

Это уравнение можно переписать в другом виде после таких же преобразований, как и при выводе уравнения ADF теста:

\[
\Delta \tilde{y}_t = -\Pi \tilde{y}_{t-1} + B_1 \Delta \tilde{y}_{t-1} + \cdots + B_{p-1} \Delta \tilde{y}_{t-p+1} + \tilde{e}_t.
\]

Здесь матрица \( \Pi = I - A_1 - A_2 - \cdots - A_p \), а матрицы \( B_1, \ldots, B_{p-1} \) выражаются через матрицы \( A_1, \ldots, A_p \) следующим образом: \( B_j = -\sum_{i=j+1}^{p} A_i \).

Доказательство того, что наличие коинтеграции связано с вырожденностью матрицы \( \Pi \), или, более точно, что количество коинтегрирующих векторов в точности равно рангу матрицы \( \Pi \), требует использования блочных матриц и останется за рамками нашего курса. Доказательство можно найти, например, для \( p = 2 \) в учебнике Джонстона и Ди Нардо. Идея доказательства заключается в стандартном сведении системы разностных уравнений порядка \( p \) для \( n \) переменных к системе разностных уравнений первого порядка, но для \( pn \) переменных. Матрица преобразованной системы обладает блочной структурой, а характеристические уравнения исходной и преобразованной систем совпадают. Эта часть доказательства и является наиболее сложной технически. Для исходного уравнения, как мы уже упоминали, характеристическое уравнение имеет вид:

\[
\det(\lambda^p I - \lambda^{p-1} A_1 - \cdots - A_p) = 0.
\]

Следовательно, существование единичного корня у исходного уравнения эквивалентно существованию нулевого корня у вышеприведенной матрицы \( \Pi \). С помощью рассуждений, аналогичных проведенным выше, мы приходим к следующему заключению. Ранг матрицы \( \Pi = I - A_1 - A_2 - \cdots - A_p \) равен числу линейно независимых коинтегрирующих векторов. Этот показатель принято называть коинтегрирующим рангом.

Итак, если между компонентами вектора \( \tilde{y}_t \) существуют коинтегрирующие соотношения, то их количество равно количеству нулевых собственных чисел матрицы \( \Pi \), причем не все коинтегрирующие соотношения дают стационарные линейные комбинации.
4.8. Коинтеграционная регрессия и тест Йохансена

Пусть у нас есть выборка объема $T$ из обеих переменных $x_t, y_t$, которые коинтегрированы. Попробуем построить обычную регрессию $x_t$ на $y_t$ методом наименьших квадратов. Как мы помним, в этом случае существует опасность получить в результате кажущуюся регрессию. Удивительная вещь заключается в следующем. Несмотря на то, что $x_t$ и $y_t$ являются процессами типа $I(1)$, оказывается, что в случае, если процессы коинтегрированы, оценка МНК коэффициентов регрессии становится состоятельной. И можно, несмотря на нестационарность переменных, применять обычные регрессионные методы.

Более того, можно показать, что скорость сходимости оценки выше, чем в случае оценки МНК для стационарных регрессий. Поскольку дисперсия случайного блуждания растет пропорционально $T$, скорость сходимости при наличии коинтеграции пропорциональна $1/T$, в то время как для стационарных регрессоров скорость сходимости пропорциональна только $1/\sqrt{T}$. Это свойство оценок МНК при наличии коинтеграции называют суперсостоятельностью.

Тест Йохансена. Основные из применяемых в настоящее время тестов наличия коинтеграции основаны на связи количества коинтегрирующих векторов с рангом матрицы $P$ в VAR-представлении многомерного временного ряда, а следовательно, с количеством нулевых собственных чисел этой матрицы. Мы рассмотрим лишь один из тестов, предложенный Йохансеном [1]. Пожалуй, этот тест является наиболее популярным и входит в большинство специализированных компьютерных пакетов, в частности в Eviews.

Первым шагом в реализации теста Йохансена является оценка модели VAR($p$) для заданного $k$-мерного временного ряда

$$\hat{y}_t = \{\hat{\mu} + t \hat{\beta}\} + A_1 \hat{y}_{t-1} + \cdots + A_p \hat{y}_{t-p} + \hat{e}_t.$$

По сравнению с ранее рассматриваемой моделью в нее добавлены векторы линейного тренда и допускается ненулевое математическое ожидание для каждой из компонент. Фигурные скобки указывают, что модель может не содержать ни одного из этих слагаемых, может включать только вектор свободных членов, а может включать оба дополнительных слагаемых. Разумеется, обычные условия для остатков модели предполагаются выполненными: остатки гомосkedастичны и не коррелированы. Первоначально Йохансен требовал также нормальности остатков, но в 1995 г. ослабил это условие до требования, чтобы остатки не очень сильно отличались от Гауссова белого шума. Мы уже рассматривали тесты для проверки этих свойств.

Порядок модели $p$ выбирают обычно с использованием информационных критериев Акаике и Шварца среди моделей, прошедших диагностику остатков. При переборе моделей включают/исключают также свободные член и линейный тренд. Возможно также проверить гипотезу о том, что порядок VAR-модели равен $p$, против альтернативной гипотезы, что ранг равен $q < p$. Для проверки используется тест отношений правдоподобия. Если вектор случайных возмущений является гауссовым белым шумом, то логарифм функции правдоподобия для $T$ наблюдений и $k$-мерной переменной принимает следующий вид [9]:

$$\ln L = const + \frac{T}{2} \ln \left| \hat{\Omega}^{-1} \right|,$$

где $\hat{\Omega}$ – оценка ковариационной матрицы остатков VAR-модели. Отношение правдоподобий поэтому принимает вид:

$$LR = -2(\ln L_q - \ln L_p) = T \left( \ln \left| \hat{\Omega}_q^{-1} \right| - \ln \left| \hat{\Omega}_p^{-1} \right| \right).$$
При справедливости нулевой гипотезы, т. е. если порядок модели равен \( p \), эта статистика распределена асимптотически как хи-квадрат с числом степеней свободы, равным \( k^2(p-q) \).

После оценки VAR-модели в тесте Йохансена рассчитывается оценка матрицы \( \Pi = I - A_1 - A_2 - \cdots - A_p \), соответствующая ECM-представлению модели

\[
\Delta \tilde{y}_t = -\Pi \tilde{y}_{t-1} + B_1 \Delta \tilde{y}_{t-1} + \cdots + B_{p-1} \Delta \tilde{y}_{t-p+1} + \tilde{e}_t.
\]

Нулевой гипотезой является то, что ранг матрицы \( \Pi \) не превышает некоторого числа \( r < k \). В качестве альтернативной гипотезы используется \( H_1: rank \Phi = k \), или \( H_1 = r + 1 \). Статистика для проверки нулевой гипотезы против первой из приведенных альтернатив имеет вид

\[-T \sum_{i=r+1}^{k} \ln(1 - \hat{\lambda}_i)\]

и называется \textit{trace statistic}. Название связано с тем, что статистика пропорциональна сумме логарифмов собственных чисел матрицы (остальные \((k-r)\) собственных чисел считаются нулевыми), т. е. следу матрицы.

Для проверки против второй альтернативной гипотезы тестовая статистика принимает вид:

\[\ln(1 - \hat{\lambda}_{r+1})\]

и называется \textit{max statistic}. Название связано с тем, что статистика пропорциональна логарифму максимального из собственных чисел матрицы, считающихся нулевыми. В приведенных выражениях через \( \hat{\lambda}_i \) обозначена оценка максимального правдоподобия \( i \)-го корня полученного Йохансеном уравнения. При этом предполагается, что корни упорядочены в порядке убывания.

В пакете Eviews применяется \textit{trace statistic}. В этом пакете последовательно перебираются значения \( r \) от 0 до \( k \). Если нулевая гипотеза \( r = 0 \) не отвергается уже на первом шаге, то процесс нестационарный и коинтеграции не существует. Если гипотеза \( r = 0 \) отвергается, на следующем шаге проверяется гипотеза \( r = 1 \). Если она не отвергается, то коинтегрирующий ранг равен 1. Если же гипотеза отвергается, то переходим к проверке нулевой гипотезы \( r = 2 \) и так далее. Стационарности исследуемого многомерного процесса соответствует отвержение нулевой гипотезы при всех \( r < k \). Распределения обоих тестовых статистик не стандартны, и их критические значения получены моделированием Монте-Карло. Таблицы критических чисел зависят от того, входят ли детерминированный тренд и свободный член в VAR-уравнение, и входит ли линейный член в коинтегрирующее соотношение. В пакете Eviews реализованы следующие пять возможностей:

- ни свободный член, ни тренд не входят ни в VAR-уравнение, ни в коинтегрирующее соотношение;
- свободный член входит в коинтегрирующее соотношение, ни свободный член, ни тренд не входят в VAR-уравнение;
- свободный член входит как в коинтегрирующее соотношение, так и в VAR уравнение;
- свободный член и тренд входят в коинтегрирующее соотношение, тренд не входит в VAR-уравнение;
- свободный член и тренд входят в коинтегрирующее соотношение, тренд входит в VAR-уравнение.
Первые два случая подразумевают отсутствие детерминированного тренда в данных, два следующих – наличие детерминированного тренда в данных, а последний допускает наличие квадратичного детерминированного тренда в данных. Нужный вариант теста обычно выбирают исходя из характера поведения данных и теоретических представлений о виде долгосрочного соотношения между переменными. Самый простой случай – это отсутствие свободного члена как в коинтегрирующем уравнении, так и в VAR-модели. Это означает, что математические ожидания всех рассматриваемых переменных равны нулю. Визуальная инспекция данных и проверка выборочных средних на равенство нулю обычно достаточно для отказа от такой спецификации теста Йохансена (первого в нашем списке). Часто коинтегрирующее уравнение содержит свободный член из теоретических соображений. Так, при моделировании производственной функции типа Кобба-Дугласа или функции потребления с постоянной склонностью к потреблению стандартной процедурой является переход к линейной в логарифмах модели. При этом масштабирующий множитель, входящий в исходную спецификацию модели, переходит при логарифмировании в свободный член (константу) уравнения долгосрочного равновесия.

Наше рассмотрение теста Йохансена до сих пор ограничивалось выяснением: коинтегрированы или нет переменные. В случае положительного ответа на поставленный вопрос следует приступить к оценке модели. Йохансен предложил не только процедуру тестирования коинтеграции, но и процедуру оценивания коинтегрирующих векторов. Такая оценка (методом максимального правдоподобия) в настоящее время является составной частью большинства компьютерных реализаций теста Йохансена. Вспомним, что коинтегрирующие векторы являются столбцами матрицы $b$ из разложения $\Pi = ab^T$. Возможность факторизации матрицы явно указывает на наличие линейных соотношений между ее элементами. В свою очередь, это означает, что учет этих ограничений при оценивании даёт более эффективные оценки коэффициентов модели.

Такой учет может быть осуществлен путём оценки модели в форме ECM с прямым включением в неё долгосрочных соотношений (обобщение процедуры Энгла-Грэнжера) либо в VAR-представлении, оценивая каждое из уравнений методом наименьших квадратов при ограничениях.

Серьезной практической проблемой является обнаружение нескольких коинтегрирующих векторов. Если существует лишь одно коинтегрирующее соотношение между компонентами вектора, его можно трактовать как долгосрочное равновесие между этими переменными. В случае нескольких коинтегрирующих соотношений существует несколько линейных комбинаций компонент, которые являются стационарными. А это означает, как мы уже отмечали, что любая линейная комбинация этих линейных комбинаций также является стационарной. И поэтому сложно выявить комбинации, которые можно содержательно интерпретировать как экономические равновесия. Можно сослаться, например, на опыт построения функции спроса на деньги Йохансеном и Джузелиусом для Дании и Финляндии. Для Дании использовалась выборка квартальных данных с I кв. 1974 г. по III кв. 1987 г. (55 наблюдений). Для Финляндии выборка включала 67 наблюдений с I кв. 1958 г. по III кв. 1984 г. Для Дании было получено одно коинтегрирующее соотношение, которое хорошо трактуется и даёт возможность построить содержательную теоретически модель спроса на деньги. А во втором случае, для Финляндии, было получено три коинтегрирующих соотношения, и интерпретация построенных соотношений в значительной степени затруднена.
4.9. Пример применения коинтеграции

Рассмотрим два ряда (рис. 19, 20).

![Рис. 19. Курс доллара к рублю, ряд Y](image1)

![Рис. 20. Цена барреля нефти в рублях, ряд X](image2)

Проверим ряды на стационарность, используя Eviews 7, с помощью Augmented Dickey-Fuller test statistic (таблица 20).

<table>
<thead>
<tr>
<th>Тест на единичный корень</th>
</tr>
</thead>
<tbody>
<tr>
<td>t-Statistic</td>
</tr>
<tr>
<td>−0.911440</td>
</tr>
</tbody>
</table>

critical values

<table>
<thead>
<tr>
<th>1 % level</th>
<th>5 % level</th>
<th>10 % level</th>
</tr>
</thead>
<tbody>
<tr>
<td>−3.442209</td>
<td>−2.866663</td>
<td>−2.569559</td>
</tr>
</tbody>
</table>

Мы видим, что −0.91440 > −3.442209, следовательно, ряд X не стационарный. Однако ряд из разностной производной является стационарным (таблица 21), так как −24.31759 меньше всех значений в столбце, который находится ниже. Следовательно, ряд X не стационарный порядка интегрирования 1. Аналогично доказывается, что ряд Y тоже обладает этими свойствами.

<table>
<thead>
<tr>
<th>Тест DF D(X)</th>
</tr>
</thead>
<tbody>
<tr>
<td>d(X) has a unit root</td>
</tr>
<tr>
<td>−24.31759</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1 % level</th>
<th>5 % level</th>
<th>10 % level</th>
</tr>
</thead>
<tbody>
<tr>
<td>−3.442231</td>
<td>−2.866673</td>
<td>−2.569564</td>
</tr>
</tbody>
</table>

Проведем тест на причинность по Грейнджеру (таблица 22). Из теста следует, что X есть причина для Y.
Тест Грейнжера

<table>
<thead>
<tr>
<th>Null Hypothesis</th>
<th>Obs</th>
<th>F-Statistic</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y does not Granger Cause X</td>
<td>540</td>
<td>4.85503</td>
<td>0.0081</td>
</tr>
<tr>
<td>X does not Granger Cause Y</td>
<td></td>
<td>0.38806</td>
<td>0.6786</td>
</tr>
</tbody>
</table>

После теста Йохансена следует, что по информационному критерию Akaike наименьшее его значение равно 8.338821 и уравнение должно удовлетворять условию: Data Trend: None, Intercept, No Trend. Таким образом, коинтеграционное уравнение должно иметь вид \( Y = C + X \).

Из таблицы 23 следует, что уравнение является значимым \( F\)-statistic = 5251.012, причем \( R^2 = 0.906752 \).

Оценка коинтегрального уравнения

<table>
<thead>
<tr>
<th>variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>167.1757</td>
<td>1.374456</td>
<td>121.6305</td>
<td>0.0000</td>
</tr>
<tr>
<td>X</td>
<td>-1.811476</td>
<td>0.024998</td>
<td>-72.46387</td>
<td>0.0000</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.906752</td>
<td>S.E. of regression</td>
<td>8.274630</td>
<td>F-statistic 5251.012</td>
</tr>
</tbody>
</table>

Искомое уравнение имеет вид: (таблица 23)

\[
Y = 167.17574183 - 1.81147603739 \cdot X.
\]

Из этого уравнения следует, что увеличение цены барреля на 1 рубль приведет к падению курса доллара на 1.81147603739 руб.
5. Задачи

1. Дан временной ряд \( x = (5, 1, 1, 3, 2, 9, 6, 2, 5, 2) \). Вычислите среднее, дисперсию (смещенную), автоковариационную и автокорреляционную матрицы.

2. Для временного ряда \( x = (7, -9, 10, -2, 21, 13, 40, 36, 67, 67) \) оцените параметры полиномиального тренда второго порядка. Постройте точечный и интервальный прогнозы по тренду на 2 шага вперед.

3. Сгенерируйте 20 рядов, задаваемых полиномиальным трендом третьего порядка \( y_t = 5 + 4t - 0.07t^2 + 0.0005t^3 \) длиной 100 наблюдений, с добавлением белого шума с нормальным распределением и дисперсией 20. Допустим, истинные значения параметров тренда неизвестны.
   A. Для 5 рядов из 20 оцените полиномиальный тренд первого, второго и третьего порядков и выберите модель, которая наиболее точно аппроксимирует сгенерированные данные.
   B. Для 20 рядов оцените полиномиальный тренд третьего порядка по первым 50 наблюдениям. Вычислите оценки параметров тренда и их ошибки. Сравните оценки с истинными значениями параметров.
   V. Проведите те же вычисления, что и в пункте (B), для 20 рядов, используя 100 наблюдений. Результаты сравните.
   Г. Используя предшествующие расчеты, найдите точечные и интервальные прогнозы на три шага вперед с уровнем доверия 95 %.

4. Дан ряд: \( x = (10, 9, 12, 11, 14, 12, 17, 14, 19, 16, 18, 21, 20, 23, 22, 26, 23, 28, 25, 30) \).
   A. Оцените модель линейного тренда. Остатки, полученные после исключения тренда, проверьте на стационарность с использованием рангового коэффициента корреляции Спирмена.
   B. Рассчитайте для остатков статистику Голдфельда-Квандта, исключив 6 наблюдений из середины ряда. Проверьте однородность выборки по дисперсии. Сравните с выводами, полученными на основе критерия Спирмена.

5. Посчитать коэффициент автокорреляции первого порядка для ряда \( x = (2, 4, 6, 8) \).

6. Ряд \( x = (0.02, 0.05, 0.06, 0.13, 0.15, 0.2, 0.31, 0.46, 0.58, 0.69, 0.78, 0.81, 0.95, 0.97, 0.98) \) характеризует долю семей, имеющих телевизор. Оцените параметры логистического тренда.

7. Пусть \( L \) – лаговый оператор. Представьте в виде степенного ряда следующие выражения:
   a) \( \frac{2}{1-0.8L} \); б) \( \frac{1.5}{1-0.9L} \); в) \( \frac{2.8}{1+0.4L} \); г) \( \frac{3}{1+0.5L} \).

8. Сгладить временной ряд \( x = (3, 4, 5, 6, 7, 11) \), используя полином первого порядка с длиной отрезка скольжения, равной трем.

9. Сгенерируйте ряд из 100 наблюдений на основе полиномиального тренда \( x_t = 5 + 4t - 0.07t^2 + 0.0005t^3 \) с добавлением белого шума с нормальным распределением и дисперсией 20. Проведите сглаживание сгенерированного ряда с помощью полинома первой степени с длиной отрезка скольжения 5 и 9. Выполните то же задание, используя полином третьей степени. Найдите отклонения исходных рядов от сглаженных рядов, полученных выше. По каждому ряду отклонений вычислите среднеквадратическую ошибку. Сделайте вывод о том, какой метод дает наименьшую среднеквадратическую ошибку.
10. Сгенерируйте ряд длинной 4000 наблюдений по модели AR(1) с параметром \( \varphi = 0.5 \) и нормально распределенной не автокоррелированной ошибкой с единичной дисперсией, предполагая, что значение ряда в момент \( t = 0 \) равно нулю. Считая, что действительности вид модели неизвестен, а задан только ряд. Оцените модели AR(1), AR(2) и модель линейного тренда. Постройте прогноз на 1 шаг вперед с помощью трех моделей: AR(1), AR(2) и модели линейного тренда и найдите ошибку прогноза. Сравните среднеквадратические ошибки прогноза по трем моделям и сделайте выводы.

11. Сгенерируйте 10 рядов длиной 20 наблюдений по модели AR(1) с авторегрессионным параметром \( \varphi_k = \frac{k-1}{10}, k = 1, ..., 10, \) с нормально распределенной не автокоррелированной ошибкой и единичной дисперсией. По каждому ряду с помощью МНК оцените модель AR(1). Постройте график отклонения оценки от истинного значения параметра в зависимости от истинного значения параметра \( \varphi_1 \). Что можно сказать по этому графику о поведении смещения оценок в зависимости от \( \varphi_1 \)? Подтверждается ли, что оценки смещены в сторону нуля и смещение тем больше, чем \( \varphi_1 \) ближе к единице?

12. Сгенерируйте 10 рядов длиной 20 наблюдений по модели MA(2) с параметрами \( \varphi_1 = 0.5, \varphi_2 = 0.3 \) и нормально распределенной не автокоррелированной ошибкой с единичной дисперсией. Для каждого из рядов постройте выборочную автокорреляционную функцию для лагов 1, 2, 3. Рассмотрите распределение коэффициентов автокорреляции и сравните их с теоретическими значениями. По каждому ряду на основе выборочных коэффициентов автокорреляции оцените модель MA(2), выбирая то решение квадратного уравнения, которое соответствует условиям обратимости процесса. Рассмотрите распределение оценок и сравните их с истинными значениями.

13. Вывести формулы для вычисления математического ожидания, дисперсии и ковариаций случайного процесса \( x_t = \mu + \varphi_1 x_{t-1} + e_t \) при условии его слабой стационарности, если \( e_t \) – белый шум с дисперсией \( \sigma^2 \).

14. Записать случайный процесс \( x_t = 0.2 + 0.6x_{t-1} + e_t \) с использованием лагового оператора в виде процесса скользящего среднего бесконечного порядка.

15. Задана модель: \( x_t = 0.25x_{t-1} + e_t \), где \( e_t \) – белый шум. Дисперсия процесса \( x_t \) равна единице. Вычислить дисперсию белого шума.

16. Прогнозирование в принятии управленческих решений способствует:
   а) принятию оптимального решения,
   б) моделированию будущих решений,
   в) предвидению вероятностных состояний внешней среды,
   г) другому.

17. Оперативный прогноз осуществляется сроком:
   а) до 1 года,
   б) до 3–6 месяцев,
   в) до 5 лет,
   г) свыше 5 лет.
18. Будут ли временным рядом следующие ряды:
a) число родившихся детей по регионам России в 2016 году,
b) экзаменационные оценки 358 группы по прогнозированию,
v) ежемесячные данные о числе проданных машин марки «Лада» в Барнауле за 2015 год.
19. Модель временного ряда имеет вид: \( x_t = 2.3 + 1.2 \cdot t + 0.5 \cdot x_{t-1} + e_t \). Какое высказывание об этом ряде является истинным?
a) ряд является стационарным,
b) ряд содержит тренд, но не имеет сезонных компонент,
v) ряд не содержит тренда,
g) содержит тренд и сезонную компоненту.
20. Модель временного ряда имеет вид: \( x_t = e_t - 0.5e_{t-1} + 0.1e_{t-2} \). Какие из следующих высказываний являются истинными?
a) ряд является не обратимым процессом MA(2),
b) ряд является нестационарным процессом ARMA(1,1),
v) ряд является обратимым процессом MA(2),
g) ряд является стационарным процессом AR(2).
21. «Белым шумом» называют
а) стационарный, «нормальный» процесс,
b) не стационарный процесс с постоянной дисперсией,
v) стационарный процесс с автокорреляцией,
g) стационарный процесс без автокорреляции и с постоянной дисперсией.
22. При экспоненциальном сглаживании величина параметра а
а) должна быть |a|<1,
b) должна быть a>1,
v) должна быть 0<a<1,
g) должна быть |a|>1.
23. Чему равна дисперсия Марковского процесса \( x_t = 0.5x_{t-1} + e_t \), если дисперсия белого шума равна 1? Изобразить график автокорреляционной функции данного процесса.
24. Для процесса \( x_t = -0.7 - 0.7x_{t-1} + e_t \), где \( e_t \) — белый шум, рассчитать частную автокорреляционную функцию, вычислить первые 6 значений автокорреляционной функции и начертить ее график.
25. Для модели AR(1): \( x_t = \mu + 0.5x_{t-1} + e_t \) показать, что частная автокорреляционная функция \( r_{part}(1) = 0.5, \ r_{part}(k) = 0 \) при \( k \geq 2 \).
26. Дана два марковских процесса: \( x_t = 0.5x_{t-1} + e_t; \) \( y_t = 0.2y_{t-1} + e_t. \) Дисперсия какого процесса больше и во сколько раз?

27. Найти математическое ожидание, дисперсию и ковариацию случайного процесса \( x_t, \) если \( x_t \) – белый шум с единичной дисперсией.
   а) \( x_t = 0.1 + 0.9x_{t-1} + e_t; \)
   б) \( x_t = -0.2x_{t-1} + e_t. \)

28. Пусть \( e_t \) – белый шум с единичной дисперсией. Найти математическое ожидание, дисперсию и ковариацию случайного процесса:
   а) \( x_t = 1 + 0.5x_{t-1} + e_t; \)
   б) \( x_t = 0.5x_{t-1} + e_t. \)

29. Проверить на стационарность следующие процессы:
   а) \( x_t - 0.4x_{t-1} - 0.4x_{t-2} = e_t; \)
   б) \( x_t + 0.4x_{t-1} - 0.4x_{t-2} = e_t; \)
   в) \( x_t - 0.4x_{t-1} + 0.4x_{t-2} = e_t. \)

30. Корни характеристического уравнения для процесса Юла (AR(2)) равны, соответственно, \( 1/5 \) и \( -1/5. \) Изобразить график автокорреляционной функции. Дать обоснование.

31. Корни характеристического уравнения, соответствующего процессу Юла, равны 10/19 и \( -10/13. \) Изобразить график автокорреляционной функции этого процесса. Ответ обосновать.

32. Коэффициенты автокорреляции первого и второго порядка в процессе Юла равны, соответственно, 0.5 и 0.4. Оценить параметры процесса. Найти дисперсию белого шума, если дисперсия процесса равна 1.

33. При каких значениях \( \varphi_2 \) следующие случайные процессы являются стационарными в широком смысле?
   а) \( x_t = x_{t-1} + \varphi_2x_{t-2} + e_t; \) б) \( x_t = -x_{t-1} + \varphi_2x_{t-2} + e_t. \)

34. Параметры \( \varphi_1 \) и \( \varphi_2 \) процесса AR(2) равны, соответственно, 0.6 и \( -0.2. \) Каковы первые три значения автокорреляционной функции?

35. Для процесса \( x_t = 0.5x_{t-1} + 0.25x_{t-2} + e_t \) коэффициенты автокорреляции первого и второго порядка равны, соответственно, 2/3 и 7/12. Найти коэффициент автокорреляции четвертого порядка.

36. Пусть процесс AR(2) \( x_t = \varphi_1x_{t-1} + \varphi_2x_{t-2} + e_t \) является стационарным в широком смысле, и \( e_t \) – белый шум с дисперсией \( \sigma^2. \) Вычислите значения частной автокорреляционной функции для лага 1 и 2.

37. По известным значениям частной автокорреляционной функции \( \rho_{part(1)} = 0.5 \) и \( \rho_{part(2)} = 2/3 \) случайного процесса найти значения коэффициентов автокорреляции первого и второго порядка.
38. Коэффициент автокорреляции первого порядка для обратимого процесса скользящего среднего первого порядка равен –0.4. Записать уравнение процесса и изобразить график его автокорреляционной функции.

39. Найти автокорреляционную функцию процесса: 
\[ x_t = e_t - 0.5x_{t-1} - 0.25x_{t-2} - 0.125x_{t-3} - 0.0625x_{t-4} - \ldots \]

40. Пусть \( e_t \) – белый шум с единичной дисперсией. Чему равна дисперсия процесса \( x_t = e_t + 0.2e_{t-1} \)? Изобразить график автокорреляционной функции.

41. Идентифицировать процесс, автокорреляционная функция которого имеет следующий вид:

а) \( \rho(1) = 0.25, \rho(k) = 0, \forall k \geq 2 \); б) \( \rho(1) = -0.4, \rho(k) = 0, \forall k \geq 2 \).

42. Для каждого из случайных процессов: а) \( x_t = e_t + 0.5e_{t-1} \); б) \( x_t = e_t - 0.5e_{t-1} \); в) \( x_t = -1 + +e_t + 0.8e_{t-1} \) рассчитать частную автокорреляционную функцию, вычислить первые 6 значений автокорреляционной функции и построить ее график.

43. Показать, что частные автокорреляционные функции следующих слабо стационарных случайных процессов совпадают:

а) \( x_t = \mu + e_t + \theta_1v_{t-1} \) и \( z_t = v_t + \theta_1v_{t-1} \),  
б) \( x_t = \mu + e_t + \theta_1e_{t-1} + \theta_2e_{t-2} + \cdots + \theta_qe_{t-q} \) и \( z_t = v_t + \theta_1v_{t-1} + \theta_2v_{t-2} + \cdots + \theta_qv_{t-q} \),  
где \( e_t \) и \( v_t \) – процессы белого шума с дисперсиями \( \sigma^2_e \) и \( \sigma^2_v \), соответственно.

44. Имеется следующий обратимый процесс: \( x_t = e_t + \theta_1e_{t-1} + \theta_2e_{t-2} \), где \( e_t \) – белый шум с дисперсией \( \sigma^2_e \). Рассчитать коэффициенты автоковариации. Записать автокорреляционную функцию для этого процесса.

45. Построить график автокорреляционной функции процесса:

а) \( x_t = e_t + 0.5e_{t-1} - 0.3e_{t-2} \); б) \( x_t = 1 + e_t - 0.4e_{t-1} + 0.4e_{t-2} \).

46. Переписать случайный процесс \( x_t = 0.5x_{t-1} + 0.5x_{t-2} + e_t - e_{t-1} + 3e_{t-2} \) с использованием лагового оператора, где \( e_t \) – белый шум. Проверить процесс на стационарность и обратимость.

47. Найти математическое ожидание, дисперсию и ковариацию случайного процесса \( x_t = 0.5x_{t-1} + e_t - 0.7e_{t-1} \), если \( e_t \) – белый шум. Построить график автокорреляционной функции.

48. Найти параметры модели ARMA(1, 1), если \( \rho(1) = 31/41, \rho(2) = 93/205 \).

49. Проверить на стационарность и обратимость процесс \( x_t = 0.6 + 0.3x_{t-1} + 0.4x_{t-2} + e_t - 0.7e_{t-1} \), где \( e_t \) – белый шум с дисперсией \( \sigma^2_e \). Представить процесс в виде AR(∞), если это возможно.

50. Определить порядок интегрирования процесса \( x_t = 1.5x_{t-1} + 0.5x_{t-2} + e_t - 0.5e_{t-1} \). Ответ обосновать.

51. Для модели \( (1 - L)(1 + 0.4L)x_t = (1 - 0.5L)e_t \) определить параметры \( p, d, q \). Является ли процесс стационарным?

52. Построить точечный прогноз на один шаг вперед, если известно, что процесс \( x_t = 0.1x_{t-1} + +e_t + 0.2e_{t-1} \), \( x_n = 10, e_n = 0.1 \).

53. Построить доверительный интервал для прогноза на два шага вперед для случайного процесса \( x_t = 0.5x_{t-1} + e_t \), если известно, что \( x_n = -1.6 \) и \( e_t \) – белый шум с единичной дисперсией.
54. Процесс с геометрическим лагом задан формулой

\[ x_t = 1 + \frac{1}{2} z_t + \frac{1}{4} z_{t-1} + \frac{1}{8} z_{t-2} + \cdots + e_t. \]

Примените к нему преобразование Койка.

55. Примените обратное преобразование Койка к модели ADL(1, 0).

56. Для процесса \( x_t = 0.1 + 0.5 x_{t-1} + 0.1 z_t + 0.2 z_{t-1} + e_t \) запишите долгосрочную зависимость между \( x \) и \( z \).
Предметный указатель

ACF, 11
AR(p), 51
LM-тест, 38
PACF, 12
Q-статистика Бокса-Льюнга, 37
Q-статистика Бокса-Пирса, 37
автокорреляционная функция, 11
автокорреляционная функция, 11
авторегрессия p-го порядка AR(p), 51
аддитивная модель, 9
белый шум, 11
gaуссовский белый шум, 11
Джордж Снедекор, 20
dисперсия на одну степень, 20
dолгосрочный мультипликатор, 43
интеграционная статистика Дарбина-Уотсона IDW, 69
интегрированные ряды порядка d, 77
интегрированный процесс d-го порядка, 68
информационный критерий Акаике, 66
информационный критерий SHварца – SIC (BIC), 66
cоррелограмма, 12
коефициент автокорреляции, 11
краткосрочный мультипликатор, 42
критерий
Бреуша-Годфри (Breusch-Godfrey), 38
Дарбина-Уотсона (DW), 38
Жарге-Бера (Jarque-Bera), 40
лаг, 11
лаговый оператор, 40
логистическая функция, 15
медианный лаг, 43
метод множителей Лагранжа (LM), 38
метод скользящих средних, 30
модель скользящего среднего MA(q), 58
мультипликативная модель, 9
относительные коэффициенты, 43
приближения Бартлетта, 40
промышленные мультипликаторы, 43
процесс DSP, 78
процесс TSP, 78
процесс с детерминированным полиномиальным трендом, 77
процесс случайного блуждания с дрейфом, 77
процессы с единичным корнем, 67
разложение Вольда, 50
разностный оператор, 41
распределение Стьюдента, 17
распределение Фишера-Снедекора, 18
распределение хи-квадрат, 17
сезонная компонента, 10
слабо стационарный временной ряд, 11
случайная компонента, 10
случайное блуждание, 54
средний лаг, 43
строго стационарный временной ряд, 11
tест Дикки-Фуллера, 80
тренд, 9, 14
условие обратимости, 60
условие стационарности AR(p), 56
характеристическое уравнение, 52
циклическая компонента, 10
частная автокорреляционная функция, 12
число степеней свободы, 19
эргодичность, 13
Библиографический список