

ЛЕКЦИОННЫЕ И МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ

Анализ временных рядов

Канторович Г.Г.

Вниманию читателей предлагается курс лекций, прочитанный студентам Государственного университета – Высшей школы экономики. Этот курс был записан студентами с помощью диктофона и затем расшифрован с использованием конспекта. Для данной публикации текст был отредактирован и значительно расширен, но в нем сохранена некоторая «живость», присущая разговорной речи, и не свойственная для письменной, академичной манеры изложения. Также сохранена разбивка на лекции, что может дать ориентир читателю-преподавателю. Поскольку материал курса планируется разместить в четырех выпусках журнала, традиционная нумерация формул для последующих ссылок не используется.

В практику отечественных высших учебных заведений курс «Анализ временных рядов» при подготовке экономистов вошел только в последнее время, и, зачастую, представляет собой «механическое» перенесение соответствующего курса для инженерных специальностей. Несмотря на обширность научной и учебной литературы по вышеуказанной тематике на иностранных языках, на русском языке отсутствует не только связное изложение, но и пригодное для использования в учебном процессе на магистерском уровне описание отдельных фрагментов предлагаемого курса, за исключением подхода Бокса–Дженкинса по построению моделей типа ARIMA. Встречающиеся в отечественной периодике статьи, анализирующие временные ряды для исследования переходной экономики России, изобилуют большим числом ошибок в применении этих методов.

В этот выпуск «Экономического журнала ВШЭ» вошла часть курса, посвященная стационарным временным рядам, наиболее полно представленная в литературе на русском языке. В последующих выпусках рассматриваются подход Бокса–Дженкинса, нестационарные временные ряды типа TS и DS, тестирование наличия единичного корня, коинтеграция временных рядов, модель коррекции отклонениями для стационарных и нестационарных регрессоров, модели векторной авторегрессии и их связь с коинтеграцией, модели с условной гетероскедастичностью.

Анализ взаимосвязей экономических данных, представленных в виде временных рядов, является необходимой составной частью современных исследований в области макроэкономической динамики, переходной экономики, эконометрики финансовых рынков. В конце 1980-х–начале 1990-х гг. было осознано, что только учет

Канторович Г.Г. – профессор, к. физ.-мат. н., проректор ГУ–ВШЭ, зав. кафедрой математической экономики и эконометрики.

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 экономистов-исследователей (аналитиков) на магистерском уровне.

Разработанные для целей экономических исследований методы анализа нестационарных случайных процессов существенно отличаются от нашедших широкое применение в технике и теории управления приемов работы со стационарными случайными процессами. В силу сложности исследуемых явлений эти методы предъявляют повышенные требования к математической, эконометрической подготовке студентов и к их знаниям современных подходов экономической теории.

Последнее обстоятельство представляется чрезвычайно важным. В технических приложениях анализ временных рядов используется преимущественно для прогнозирования и сопровождается значительной долей так называемого «data mining». В современной экономике значительно большее место занимает оценивание динамических моделей, отражающих краткосрочные и долгосрочные связи между экономическими переменными.

Лекция 1

Уже в начальном курсе эконометрики при изучении автокорреляции рассматриваются соотношения вида $x_t = \rho x_{t-1} + \varepsilon_t$. Значения некоторой экономической переменной зависят от ее же значений в предыдущий момент времени, от ее значений с лагом – сдвигом по времени на один шаг назад. Включение переменных с лагом в эконометрические соотношения означает существенное изменение в «философии» моделирования. Если в обычных эконометрических моделях значения одной переменной зависят от одновременных значений других переменных, то есть от текущего состояния экономической системы, то наличие переменных с лагом означает, что поведение системы определяется не только ее текущим состоянием, но и траекторией, по которой система пришла в это состояние. С математической точки зрения эконометрическая модель такого типа представляет собой не функцию объясняющих переменных, а функционал от траектории (траекторий) экономических переменных. Поэтому элементами таких моделей являются траектории: множества данных $\{x_t, t \in T\}$, где T – некоторое счетное или континуальное множество. Моделирование зависимостей вида $y_t = f(x_t, \varepsilon_t)$, где x и y есть данные типы траекторий, приводит к ситуации, когда обычные приемы регрессионного анализа не дают приемлемых оценок параметров. Курс «Анализ временных рядов» посвящен изучению методов построения и анализа таких зависимостей.

История использования динамических эконометрических моделей началась, пожалуй, с *Луи Башелье*, который в 1900 г. в своей диссертации [1] описал динамику поведения французских государственных облигаций, схожую с броуновским движением, известным из физики.

В 1984 г. Нельсоном и Кангом [5] было показано, что некоторые типы траекторий (в частности связанные со случайным блужданием) приводят к тому, что

регрессии, получаемые обычным методом наименьших квадратов, являются *кажущимися* (*spurious*). В 1987 г. Нельсон и Плоссер [6] показали, что почти все исторические макроэкономические ряды США относятся именно к этому типу. Другими словами, коэффициенты регрессии являются значимыми, стандартные тесты регрессионного анализа не диагностируют нарушений предпосылок классической модели, но, тем не менее, никакой зависимости между экономическими показателями нет. Позже мы обсудим, как они это показали и что следует из их результатов. Дальнейшие исследования установили, что исторические ряды макроэкономических данных США относятся именно к тому классу траекторий, который порождает кажущиеся регрессии. Эти факты заставили пересмотреть все до тех пор полученные эконометрические результаты в области анализа динамических моделей.

Случайные процессы и временные ряды

В курсе эконометрики ряд содержательных задач приводит к уравнениям, которые естественно назвать *авторегрессионными* (*autoregressive*). В частности, при устранении автокорреляции было получено уравнение $Y_t = \alpha + \beta \cdot Y_{t-1} + \gamma \cdot X_t + \varepsilon_t$. В нем в качестве объясняющей переменной появилась переменная Y с запаздыванием, т.е. регрессия переменной на саму себя – именно поэтому такое уравнение и назвали авторегрессионным.

Кроме того, в курсе эконометрики рассматривались *модели с распределенными лагами* (*distributed lag models*), в которых объясняющая переменная X также присутствует с запаздыванием: $Y_t = \alpha + \beta \cdot Y_{t-1} + \gamma_0 \cdot X_t + \gamma_1 \cdot X_{t-1} + \dots + \varepsilon_t$. В математических терминах можно записать: $Y_t = f(Y_{t-\tau}, X_{t-\tau}; t \in \{1, T\}, \tau \in [0, t])$. Важно, что теперь в модель входит не только по одному наблюдаемому значению переменных X и Y , а совокупность наблюдаемых значений (траектория). Причем по самому смыслу мы не можем считать, что последовательные значения каждой из переменных независимы между собой в статистическом смысле. Именно характер их зависимости заставляет нас рассматривать такую модель. Некоторые частные случаи такого рода моделей уже рассматривались в курсе эконометрики, а теперь наша задача – изучить общую базу их построения. Начнем с определения основных понятий.

Мы будем называть *временным рядом* (*time series*) совокупность наблюдений экономической величины в различные моменты времени. При этом наблюдение может характеризовать экономическую величину в данный момент времени, то есть быть типа запаса (например цена, ставка процента), или – характеризовать промежуток времени, то есть быть типа потока (например ВВП, продукция промышленности, поступления налогов). Как обычно в эконометрике, мы будем рассматривать временной ряд как выборку из последовательности случайных величин X_t , где t принимает целочисленные значения от 1 до T . Иногда мы будем рассматривать промежуток времени от 0 до T . Совокупность случайных величин $\{X_t, t \in [1, T]\}$ мы будем называть *дискретным случайным или стохастическим процессом*. Поскольку в нашем курсе все случайные процессы будут дискретными, в дальнейшем это слово будем опускать.

Иногда говорят, что стохастический процесс «для каждого случая» является некоторой функцией времени, что позволяет рассматривать процесс как случай-

ную функцию времени $X(t)$. При каждом фиксированном t значение стохастического процесса рассматривается просто как случайная величина. Эти два эквивалентных подхода позволяют рассматривать стохастический процесс как функцию двух разнородных величин, случая и момента времени: $X(\omega, t)$. При фиксированном случае у нас есть некоторая последовательность значений первой случайной величины, второй, третьей и так далее, которую мы будем называть *реализацией случайного процесса*. Говорят, что наблюдаемый временной ряд является реализацией стохастического процесса, или временной ряд порождается стохастическим процессом. Часто имеет смысл трактовать последовательность $\{X_t, t \in [1, T]\}$ как подпоследовательность бесконечной последовательности $\{X_t, t = 0, \pm 1, \pm 2, \dots\}$ и именно эту последовательность назвать стохастическим процессом, порождающим наблюдаемые данные. Как правило, в дальнейшем, если не оговорено обратное, в курсе под стохастическим процессом понимается бесконечная последовательность $\{X_t, t = 0, \pm 1, \pm 2, \dots\}$. Если t пробегает непрерывный отрезок времени, а иногда аналитически удобнее работать с непрерывным временем, то $X(\omega, t)$ называют *случайным процессом с непрерывным временем*. Принято обозначать значения реализации и стохастический процесс одними и теми буквами, например X_t . Обычно это не приводит к недоразумениям. В экономике мы чаще всего встречаемся, конечно, с дискретным временем. Примеры временных рядов: ВВП с 1921 по 1991 гг. в некоторой стране, ежемесячные значения инфляции, поквартальный объем производства, фондовый индекс или курс некоторой акции.

Поскольку случайный дискретный процесс $X(\omega, t)$ представляет собой совокупность случайных величин, то его наиболее полной статистической характеристикой является совместная функция распределения или функция плотности распределения (если плотность существует). Когда у нас было 2 случайных величины, мы говорили, что такой характеристикой является двумерная *функция распределения (или плотности распределения)*. При рассмотрении временного ряда число случайных величин велико и может быть бесконечным. Поэтому, строго говоря, чтобы задать все вероятностные свойства этой конструкции, нам нужна совокупность функций распределения, а именно одномерная функция распределения, двумерная функция распределения и так далее: $f_1(x_{t_1}); f_2(x_{t_1}, x_{t_2}); f_3(x_{t_1}, x_{t_2}, x_{t_3}); \dots$

Индексы у величин x_{t_1} и x_{t_2} означают, что одна случайная величина рассматривается в момент t_1 , вторая – в момент t_2 и так далее, и у них есть совместная функция распределения. Если мы возьмем другие моменты времени, то, вообще говоря, функция распределения будет другая. Такая совокупность функций распределения полностью характеризует случайный процесс. Эта совокупность функций согласована между собой в следующем смысле. Каждую функцию распределения размерности n можно получить из функции распределения размерности $n+1$. Для этого надо проинтегрировать функцию большей размерности по всем значениям одной из переменных.

Стоит задуматься, как мы дальше будем использовать эту математическую конструкцию? С точки зрения построения моделей наша эконометрическая ситуация осталась неизменной. А именно, как правило, имеется одна единственная реализация временного ряда, и ясно, что говорить об оценивании совокупности

всех функций распределения вообще никогда не приходится. И, во-вторых, пока на интуитивном уровне, если процесс ведет себя так, что его основные статистические характеристики со временем меняются, то мы по короткому кусочку наших наблюдений вообще ничего не сможем сказать о нем. Или нам что-то еще нужно знать дополнительно, сверх наблюдений. Поэтому, чтобы несколько снять остроту этой проблемы, мы будем говорить о более узком классе случайных процессов, которые мы назовем *стационарными случайными процессами*.

Объектом нашего рассмотрения будут не только стационарные процессы, хотя это один из самых важных классов. Под стационарностью мы будем понимать, что у случайного процесса некоторые свойства не меняются с течением времени. В соответствии с этим мы будем рассматривать 2 типа стационарности.

Строгая стационарность, или стационарность в узком смысле.

Мы будем называть случайный процесс *строго стационарным*, если сдвиг во времени не меняет ни одну из функций плотности распределения. Это значит, что если ко всем моментам времени прибавить некоторую (целочисленную) величину, то сама функция плотности не изменится, $f_n(x_1, \dots, x_n) = f_n(x_1+\Delta, \dots, x_n+\Delta)$ для всех n , моментов времени t_1, \dots, t_n и целочисленных Δ .

Можно сформулировать это определение не только в терминах плотности распределения, но и в терминах функций распределения, если плотности распределения не существуют. Иногда этот тип стационарности называют *сильным*.

Следствия.

Поскольку при каждом фиксированном t случайный процесс есть случайная величина, то для каждой из них можно рассматривать те характеристики, с которыми мы работали гораздо чаще, чем с функциями распределения, а именно математическое ожидание, дисперсию и так далее. Разумеется, они не обязаны существовать, поэтому в дальнейшем предполагаем сходимости соответствующих интегралов, не оговаривая это всякий раз отдельно.

По определению математическое ожидание непрерывной случайной величины X_t равно $E\{X_t\} = \int_{-\infty}^{+\infty} z \cdot f_1(z) dz = \mu$, если этот интеграл сходится. Если процесс стационарный, то какой бы момент времени t мы не взяли, подынтегральное выражение не меняется, поэтому математическое ожидание не зависит от времени.

1-ое следствие: если процесс строго стационарный, то его математическое ожидание не зависит от времени. Обратите внимание, что если свойства стационарности нет, то в различные моменты времени может быть разное по величине математическое ожидание.

Аналогичный результат справедлив для дисперсии стационарного процесса:

$$\text{Var}(X_t) = V(X_t) = \sigma^2.$$

2-ое следствие: дисперсия строго стационарного процесса в каждый момент времени одинакова.

Поскольку значения временного ряда в различные значения времени являются зависимыми между собой, то имеет смысл рассмотреть величину $\text{Cov}(X_{t_1}, X_{t_2})$. $\text{Cov}(X_{t_1}, X_{t_2}) = \iint (x_1 - \mu) \cdot (x_2 - \mu) \cdot f_2(x_1, x_2) \cdot dx_1 \cdot dx_2$. В этом интеграле мы можем сделать замену переменных и увидим, что, благодаря свойству стационарности, интеграл не изменится при сдвиге времени на τ вперед или на τ на-

зад. Следовательно, ковариация может рассматриваться как функция не двух переменных: t_1 и t_2 , а – единственной переменной: разности $(t_1 - t_2)$. Совокупность значений ковариаций при всевозможных значениях расстояния между моментами времени называется *автоковариационной функцией* случайного процесса.

3-е следствие: автоковариационная функция стационарного временного ряда зависит только от разности моментов времени $(t_1 - t_2)$.

Используя для автоковариационной функции обозначение γ , получим $\gamma(0) = \text{Cov}(X_{t_1}, X_{t_1}) = \sigma^2$. Мы можем рассматривать функцию $\gamma(\tau)$ как всевозможные значения автоковариаций, где τ пробегает целочисленные значения от $-\infty$ до ∞ . Поскольку очевидно, что эта функция четная, достаточно рассматривать только неотрицательные значения τ . Можно стандартным образом рассчитать коэффициент корреляции между разделенными на τ значениями временного ряда $\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)}$.

(Коэффициент корреляции – это ковариация, разделенная на корень из произведения двух дисперсий, но поскольку дисперсия постоянна, то мы получаем просто σ^2 или $\gamma(0)$.) Это выражение определяет *автокорреляционную функцию* временного ряда. Она показывает, насколько статистически зависимы значения временного ряда при различных сдвигах времени, то есть с разницей в год (для годовых данных) или в два года и так далее.

Итак, результаты, которые мы получили, заключаются в следующем.

Если случайный процесс стационарен в узком смысле, то у него:

- математическое ожидание не зависит от времени;
- дисперсия не зависит от времени;
- автоковариационная и автокорреляционные функции: а) зависят только от сдвига, от разности моментов времени; б) являются четными функциями.

Используем эти свойства для второго определения стационарности.

Слабая стационарность, или стационарность в широком смысле.

Если случайный процесс таков, что у него математическое ожидание и дисперсия существуют и не зависят от времени, а автокорреляционная (автоковариационная) функция зависит только от разности значений $(t_1 - t_2)$, то такой процесс мы назовем *стационарным в широком смысле*, или *слабо стационарным*. Существует еще одно название процесса с такими свойствами – стационарный в ковариациях случайный процесс. Всякий строго стационарный процесс является слабо стационарным, обратное, вообще говоря, не верно, например третий центральный момент может зависеть от времени. Не зря эти термины подобраны так, а не наоборот. Для очень важного класса нормальных процессов два эти определения стационарности эквивалентны. Мы назовем случайный процесс нормальным или гауссовым, если все его многомерные функции распределения нормальны. Поскольку многомерные функции распределения нормального вектора любого порядка полностью определяются вектором математических ожиданий и ковариационной матрицей, то для гауссова случайного процесса из слабой стационарности следует сильная стационарность.

И последнее свойство автокорреляционной или автоковариационной функции. Значения автокорреляционной функции не могут быть произвольными. Если

мы рассмотрим квадратичную форму вида: $\sum_{i,j=1}^N z_i \gamma(i-j) z_j$ или $\sum_{i,j=1}^N z_i \rho(i-j) z_j$, то

она будет неотрицательно определенной для любого натурального числа N . Доказательство непосредственно следует из того факта, что квадратичная форма

$$\sum_{i,j=1}^N z_i \gamma(i-j) z_j$$

равна дисперсии линейной комбинации $\sum_{i=1}^N z_i X_i$. Когда мы будем строить

оценки автокорреляционной функции, то нам надо будет построить их так, чтобы это свойство выполнялось. График функции $\rho(\tau)$ носит название *коррелограммы*. Коэффициент корреляции по модулю меньше единицы, поэтому $|\rho(\tau)| \leq 1$. Иногда этот график представляют в виде, как если бы функция $\rho(\tau)$ была определена для всех, а

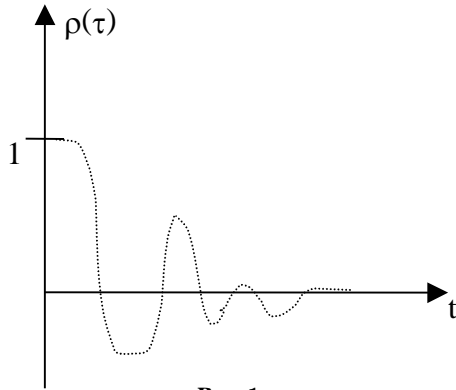


Рис. 1.

не только целочисленных τ (см. рис. 1).

После того, как мы познакомились с основными понятиями и характеристиками случайных процессов, приведем несколько простейших примеров, с которыми часто будем иметь дело.

1. *Процесс ε_t , удовлетворяющий условиям теоремы Гаусса-Маркова* Это означает, что $E(\varepsilon_t) = 0$; $Var(\varepsilon_t) = \sigma^2$; $Cov(\varepsilon_i, \varepsilon_j) = 0$ при $i \neq j$. Этот процесс заведомо слабо стационарен или стационарен в широком смысле. Мы будем называть этот процесс *белым шумом (white noise)*. Если добавить еще, что случайные величины ε_i распределены в совокупности нормально, то процесс стационарен также и в узком смысле. Мы будем называть его гауссовым *белым шумом* и писать $\varepsilon_e \sim WN(0, \sigma^2)$. Белый шум играет очень важную роль при анализе временных рядов и порождает более сложные процессы. Иногда то, что мы определили, могут назвать слабым белым шумом, и назвать сильным белым шумом, если заменить требование некоррелированности требованием независимости.

2. *Процесс случайного блуждания (Random walk)*. Иногда его называют броуновским движением. Это процесс, который задается следующим образом: $X_t = X_{t-1} + \varepsilon_t$, где ε_t – белый шум. Этот процесс можно рассматривать как авторегрессию с коэффициентом 1. Сам термин «случайное блуждание» появился в начале века в связи с шуточной по формулировке задачей: если в чистое поле выпустить пьяного, то где его следует искать через некоторое время? Результат – если пьяный блуждает случайно, то его надо искать на том же месте, то есть в среднем пьяный останется на том же месте.

Вычислим математическое ожидание, дисперсию и автокорреляционную функцию случайного блуждания при условии, что есть некоторая начальная точка X_0 . $X_t = X_{t-1} + \varepsilon_t$ – это обычное разностное уравнение. Легко записать его об-

щее решение: $X_t = X_{t-2} + \varepsilon_t + \varepsilon_{t-1} = X_{t-3} + \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \dots = X_0 + \sum_{\tau=0}^t \varepsilon_{t-\tau}$. Мы выпи-

ли решение в явном виде не через предыдущие значения, а только через начальное значение и правые части, в которых стоит белый шум.

$$E\{X_t\} = E(X_0) + E\left(\sum_{\tau=0}^t \varepsilon_{t-\tau}\right) = X_0 + \sum E(\varepsilon_{t-\tau}) = X_0 + 0 = const, \text{ то есть математическое}$$

ожидание удовлетворяет условию стационарности. Подсчитаем дисперсию.

$$V(X_t) = E\left\{\left(\sum_{\tau=0}^t \varepsilon_{t-\tau}\right)^2\right\}. \text{ Если мы раскроем скобки, то удвоенные произведения после}$$

взятия математического ожидания будут равны 0, и останется только математическое ожидание суммы квадратов. Учитывая свойства дисперсии белого шума, получим $V(X_t) = t \cdot \sigma_\varepsilon^2$. Сомножитель t показывает, что дисперсия случайного блуждания меняется со временем, более того, она растет пропорционально времени. Процесс случайного блуждания является нестационарным, даже в широком смысле, так как дисперсия не постоянна, она меняется во времени.

Обратим внимание на следующую интересную особенность. Мы могли это же уравнение переписать в виде $X_t - X_{t-1} = \varepsilon_t$, или $\Delta X_t = \varepsilon_t$, введя стандартное обозначение для приращения величины X_t . Приращение, или первую разность (*first difference*) ΔX_t , можно рассматривать как другой временной ряд, обозначим его через z_t , тогда $z_t = \varepsilon_t$. Таким образом, если мы возьмем первую разность нестационарного ряда, то в результате можем получить новый ряд, который будет стационарным. Здесь это очень просто, но на самом деле это один из самых распространенных способов приведения временного ряда к стационарному.

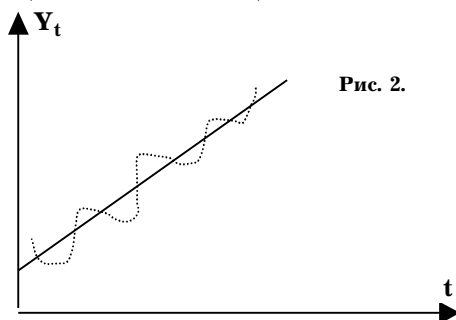
Интуитивно ясно, что со стационарным рядом работать гораздо проще, чем с нестационарным. Но далеко не всегда экономические показатели ведут себя стационарным образом. Из макроэкономики, да и из собственной жизни нам знакомы тренд в экономических данных, сезонное и циклическое поведение экономических показателей. Не слишком ли узок класс стационарных рядов?

Выделим из исходного временного ряда две составляющие: $Y_t = f(t) + X_t$, где X_t – случайная, недетерминированная часть, $f(t)$ – детерминированная часть. Под «детерминированной» составляющей мы будем понимать составляющую, которую можно точно (с нулевой дисперсией) предсказать, используя линейную комбинацию всех (или части) предыдущих значений ряда.

Давайте рассмотрим несколько примеров, чтобы пояснить, что имеется в виду. Предположим, наш случайный ряд имеет следующий вид: $Y_t = \alpha + \beta \cdot t + \varepsilon_t$. Мы уже рассматривали такие модели, это регрессия на время t . Здесь $\alpha + \beta \cdot t$ – детерминированная часть.

Если все это изобразить на графике, то получим тренд, вокруг которого совершается некоторое случайное движение.

Если ε_t является стационарным процессом с нулевым математическим ожида-



нием, то $E(Y_t) = \alpha + \beta \cdot t$. Здесь математическое ожидание зависит от t , но составляющая ε является стационарной.

В 1938 г. Wold [7] доказал следующий фундаментальный результат.

Декомпозиция или теорема Вольда (Wold decomposition).

Мы только сформулируем этот результат и не будем его доказывать. Вольд доказал, что чисто недетерминированный стационарный в широком смысле случайный процесс может быть представлен в следующем виде: $X_t - \mu = \sum_{\tau=0}^{\infty} \psi_{\tau} \cdot \varepsilon_{t-\tau}$,

где μ_t – математическое ожидание этого процесса, а ε_j – белый шум с конечными математическим ожиданием и дисперсией. То есть всякий слабо стационарный процесс представляется в виде линейной комбинации белых шумов, с разными весовыми коэффициентами. Так как это выражение линейное, его часто называют линейным фильтром. Как бы белый шум пропустили через линейный фильтр. И это очень удобно, это значит, что без потери общности мы можем ограничиться удобным линейным представлением и все про стационарные процессы изучить.

Разумеется, для того, чтобы это выражение имело смысл, надо потребовать выполнения условия сходимости, причем сходимости по вероятности, потому что суммируются случайные величины. Это условие очень простое, мы его выпишем и больше к нему возвращаться не будем: $\sum_{i=0}^{\infty} |\psi_i| < \infty$. Поскольку реализации белого шума не наблюдаемы, весовые коэффициенты определены с точностью до множителя. Поэтому договорились без потери общности считать $\psi_0 = 1$. Чем больше весовой коэффициент ψ_{τ} , тем больше влияние случайного возмущения в момент $t-\tau$ на текущий момент t . Разумеется, бесконечное число слагаемых порождает технические проблемы. К счастью, оказалось, что во многих случаях достаточно рассматривать не общее представление Вольда, а его частные случаи, когда число слагаемых конечно.

Процессы скользящего среднего (МА)¹⁾

Стохастический процесс называется процессом скользящего среднего порядка q , если в разложении Вольда присутствуют только q слагаемых. То есть: $MA(q): x_t = \sum_{\tau=0}^q \psi_{\tau} \cdot \varepsilon_{t-\tau}$ (для упрощения записи мы здесь обозначаем через $x_t = X_t - \mu$ отклонение процесса от его математического ожидания). Эквивалентно этот ряд можно представить в виде: $x_t = \varepsilon_t + \psi_1 \cdot \varepsilon_{t-1} + \dots + \psi_q \cdot \varepsilon_{t-q}$. Название «скользящее среднее» объясняется тем, что текущее значение случайного процесса определяется взвешенным средним q предыдущих значений белого шума. Процедуру скользящего среднего часто используют для того, чтобы сгладить данные, которые сильно колеблются.

¹⁾ МА = moving average.

Давайте посмотрим свойства этого ряда.

Процесс стационарен просто потому, что он есть частный случай разложения Вольда. Математическое ожидание: $E(x_t) = 0$. Дисперсия: $V(x_t) = \sigma^2 \sum_{i=1}^q \psi_i^2$.

Итак, мы видим, что ни математическое ожидание, ни дисперсия не зависят от времени. Перемножая значения для различных моментов времени и беря математическое ожидание, получим $Cov(x_t, x_{t+\tau}) = \sigma^2 \sum_{i=0}^{q-k} \psi_i \psi_{i+\tau}$ для $\tau = 0, 1, 2, \dots, q$. Для $\tau > q$ $Cov(x_t, x_{t+\tau}) = 0$.

Лекция 2

В прошлой лекции мы остановились на рассмотрении процессов скользящего среднего, которые мы определили как процессы, для которых сумма ряда в разложении Вольда простирается не до бесконечности, а до некоторого конечного целого числа q . Для обозначения коэффициентов конечного ряда MA(q) мы будем

использовать иную букву β : $X_t = \sum_{\tau=0}^q \beta_\tau \cdot \varepsilon_{t-\tau}$. Это означает, что ψ_i до $i=q$ включительно равны β_i , а остальные ψ_i равны нулю. Тогда выражение для ковариационной функции этого процесса принимает вид: $Cov(X_t, X_{t+\tau}) = \gamma(\tau) = \sigma_\varepsilon^2 \sum_{i=0}^{q-\tau} \beta_i \cdot \beta_{i+\tau}$,

где $\tau \leq q$, а $\sigma_\varepsilon^2 = Var(\varepsilon_t)$. Другими словами, между значениями временного ряда, достаточно далеко отстоящими друг от друга, отсутствует корреляционная связь. Из полученных выражений для ковариаций и дисперсии непосредственно следует выражение для автокорреляционной функции: $\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)}$.

Для изучения свойств временных рядов удобно использовать оператор сдвига L .

Определим $LX_t = X_{t-1}$, то есть действие оператора сдвига на временной ряд дает значение временного ряда в предыдущий момент времени. Естественным образом последовательное применение оператора сдвига p раз дает значение временного ряда в момент времени на p периодов ранее, что позволяет ввести степень оператора сдвига: $L^p X_t = L(L(L\dots))X_t = X_{t-p}$. Иногда удобно использовать нулевую степень оператора лага: $L^0 X_t = X_t$, играющую роль единичного оператора. В дальнейшем мы не будем специально оговаривать разницу между умножением на единицу, единичным оператором и нулевой степенью оператора сдвига, используя по обстоятельствам тот из них, который удобнее. Также естественным образом вводится умножение оператора сдвига на число и операции сложения и умножения степеней этого оператора.

С помощью этого оператора можно записать процесс скользящего среднего следующим образом:

$$X_t = (1L^0 + \beta_1 L + \beta_2 L^2 + \dots + \beta_q L^q) \cdot \varepsilon_t = (1 + \beta_1 L + \beta_2 L^2 + \dots + \beta_q L^q) \cdot \varepsilon_t = \beta_q(L) \varepsilon_t.$$

Напомним, что коэффициент при $L^0 \varepsilon_t$ или ε_t , то есть когда $\tau = 0$, всегда равен 1 из условия нормировки. Здесь через $\beta_q(L)$ обозначен операторный полином.

Заметим, что если у нас есть два операторных полинома $\psi(L)$ и $\varphi(L)$, то естественным образом для них можно определить арифметические операции сложения, вычитания и умножения на число. Определим суперпозицию этих операторных полиномов, то есть результат последовательного воздействия на X_t оператора φ , а потом на полученный результат – еще и оператора ψ . Непосредственно видно, что применение суперпозиции этих операторов к процессу X_t приводит к тому же результату, что и применение к процессу X_t произведения полиномов $\psi(L)$ и $\varphi(L)$: $\psi(L)[\varphi(L)X_t] = [\psi(L)\varphi(L)]X_t$. Это, в частности, означает, что мы можем разложить операторный многочлен на множители, используя корни уравнения $\varphi(L) = 0$. Для этого вспомним, что любой полином степени p с действительными коэффициентами имеет p комплексных корней, среди которых могут быть равные по величине. Этот факт называется *основной теоремой алгебры*, которую связывают с именем Гаусса. То есть любой операторный многочлен можно представить в виде: $(1 + \beta_1 L + \dots + \beta_q L^q) = \beta_q \prod_{i=1}^q (L - Z_i)$, где Z_i – корни уравнения

$$1 + \beta_1 Z + \dots + \beta_q Z^q = 0.$$

Для операторного многочлена $\varphi(L)$, действующего на некоторый процесс X_t , определим *обратный оператор* таким образом, чтобы суперпозиция оператора и обратного к нему была эквивалентна единичному оператору, то есть умножению на 1. Другими словами, если $\varphi(L)X_t = Y_t$, то действие на Y_t обратного оператора должно дать X_t : $[\varphi(L)]^{-1}Y_t = X_t = \underbrace{[\varphi(L)]^{-1}\varphi(L)}_{=1}X_t$. Это как с двумя матрицами: произведение прямой матрицы на обратную дает единичную матрицу.

Рассмотрим полином 1-ого порядка. При его умножении на обратный оператор должен получиться единичный оператор, то есть единица. Заметим, что $(1 - \alpha L) \cdot (1 + \alpha L + \alpha^2 L^2 + \dots + \alpha^p L^p + \dots) = 1$, если бесконечный ряд в скобках существует, имеет смысл. Тогда можно написать: $[1 - \alpha L]^{-1} = \frac{1}{1 - \alpha L} = 1 + \alpha L + \dots + \alpha^p L^p + \dots$, если правая часть имеет смысл. Рассмотрим конечную сумму вместо бесконечного ряда: $(1 - \alpha L) \cdot (1 + \alpha L + \alpha^2 L^2 + \dots + \alpha^p L^p) = 1 - \alpha^{p+1} L^{p+1}$. Действие этого оператора на случайный процесс дает: $(1 - \alpha^{p+1} L^{p+1})X_t = X_t - \alpha^{p+1} X_{t-p-1}$. При $p \rightarrow \infty$ правая часть выражения стремится к X_t , если наблюдаемые значения случайного про-

цесса ограничены и $|\alpha| < 1$. Следовательно, условие существования обратного оператора для полинома первого порядка имеет вид²⁾ $|\alpha| < 1$.

Используя понятие обратного оператора, попробуем выяснить, когда можно выразить ε через X . То есть, при каких условиях $\varepsilon_t = \varphi(L)X_t$. Разложив полином в

МА(q) представлении на множители, получим: $X_t = \beta_q \prod_{i=1}^q (L - Z_i) \varepsilon_t$. Для того, что-

бы использовать полученные условия обратимости линейного оператора, удобнее использовать корни другого *характеристического уравнения*, а именно: $\lambda^q + \beta_1 \lambda^{q-1} + \dots + \beta_q = 0$. Обозначим их через π_i . Такая запись характеристического уравнения совпадает с принятой в теории разностных линейных уравнений с постоянными коэффициентами.

В чем разница между предыдущей записью и этой? Обычно разностные уравнения содержат текущий и последующие члены последовательности. А во временных рядах используются: текущий и предыдущие члены. Поэтому есть два способа записи характеристического уравнения. Первый, $\lambda^q + \beta_1 \lambda^{q-1} + \dots + \beta_q = 0$, что соответствует общему правилу для разностных уравнений. Либо пишут наоборот: $1 + \beta_1 Z + \dots + \beta_q Z^q = 0$. Корни этих уравнений связаны между собой очень просто. Любой корень одного уравнения равен 1, деленной на соответствующий корень другого уравнения, и наоборот: $Z_i = 1/\pi_i$. Отсюда получим:

$$X_t = \beta_q \prod_{i=1}^q \left(L - \frac{1}{\pi_i}\right) \varepsilon_t = \prod_{i=1}^q (1 - \pi_i L) \varepsilon_t.$$

При выводе использован один из результатов теоремы Виета: произведение корней характеристического уравнения равно $\beta_q (-1)^q$.

В выражении $\prod_{i=1}^q (1 - \pi_i L) \varepsilon_t$ каждый сомножитель – это просто линейное выражение $(1 - \pi_i L)$. Если формально выразить ε_t через X , не обращая внимания

на то, что L – это оператор, получится $\varepsilon_t = \frac{1}{\prod_{i=1}^q (1 - \pi_i L)} X_t$. Оказывается, этому вы-

ражению можно придать смысл. Если все корни характеристического уравнения действительны и различны по величине, операторную дробь, знаменатель которой разложен на произведение одночленов первой степени относительно L , можно

представить в виде суммы элементарных дробей: $\frac{1}{\prod_{i=1}^q (1 - \pi_i L)} = \frac{A_1}{1 - \pi_1 L} + \frac{A_2}{1 - \pi_2 L} + \dots$

²⁾ Тот же результат можно получить, используя норму в пространстве операторов сдвига. При любом разумном определении нормы случайного процесса X_t , норма процесса X_t равна норме процесса X_{t-1} , поэтому оператор сдвига имеет норму, равную единице: $\|L\| = 1$.

Такое разложение применяется при интегрировании рациональных дробей и должно быть вам знакомо из курса математического анализа. Каждое слагаемое напоминает выражение для формулы бесконечно убывающей геометрической прогрессии. В данном случае A_i играет роль первого члена этой прогрессии, а вместо множителя бесконечно убывающей геометрической прогрессии у нас стоит $\pi_i L$. Если $|\pi_i| < 1$, то первую элементарную дробь можно разложить в сумму бесконечно убывающей геометрической прогрессии или в бесконечный степенной ряд

$$\frac{A_i}{1 - \pi_i L} = A_i (1 + \pi_i L + \pi_i^2 L^2 + \dots + \pi_i^k L^k + \dots).$$

Если аналогичное условие выполнено для каждого из корней, то получается, что ε_t равно сумме q таких бесконечных разложений. Произведя приведение подобных членов, получим некоторый бесконечный многочлен от L с какими-то – можно явно выписать их значения – коэффициентами. Таким образом, получим выражение следующего вида: $(\alpha_0 + \alpha_1 L + \alpha_2 L^2 + \dots + \alpha_k L^k + \dots) X_t$, где коэффициенты α_i сложным образом будут выражены через характеристические корни π . Если это все выполнено, то ε выражено через текущие и предыдущие значения X .

Проведенные преобразования справедливы, если каждая дробь типа $\frac{A_i}{1 - \pi_i L}$ может трактоваться как сумма бесконечно убывающей геометрической прогрессии, то есть мы возвращаемся к условию, что все корни по модулю строго меньше единицы: $|\pi_i| < 1$.

Это условие называется *условием обратимости (invertability)* процесса МА. При этом выражение для ε_t будет бесконечным, но сходящимся операторным полиномом, действующим на X_t . Это значит, что ε_t будет выражено через уходящие в бесконечно далекое прошлое значения X . Можно показать, что условие обратимости сохраняет свой вид, а именно: $|\pi_i| < 1$ и для случая комплексных и/или кратных корней.

Напомним, что существуют два различных способа записи характеристического уравнения. Эти способы эквивалентны, меняется лишь форма условия существования обратного оператора: в первой записи условие – все характеристические корни по модулю меньше 1, а если характеристическое уравнение записать во втором виде, то условие сходимости – все характеристические корни по модулю больше 1. Здесь и в дальнейшем будет использоваться первый способ записи характеристического уравнения.

После преобразования процесс скользящего среднего принимает следующий вид: $(\alpha_0 + \alpha_1 L + \alpha_2 L^2 + \dots) X_t = \varepsilon_t$. Если выразить X_t через все его предыдущие значения и ε_t и для нормировки разделить на α_0 , то получается эквивалентное представление $X_t = -\frac{\alpha_1}{\alpha_0} X_{t-1} - \dots - \frac{\alpha_k}{\alpha_0} X_{t-k} - \dots + \frac{1}{\alpha_0} \varepsilon_t$. Тот же самый процесс МА(q) пред-

ставлен таким образом, что текущее значение X_t выражено через текущее значение ε_t , а вместо q предыдущих значений ε появляется бесконечный ряд прошлых значений X_{t-k} . На первый взгляд этот процесс не напоминает представление Вольда, но в то же время он по построению является стационарным случайным процессом. Обобщая это представление, переходим к рассмотрению нового класса процессов, которые при некоторых условиях могут быть стационарными, а именно к процессам авторегрессии.

Если значение случайного процесса определяется линейной комбинацией конечного числа его предыдущих значений и добавлением белого шума, то такой процесс называется процессом *авторегрессии (autoregression) порядка p* , и его общее уравнение имеет вид: $X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \dots + \alpha_p X_{t-p} + \varepsilon_t$, где ε_t – белый шум.

В этих терминах процесс, обратный к $MA(q)$, может быть обозначен как $AR(\infty)$. Однако у нас нет гарантии, что при любых коэффициентах $\alpha_1, \alpha_2, \dots, \alpha_p$ этот процесс будет стационарным. Для того чтобы он был стационарным, нужно, чтобы он был представим в виде разложения Вольда, чтобы его можно было перевести в $MA(q)$ представление, имеющее смысл.

Перепишем уравнение процесса в следующем виде:

$$\underbrace{X_t - \alpha_1 X_{t-1} - \alpha_2 X_{t-2} - \dots - \alpha_p X_{t-p}}_{\alpha(L)X} = \varepsilon_t.$$

Обозначим через $\alpha_p(L) = 1 - \alpha_1 L - \alpha_2 L^2 - \dots - \alpha_p L^p$ операторный полином, действующий на X_t . В некотором смысле получено «зеркальное отображение» процесса $MA(q)$. Теперь операторный полином действует на X , а не на ε , и результат равен ε_t . Полученное выражение представляет собой разностное уравнение относительно X со случайной правой частью ε_t . Общее решение этого уравнения состоит из общего решения однородного уравнения (когда нет ε_t) плюс частное решение полного (неоднородного) уравнения, зависящее от ε_t . Общее решение однородного уравнения имеет следующий вид: $X_t = \sum C_i(t)(\pi_i)^t$, где π_i – различные по величине корни характеристического уравнения (может быть, комплексные), которое имеет вид: $\lambda^p - \alpha_1 \lambda^{p-1} - \dots - \alpha_p = 0$, а $C_i(t)$ – полиномы, степень которых на единицу меньше кратности соответствующего корня. Частное решение неоднородного уравнения выписывается через ε_t , и мы обсудим его позже.

Решение однородного уравнения не будет уходить в бесконечность, будет устойчивым при условии, что корни характеристического уравнения по модулю будут меньше 1: $|\pi_i| < 1$. Но именно при этом условии существует оператор, обратный оператору $\alpha(L)$, то есть имеет смысл выражение: $X_t = \frac{1}{\alpha(L)} \varepsilon_t$. Следова-

тельно, процесс X_t принимает вид, соответствующий теореме Вольда: $\sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}$, из чего следует, что этот ряд является стационарным. Если же при некотором i

$|\pi_i| \geq 1$, то решение не стремится к нулю или даже уходит в бесконечность, и ни о какой стационарности речи быть не может.

В результате условие, что все корни характеристического уравнения для процесса AR(p) лежат внутри единичного круга, является необходимым и достаточным для того, чтобы ряд был стационарным.

Тут есть некая асимметрия. Процессы AR(p) и MA(q) в чем-то похожи. Но процесс MA(q) всегда стационарен, а условия обратимости обеспечивают его некоторое полезное свойство, не затрагивая фундаментального свойства стационарности. Для AR(p) это условие очень жесткое: либо он стационарен и сводится к MA(∞), либо он вообще не стационарен, и тогда он выпадает (пока) из нашего рассмотрения.

В литературе существует немного разное понимание обозначения AR(p). Это различие обычно ничему не препятствует, но я хочу обратить на него ваше внимание. Иногда называют процессом авторегрессии только такой процесс AR(p), для которого выполнено условие стационарности. А иногда AR(p) называют любой процесс авторегрессии, даже если он не стационарен.

После того, как мы выяснили условие стационарности, рассмотрим свойства этого процесса. Начнем с математического ожидания. Если процесс стационарный и его можно выписать в виде сходящегося разложения Вольда, то вопрос о математическом ожидании решается просто, оно нулевое: $E(X_t) = 0$. Автоковариационную функцию можно вывести, используя представление процесса в виде MA(∞) с помощью операторных полиномов, но это довольно кропотливая и не очень интересная работа. Можно получить нужный результат другим образом.

Умножим обе части выражения $X_t = \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} + \varepsilon_t$ на $X_{t-\tau}$ и возьмем математическое ожидание. Поскольку математическое ожидание процесса равно 0, получим уравнение для значений автоковариационной функции. Если $\tau > p$ то получаем следующее соотношение

$$\gamma(\tau) = \alpha_1 \gamma(\tau-1) + \alpha_2 \gamma(\tau-2) + \dots + \alpha_p \gamma(\tau-p) + E\{X_{t-\tau} \cdot \varepsilon_t\}.$$

В левой части равенства мы использовали четность автоковариационной функции: $\gamma(\tau) = \gamma(-\tau)$. Так как процесс стационарный, то $X_{t-\tau}$ представим в виде бесконечной суммы $\varepsilon_{t-\tau}$ и более ранних значений ε . Все ε , входящие в определение $X_{t-\tau}$, не совпадают по времени ε_t , они все одновременные. Поэтому $E\{X_{t-\tau} \cdot \varepsilon_t\} = 0$, потому что значения белого шума не коррелируют между собой. Другими словами, получается следующее выражение: $\gamma(\tau) = \alpha_1 \gamma(\tau-1) + \dots + \alpha_p \gamma(\tau-p)$.

Если рассматривать его как разностное уравнение, то оно совпадает с однородным уравнением для процесса X_t . То есть значение автоковариационной функции для $\tau > p$ удовлетворяет точно тому же разностному уравнению. Поскольку разностное уравнение устойчиво, то значения автоковариационной функции убывают по абсолютной величине, по крайней мере, со значения с индексом $(p+1)$.

Для $\tau < p$ эти уравнения надо модифицировать. Вместо рекуррентного соотношения для первых p значений автоковариационной функции получается система p линейных уравнений с p неизвестными.

В качестве примера рассмотрим процесс AR(1). Его характеристическое уравнение имеет вид: $\lambda - \alpha = 0$. Условие стационарности: $|\alpha| < 1$. Для автоковариационной функции получаем:

$$\begin{aligned}\gamma(1) &= \alpha\gamma(0) \\ \gamma(2) &= \alpha\gamma(1) = \alpha^2\gamma(0) \\ &\dots\dots\dots \\ \gamma(\tau) &= \alpha^\tau\gamma(0)\end{aligned}$$

Теперь очень хорошо видно, почему важно условие $|\alpha| < 1$.

Чтобы вычислить $\gamma(0)$, надо найти дисперсию. Но если мы интересуемся только автокорреляционной функцией, то, разделив наши соотношения на $\gamma(0)$, получим $\rho(\tau) = \alpha^\tau$.

Следовательно, для процесса AR(1) автокорреляционная функция $\rho(\tau)$ ведет себя схематически следующим образом:

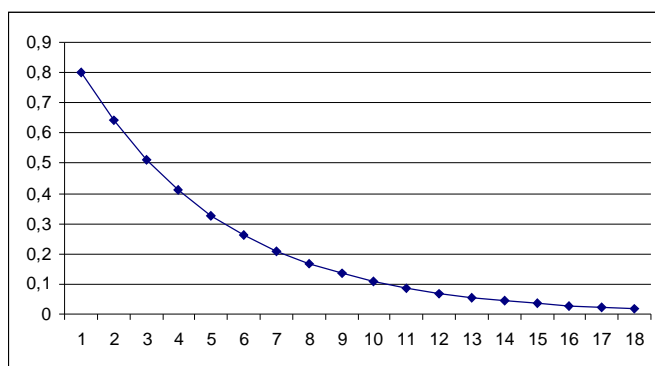


Рис. 1. Значения автокорреляционной функции процесса AR(1) при значении коэффициента равном 0,8

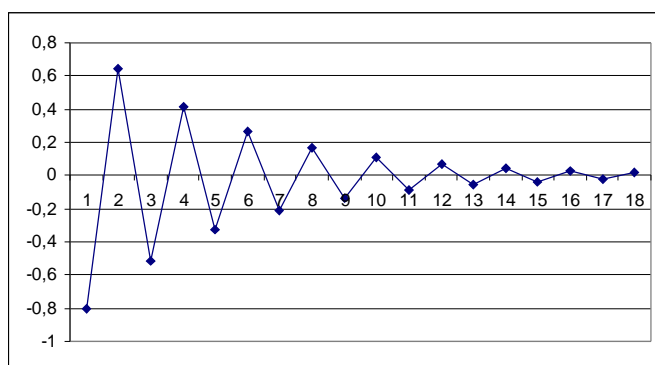


Рис. 2. Значения автокорреляционной функции процесса AR(1) при значении коэффициента равном -0,8

Лекция 3

В прошлой лекции мы установили, что если процесс AR стационарен, то он кроме конечного представления AR(p) имеет бесконечное представление MA(∞). И если выполнено условие обратимости, то конечный процесс MA(q) имеет бесконечное представление AR(∞). Такой дуализм представлений временного ряда можно рассматривать как обобщение преобразования Койка, которое применяется в эконометрике для борьбы с автокорреляцией первого порядка. Существенное отличие процесса AR от процесса MA состоит в том, что для процесса MA автокорреляционная функция после $i > q$ становилась равной нулю, обрывалась.

Рассмотрим процесс AR(2). Это процесс вида $X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \varepsilon_t$. Посмотрим условия стационарности этого процесса. Характеристическое уравнение принимает вид $\lambda^2 - \alpha_1 \lambda - \alpha_2 = 0$ (как уже отмечалось, иногда характеристическое уравнение пишут в виде $1 - \alpha_1 Z - \alpha_2 Z^2 = 0$).

Корни этого уравнения: $\lambda_{1,2} = \frac{\alpha_1 \pm \sqrt{\alpha_1^2 + 4\alpha_2}}{2}$.

Условие стационарности при выбранной форме записи характеристического уравнения: $|\lambda_{1,2}| < 1$. В принципе, этого условия достаточно, но его можно выразить в терминах коэффициентов процесса AR(2) и свести к 3-м следующим неравенствам: $\alpha_1 + \alpha_2 < 1$; $\alpha_1 - \alpha_2 > 1$; $|\alpha_2| < 1$.

Предположим, что условия стационарности выполнены, и тогда имеют смысл математические ожидания от левой и правой части после умножения их на X в какой-то другой момент времени. Умножаем на X_{t-1} , и получаем: $E\{X_{t-1} \cdot X_t\}$. Поскольку процесс стационарен, его можно представить в виде взвешенной суммы текущего и предыдущих значений белого шума (декомпозиция Вольда). Следовательно, математическое ожидание процесса равно нулю, и $E\{X_{t-1} \cdot X_t\}$ равно автоковариации $\gamma(1)$. Давайте все запишем сразу в терминах γ . Получаем:

$$E\{X_{t-1} \cdot X_t\} = \gamma(1) = \alpha_1 \gamma(0) + \alpha_2 \gamma(-1).$$

Поскольку автоковариационная функция симметрична, то знак «минус» можно заменить на знак «плюс».

Перейдем к автокорреляционной функции: $\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)}$. Иногда ее обозначают асf. Получаем более удобное соотношение: $\rho(1) = \alpha_1 + \alpha_2 \rho(1)$. Для упрощения записи будем использовать индекс вместо аргумента у автоковариационной и автокорреляционной функций.

Отсюда следует, что $\rho(1) = \rho_1 = \frac{\alpha_1}{1 - \alpha_2}$. В результате

аналогичных действий получаем: $\rho_2 = \rho(2) = \alpha_1 \rho_1 + \alpha_2 = \frac{\alpha_1^2}{1 - \alpha_2} + \alpha_2$. При умножении

на X_{t-3} получаем: $\rho_3 = \alpha_1 \rho_2 + \alpha_2 \rho_1$. Если так делать дальше, то получим:

$\rho_n = \alpha_1 \rho_{n-1} + \alpha_2 \rho_{n-2}$. Это то, что мы отметили в прошлой лекции в общем виде, начиная с номера большего, чем p , значения автоковариационной функции подчиняются разностному уравнению. Совокупность таких разностных уравнений для значений автокорреляционной функции при различных n носит название уравнений Юла-Уокера (*Yule-Walker equations*). Решение уравнения в случае разных по величине корней имеет следующий вид: $\rho_n = C_1(\pi_1)^n + C_2(\pi_2)^n$, где π_1 и π_2 – характеристические корни. Для совпадающих корней решение имеет вид:

$$\rho_n = C_1(\pi_1)^n + C_2 t(\pi_1)^n.$$

Исходя из условий стационарности, корни по модулю меньше 1, решение разностного уравнения устойчиво, поэтому это выражение затухающее. Если π_1 и π_2 – действительные числа, то все записано в терминах действительных чисел. Если корни комплексные, то тогда они обязательно комплексно сопряженные, и помимо степени в каждом из слагаемых появляется сомножитель в виде косинуса или синуса некоего аргумента. Но характер асимптотического поведения решения от этого не меняется. Выражение в степени n – это модуль комплексного числа, и если он меньше 1, то все равно это будут затухающие экспоненты, ну, может быть, умноженные на синусы и косинусы. Чтобы определить значения неопределенных констант C_1 и C_2 , нам нужны начальные условия. Уже упоминавшиеся «нестандартные» уравнения для значений ρ_1 и ρ_2 служат для их определения. Мы выразили ρ_1 и ρ_2 через коэффициенты нашего авторегрессионного уравнения α_1, α_2 . Причем мы заметили, что они выражаются не так, как все остальные значения автокорреляционной функции, которые определяются просто общей рекуррентной схемой, разностным уравнением.

Во многих учебниках по анализу временных рядов приведены примеры реализаций процессов AR(2) и соответствующие коррелограммы. Однако при сегодняшней компьютерной технике гораздо более содержательно построить эти примеры самому. Постройте, например, в Excel'е с помощью датчика псевдослучайных чисел реализацию процесса AR(2) и – по вышеприведенным формулам – значения автокорреляционной функции. Проведите генерацию 10–20 реализаций, после чего измените значения параметров α_1 и α_2 и повторите генерацию реализаций. Основной вывод, который можно сделать по результатам такого моделирования, состоит в том, что по реализации случайного процесса визуально ничего нельзя сказать ни о его типе, ни о численных значениях коэффициентов. Если в курсе эконометрики, например по графику остатков регрессии, можно было сделать какой-то качественный вывод или предположение о наличии автокорреляции или гетероскедастичности, то при анализе временных рядов, как правило, мы этого сделать не можем.

Вернемся к общему стационарному авторегрессионному уравнению

$$X_t = \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} + \varepsilon_t.$$

Начиная с номера $p+1$ и дальше автокорреляционная функция – сумма затухающих экспонент, может быть умноженных на синусы, косинусы или полиномы от времени, но все равно это затухающие экспоненты. А для значения с индексами от 1 до p получаем следующую систему уравнений:

$$\begin{aligned} \gamma_1 &= \alpha_1 \gamma_0 + \alpha_2 \gamma_1 + \dots + \alpha_p \gamma_{p-1} \\ \gamma_2 &= \alpha_1 \gamma_1 + \alpha_2 \gamma_0 + \dots + \alpha_p \gamma_{p-2} \\ &\dots\dots\dots \\ \gamma_p &= \alpha_1 \gamma_{p-1} + \alpha_2 \gamma_{p-2} + \dots + \alpha_p \gamma_0. \end{aligned}$$

Разделив на дисперсию, получим систему линейных уравнений относительно первых p значений автокорреляционной функции. Следовательно, если ряд стационарен, то обязательно, начиная с некоторого момента, автокорреляционная функция начинает затухать, как сумма каких-то экспонент, может быть умноженных на синусы и косинусы. Если поведение автокорреляционной функции не такое, то она не может быть автокорреляционной функцией стационарного процесса.

У нас появилось важное качественное различие между стационарным и нестационарным процессом. Если процесс стационарный, будь он AR(p), будь он MA(q), начиная с некоторого момента автокорреляционная функция начинает затухать или совсем исчезает. Если же процесс не стационарный, то это не так.

Наши два примера позволяют прийти еще к следующему выводу. Ясно, что порядок авторегрессионной модели имеет большое значение. То есть для нас важно, является исследуемый процесс процессом типа AR(1) или типа AR(2), ведь нам надо еще оценивать коэффициенты. И здесь нас ждет очень неприятное открытие. Если в моделях скользящего среднего q – качественно понятный параметр, после которого автокорреляционная функция обрывается; то в случае модели AR(p) такого не происходит. Поэтому хотелось бы получить характеристику, качественно ведущую себя сходным образом, указывающую на точный порядок p авторегрессионного уравнения.

Такой характеристикой является *частная автокорреляционная функция* (*partial autocorrelation function*). Стандартная английская аббревиатура – *pacf*. В

определении автокорреляционной функции $\rho(\tau) = \frac{1}{\text{Var}(X_t)} E\{(X_t - \mu)(X_{t-\tau} - \mu)\}$ вхо-

дит ковариация между значениями процесса, отстоящими на τ шагов по времени друг от друга. Однако на поведение процесса AR(p) статистически влияет не только его значение в момент, отстоящий на τ единиц назад, но и все промежуточные значения процесса между моментами t и $t-\tau$. Поэтому можно поставить другой вопрос. Какова линейная статистическая зависимость между значениями процесса в эти моменты, если мы устраним или элиминируем влияние всех промежуточных значений? Какова «чистая» взаимосвязь между этими значениями? То есть, если мы уже учли влияние всех промежуточных значений, то какое дополнительное частное влияние оказывает значение момента $t-\tau$. Коэффициент корреляции при элиминировании промежуточных значений называется частным коэффициентом корреляции. В курсе эконометрики подобная конструкция встречалась при рассмотрении излишних и не включенных в модель переменных. Рассматривалось влияние последней добавленной переменной с учетом того, что все предыдущие уже включены.

Обозначим через φ_{kk} – k -е значение частной автокорреляционной функции. По определению φ_{kk} – это коэффициент корреляции между X_{t-k} и X_t (или между x_{t-k} и x_t) за вычетом той части x_t , которая линейно объяснена промежуточными

лагами $x_{t-1}, x_{t-2}, \dots, x_{t-k-1}$. Эта объясняющая линейная комбинация называется *оптимальным линейным предиктором*, если она реализует минимальную среднеквадратичную ошибку предсказания (прогноза) x_t . Более строго, нас интересует коэффициент корреляции

$$\text{Corr}(x_t - \varphi_{k1}x_{t-1} - \varphi_{k2}x_{t-2} - \dots - \varphi_{kk-1}x_{t-k-1}, x_{t-k}),$$

где $\varphi_{k1}, \varphi_{k2}, \dots, \varphi_{kk-1}$ – коэффициенты линейной комбинации, обеспечивающей минимальную среднеквадратичную ошибку предсказания

$$\min E\{(x_t - \varphi_{k1}x_{t-1} - \varphi_{k2}x_{t-2} - \dots - \varphi_{kk-1}x_{t-k-1})^2\}.$$

Это выражение аналогично целевой функции метода наименьших квадратов с заменой суммирования по наблюдениям на взятие математического ожидания. Искомый коэффициент корреляции между оптимальным линейным предиктором и x_{t-k} является просто коэффициентом теоретической регрессии предиктора на x_{t-k} . Но тогда он является коэффициентом при последнем регрессоре в следующей теоретической регрессии $x_t = \varphi_{k1}x_{t-1} + \varphi_{k2}x_{t-2} + \dots + \varphi_{kk}x_{t-k} + \varepsilon_t$. Регрессия строится на k предыдущих значений процесса, определяются k коэффициентов, но соответствующее значение частной автокорреляционной функции дает лишь последний коэффициент этой регрессии. Подчеркнем, что речь идет не о выборочной, а о теоретической регрессии, то есть об условном математическом ожидании. Кроме того, следует иметь в виду, что эта регрессия не связана с тем, определяется ли процесс авторегрессионным соотношением или нет. Очевидно, что если исследуемый процесс принадлежит к виду $AR(p)$, то $\varphi_{pp} = \alpha_p$. Более того, для такого процесса текущее значение оптимально (в смысле сформулированного критерия) выражается через ровно p предыдущих значений, и учет более ранних значений уже не может улучшить прогноз. Это означает, что для процесса $AR(p)$ $\varphi_{kk} = 0$ при $k > p$. Далее мы получим этот важнейший результат из уравнений Юла-Уокера.

Рассмотрим несколько простых примеров расчета значений частной автокорреляционной функции.

Процесс $AR(1)$. $X_t = \alpha X_{t-1} + \varepsilon_t$. Рассмотрим регрессию: $X_t = \varphi_{11}X_{t-1} + \varepsilon_t$. Умножаем уравнение на X_{t-1} , берем математическое ожидание от обеих частей и делим результат на $\gamma(0)$, получаем: $\rho_1 = \varphi_{11} = \alpha$, так как математическое ожидание процесса равно нулю.

Чтобы получить φ_{22} , строим теоретическую регрессию

$$X_t = \varphi_{21}X_{t-1} + \varphi_{22}X_{t-2} + \varepsilon_t.$$

Как и раньше, умножаем обе части на X_{t-1} , берем математическое ожидание, делим на $\gamma(0)$. Получаем: $\frac{E\{X_t \cdot X_{t-1}\}}{\gamma(0)} \Rightarrow \rho_1 = \varphi_{21} \cdot 1 + \varphi_{22} \cdot \rho_1$. В это соотношение входит

2 неизвестных: φ_{21} и φ_{22} . Нужно получить еще одно соотношение, чтобы определить их. Поэтому умножаем обе части исходного уравнения на X_{t-2} и производим

ту же самую операцию. Получится $\rho_2 = \varphi_{21}\rho_1 + \varphi_{22}$. Получаем систему двух линейных уравнений относительно φ_{21} и φ_{22} , а коэффициенты этой системы уравнений выражены через ρ_1 и ρ_2 , связь которых с коэффициентами исходного уравнения процесса нам уже известна. Другими словами, наш подход дает связь между коэффициентами автокорреляционной функции и частной автокорреляционной функции. В данном случае у нас всего два уравнения, можно решать их по-разному. Но, вообще говоря, нас интересует только один единственный коэффициент φ_{22} , поэтому наиболее экономно будет использовать правило Крамера.

$$\varphi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{1 - \rho_1^2}$$
. Кстати, обратите внимание, поскольку ρ_1 – это значение автокорреляционной функции, то оно меньше 1, то есть в знаменателе 0 не будет. Для процесса AR(1) $\rho_1 = \alpha$; $\rho_2 = \alpha^2$. Поэтому: $\varphi_{22} = \frac{\alpha^2 - \alpha^2}{1 - \alpha^2} = 0$, как мы и ожидали.

Для того чтобы посчитать φ_{33} , запишем теоретическую регрессию $X_t = \varphi_{31}X_{t-1} + \varphi_{32}X_{t-2} + \varphi_{33}X_{t-3} + \varepsilon_t$. Аналогично предыдущему получаем систему уравнений:

$$\begin{cases} \rho_1 = \varphi_{31} \cdot 1 + \varphi_{32}\rho_1 + \varphi_{33}\rho_2 \\ \rho_2 = \varphi_{31}\rho_1 + \varphi_{32} \cdot 1 + \varphi_{33}\rho_1 \\ \rho_3 = \varphi_{31}\rho_2 + \varphi_{32}\rho_1 + \varphi_{33} \cdot 1. \end{cases}$$

Вновь нас интересует только коэффициент φ_{33} . Воспользуемся правилом Крамера, но, вооруженные опытом предыдущего, считаем только определитель в числителе.

$$\Delta_3 = \begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}.$$

Теперь можно выразить определитель в числителе для произвольного k . Система линейных уравнений, связывающая значения автокорреляционной и частной автокорреляционной функций, фактически является системой *уравнений Юла-Уокера* для теоретической регрессии с неизвестными и подлежащими определению значениями частной автокорреляционной функции:

$$\begin{cases} \rho_1 = \varphi_{k1} \cdot 1 + \varphi_{k2}\rho_1 + \varphi_{k3}\rho_2 + \dots + \varphi_{kk}\rho_{k-1} \\ \rho_2 = \varphi_{k1}\rho_1 + \varphi_{k2} \cdot 1 + \varphi_{k3}\rho_1 + \dots + \varphi_{kk}\rho_{k-2} \\ \dots \\ \rho_k = \varphi_{k1}\rho_{k-1} + \varphi_{k2}\rho_{k-2} + \varphi_{k3} \cdot \rho_{k-3} + \dots + \varphi_{kk} \cdot 1. \end{cases}$$

Искомый определитель имеет вид:

$$\Delta_k = \begin{vmatrix} 1 & \rho_1 & \dots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \dots & \dots & \rho_2 \\ \vdots & \dots & \dots & \dots & \vdots \\ \rho_{k-1} & \dots & \dots & \dots & \rho_k \end{vmatrix}.$$

На главной диагонали все элементы кроме последнего будут равны 1. Связь между коэффициентами φ_{kk} и ρ_1, \dots, ρ_k , то есть между автокорреляционной и частной автокорреляционной функциями, выражается следующим соотношением:

$\varphi_{kk} = \frac{\Delta_k}{\Delta}$, где Δ – определитель системы уравнений

$$\Delta = \begin{vmatrix} 1 & \rho_1 & \dots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \dots & \dots & \rho_{k-2} \\ \vdots & \dots & \dots & \dots & \vdots \\ \rho_{k-1} & \dots & \dots & \dots & 1 \end{vmatrix}.$$

Зная коэффициенты автокорреляционной функции, можно по уравнениям Юла-Уокера пересчитать коэффициенты частной автокорреляционной функции, и наоборот.

Посмотрим, чему будет равно Δ_3 . Для случая AR(1) получаем следующее выражение:

$$\begin{vmatrix} 1 & \alpha & \alpha \\ \alpha & 1 & \alpha^2 \\ \alpha^2 & \alpha & \alpha^3 \end{vmatrix} = 0.$$

Определитель, как и ожидалось, равен 0, потому что третий столбец пропорционален первому, причем коэффициент пропорциональности равен коэффициенту уравнения процесса. Если считать четвертое, пятое и так далее значения частной автокорреляционной функции, то все они равны 0. Поэтому для процесса AR(1) значения частной автокорреляционной функции, начиная со второго, равны 0. Получен индикатор того, что исследуемый процесс точно является процессом AR(1).

В общем случае процесса AR(p): если $k > p$, то число столбцов в определителе Δ_k больше, чем порядок авторегрессии. В этом случае разностное уравнение для ρ_1, \dots, ρ_k показывает, что начиная с $k > p$, каждое ρ выражается одной и той же линейной комбинацией предыдущих значений. Как только число столбцов больше p , то каждый столбец с номером большим k является линейной комбинацией предыдущих столбцов. Причем коэффициенты этой линейной комбинации – это коэффициенты уравнения процесса $\alpha_1, \alpha_2, \dots, \alpha_p$. Таким образом, последний столбец всегда есть линейная комбинация p предыдущих столбцов, как только $k > p$. Поэтому определитель такой матрицы обязан обратиться в нуль.

Общий вывод: частная автокорреляционная функция авторегрессионного процесса AR(p), равна 0 для $k > p$, и, вообще говоря, не равна 0 при $k \leq p$.

В результате частная автокорреляционная функция для процесса AR играет точно такую же качественную роль, как автокорреляционная функция для процесса MA. Она обращается в нуль, как только $k > p$. Мы это доказали в общем случае и продемонстрировали на процессах первого и второго порядка.

Теперь делаем следующий шаг. Переходим к комбинации двух различных типов процессов AR(p) и MA(q). Рассмотрим общий процесс авторегрессии-скользящего среднего, который носит название: ARMA – авторегрессия-скользящее среднее. Мы назовем так процесс следующего вида: $\alpha_p(L)X_t = \beta_q(L)\varepsilon_t$ или в развернутом виде: $X_t = \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \dots + \beta_q \varepsilon_{t-q}$.

Выведем условия стационарности процесса ARMA. Мы уже установили, что если все корни полинома $\alpha_p(L)$ по модулю меньше единицы, то существует обратный оператор, и мы можем записать: $X_t = [\alpha_p(L)]^{-1} \beta_q(L) \varepsilon_t$. Обратный оператор можно разложить в сумму элементарных дробей, каждую представить как бесконечно убывающую геометрическую прогрессию, то есть бесконечный операторный полином. При умножении на конечный полином мы получим вновь бесконечный полином. Выражение имеет смысл, если все характеристические корни полинома $\alpha_p(L)$ по модулю меньше единицы. А тогда полученное выражение является разложением Вольда, и процесс стационарен. То есть вся эта цепочка приводит к тому, что *стационарность процесса ARMA определяется только его AR-частью*. Поэтому условия те же самые, что у процесса AR. Процесс ARMA стационарен, если корни характеристического уравнения AR-части по модулю меньше единицы. Точно так же, *условия обратимости процесса*, то есть возможность выразить ε_t через X_t , то есть существование выражения: $\varepsilon_t = [\beta_q(L)]^{-1} \alpha_p(L) X_t$, *полностью определяются условиями обратимости MA-части*. Если MA-часть обратима, то и весь процесс обратим.

Итак, если процесс ARMA стационарный, то он имеет обязательно MA(∞) представление, как любой стационарный процесс по теореме Вольда, но он имеет и конечное представление ARMA(p,q). Он может иметь еще и бесконечное AR(∞) представление. То есть мы видим, что ARMA(p,q) может быть очень удобным представлением одного и того же процесса. Поэтому, если можно «согнуть» процесс в ARMA, то он определяется всего $p+q$ параметрами.

Очевидно, что математическое ожидание стационарного процесса ARMA равно нулю: $E\{X_t\} = 0$.

Чтобы выяснить, как ведет себя автокорреляционная и частная автокорреляционная функции процесса ARMA, начнем с простой ситуации, а именно с процесса ARMA(1,1). В операторной записи: $(1 - \alpha L)X_t = (1 + \beta L)\varepsilon_t$, или в обычном виде:

$$X_t = \alpha X_{t-1} + \varepsilon_t + \beta \varepsilon_{t-1}.$$

Условие стационарности принимает вид $|\alpha| < 1$, условие обратимости – $|\beta| < 1$. Для вычисления дисперсии процесса удобно использовать MA(∞) представление,

соответствующее теореме Вольда, так называемое представление линейного фильтра.

$$X_t = \frac{1 + \beta L}{1 - \alpha L} \varepsilon_t = (1 + \beta L)(1 + \alpha L + \alpha^2 L^2 + \dots) \varepsilon_t = [1 + (\alpha + \beta)L + \alpha(\alpha + \beta)L^2 + \dots] \varepsilon_t.$$

$$\text{Тогда } \text{Var}(X_t) = (1 + (\alpha + \beta)^2 + \alpha^2(\alpha + \beta)^2 + \dots) \sigma_\varepsilon^2 = \left[1 + \frac{(\alpha + \beta)^2}{1 - \alpha^2} \right] \sigma_\varepsilon^2 = \frac{1 + \beta^2 + 2\alpha\beta}{1 - \alpha^2} \sigma_\varepsilon^2.$$

Чтобы посчитать первую автокорреляцию, умножим выражение

$$X_t = \alpha X_{t-1} + \varepsilon_t + \beta \varepsilon_{t-1} \text{ на } X_{t-1}$$

и возьмем математическое ожидание от обеих частей. Получим

$$\gamma_1 = \alpha \gamma_0 + \beta \text{Cov}(\varepsilon_{t-1}, X_{t-1}).$$

Умножив выражение $X_{t-1} = \alpha X_{t-2} + \varepsilon_{t-1} + \beta \varepsilon_{t-2}$ на ε_{t-1} и взяв математическое ожидание, получаем $\text{Cov}(\varepsilon_{t-1}, X_{t-1}) = \sigma_\varepsilon^2$. После подстановки получаем $\gamma_1 = \alpha \gamma_0 + \beta \sigma_\varepsilon^2$.

$$\text{Окончательно: } \rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{(\alpha + \beta)(1 + \alpha\beta)}{1 + \beta^2 + 2\alpha\beta}.$$

Последующие значения автокорреляционной функции получить проще. Если умножить X_t на X_{t-2} и взять математическое ожидание, получим: $\gamma_2 = \alpha \gamma_1$. Для всех значений γ с индексом большим, чем порядок МА-части, получаем, что из равенства $\gamma_{k+1} = \alpha \gamma_k$ следует равенство $\rho_{k+1} = \alpha \rho_k$. То есть мы получили то же самое соотношение, которое имели для «чистой» модели AR(1). Мы называли это уравнениями Юла-Уокера для автокорреляционной функции. Другими словами, мы установили, что, начиная со второй, автокорреляции ARMA(1,1) ведут себя так же, как автокорреляции AR(1), но первые автокорреляции этих процессов различаются.

Для AR(1), автокорреляции имели вид $\rho_i = \alpha^i$. Для процесса ARMA(1,1)

$$\rho_1 \neq \alpha_1, \quad \rho_1 = \frac{(\alpha + \beta)(1 + \alpha\beta)}{1 + \beta^2 + 2\alpha\beta}.$$

Однако, начиная со второго номера, значения автокорреляционной функции все равно убывают экспоненциально. Для процесса ARMA(p,q) справедливо аналогичное утверждение, если, конечно, процесс стационарен. Первые p значений автокорреляционной функции определяются через коэффициенты AR и МА-частей, а потом значения автокорреляционной функции выражаются в виде суммы экспоненциально затухающих слагаемых. Вывод очень важен для дальнейшего.

Для процесса ARMA(p,q), применяя тот же метод умножения уравнения процесса на значения X_{t-i} и последующего взятия математического ожидания, получим, что для $k > \max(p, q + 1)$ автокорреляции ρ_k определяются разностным уравнением, соответствующим AR-части. А все предыдущие значения ρ_k могут быть выражены через коэффициенты $\alpha_1, \alpha_2, \dots, \alpha_p, \beta_1, \beta_2, \dots, \beta_q$ как решения системы линейных уравнений, полученной способом аналогичным процессу ARMA(1,1). Следовательно, начиная с номера $\max(p, q + 1)$, автокорреляционная функция «затухает» в том же смысле, что и для процесса AR. Таким же качественным поведением

характеризуется и частная автокорреляционная функция процесса ARMA(p,q). Интуитивно такое свойство понятно. Мы говорили, что для MA(q) автокорреляционная функция для номеров, больших q, просто равна нулю. Поэтому влияние MA-части при $k > q$ как бы прекращается, и дальше «работает» только AR(p). Наоборот, частная автокорреляционная функция для процесса AR(p) для k, больших чем p, равна нулю. То есть AR(p) перестает влиять на частную автокорреляционную функцию, и остается только влияние MA(q).

Резюмируя, получаем, что если процесс относится к типу ARMA(p,q), то, начиная с некоторого номера (причем этот номер важен, он нам говорит о величине p и q), и автокорреляционная, и частная автокорреляционная функции ведут себя как сумма затухающих экспонент, если ряд стационарен.

Хотелось бы обратить внимание на еще один побочный результат. Для процесса AR(1) получено выражение $\rho_1 = \alpha$. Сравним эту первую автокорреляцию с

аналогичной характеристикой процесса ARMA(1,1): $\rho_1 = \frac{(\alpha + \beta)(1 + \alpha\beta)}{1 + \beta^2 + 2\alpha\beta}$. В длин-

ных финансовых рядах было замечено, что, как правило, $\beta < 0$ и $\alpha > |\beta|$. В этом случае первая автокорреляция для процесса ARMA(1,1) будет заметно меньше, чем для процесса AR(1) с тем же коэффициентом α .

Процессы ARMA, рассматриваемые до сих пор, являлись строго недетерминированными, в частности имели нулевое математическое ожидание. Если в уравнение процесса ARMA $\alpha_p(L)x_t = \beta_q(L)\varepsilon_t$ подставить выражение $x_t = X_t - \mu$, то получим: $\alpha_p(L)(X_t - \mu) = \beta_q(L)\varepsilon_t$. Раскрывая скобки и используя очевидное равенство $\alpha_p(L)\mu = \mu(1 - \alpha_1 - \alpha_2 - \dots - \alpha_p) = \alpha_p(1)\mu$, получаем $\alpha_p(L)Y_t = \underbrace{\alpha_p(1)\mu}_{\theta} + \beta_q(L)\varepsilon_t$, где

θ – некоторая константа. Таким образом, введением свободного члена в уравнение мы учитываем ненулевое математическое ожидание. Тогда математическое ожидание процесса будет равно $\mu = \frac{\theta}{\alpha_p(1)}$. Поэтому мы можем без потери общности

рассматривать процесс с ненулевым, но постоянным математическим ожиданием.

Лекция 4

Теперь нам осталось сделать один небольшой шаг. Мы познакомились с моделью стационарного процесса ARMA(p,q). В то же время один из примеров, который мы рассмотрели, был нестационарный ряд случайного блуждания. Его уравнение имело вид: $X_t = X_{t-1} + \varepsilon_t$. Мы видели, что если взять первую разность, равную $\Delta X_t = X_t - X_{t-1} = (1-L)X_t$, то уравнение сведется к следующему: $\Delta X_t = y_t = \varepsilon_t$. То есть в первых разностях ряд станет стационарным. Это был первый пример, когда операции взятия разности приводят к стационарному ряду. Можно заметить, что такой подход приводит к стационарности не только случайное блуждание.

Рассмотрим ряд вида $X_t = \alpha + \beta t + \gamma t^2 + \varepsilon_t$. 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:

$$\Delta X_t = \alpha + \beta t + \gamma t^2 + \varepsilon_t - \alpha - \beta(t-1) - \gamma(t-1)^2 - \varepsilon_{t-1} = \beta + 2\gamma t - \gamma + (\varepsilon_t - \varepsilon_{t-1}).$$

Степень полинома, описывающего тренд, понизилась на единицу. Если провести взятие второй разности, то останется $\Delta^2 X_t = \Delta X_t - \Delta X_{t-1} = 2\gamma + (\varepsilon_t - 2\varepsilon_{t-1} + \varepsilon_{t+2})$, 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. Подметив это свойство, Бокс и Дженкинс [2]³⁾ предложили выделить класс нестационарных рядов, которые взятием последовательных разностей можно привести к стационарному виду, а именно к виду ARMA. Если ряд после взятия d последовательных разностей приводится к стационарному, то мы вслед за Боксом и Дженкинсом назовем этот ряд ARIMA(p,d,q). Так принято, что d вставляется в середину. Сокращение I (от английского – *Integrated*) означает интегрированный. Операция, обратная к взятию последовательной разности, – это суммирование. Бесконечно малая разность называется дифференциалом. Обратный переход от дифференциала называется интегрированием, и этот термин используется в аббревиатуре процесса, хотя имеется в виду суммирование.

ARIMA – процесс авторегрессии – интегрированного скользящего среднего. При этом p – параметр AR-части, d – степень интеграции, и q – это параметр MA-части. В операторном виде ARIMA (p,d,q) записывается как:

$$\alpha_p(L)\Delta^d x_t = \beta_q(L)\varepsilon_t.$$

Или по-другому $\underbrace{\alpha_p(L)(1-L)^d}_{p+d} x_t = \beta_q(L)\varepsilon_t$. Этот процесс нестационарный, потому

что здесь не выполняется условие, что все корни характеристического уравнения по модулю меньше единицы. Но, если обозначить $(1-L)^d x_t = y_t$, то y_t – это стационарный процесс.

Основная заслуга Бокса и Дженкинса не в том, что они это придумали, а в том, что они сделали этот подход лет 25–30 назад весьма популярным и ввели в практику программы для применения этого подхода в компьютерный пакет научных программ для системы IBM-360, распространенной в 1970–1980 гг. по всему миру.

Подход Бокса-Дженкинса

До сих пор мы рассматривали случайные процессы, в частности ARIMA, как некоторую математическую конструкцию. При моделировании экономических про-

³⁾ Первое издание этой книги вышло в 1970 г. Впервые модели типа ARMA были рассмотрены в [4].

цессов мы встречаемся с обратной ситуацией, когда у нас есть реализация ряда, и надо подобрать соответствующую теоретическую конструкцию, которая могла бы породить такую реализацию, то есть построить модель экономического процесса. Пользуясь терминологией Д. Хендри, будем иногда называть такую модель *DGP (Data Generating Process)*. Бокс и Дженкинс предложили следующий подход к выбору модели типа ARIMA по наблюдаемой реализации временного ряда.

Прежде чем описывать этот метод, надо ввести еще одно понятие, которое необходимо иметь в виду, занимаясь оцениванием или подбором модели, порождающей реализацию временного ряда. 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 $E\{X_t\}$ 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, возможностью 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 усреднения по времени (по реализации). Например, мы хотим оценить математическое ожидание. Мы должны взять всевозможные 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 [3] $(T^{-1} \sum_{k=1}^T \gamma_k) \xrightarrow{T \rightarrow \infty} 0$.

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 подразумевает эргодичность. Важно понимать, что одной стационарности при статистической обработке реализаций недостаточно.

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, «остационарить» ряд.

2. После этого мы получаем временной ряд Y_t , к которому нужно подобрать уже ARMA(p,q). Исходя из поведения автокорреляционной (SACF) и частной автокорреляционной функций (SPACF)⁴, 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 $\alpha_1, \alpha_2, \dots, \alpha_p, \beta_1, \beta_2, \dots, \beta_q$ 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.

Бокс и Дженкинс применили этот подход ко многим временным рядам, которые были известны в то время, как к финансовым, так и к макроэкономическим. Правда, надо сказать, что макроэкономические ряды тогда были, в основном, короткие. Бокс и Дженкинс смогли построить модели типа ARIMA для всех исследуемых рядов и установили, в частности, что практически все экономические процессы описываются моделями с относительно небольшими величинами параметров p и q , а параметр d обычно не превышает 2. К тому же оказалось, что точность прогнозирования по моделям ARIMA оказалась выше, чем давали в то время эконометрические модели. Бокс и Дженкинс приложили много усилий, пропагандируя использование моделей типа ARIMA, которые дают удобное и компактное описание очень многих процессов.

Если исследуемый ряд нестационарный, то его автокорреляционная функция не будет убывать. Если ряд стационарен, то мы знаем, что, начиная с какого-то номера, теоретические автокорреляции будут убывать. Поэтому можно рассчитать их оценки – выборочные автокорреляции, и посмотреть, убывают они или нет. Если ряд окажется стационарным, перейти к определению параметров p и q . Если нет, то надо построить ряд первых разностей и проверить на стационарность его.

Рассмотрим, например, процесс $X_t = \alpha X_{t-1} + \varepsilon_t, \varepsilon_t \sim WN(0, \sigma^2)$. При $\alpha > 1$ взятие первой разности не поможет сделать ряд стационарным. Это нестационарный ряд взрывного типа, оценки его «автокорреляционной» функции растут с увеличением сдвига во времени. Если же $\alpha = 1$, то ряд представляет собой случайное блуждание и после взятия первой разности он станет стационарным.

Переходим к оцениванию по выборке статистических характеристик исследуемого процесса в предположении его эргодичности. В качестве оценки математического ожидания применяется статистика $\bar{X} = \frac{1}{T} \sum_{t=1}^T X_t$, то есть обычное сред-

⁴) S – от слова выборочный. Полная аббревиатура означает Sample Autocorrelation Function и Sample Partial Autocorrelation Function соответственно.

нее по выборке. В качестве оценки дисперсии процесса обычно принимается следующая величина: $S^2 = T^{-1} \sum (X_t - \bar{X})^2$. Обратите внимание, делитель не $(T-1)$, как привычно для обработки независимых наблюдений, а T . Для оценки коэффициен-

та теоретической автокорреляции используем $r_k = \frac{\sum_{t=k+1}^T (X_t - \bar{X})(X_{t-k} - \bar{X})}{T \cdot S^2}$. Это озна-

чает, что для расчета выборочных ковариаций используется одинаковый делитель – T , что, в случае независимых наблюдений, дает смещенные оценки соответствующих теоретических ковариаций. Если бы мы требовали несмещенности оценок, то у нас бы были разные делители при различных k . Кроме того, при увеличении k уменьшается объем выборки, пригодный для расчета оценки соответствующего коэффициента корреляции. На практике, по рекомендации, идущей еще от Бокса и Дженкинса, не рассчитываются оценки r_k для $k > T/4$.

Одним из следствий выбора именно таких оценок значений автокорреляционной функции является гарантированная положительная определенность выборочной корреляционной матрицы. Ранее мы отмечали наличие такого свойства у теоретической автокорреляционной функции. При одинаковом делителе у оценок автокорреляционной функции при различных k положительная определенность выборочной корреляционной матрицы гарантирована. Поэтому мы предпочитаем пусть смещенную оценку, но гарантирующую это свойство. Во-вторых, при дополнительном предположении о нормальности распределения значений процесса именно эта оценка совпадает с оценкой метода максимального правдоподобия. Поэтому мы жертвуем здесь несмещенностью, тем более, что у нас обычно T – относительно большое, добываясь 2-х вещей: 1) приближаясь к методу максимального правдоподобия, а он здесь основной, очень важный метод; 2) добываясь положительной определенности соответствующей автокорреляционной матрицы оценок

$$\begin{pmatrix} 1 & r_1 & r_2 & & \\ r_1 & 1 & & & \\ r_2 & & \ddots & & \\ & & & \ddots & \\ & & & & \ddots \end{pmatrix}.$$

В качестве оценок частной автокорреляционной функции $\hat{\phi}_{kk}$ принимаются значения выборочной частной автокорреляционной функции SPACF, которые получаются как коэффициенты выборочной регрессии текущего центрированного значения ряда на k его предыдущих значений

$$(X_t - \bar{X}) = \varphi_{k1}(X_{t-1} - \bar{X}) + \dots + \varphi_{kk}(X_{t-k} - \bar{X}) + \varepsilon_t.$$

Коэффициент $\hat{\phi}_{kk}$ при последнем члене и есть оценка коэффициента φ_{kk} частной автокорреляционной функции. То, что полученная оценка отражает статистическую линейную связь между x_t и x_{t-k} , очищенную от влияния промежуточных значений, подтверждает замечательная теорема Фриша-Вау (1933), которая заключается в следующем.

Теорема Фриша-Вау. Пусть Y – объясняемая переменная, а объясняющие переменные разбиты на две группы: $X_1, \dots, X_k, Z_1, \dots, Z_l$. Обозначим через u, v_1, \dots, v_k остатки метода наименьших квадратов (МНК) от регрессий Y, X_1, \dots, X_k соответственно на совокупность Z_1, \dots, Z_l . Тогда оценки МНК коэффициентов регрессии u_k на v_1, \dots, v_k совпадают с оценками при переменных X_1, \dots, X_k в регрессии Y на всю совокупность $X_1, \dots, X_k, Z_1, \dots, Z_l$.

В математике многие результаты можно выразить в различных терминах. Так теорема Фриша-Вау выражает в терминах коэффициентов регрессий просто формулу обращения блочной матрицы.

Свойства выборочных моментов процесса

Для того, чтобы установить свойства выборочных оценок, введенных ранее, требуются довольно громоздкие и утонченные математические выкладки. Поэтому наиболее сложные результаты мы приведем без доказательства.

Выборочное среднее очевидно является несмещенной оценкой математического ожидания процесса, если он стационарен. Легко выразить дисперсию выборочного среднего:
$$\text{var}(\bar{X}) = T^{-2} \sum_{i=1}^T \sum_{j=1}^T \gamma_{|i-j|} = T^{-2} \sum_{i-j=-T}^T (T - |i-j|) \gamma_{i-j} = T^{-1} \sum_{k=-T}^T \left(1 - \frac{|k|}{T}\right) \gamma_k.$$

Если теоретические ковариации стремятся к нулю при увеличении k , то $\text{var}(\bar{X}) \xrightarrow{T \rightarrow \infty} 0$, и наша оценка будет состоятельной. Если процесс X_t является гауссовым, то и оценка \bar{X} распределена нормально. Более того, эта оценка является суммой случайных величин, более или менее большого количества, и при каких-то достаточно слабых условиях здесь будет работать Центральная Предельная Теорема (ЦПТ). Можно строго математически показать, что для процесса типа ARMA оценка \bar{X} распределена асимптотически нормально, даже для негауссова процесса.

Получение асимптотического распределения оценок r_k сопряжено с большими математическими выкладками, хотя и основано на тех же идеях, что и полученное выше. Сформулируем основные результаты без вывода. Если теоретические автокорреляции $\rho_k = 0$ для всех $k > 0$, то $\text{var}(r_k) \xrightarrow{T \rightarrow \infty} \frac{1}{T}$. Это более точный результат, чем просто сказать, что дисперсия стремится к нулю [2]. И нам очень важно, что при больших T дисперсия ведет себя, как $\frac{1}{T}$. Более точный результат

для больших T говорит следующее: $\sqrt{T} \cdot r_k \stackrel{as}{\sim} N(0,1)$. При достаточно больших T 95-процентный доверительный интервал составляет ± 2 стандартных отклонения. Если мы хотим, например, проверить гипотезу $H_0: \rho_{10} = 0$ против обычной альтернативной гипотезы, надо посчитать r_{10} и посмотреть, попало ли оно в $\pm 2 \frac{1}{\sqrt{T}}$. Ес-

ли выборочное значение r_{10} попадает в этот интервал, то гипотезу H_0 не отвергаем на уровне значимости 5%, если выходим за эту границу, то отвергаем. Эта оценка может быть улучшена. Более точная асимптотическая оценка для дисперсии оценки r_k в предположении, что все $\rho_k = 0$ для $k > q$ дается формулой

$\text{var}(\bar{r}_k) = \frac{1}{T}(1 + 2\rho_1^2 + \dots + 2\rho_q^2)$ для $k > q$. Гипотеза $H_0: \rho_k = 0$ для $k > q$ означает, что процесс есть $MA(q)$. Уточненная оценка означает, что дисперсия r_k асимптотически стремится не просто к $1/T$, а к этой величине с некоторым дополнительным множителем вида $(1 + 2\rho_1^2 + \dots + 2\rho_q^2)$. Этот теоретический результат используется для оценки дисперсий путем замены ρ на выборочное значение r . Для оценки дисперсии $\text{var}(r_1)$ по-прежнему используется $1/T$. Для оценки $\text{var}(r_2)$ используется $T^{-1}(1 + 2r_1^2)$. Для оценки $\text{var}(r_3)$ используется $T^{-1}(1 + 2r_1^2 + 2r_2^2)$. И так далее.

Если временной ряд порожден процессом $AR(p)$, то для $k > p$ оценки значений частной автокорреляционной функции $\hat{\phi}_{kk}$ имеют то же асимптотически нормальное распределение:

$$\sqrt{T} \cdot \hat{\phi}_{kk} \stackrel{as}{\sim} N(0,1).$$

Теперь первый шаг подхода Бокса–Дженкинса заключается в следующем. Только учтите, что свойства всех оценок получены в предположении, что ε является белым шумом. Нормальность белого шума нам здесь не нужна, асимптотическая нормальность оценок обеспечивается за счет Центральной Предельной Теоремы. По реализации временного ряда рассчитываем оценки автокорреляционной и частной автокорреляционной функций, проверяем стационарность, при необходимости переходим к ряду первых разностей.

Специализированные компьютерные программы сегодня устроены следующим образом. Вы говорите: «Хочу исследовать ряд, построить статистику». Программа выдает соответствующие статистики, вы смотрите, они вам не нравятся. Вы, ничего не меняя, говорите: сделай то же самое для ряда разностей. Там это встроено, вам не надо вручную это программировать. Программа строит статистики для первых разностей, вторых. А ваше дело – пока визуально смотреть и решать, достаточно брать конечные разности или нет. Это первый шаг. После выбора стационарного ряда, вы смотрите, с какого номера начинается убывание по абсолютной величине выборочных автокорреляционной и частной автокорреляционной функций. И, исходя из этого, делаете предположение о возможных значениях параметров p и q . Далее можно переходить ко второму этапу: процедуре оценки коэффициентов.



Автор выражает глубокую благодарность своим студентам, которые не только терпеливо слушали и сдавали этот курс, но и внесли большой вклад в создание этого текста. Я также признателен профессору ГУ–ВШЭ Эмилию Борисовичу Ершову за ценные замечания по содержанию и тексту лекций. Разумеется, все ошибки, неточности и неудачные объяснения остаются на совести автора.

* *
*

СПИСОК ЛИТЕРАТУРЫ

1. *Bachelier L.* «Theorie de la Speculation», Annales de l'Ecole Normal Superieure. 1900. Series 3, 17, 21–86.
2. *Box G. E. P.* and *Jenkins G. M.* Time Series Analysis, Forecasting and Control, rev. Ed., San Francisco: Holden-Day, 1976.
3. *Davidson R.* and *MacKinnon J. G.* Estimation and Inference in Econometrics. Oxford University Press, 1992.
4. *Quenouille M. H.* The Analysis of Multiple Time Series London: Charles Griffin, 1957.
5. *Nelson C. R.* and *Kang H.* Pitfalls in the Use of Time as an Explanatory Variable in Regression // Journal of Business and Economic Statistics. Vol. 2. January 1984. P. 73–82.
6. *Nelson C. R.* and *Plosser C. I.* Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implication // Journal of Monetary Economics 1982. 10. P. 139–62.
7. *Wold H.* A Study in the Analysis of Stationary Time Series. Stockholm: Almqvist and Wiksel, 1938.