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: $Y_t = \alpha + \beta \cdot Y_{t-1} + \gamma_0 \cdot X_t + \gamma_1 \cdot X_{t-1} + ... + \varepsilon_t$. 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 наблюденных 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 $\{X_t, t = 0,\pm1,\pm2,...\}$ и именно эту последовательность назвать стохастическим процессом, порождающим наблюденные данные. Как правило, в дальнейшем, если не оговорено обратное, в курсе под стохастическим процессом понимается бесконечная последовательность $\{X_t, t = 0,\pm1,\pm2,...\}$. 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, помесячные 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: $f_1(x_{t_1}); f_2(x_{t_1}, x_{t_2}); f_3(x_{t_1}, x_{t_2}, x_{t_3}); ...$ 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 переменной.

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, $f_n(x_{t_1},...x_{t_n}) = f_n(x_{t_1+\Delta},...x_{t_n+\Delta})$ для всех n, моментов времени $t_1,...t_n$ 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 сходимость 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: $Var(X_t) = V(X_t) = \sigma^2$.

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 $Cov(X_{t_1}, X_{t_2})$. $Cov(X_{t_1}, X_{t_2}) = \iint (x_{t_1} - \mu) \cdot (x_{t_2} - \mu) \cdot f_2(x_{t_1}, x_{t_2}) \cdot dx_{t_1} \cdot dx_{t_2}$. 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 $\gamma(0) = Cov(X_{t_1}, X_{t_1}) = \sigma^2$. Мы можем рассматривать функцию $\gamma(\tau)$ как всевозможные значения автоковариаций, где τ пробегает целочисленные значения от -∞ 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 $|\rho(\tau)| \le 1$. Иногда этот график представляют в виде, как если бы функция $\rho(\tau)$ была определена для всех, а не только целочисленных τ (см. рис. 1).

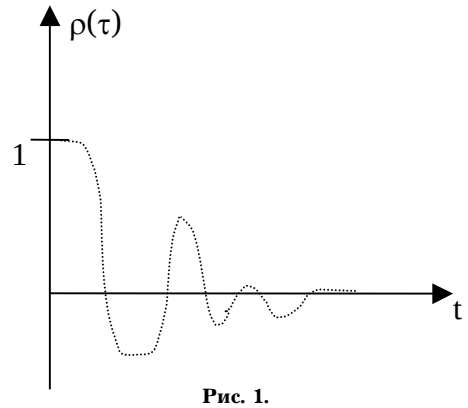

Рис. 1.

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

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

$E\{X_t\} = E(X_0) + E(\sum_{\tau=0}^{t} \varepsilon_{t-\tau}) = X_0 + \sum E(\varepsilon_{t-\tau}) = X_0 + 0 = const$, то есть математическое ожидание удовлетворяет условию стационарности. Подсчитаем дисперсию. $V(X_t) = E\{(\sum_{\tau=0}^{t} \varepsilon_{t-\tau})^2\}$. 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.

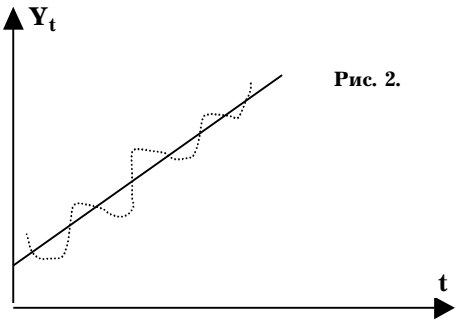

Рис. 2.

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-τ 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 (MA)[1)]

Стохастический процесс называется процессом скользящего среднего порядка q, если в разложении Вольда присутствуют только q слагаемых. То есть: $MA(q): x_t = \sum_{\tau=0}^{q} \psi_t \cdot \varepsilon_{t-\tau}$ (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: $x_t = \varepsilon_t + \psi_1 \cdot \varepsilon_{t-1} + ... + \psi_q \cdot \varepsilon_{t-q}$. 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)] MA = 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 $Cov(x_t, x_{t+\tau}) = \sigma^2 \sum_{i=0}^{q-k} \psi_i \psi_{i+r}$ для τ = 0, 1, 2,...*q*. Для τ>*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 $\tau \le q$, 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: $L^p X_t = L(L(L....)X_t = X_{t-p}$. 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:

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

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: $(1 + \beta_1 L + ... + \beta_q L^q) = \beta_q \prod_{i=1}^{q} (L - Z_i)$, где Z_i – корни уравнения $1 + \beta_1 Z + ... + \beta_q Z^q = 0$.

Для операторного многочлена φ(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 $(1 - \alpha L) \cdot (1 + \alpha L + \alpha^2 L^2 + ... + \alpha^p L^p + ...) = 1$, если бесконечный ряд в скобках существует, имеет смысл. Тогда можно написать: $[1 - \alpha L]^{-1} = \frac{1}{1 - \alpha L} = 1 + \alpha L + ... + \alpha^p L^p + ...$, если правая часть имеет смысл. Рассмотрим конечную сумму вместо бесконечного ряда: $(1 - \alpha L) \cdot (1 + \alpha L + \alpha^2 L^2 + ... + \alpha^p L^p) = 1 - \alpha^{p+1} L^{p+1}$. 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 $p \to \infty$ 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 вид[2)] $|\alpha|<1$.

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 MA(q) представлении на множители, получим: $X_t = \beta_q \prod_{i=1}^{q}(L - Z_i)\varepsilon_t$. Для того, чтобы использовать полученные условия обратимости линейного оператора, удобнее использовать корни другого *характеристического уравнения*, а именно: $\lambda^q + \beta_1\lambda^{q-1} + ... + \beta_q = 0$. 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, $\lambda^q + \beta_1\lambda^{q-1} + ... + \beta_q = 0$, что соответствует общему правилу для разностных уравнений. Либо пишут наоборот: $1 + \beta_1 Z + ... + \beta_q Z^q = 0$. 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:

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

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 $\varepsilon_t = \frac{1}{\prod(1-\pi_i L)} X_t$. Оказывается, этому выражению 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: $\frac{1}{\prod(1-\pi_i L)} = \frac{A_1}{1-\pi_1 L} + \frac{A_2}{1-\pi_2 L} + ...$

[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 $|\pi_1| < 1$, то первую элементарную дробь можно разложить в сумму бесконечно убывающей геометрической прогрессии или в бесконечный степенной ряд

$$\frac{A_1}{1-\pi_1 L} = A_1(1+\pi_1 L+\pi_1^2 L^2+...+\pi_1^k L^k+...).$$

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: $(\alpha_0+\alpha_1 L+\alpha_2 L^2+...+\alpha_k L^k+...+)X_t$, 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 $\frac{A_1}{1-\pi_i L}$ 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 MA. 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: $(\alpha_0+\alpha_1 L+\alpha_2 L^2+...)X_t=\varepsilon_t$. Если выразить X_t через все его предыдущие значения и ε_t и для нормировки разделить на α_0, то получается эквивалентное представление $X_t = -\frac{\alpha_1}{\alpha_0}X_{t-1} - ... - \frac{\alpha_k}{\alpha_0}X_{t-k} - ... + \frac{1}{\alpha_0}\varepsilon_t$. Тот же самый процесс MA(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: $X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + ... + \alpha_p X_{t-p} + \varepsilon_t$, где ε_t – белый шум.

В этих терминах процесс, обратный к MA(q), может быть обозначен как AR(∞). Однако у нас нет гарантии, что при любых коэффициентах $\alpha_1, \alpha_2, ..., \alpha_p$ 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:

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

Обозначим. через $\alpha_p(L) = 1 - \alpha_1 L - \alpha_2 L^2 - ... - \alpha_p L^p$ 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: $\lambda^p - \alpha_1 \lambda^{p-1} - ... - \alpha_p = 0$, 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: $X_t = \frac{1}{\alpha(L)}\varepsilon_t$. Следовательно, 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 $X_t = \alpha_1 X_{t-1} + ... + \alpha_p X_{t-p} + \varepsilon_t$ 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

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

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 разновременные. Поэтому $E\{X_{t-\tau} \cdot \varepsilon_t\} = 0$, потому что значения белого шума не коррелируют между собой. Другими словами, получается следующее выражение: $\gamma(\tau) = \alpha_1 \gamma(\tau - 1) + ... + \alpha_p \gamma(\tau - p)$.

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:

$$\gamma(1) = \alpha\gamma(0)$$
$$\gamma(2) = \alpha\gamma(1) = \alpha^2\gamma(0)$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$\gamma(\tau) = \alpha^\tau\gamma(0)$$

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:

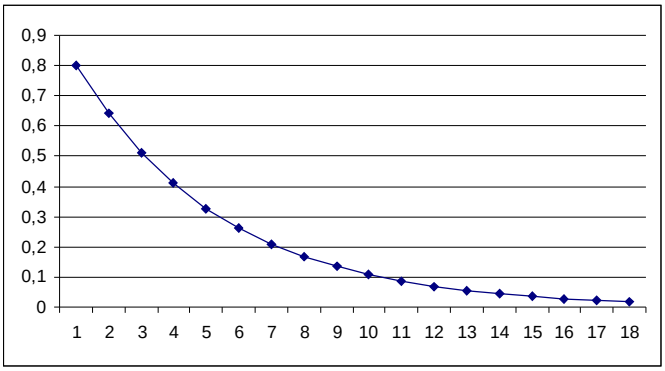

Рис. 1. Значения автокорреляционной функции процесса AR(1) при значении коэффициента равном 0,8

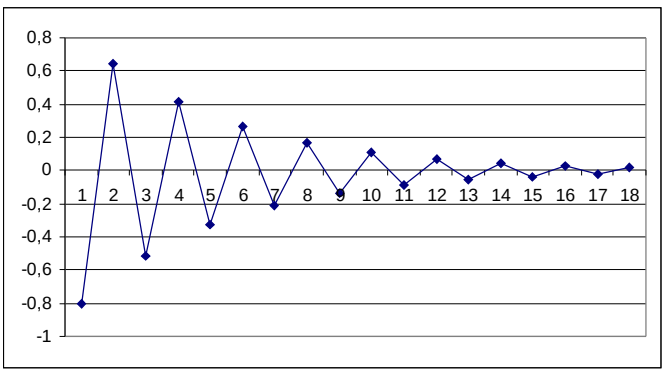

Рис. 2. Значения автокорреляционной функции процесса AR(1) при значении коэффициента равном −0,8

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 acf. 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 авторегрессионого 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 авторегрессионому уравнению

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

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:

$$\gamma_1 = \alpha_1\gamma_0 + \alpha_2\gamma_1 + ... + \alpha_p\gamma_{p-1}$$
$$\gamma_2 = \alpha_1\gamma_1 + \alpha_2\gamma_0 + ... + \alpha_p\gamma_{p-2}$$
$$.....$$
$$\gamma_p = \alpha_1\gamma_{p-1} + \alpha_2\gamma_{p-2} + ... + \alpha_p\gamma_0.$$

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 авторегрессионой 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 определение автокорреляционной функции $\rho(\tau) = \frac{1}{Var(X_t)} E\{(X_t - \mu)(X_{t-\tau} - \mu)\}$ входит 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-τ. 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 $x_{t-1}, x_{t-2}, ..., x_{t-k-1}$. Эта объясняющая линейная комбинация называется *оптимальным линейным предиктором*, если она реализует минимальную среднеквадратичную ошибку предсказания (прогноза) x_t. Более строго, нас интересует коэффициент корреляции

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

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

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

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 $x_t = \varphi_{k1}x_{t-1} + \varphi_{k2}x_{t-2} + ... + \varphi_{kk}x_{t-k} + \varepsilon_t$. 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 авторегрессионым 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 поразному. Но, вообще говоря, нас интересует только один единственный коэффициент φ_{22}, поэтому наиболее экономно будет использовать правило Крамера. $\varphi_{22} = \dfrac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{1-\rho_1^2}$. 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: $\varphi_{22} = \dfrac{\alpha^2 - \alpha^2}{1-\rho_1^2} = 0$, 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:

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

Искомый определитель имеет вид:

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

На главной диагонали все элементы кроме последнего будут равны 1. Связь между коэффициентами φ_{kk} и $\rho_1,...\rho_k$, то есть между автокорреляционной и частной автокорреляционной функциями, выражается следующим соотношением: $\varphi_{kk} = \frac{\Delta_k}{\Delta}$, где Δ – определитель системы уравнений

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

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 $\rho_1,...\rho_k$ 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 $\alpha_1,\alpha_2,...,\alpha_p$. 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 авторегрессионого процесса AR(p), равна 0 для $k>p$, и, вообще говоря, не равна 0 при $k \le p$.

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: $X_t = \alpha_1 X_{t-1} + ... + \alpha_p X_{t-p} + \varepsilon_t + \beta_1\varepsilon_{t-1} + ... + \beta_q\varepsilon_{t-q}$.

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 «согнать» 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.

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

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

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

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

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 $Cov(\varepsilon_{t-1}, X_{t-1}) = \sigma_\varepsilon^2$. После подстановки получаем $\gamma_1 = \alpha\gamma_0 + \beta\sigma_\varepsilon^2$. 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 MA-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 $\rho_i = \alpha_1^i$. Для процесса ARMA(1,1) $\rho_1 \neq \alpha_1$, $\rho_1 = \frac{(\alpha+\beta)(1+\alpha\beta)}{1+\beta^2+2\alpha\beta}$. 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 MA-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 $\alpha_1, \alpha_2, ..., \alpha_p, \beta_1, \beta_2, ..., \beta_q$ 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 длинных 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 $\alpha_p(L)\mu = \mu(1-\alpha_1-\alpha_2-...-\alpha_p) = \alpha_p(1)\mu$, 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 крайне 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] $(T^{-1}\sum_{k=1}^{T}\gamma_k)\xrightarrow[T\to\infty]{}0$.

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 $\alpha_1,\alpha_2,...,\alpha_p,\beta_1,\beta_2,...,\beta_q$ 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 $\overline{X} = \frac{1}{T}\sum_{t=1}^{T} X_t$, 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: $S^2 = T^{-1}\sum(X_t - \overline{X})^2$. Обратите внимание, делитель не (*T*-1), как привычно для обработки независимых наблюдений, а *T*. Для оценки коэффициента теоретической автокорреляции используем $r_k = \frac{\sum_{t=k+1}^{T}(X_t - \overline{X})(X_{t-k} - \overline{X})}{T \cdot S^2}$. Это означает, 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, добиваясь 2-х вещей: 1) приближаясь к методу максимального правдоподобия, а он здесь основной, очень важный метод; 2) добиваясь положительной определенности соответствующей автокорреляционной матрицы оценок

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

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

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

Коэффициент $\hat{\varphi}_{kk}$ 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: $X_1,...,X_k,Z_1,...Z_l$. Обозначим через $u,v_1,...,v_k$ остатки метода наименьших квадратов (МНК) от регрессий $Y,X_1,...,X_k$ соответственно на совокупность $Z_1,...Z_l$. Тогда оценки МНК коэффициентов регрессии u_k на $v_1,...,v_k$ совпадают с оценками при переменных $X_1,...,X_k$ в регрессии Y на всю совокупность $X_1,...,X_k,Z_1,...Z_l$.

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: $\mathrm{var}(\overline{X}) = T^{-2}\sum_{i=1}^{T}\sum_{j=1}^{T}\gamma_{|i-j|} = T^{-2}\sum_{i-j=-T}^{T}(T-|i-j|)\gamma_{i-j} = T^{-1}\sum_{k=-T}^{T}(1-\left|\frac{k}{T}\right|)\gamma_k$.

Если теоретические ковариации стремятся к нулю при увеличении k, то $\mathrm{var}(\overline{X}) \xrightarrow[T\to\infty]{} 0$, и наша оценка будет состоятельной. Если процесс X_t является гауссовым, то и оценка $\overline{X}$ распределена нормально. Более того, эта оценка являются 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 $\overline{X}$ 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 $\mathrm{var}(r_k) \xrightarrow[T\to\infty]{} \frac{1}{T}$. 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: $\sqrt{T}\cdot r_k \overset{as}{\sim} N(0,1)$. 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 $\mathrm{var}(r_k) = \frac{1}{T}(1+2\rho_1^2+...+2\rho_q^2)$ 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 $(1+2\rho_1^2+...+2\rho_q^2)$. Этот теоретический результат используется для оценки дисперсий путем замены ρ на выборочное значение r. Для оценки дисперсии $\mathrm{var}(r_1)$ по-прежнему используется $1/T$. Для оценки $\mathrm{var}(r_2)$ используется $T^{-1}(1+2r_1^2)$. Для оценки $\mathrm{var}(r_3)$ 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 $\hat{\varphi}_{kk}$ имеют то же асимптотически нормальное распределение: $\sqrt{T}\cdot\hat{\varphi}_{kk} \overset{as}{\sim} N(0,1)$.

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 Эмилю 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.

ЛЕКЦИОННЫЕ И МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ

Анализ временных рядов

Канторович Г.Г.

В этом номере продолжается публикация курса лекций «Анализ временных рядов». В прошлом номере были рассмотрены основные определения понятия стационарных случайных процессов, их характеристики и свойства, а также класс стационарных случайных процессов типа ARMA и нестационарных – типа ARIMA. Проанализирован подход Бокса–Дженкинса к идентификации временных рядов. В настоящем номере продолжается рассмотрение подхода Бокса–Дженкинса, в частности, оценивание параметров типа ARIMA и прогнозирование с помощью этих моделей, приведены примеры применения подхода Бокса–Дженкинса.

Рассмотрены особенности поведения некоторых нестационарных временных рядов, нестационарные ряды типа TS и DS, тест Дикки–Фуллера для проверки гипотезы о типе ряда.

Лекция 5

Оценивание коэффициентов моделей типа ARMA

Для оценки коэффициентов модели AR(p) $X_t = \alpha_1 X_{t-1} + ... + \alpha_p X_{t-p} + \varepsilon_t$ можно применить обычный метод наименьших квадратов (МНК). Поскольку регрессоры относятся к предыдущим моментам времени, а ε_t – белый шум, то корреляция регрессоров со случайным возмущением ε_t отсутствует. МНК не дает несмещенных оценок, но дает состоятельные оценки, несмотря на то, что присутствует стохастический регрессор. Разумеется, мы должны тщательно проверить по остаткам, действительно ли наши предположения о том, что ε_t – белый шум, выполнены. Если дополнительно белый шум является гауссовым, то значения X_t распределены нормально, а оценки коэффициентов, произведенные МНК, состоятельны и асимптотически нормальны. Подчеркнем, что нормальность оценок достигается только асимптотически.

Если данные имеют ненулевое выборочное среднее, то можно либо вычесть это среднее из данных и строить регрессию без свободного члена, либо просто строить регрессию со свободным членом θ. Другими словами, для оценки математического ожидания процесса AR(p) можно использовать две статистики: уже упомя-

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

нутую в прошлой лекции $\overline{X} = \frac{1}{T}\sum_{t=1}^{T} X_t$ и $\hat{\mu} = \frac{\hat{\theta}}{1-\hat{\alpha}_1 - ...\hat{\alpha}_p}$, в которой использованы оценки МНК. Для гауссова процесса ε_t обе оценки состоятельны и асимптотически нормальны. Более того, обе они асимптотически независимы от оценок параметров модели, полученных МНК.

Для моделей скользящего среднего невозможно аналитически выразить остаточную сумму квадратов $\sum e_i^2$ через значения реализации X_t и параметры модели, а следовательно, применить МНК. Для оценки параметров существуют 2 возможности, по-разному реализованные сегодня в специализированных компьютерных программах.

Первая возможность состоит в применении метода максимального правдоподобия. В предположении нормальности ошибки ε_t выражаем ковариационную матрицу ошибок (ее элементы – значения автокорреляционной функции) через параметры β_1, β_2, ..., β_q обратимой модели MA(q). Функция правдоподобия для нормально распределенного вектора $\vec{X}$ имеет вид:

$L(\vec{\beta}, \sigma_\varepsilon^2 \mid \vec{X}) = \frac{1}{(\sqrt{2\pi\sigma_\varepsilon^2})^T}(\sqrt{\det \Sigma_X})^{-1} \exp(-\frac{\vec{x}' \Sigma_X^{-1} \vec{x}}{2\sigma_\varepsilon^2})$, где $\vec{x}$ – отклонения от среднего.

Здесь через Σ_X обозначена ковариационная матрица процесса $\vec{X}$. Элементы этой матрицы выражаются, как мы видели, через параметры модели β_1, β_2, ..., β_q, поэтому процедура численной оптимизации позволяет найти оценки метода максимального правдоподобия, которые будут обладать обычными свойствами состоятельности и асимптотической нормальности. Кроме того, оценки параметров модели MA(q) и оценка дисперсии случайного возмущения асимптотически независимы.

Второй подход, позволяющий облегчить вычислительные затраты, применили Бокс и Дженкинс [2]. Они предложили использовать процедуру нелинейной оптимизации: процедуру поиска на сетке (grid-search procedure). Рассмотрим идею этого подхода на примере процесса MA(2). Найдем сначала оценку математического ожидания процесса $\overline{X}$ в предположении о его стационарности и эргодичности. Затем выберем некоторые значения (β_1, β_2), например (0,5; 0,5). Исходя из уравнения процесса, X_1 выражается через предыдущие ошибки, но они не известны. Поэтому будем считать, что $X_1 = \overline{X} + e_1$. Это дает нам $e_1 = X_1 - \overline{X}$. Затем полагаем $e_2 = X_2 - \overline{X} - \beta_1 e_1$, $e_3 = X_3 - \overline{X} - \beta_1 e_2 - \beta_2 e_1$ и так далее.

Теперь можно вычислить сумму квадратов отклонений $\sum e_i^2 = S(\beta_1, \beta_2)$. Поскольку она посчитана для выбранных значений параметров (β_1, β_2), мы можем рассматривать ее как функцию этих переменных. Далее какой-либо численной процедурой, например поиском на сетке, можно перебирать комбинации β_1 и β_2 или численно искать минимум этой функции: $\min_{\beta_1, \beta_2} \sum e_i^2$. Коэффициенты (β_1^*, β_2^*),

которые обеспечивают минимум выражения $\sum e_i^2$, и будут оценками коэффициентов модели MA(2).

В общем случае модели MA(q) мы фиксируем начальные значения $\beta_1^0,...\beta_q^0$ и определяем остатки через наблюдения: $e_1 = X_1 - \overline{X}$, $e_2 = X_2 - \overline{X} - \beta_1^0 e_1$ и так далее. Процедура получения остатков достаточно проста, все остатки с отрицательными и нулевым индексами в используемых выражениях заменяем нулями. Затем находим $\min\limits_{\vec{\beta}} \sum e_i^2$. Численным поиском находим оценки параметров. Вопросы сходимости этой процедуры и влияния выбора начальных значений на результат мы оставим за рамками рассмотрения. Упомянем только, что показана асимптотическая эквивалентность этой процедуры оцениванию методом максимального правдоподобия.

Аналогичная процедура может быть использована для оценки параметров модели ARMA(p,q). Для их оценивания может применяться комбинация метода наименьших квадратов с поиском на сетке. Чтобы проиллюстрировать, как это делается, рассмотрим модель ARMA(2,2). Пусть уравнение модели имеет вид

$$(1-\alpha_1 L-\alpha_2 L^2)X_t = (1+\beta_1 L+\beta_2 L^2)\varepsilon_t .$$

Его можно переписать в виде $X_t = \dfrac{(1+\beta_1 L+\beta_2 L^2)\varepsilon_t}{1-\alpha_1 L-\alpha_2 L^2}$. Введем вспомогательный случайный процесс $Z_t = \dfrac{1}{1-\alpha_1 L-\alpha_2 L^2}\varepsilon_t$. Тогда процессы X_t и Z_t связаны соотношением $X_t = (1+\beta_1 L+\beta_2 L^2)Z_t$, которое напоминает уравнение MA(2), только вместо ε_t стоит Z_t. Определим «наблюдения» Z_t через наблюдения X_t, сконструировав Z_t так же, как ранее остатки модели MA, т.е. значения Z_t, которые еще не определены (с нулевыми и отрицательными индексами), полагаем равными нулю. Тогда:

$$\begin{aligned} Z_1 &= X_1 \\ Z_2 &= X_2 - \beta_1 Z_1 \\ Z_3 &= X_3 - \beta_1 Z_2 - \beta_2 Z_1 \\ &\cdots \end{aligned}$$

Например, из выражения $X_2 = Z_2 + \beta_1 Z_1 + \underbrace{\beta_2 Z_0}_{=0}$ следует второе из соотношений. Далее значения Z_t связаны с остатками e_t исходной модели следующими соотношениями: $Z_t - \alpha_1 Z_{t-1} - \alpha_2 Z_{t-2} = e_t$. Относительно процесса Z_t модель стала AR(2). Зная «реализацию» Z_t для выбранных значений коэффициентов β_1, β_2, можно оценить коэффициенты α_1 и α_2 с помощью обыкновенного МНК. В результате находим оценки коэффициентов $(\tilde{\alpha}_1, \tilde{\alpha}_2, \tilde{\beta}_1, \tilde{\beta}_2)$, но получены они по-разному.

Оценки $(\tilde{\beta}_1,\tilde{\beta}_2)$ заданы как начальные значения. А потом, исходя из них, построены оптимальные оценки $(\tilde{\alpha}_1,\tilde{\alpha}_2)$. Применяя численный метод оптимизации (например поиск на сетке) оцениваем значения параметров, обеспечивающие минимум остаточной суммы квадратов $\sum e_i^2$.

Если белый шум гауссов, для оценивания модели ARMA(p,q) можно также применять метод максимального правдоподобия. Общая схема его применения такова. Выражаем значения автокорреляционной функции процесса через параметры модели, формируем ковариационную матрицу порядка T, записываем функцию правдоподобия для имеющейся выборки $X_1,X_2,...,X_T$, решаем (как правило, численно) систему уравнений правдоподобия относительно оценок коэффициентов. Реальные алгоритмы метода максимального правдоподобия, реализованные в специализированных компьютерных пакетах, отличаются от этой схемы, но принципиально она осуществима и отражает логику получения оценок метода максимального правдоподобия.

Напомним процедуру, которую предложили Бокс и Дженкинс. На первом шаге мы определяем, является ли ряд стационарным. Для этого смотрим на поведение выборочной автокорреляционной функции и выборочной частной автокорреляционной функции. Если их поведение не свидетельствует о стационарности исследуемого процесса, преобразуем исходный ряд взятием последовательных разностей и смотрим на поведение тех же функций для преобразованного ряда. Добившись стационарности, мы тем самым определили параметр интеграции d временного ряда. После этого, анализируя для стационарного ряда поведение выборочной автокорреляционной и частной автокорреляционной функций, подбираем параметры p и q. Все это вместе принято называть *идентификацией модели*. Это некая смесь точных знаний, интуиции и искусства. Затем, после определения (p,d,q), мы тем или иным способом оцениваем коэффициенты соответствующей ARIMA модели.

Прежде чем рассмотреть примеры применения этого подхода, обсудим, как оценить качество полученной модели. Конечно, надо проверить, хорошо ли легла модель на данные, т.е. хорошо ли мы подогнали модель. А также – не нарушены ли нами некоторые предположения о характере случайного возмущения ε_t, что заставит нас сменить, возможно, методы оценивания.

Диагностика модели ARMA

Ясно, что исходной информацией для диагностики служат остатки модели. Мы предполагали, что случайное возмущение является белым шумом, поэтому, прежде всего, надо проверять некоррелированность остатков. Итак, проверке подлежат, в первую очередь, 2 момента. Первый – это качество модели. Нам нужен индикатор типа критерия множественной детерминации R^2. И второй – нам надо обязательно проверить некоррелированность остатков, иначе оценивать AR(p)-часть методом наименьших квадратов нельзя, так как мы получаем несостоятельные оценки.

Как правило, при построении моделей временных рядов критерии качества подгонки моделей применяются для сравнения моделей между собой. Поскольку оценки коэффициентов проводятся путем оптимизации, фактически речь идет о

выборе порядка модели, т.е. о сравнении моделей с различным числом параметров. Абсолютные критерии, типа стандартного коэффициента множественной детерминации R^2, не применяются. Из значительного числа критериев познакомимся с двумя наиболее часто применяемыми. Более подробно о критериях подгонки моделей можно прочитать в известной монографии [5].

Наиболее распространенными в настоящее время является предложенный Акаике в 1974 г. [1] критерий **AIC** (Akaike information criterion) и предложенный Шварцем в 1978 г. [14] **BIC** (Bayesian information criterion). Оба эти критерия построены примерно одинаковым способом. Информационный критерий AIC для модели ARMA(p,q) выглядит следующим образом:

$$AIC(p,q) = \ln\hat{\sigma}^2 + 2\frac{p+q}{T},$$

где T – число наблюдений. А критерий Шварца имеет несколько другой вид: $BIC(p,q) = \ln\hat{\sigma}^2 + \ln T\frac{(p+q)}{T}$. Разумеется, $\hat{\sigma}^2 = \frac{RSS}{T-p-q}$. Обратите внимание, как построены эти критерии: логарифмы остаточной суммы квадратов плюс штраф за уменьшение числа степеней свободы. Нам нужно так подобрать значения параметров p и q, чтобы получить минимальное значение каждого из критериев, в то время как по критерию R^2 мы отдавали предпочтение модели с большим его значением. Обратите внимание, что в отличие от коэффициента детерминации ничего нельзя сказать ни о диапазоне изменения, ни о знаке критериев Акаике и Шварца, поскольку численная величина оценки σ^2 зависит от единиц измерения.

Критерий Акаике базируется на обобщении принципа максимального правдоподобия. Приведенное выражение подразумевает, что случайное возмущение является гауссовым. Шибата [13] показал, что для процессов AR критерий AIC переоценивает порядок модели, следовательно, оценка порядка модели на основании этого критерия несостоятельна.

Критерий Шварца BIC основан на байесовском подходе и имеет более фундаментальное теоретическое обоснование. Показано, что оценка порядка модели по этому критерию является состоятельной. Тем не менее, на практике чаще используется информационный критерий AIC. Существует масса работ, в которых сравнивается применение этих критериев по отношению к разным моделям, но окончательного вывода пока не сделано. На практике разные критерии могут привести к выбору различных моделей. Есть некоторый накопленный опыт по поводу того, к каким типам моделей приводит один критерий и к каким приводит другой. Если вы публикуете какие-то исследования, считается вполне нормальным просто указать, какой критерий вы применяете без специального обоснования вашего выбора. Популярный пакет для анализа временных рядов *Econometric views* рассчитывает оба эти критерия.

Теперь перейдем к проверке автокорреляции остатков. В этом вопросе существует расхождение между тем, как следовало бы поступать с теоретической точки зрения и тем, что делается практически. Еще в 1970 г. Бокс и Пирс предложили *статистический критерий для проверки автокорреляции временного ряда*, который сегодня принято называть *Q*-статистикой*. Тест Бокса–Пирса проверяет гипотезу о совместном равенстве нулю всех автокорреляций временного ряда до порядка m включительно, т.е. гипотезу H_0: $\rho_1 = \rho_2 = ... = \rho_m$ против альтернативной

гипотезы H_1: $\sum_{i=1}^{m} \rho_i^2 > 0$. Бокс и Пирс показали, что при увеличении длины выборки статистика $Q^* = T\sum_{k=1}^{m} r_k^2$ имеет асимптотическое распределение χ_m^2. Этот тест было предложено применять для проверки наличия автокорреляции остатков модели типа ARMA(p,q), при этом число степеней свободы уменьшается на (p+q).

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 Льюнг [6] в 1978 г. предложили использовать для тех же целей улучшенную Q-статистику $Q = (T+2)T\sum_{i=1}^{m}(T-i)^{-1} r_i^2$, которую, как и статистику Бокса–Пирса, часто называют портманто-статистикой (portmanteau-statistics). По сравнению со статистикой Бокса–Пирса различным слагаемым приданы разные веса. Авторы показали, что эта статистика имеет то же асимптотическое распределение χ_m^2, но лучше им аппроксимируется при конечном числе наблюдений. К сожалению, и статистика Бокса–Льюнга 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–Льюнга входит во все специализированные пакеты, множество исследователей ею пользуются, хотя она теоретически не состоятельна.

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 Бройша–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. Мы знакомились с этим тестом еще в курсе эконометрики, сформулируем его применительно к анализу автокорреляции остатков.

Пусть рассматривается модель множественной регрессии

$$Y_t = \beta_0 + \beta_1 X_1 + ... + \beta_k X_k + \varepsilon_t,$$

где $X_1,...X_k$ – разные регрессоры, в том числе, возможно, и лаговые значения как объясняющих, так и объясняемой переменных. Проверим предположение, что ε_t

подчиняется авторегрессионой схеме порядка p, т.е. задается уравнением $\varepsilon_t = \gamma_1 \varepsilon_{t-1} + ... + \gamma_p \varepsilon_{t-p} + u_t$, где u_t – белый шум. В терминах коэффициентов модели основная и альтернативная гипотезы принимают вид:

$$H_0 : \gamma_1 = ... = \gamma_p = 0$$
$$H_1 : \gamma_1^2 + ... + \gamma_p^2 > 0 .$$

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

$$Y_t = \beta_0 + \beta_1 X_1 + ... + \beta_k X_k + \varepsilon_t .$$

Обозначим ее остатки через e_t.

1. Строится регрессия либо той же объясняемой переменной Y_t, либо остатков e_t на старые регрессоры и остатки с лагом до p включительно (то есть в качестве дополнительных объясняющих переменных используем $e_{t-1}, e_{t-2}, ..., e_{t-p}$).

2. Проверяется гипотеза о том, что группа дополнительных переменных является излишней. Если в качестве объясняемой переменной используются остатки, то статистика TR^2, где T – число наблюдений, а R^2 – коэффициент множественной детерминации регрессии из пункта 1, имеет асимптотическое (при увеличении числа наблюдений) распределение χ^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, напомним, она применима только при нормальном распределении случайного члена. Применение же теста множителей Лагранжа не требует нормальности распределения, но «работает» только асимптотически. Современные эконометрические 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 TR^2, верную лишь асимптотически.

Для проверки нормальности существует множество тестов. Мы рассмотрим *тест Харке–Бера (Jarque–Bera).* Этот тест выглядит следующим образом. Он вычисляет выборочные значения для коэффициентов асимметрии $S = \frac{1}{T\sigma^3}\sum (e_t - \bar{X})^3$ и эксцесса $K = \frac{1}{T\sigma^4}\sum (e_t - \mu)^4$, где $\bar{X}$ – выборочное среднее, а σ – выборочное среднеквадратичное остатков модели. При условии нормальности остатков, статистика Харке–Бера $\frac{(T-p-q-1)}{6}[S^2 + \frac{1}{4}(K-3)^2]$ имеет χ^2 распределение с двумя степенями свободы.

Лекция 6

Рассмотрим несколько примеров применения подхода Бокса–Дженкинса к временным рядам, представляющим экономические переменные.

Одним из наиболее часто анализируемых финансовых временных рядов является индекс реальной годовой доходности, рассчитываемый компанией Standard & Poors по 500 акциям для США (S&P-500). Рассмотрим ряд для периода 1872–1995 гг. [9]. Среднее значение временного ряда составляет 3,08 американских цента за год.

Выборочные значения автокорреляционной функции r_k и соответствующие значения среднеквадратичных ошибок этих оценок приведены в табл. 6.1. В третьем столбце приведены значения статистики Бокса–Льюнга и соответствующие им критические значения вероятностей.

Таблица 6.1.

Номер лага	r_k	s.e.(r_k)	$Q(k)$
1	0,043	0,093	0,24 (0,62)
2	-0,169	0,093	3,89 (0,14)
3	0,108	0,093	5,40 (0,14)
4	-0,057	0,094	5,83 (0,21)
5	-0,117	0,094	7,61 (0,18)
6	0,030	0,094	7,73 (0,26)
7	0,096	0,094	8,96 (0,25)
8	-0,076	0,096	9,74 (0,28)
9	-0,000	0,097	9,74 (0,37)
10	0,086	0,097	10,76 (0,38)
11	-0,038	0,099	10,96 (0,45)
12	-0,148	0,099	14,00 (0,30)

Ряд содержит 123 точки. Это довольно большое число, поэтому для проверки значимости можно пользоваться нормальным распределением. Совершенно очевидно, что ни одно из значений r_k не значимо на традиционных уровнях значимости. Это видно просто по обычной t-статистике. Отношение оценки к ее среднеквадратическому отклонению, как правило, даже меньше единицы.

Тот же вывод можно сделать по статистикам Q, все их значения не позволяют отвергнуть гипотезу о нулевых автокорреляциях. Наименьшее из значений критической вероятности p-value составило 0,14. Это значит, что при любом уровне значимости меньше чем 0,14 мы не можем отвергнуть гипотезу об отсутствии автокорреляции, т.е. о совместном равенстве нулю всех ρ_k. Отсюда мы делаем первый вывод, что временной ряд годовых значений индекса S&P-500 является стационарным. Во-вторых, можно идентифицировать модель ARMA. Если бы у ряда была MA-часть, то мы знаем, что $\rho_k = 0$, начиная с некоторого номера. У нас уже $\rho_1 = 0$, т.е. MA-части нет. Если бы была AR-часть, то значения ρ_k должны были бы экспоненциально убывать, что не наблюдается. Итак, мы установили,

что наш процесс имеет вид $X_t = \mu + e_t$, другими словами – это просто белый шум, или ARMA(0,0). Это давно известный результат. Обычно его трактуют как справедливость игры на бирже, т.е. полную непредсказуемость изменения доходности акций.

Рассмотрим, следуя Т. Миллсу [8], пример другого временного ряда. Объектом рассмотрения является так называемый interest rate spread, т.е. разница между ставками по коротким и длинным бумагам. В качестве длинных бумаг для Великобритании выбраны 20-летние бумаги (UK gilts), а в качестве коротких – бумаги на 91 день (91 day Treasury Bills). Данные являются квартальными, период наблюдения с I квартала 1952 г. по IV квартал 1988 г.

Для этого ряда были рассчитаны 12 значений выборочных автокорреляционной и частной автокорреляционной функций (см. табл. 6.2).

Таблица 6.2.

Номер лага	r_k	s.e. (r_k)	$\hat{\Phi}_{kk}$	s.e. ($\hat{\Phi}_{kk}$)
1	0,829	0,082	0,829	0,082
2	0,672	0,126	-0,048	0,082
3	0,547	0,148	0,009	0,082
4	0,435	0,161	-0,034	0,082
5	0,346	0,169	0,005	0,082
6	0,279	0,174	0,012	0,082
7	0,189	0,176	-0,116	0,082
8	0,154	0,178	0,114	0,082
9	0,145	0,179	0,047	0,082
10	0,164	0,180	0,100	0,082
11	0,185	0,181	0,028	0,082
12	0,207	0,182	0,038	0,082

Вплоть до 9-ого лага значения выборочной автокорреляционной функции r_k убывают по величине, а потом они начинают немного расти. Можно сделать заключение, что этот процесс абсолютно точно не является белым шумом. В то же время сомнений в его стационарности нет. Пять первых автокорреляций статистически значимы, остальные нет. Похоже, что ряд может иметь порядок авторегрессии 1 или больше, возможно, увеличение значений функции после 9-ого лага вызвано присутствием пары комплексных корней в характеристическом уравнении.

Для проверки наличия автокорреляций у исходного ряда посмотрим значения портманто-статистик: $Q(12)=305{,}0$, а $Q^*(12)=295{,}2$. Поскольку известно, что математическое ожидание случайной величины, распределенной по χ^2_k, равно k, то даже без таблиц понятно, что, используя Q статистику, мы должны отклонить гипотезу о том, что первые 12 теоретических автокорреляций совместно равны нулю.

Стоит посмотреть на поведение выборочной частной автокорреляционной функции. Единственное статистически значимое значение выборочной частной автокорреляционной функции – первое. Можно сделать вывод, что это достаточно веское свидетельство в пользу модели AR(1).

Раз ряд стационарен, то d=0. Следовательно, модель, скорее всего, имеет вид ARIMA(1,0,0). Конечно, это не единственная возможность. Давайте оценим эту модель. Модель AR(1) оценивается просто методом наименьших квадратов, что дает следующие результаты (в скобках под оценками коэффициентов – их стандартные ошибки): $X_t = \underset{(0,098)}{0,176} + \underset{(0,045)}{0,856}\, X_{t-1} + e_t$; $\hat{\sigma} = 0,870$. Мы видим, что interest rate spread уже имеет смысл прогнозировать, так как некоторая информация о его будущем содержится в его текущем значении. Модель содержит значимый свободный член, это означает, что у ряда есть ненулевое математическое ожидание. Его оценка равна $\hat{\mu} = \dfrac{0,176}{1-0,856} = 1,227$, стандартная ошибка этой оценки составляет 0,506. Q-статистика для ряда остатков этой модели составила 13,21, что очевидно свидетельствует об отсутствии автокорреляций, хотя мы помним, что теоретически этот тест здесь неприменим. Впрочем, тест множителей Лагранжа приводит к тому же выводу.

Наш подход все время строился на том, что мы подбирали наилучшую модель. Есть альтернативный подход. Если мы можем один и тот же процесс записать моделями достаточно близкими по характеристикам, то будем формировать так называемый портфель моделей (набирать группу моделей). В нашем случае в первую очередь имело бы смысл рассмотреть модели ARMA(1,1) и AR(2). После оценивания получим следующие модели:

$$AR(2): X_t = \underset{(0,101)}{0,197} + \underset{(0,054)}{0,927}\, X_{t-1} - \underset{(0,084)}{0,079}\, X_{t-2} + e_t; \qquad \hat{\sigma} = 0,869.$$

Обратите внимание, оценка остаточной суммы квадратов стала чуть меньше, но не очень существенно. Коэффициент при втором регрессоре незначимый.

$$ARMA(1,1): X_t = \underset{(0,104)}{0,213} + \underset{(0,051)}{0,831}\, X_{t-1} + e_t + \underset{0,095}{0,092}\, e_{t-1}; \qquad \hat{\sigma} = 0,870.$$

И в этой модели уменьшения оценки дисперсии ошибки не произошло, хотя мы добавили параметр. Естественно, что мы отдаем предпочтение AR(1). Метод множителей Лагранжа, примененный к каждой из двух моделей, показал отсутствие автокорреляции остатков.

Приведем еще один пример, который показывает, что отбор лучшей модели по критериям AIC и BIC может привести к различным результатам. Рассмотрим ряд месячных значений индекса деловой активности для Великобритании FTA All Share (Financial Times-Actuaries) за период с 1965 по 1990 гг. [8]. Ряд насчитывает около 300 точек, т.е. достаточно длинный. Вначале рассчитаем значения выборочных автокорреляционной и частной автокорреляционной функций (см. табл. 6.3).

Кроме того, была посчитана Q-статистика, $Q(12)$=19,3. Аргумент 12 указывает количество слагаемых, включаемых в сумму, т.е. количество автокорреляций, совместное равенство которых нулю проверяется. Нулевая гипотеза для этой статистики заключается в том, что вплоть до лага с номером 12 истинные значения ρ_k равны нулю. Эквивалент p-value для $Q(12)$ составил 0,08. Следовательно, на уровне значимости 5% мы не можем отвергнуть нулевую гипотезу об отсутствии ненулевых значений автокорреляционной функции случайного члена вплоть до порядка, равного 12.

Таблица 6.3.

Количество лагов	r_k	s.e. (r_k)	$\hat{\Phi}_{kk}$	s.e. ($\hat{\Phi}_{kk}$)
1	0,148	0,057	0,148	0,057
2	-0,061	0,059	0,085	0,057
3	0,117	0,060	0,143	0,057
4	0,067	0,061	0,020	0,057
5	-0,082	0,062	-0,079	0,057
6	0,013	0,062	0,034	0,057
7	0,041	0,063	0,008	0,057
8	-0,011	0,064	0,002	0,057
9	0,087	0,064	0,102	0,057
10	0,021	0,064	-0,030	0,057
11	-0,008	0,064	0,012	0,057
12	0,026	0,064	0,064	0,057

Из табл. 6.3 трудно сразу выявить какое-то определенное поведение r_k или $\hat{\Phi}_{kk}$, нет убывания значений по абсолютной величине начиная с некоторого номера. Значения r_k для лагов 1 и 3 значимы на 5-процентном уровне значимости, если мы применим критические значения нормального распределения t=1,96. Во всяком случае, трудно увидеть в этих данных свидетельства в пользу точных значений параметров p и q. Пожалуй, имеет смысл рассматривать p и q не больше 3, поскольку значения с большим лагом незначимы. Итак, $p \le 3, q \le 3$. Мы не пытаемся определить точно p и q, а выбираем некие их максимально возможные значения.

Поскольку значения выборочных автокорреляционной и частной автокорреляционной функций быстро становятся незначимыми, есть основания считать ряд стационарным. В принципе поведение r_k и $\hat{\Phi}_{kk}$ соответствует схеме поведения стационарного ряда. У нас есть некоторое количество значимых выборочных значений, правда, не видно, чтобы они экспоненциально убывали, но, по крайней мере, стационарные модели с p и q меньше или равным 3 вполне разумные «кандидаты» на более подробное исследование.

Один из популярных подходов заключается в том, что строятся все модели для $p \le 3$ и $q \le 3$. Здесь их совсем немного. Поэтому составим сводную таблицу для значений критериев AIC и BIC в зависимости от параметров модели.

Таблица 6.4.

Критерий AIC				
p	q			
	0	1	2	3
0	3,701	3,684	3,685	3,688
1	3,689	3,685	3,694	3,696
2	3,691	3,695	3,704	3,699
3	**3,683**	3,693	3,698	3,707

Продолжение таблицы

Критерий BIC				
p	q			
	0	1	2	3
0	3,701	**3,696**	3,709	3,724
1	3,701	3,709	3,730	3,744
2	3,715	3,731	3,752	3,759
3	3,719	3,741	3,758	3,779

Как видно, по критерию AIC, минимальный показатель у модели ARMA(3,0). По критерию Шварца наилучшей оказалась другая модель ARMA(0,1), т.е. скользящее среднее первого порядка.

Рассмотрим оцененные модели, отобранные по обоим критериям.

Для модели ARMA(3,0) было получено следующее регрессионное уравнение (в скобках – среднеквадратические ошибки оценок соответствующих коэффициентов):

$$\underset{s.e}{X_t} = \underset{}{4{,}41} + \underset{(0{,}062)}{0{,}173}\, X_{t-1} - \underset{(0{,}057)}{0{,}108}\, X_{t-2} + \underset{(0{,}057)}{0{,}144}\, X_{t-3} + e_t; \quad \hat{\sigma} = 6{,}25.$$

Для модели ARMA(0,1) было получено:

$$X_t = \underset{(0{,}35)}{5{,}58} + \underset{(0{,}056)}{0{,}186}\, e_{t-1} + e_t; \quad \hat{\sigma} = 6{,}29\,.$$

Прежде всего отметим, что все коэффициенты в обеих моделях значимы. Во-вторых, полученные модели выглядят весьма разными. Давайте посмотрим, так ли это. Начнем с оценок математических ожиданий. Для модели ARMA(0,1) оценка математического ожидания, очевидно, равна $\hat{\mu} = 5{,}58$, а для модели ARMA(3,0) оценка математического ожидания будет равна

$$\hat{\mu} = \frac{4{,}41}{1 - 0{,}173 + 0{,}108 - 0{,}144} = 5{,}58\,.$$

Видно, что по оценке математического ожидания эти модели эквивалентны.

Так как ARMA(0,1) – это, по сути, есть модель MA(1), и коэффициент 0,186 меньше 1, эта модель является обратимой. Поэтому она эквивалентна следующему выражению:

$$\varepsilon_t = \theta_0 - 0{,}186X_{t-1} + (0{,}186)^2 X_{t-2} - (0{,}186)^3 X_{t-3} + \ldots = \theta_0 - 0{,}186X_{t-1} + 0{,}035X_{t-2} - 0{,}006X_{t-3} + \ldots$$

Теперь можно сравнить соответствующие коэффициенты: 0,173 и 0,186 (это первые коэффициенты соответствующих моделей); 0,035 и 0,108 и, наконец, последние – 0,006 и 0,144. Если первые коэффициенты более или менее соответствуют друг другу, то последующие уже не очень. Это означает, что краткосрочная динамика моделей различна. Но если сложить все коэффициенты, то эти суммы оказываются достаточно близкими. Одна равна 0,843, а вторая – 0,791. Сумма коэффициентов модели с распределенными лагами характеризует полное или долгосрочное поведение. Поэтому долгосрочное поведение этих двух моделей не очень отличается друг от друга.

Как уже упоминалось, была предложена другая идея, идея использования портфеля моделей. Т.е. вместо требования выбора одной единственной модели мы выбираем набор, портфель моделей, которыми можно описывать исследуемый временной ряд.

Чтобы сформировать этот портфель, был предложен эвристический подход, т.е. без строгого обоснования, просто из разумных соображений. Poskitt и Tremayne в 1987 г. [12] предложили рассматривать в качестве меры близости моделей следующую величину: $R = \exp^{\{-\frac{1}{2}T[AIC(p_1,q_1)-AIC(p,q)]\}}$. Здесь параметры ($p_1$, q_1) характеризуют наилучшую по рассматриваемому критерию модель. А для всех остальных моделей с параметрами (p, q) мы рассчитываем величину R. Никакого содержательного смысла в ней нет, это просто мера «расстояния» между моделями. Poskitt и Tremayne предложили включать модель в портфель, если $R < \sqrt{10}$.

Применим этот подход к нашему примеру, посмотрим к каким моделям мы приходим. Критерий AIC отбирает в нашем случае 6 моделей:

$$\{(3,0);(0,1);(0,2);(1,1);(0,3);(1,0)\},$$

которые разумно рассматривать.

А вот критерий BIC выбрал всего 3 модели: $\{(0,1);(1,0);(0,0)\}$. Любопытно, что по критерию BIC даже модель (0,0) включается в портфель, т.е. даже модель чистого белого шума не так уж плоха. Все модели очень простые (с малым числом коэффициентов), как мы уже отмечали у критерия BIC более серьезные штрафы за потерю степеней свободы.

Отобранные модели можно использовать для прогнозирования и для поиска взаимосвязей между переменными. Это более или менее жесткий отбор «кандидатов».

Лекция 7

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 ε;

– 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 математической ожидание: $E\{X_{T+h}\big|X_1,...X_T\}$. Такое условное математическое ожидание будет гарантировать получение MSE, или иногда ее обозначают MMSE, т.е. Minimum mean square error.

Начнем с модели MA(q): $X_t = \theta + \varepsilon_t + \beta_1\varepsilon_{t-1} + ... + \beta_q\varepsilon_{t-q}$. Мы полагаем, что коэффициенты модели точно известны, и что имеются значения X_t для $t \in [1,T]$. Очевидно, что безусловным точечным прогнозом для любого момента времени будет математическое ожидание процесса, т.е. θ. Условным прогнозом для момента времени T+1 будет условное математическое ожидание

$$\widetilde{X}_{T+1} = E\{\theta + \varepsilon_{T+1} + \beta_1\varepsilon_T + ... + \beta_q\varepsilon_{T-q+1}\big|X_1,..., X_T\}.$$

Среди случайных величин ε, 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 ε_{T+1}, не равны нулю. Рассмотрим сначала $E\{\varepsilon_T\big|X_1,...X_T\}$, потому что схема одна и та же. Это математическое ожидание – остаток между наблюдением и расчетом, прогнозом по модели, т.е. $e_T = X_T - \widetilde{X}_T$. Поэтому условные математические ожидания от всех предыдущих значений случайной составляющей надо заменить соответствующими остатками. Точно так же конструируется прогноз не на 1, а на 2 и вообще на h шагов вперед. Все последующие ε 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 θ, т.е. условный прогноз совпадает с безусловным.

Рассмотрим условную дисперсию ошибки прогноза на 1 шаг:

$$\mathrm{var}(X_{T+1}-\widetilde{X}_{T+1}\Big|X_1,...X_T)=E\{((\theta+\varepsilon_{T+1}+\beta_1\varepsilon_T+...+\beta_q\varepsilon_{T-q+1}-\theta-\beta_1 e_T+...+\beta_q e_{T-q+1})\Big|X_1,...X_T)^2\}=\sigma_\varepsilon^2.$$

Аналогично дисперсия прогноза на 2 шага равна $\mathrm{var}(X_{T+2}-\widetilde{X}_{T+2}\Big|X_1,...X_T)=(1+\beta_1^2)\sigma_\varepsilon^2$, а дисперсия прогноза на h шагов составит $(1+\beta_1^2+...+\beta_h^2)\sigma_\varepsilon^2$ 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):

$$X_t=\theta+\alpha_1X_{t-1}+\alpha_2X_{t-2}+...+\alpha_pX_{t-p}+\varepsilon_t.$$

Для прогноза на 1 шаг вперед можно записать:

$$\widetilde{X}_{T+1}=E\{X_{T+1}\Big|X_1,...X_T\}=E\{\theta+\alpha_1X_T+...+\alpha_pX_{T-p}+\varepsilon_{T+1}\Big|X_1,...X_T\}=\theta+\alpha_1X_T+...+\alpha_pX_{T-p}.$$

Мы просто подставляем в уравнение модели p предыдущих значений реализации временного ряда. Для прогноза на 2 шага вперед получаем:

$$E\{X_{T+2}\Big|X_1,...X_T\}=E\{\theta+\alpha_1X_{T+1}+...+\alpha_pX_{T-p+2}+\varepsilon_{T+2}\Big|X_1,...X_T\}.$$

Разумеется, математическое ожидание от случайной ошибки ε опять даст 0, условные математические ожидания от $X_1, X_2,..., X_T$ равны самим этим значениям, но в это выражение входит условное математическое ожидание от X_{T+1}, полученное на предыдущем шаге. Можно подставить его выражение и получить развернутую формулу через значения реализации, но обычно это не нужно. Удобнее рассматривать рекуррентное соотношение, связывающее последовательные значения прогноза. Это соотношение является линейным разностным уравнением порядка p, и его решение стремится при увеличении t к величине $\dfrac{\theta}{1-\alpha_1-...-\alpha_p}$, т.е. опять к безусловному прогнозу.

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 $\widetilde{X}_{T+1}=\alpha_1X_T+\alpha_2X_{T-1}$, а $X_{T+1}=\alpha_1X_T+\alpha_2X_{T-1}+\varepsilon_{T+1}$. Очевидно, что дисперсия ошибки прогноза на 1 шаг равна σ_ε^2. Для прогноза на 2 шага соответственно получаем:

$$\widetilde{X}_{T+2}=\alpha_1\widetilde{X}_{T+1}+\alpha_2X_T=\alpha_1(\alpha_1X_T+\alpha_2X_{T-1})+\alpha_2X_T,$$
$$X_{T+2}=\alpha_1X_{T+1}+\alpha_2X_T+\varepsilon_{T+2}=\alpha_1(\alpha_1X_T+\alpha_2X_{T-1}+\varepsilon_{T+1})+\alpha_2X_T+\varepsilon_{T+2}.$$

Дисперсия ошибки прогноза на 2 шага равна: $(1+\beta_1^2)\sigma_\varepsilon^2$. Для прогноза на 3 шага получим:

$$\widetilde{X}_{T+3}=\alpha_1\widetilde{X}_{T21}+\alpha_2\widetilde{X}_{T+1}=\alpha_1(\alpha_1(\alpha_1X_T+\alpha_2X_{T-1})+\alpha_2X_T)+\alpha_2(\alpha_1X_T+\alpha_2X_{T-1}),$$
$$X_{T+3}=\alpha_1X_{T+2}+\alpha_2X_{T+1}+\varepsilon_{T+3}=$$
$$=\alpha_1(\alpha_1(\alpha_1X_T+\alpha_2X_{T-1}+\varepsilon_{T+1})+\alpha_2X_T+\varepsilon_{T+2})+\alpha_2(\alpha_1X_T+\alpha_2X_{T-1}+\varepsilon_{T+1})+\varepsilon_{T+3}.$$

Дисперсия ошибки прогноза на 3 шага равна $(1+\beta_1^2+\beta_2^2+2\beta_1^2\beta_2+\beta_1^4)\sigma_\varepsilon^2$. 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 $X_t=\theta+\varepsilon_t+\psi_1\varepsilon_{t-1}+\ldots+\psi_q\varepsilon_{t-q}+\ldots$, хотя и с бесконечным числом слагаемых. Тогда дисперсия ошибки прогноза на h шагов выражается очевидной формулой $\sigma_\varepsilon^2\sum_{i=0}^{h-1}\psi_i^2$ $(\psi_0=1)$.

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,T] – наблюденные 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 что полученной формулой.

В качестве примера рассмотрим модель ARMA(1,1) $X_t=\alpha X_{t-1}+\varepsilon_t+\beta\varepsilon_{t-1}$. Тогда $X_t=(1-\alpha L)^{-1}(1+\beta L)\varepsilon_t=(1+\alpha L+\alpha^2L^2+\ldots)(1+\beta L)\varepsilon_t=\varepsilon_t+\eta_1\varepsilon_{t-1}+\eta_2\varepsilon_{t-2}+\ldots$, где $\eta_1=\alpha+\beta$, $\eta_2=\alpha(\alpha+\beta),\ldots,\eta_k=\alpha^{k-1}(\alpha+\beta)$. Начиная со второго коэффициенты убывают в геометрической прогрессии. Теперь легко посчитать дисперсию ошибки прогноза на h шагов: $\sigma_\varepsilon^2\sum_{i=0}^{h-1}\eta_i^2=\sigma_\varepsilon^2(1+\sum_{i=0}^{h-1}\alpha^{2i}(\alpha+\beta)^2)=\sigma_\varepsilon^2(1+\frac{(\alpha+\beta)(1-\alpha^{2h})}{1-\alpha^2})$. Сравнив с результатом, полученным в лекции 3, видим, что и для этой модели дисперсия ошибки прогноза асимптотически равна дисперсии временного ряда.

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 нереалистического 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 общеупотребительных приемов является разбиение имеющейся реализации на 2 части. По первым n наблюдениям выбирается и оценивается модель, а по последним (T-n) наблюдениям проводится сравнение наблюденных и рассчитанных по модели значений. Такая процедура иногда называется постпрогнозом. Сравнивая прогнозную силу моделей, отобранных по критерию Поскитта–Тремейна, мы выбираем «наилучшую».

Нестационарные временные ряды

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

1 тип: процесс с детерминированным полиномиальным трендом.

$X_t = P_k(t)$, где $P_k(t)$ – полином степени k от t, а ε_t – стационарный процесс, не обязательно белый шум. Если ограничиться рассмотрением только линейного тренда $X_t = \alpha + \beta t + \varepsilon_t$, то можно записать: $\Delta X_t = X_t - X_{t-1} = (1-L)X_t = \beta + (\varepsilon_t - \varepsilon_{t-1})$. Поскольку ε_t – стационарный процесс, то его первая разность также стационарный процесс, хотя если ε_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: $X_t = \mu + X_{t-1} + \varepsilon_t$. В этом случае $\Delta X_t = \mu + \varepsilon_t$, и процесс X_t называется *случайным блужданием с дрейфом*. Мы можем записать решение этого разностного уравнения в следующем виде:

$$X_t = \mu t + \sum_{j=0}^{t-1} \varepsilon_{t-j}.$$

Поэтому, 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 $\Delta X_t = \varepsilon_t$, а мы строим регрессию на время вида $X_t = \alpha + \beta t + \varepsilon_t$? Можно интуитивно догадаться, что мы получим значимый по

t-статистике коэффициент β. 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. Для краткости 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 $X_t = \alpha + \beta t + \varepsilon_t$, 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 (diferencing stationary process)*. Вид этого процесса таков: $X_t = X_{t-1} + \varepsilon_t$. Иногда такой процесс называют DS.

Сведем же рассмотренные нами различия в свойствах этих процессов в табл. 7.1.

Таблица 7.1.

Процесс TSP	Процесс DSP
• Не стационарен из-за непостоянного тренда.	• Не стационарен из-за непостоянной дисперсии.
• 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.	• Так как в явном решении стоит сумма всех предыдущих ε, то шоки помнятся все время. Это процесс с бесконечной памятью. Экономически это не очень понятно, шоки не должны сказываться постоянно.

В 1984 г. была опубликована работа Нельсона и Канга [10], в которой они задались вопросом: что произойдет, если процесс является DSP, а мы будем выделять линейный тренд, как для процесса типа TSP? 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 $R^2 \approx 0{,}44$ независимо от размера выборки, т.е. R^2 значим даже для малого количества точек наблюдения.

2. Если дрейф присутствует ($\mu \neq 0$), то $R^2 > 0{,}44$ и зависит от размера выборки. Это интуитивно можно понять, потому что как только $\mu \neq 0$, появляется свой тренд – нестационарное среднее. Также оказалось, что $R^2 \to 1$ при $T \to \infty$, т.е. для процесса DSP с дрейфом $R^2 \to 1$ при увеличении объема выборки. Это явление значимого коэффициента тренда, когда в истинном уравнении тренд отсутствует, было названо «кажущиеся тренды» (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 $\rho_1 = 1 - \frac{10}{T}$.

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:

$$\begin{aligned} Y_t &= Y_{t-1} + \upsilon_t \\ X_t &= X_{t-1} + \varepsilon_t \end{aligned},$$

где υ_t и ε_t – независимые белые шумы, и построить регрессию $Y_t = a + bX_t + u_t$, 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 зависимы 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 связана динамика денежной массы и инфляции, но мы не учли, что оба процесса – DSP, 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 [11], 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. Раньше мы рассматривали для этого поведение выборочной автокорреляционной функции.

В прошлой лекции мы рассматривали ряд значений спрэда в Великобритании и пришли к выводу, что он описывается моделью AR(1). Если взять первые разности этого ряда, то он будет также хорошо описываться моделью ARIMA(0,1,0). А это очень разные ряды. Один из них (TS с нулевым коэффициентом при трен-

де) является стационарным, а второй – нестационарным. Визуальное же исследование поведения выборочной автокорреляционной функции, т.е. применение подхода Бокса–Дженкинса не позволяет надежно выбрать, к какому из типов относится исследуемый ряд.

Мы нуждаемся в методе, позволяющем формально проводить различие между рядами типа TSP и DSP.

Эти 2 ряда описываются следующими моделями:

TSP	DSP
$X_t = \alpha + \beta t + \varepsilon_t$	$X_t = \mu + X_{t-1} + \varepsilon_t$

Попробуем рассмотреть модель: $X_t = \alpha + \rho X_{t-1} + \beta t + \varepsilon_t$. Она вобрала черты обеих моделей, и гипотезы о характере ряда можно записать в виде простых гипотез о ее параметрах.

H_0: ряд является DS $\Rightarrow \begin{cases} \rho = 1 \\ \beta = 0 \end{cases}$.

H_1: ряд является TS $\Rightarrow |\rho| < 1$ (тогда ε будет не просто белый шум, а некоторый стационарный ряд. Вообще-то мы это имели в виду, когда определяли TS, предполагая, что он приводится к стационарному с помощью выделения тренда).

Нулевая гипотеза относится к классу общих линейных гипотез. При традиционном подходе для ее проверки нужно оценить 2 регрессии: $X_t = \alpha + \rho X_{t-1} + \beta t + \varepsilon_t$ и $X_t = \alpha + X_{t-1} + \varepsilon_t$. И затем проверить значимость разности остаточных сумм квадратов, используя F-статистику.

Тест, к рассмотрению которого мы сейчас переходим, появился сначала в более простой версии этой модели: $X_t = \alpha + \rho X_{t-1} + \varepsilon_t$, т.е. без включения линейного тренда. В этом случае гипотезы принимают вид:

$$H_0: \ \rho = 1 \Rightarrow \text{DS}$$
$$H_1: \ \rho < 1 \Rightarrow \text{TS}.$$

Теперь можно проверить гипотезу о том, что $\rho = 1$ при помощи t-статистики. Уравнение можно переписать в другом виде. После вычитания из обеих частей X_{t-1}, получим $\Delta X_t = \alpha + (\rho - 1) X_{t-1} + \varepsilon_t$. Пусть $\rho - 1 = \gamma$, тогда проверяемые гипотезы примут вид:

$$H_0: \ \gamma = 0$$
$$H_1: \ \gamma < 0.$$

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 $\frac{\hat{\gamma}}{s.e.(\hat{\gamma})}$ не является распределени-

ем 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 $\frac{\hat{\gamma}}{s.e.(\hat{\gamma})}$ можно выразить через стохастические интегралы от винеровского случайного процесса, но мы этого делать не будем. Критические точки этого распределения приходится рассчитывать численно, используя симуляционные процедуры Монте-Карло. Впервые это распределение было выведено и затабулировано в работе Дикки и Фуллера [3] и носит их имя. Тест, использующий для проверки типа нестационарности это распределение, при условии $\gamma = 0$, т.е. когда процесс принадлежит типу DS, называется тестом Дикки–Фуллера, и обозначается как DF-тест**.** При условии, что нулевая гипотеза о том, что $\gamma = 0$, выполнена, мы имеем процесс типа случайного блуждания (DSP). Именно для этого случая не работает *t*-статистика.

Позже MacKinnon [7] расширил эти таблицы, рассмотрел некоторые другие случаи и предложил аппроксимирующие формулы для быстрого расчета критических точек в компьютерных программах. С помощью моделирования Дикки и Фуллер 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: $X_t = \alpha + \beta t + (\rho - 1)X_{t-1} + \varepsilon_t$. В этом случае мы имеем следующие гипотезы:

$$H_0 : \rho = 1; \quad \beta = 0$$
$$H_1 : \rho < 1.$$

Сравним, как ведет себя *DF*-статистика и *F*-статистика. Нас интересует критическое значение при заданом числе наблюдений и для заданного уровня значимости. Если T – количество наблюдений, то в зависимости от него получим следующую табл. 7.2 для уровня значимости 5%, в которой *F*-статистика – это фактически $F_{2,T-3}$.

Таблица 7.2.

T	DF	*F*-статистика
25	7,24	3,42
50	6,73	3,20
100	6,49	3,10
∞	6,25	3,00

На рис. 7.1 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 (а; в).

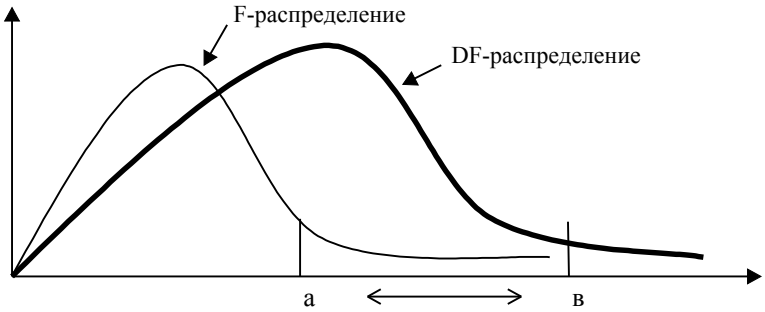

Рис. 7.1.

Кроме того, оказалось, что аналитическое выражение, а следовательно, и форма DF-распределения зависят от того, включены ли в модель регрессоры $\alpha t, \beta t$ 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 [4]. Рассматриваются данные с I квартала 1950 г. по IV квартал 1977 г. Специфицируем модель в следующем виде:

$$Y_t = \beta_0 + \beta_1 \cdot t + \alpha_1 \cdot Y_{t-1} + \alpha_2 (Y_{t-1} - Y_{t-2}) + \varepsilon_t .$$

Естественно, мы предполагаем, что $\varepsilon_t \sim WN(0, \sigma^2)$, т.е. белый шум.

Эта модель была оценена методом наименьших квадратов, но перед оцениванием из обеих частей вычли Y_{t-1}. Это стандартный прием, тогда можно сравнивать коэффициент с нулем, а не с единицей. А в левой части уравнения оказывается объясняемая переменная ΔY. Поэтому были оценены 2 следующие модели, соответствующие нулевой и альтернативной гипотезе (в скобках – стандартные ошибки оценок коэффициентов).

Первая модель: $\underset{s.e.}{Y_t - Y_{t-1} = \Delta Y_t} = \underset{(0,13)}{0,52} + \underset{(0,00034)}{0,0012}\, t - \underset{(0,033)}{0,119}\, Y_{t-1} + \underset{(0,081)}{0,498}(Y_{t-1} - Y_{t-2})$;

$$RSS = 0,056448 \ (RSS_{UR}) .$$

***Вторая модель*:** Пусть $\beta_1 = 0; \alpha_1 = 1$. МНК дает следующую модель: $\underset{s.e.}{Y_t - Y_{t-1}} = \underset{(0,0025)}{0,0054} + \underset{(0,083)}{0,447}(Y_{t-1} - Y_{t-2})$. Соответственно $RSS = 0,063211\,(RSS_R)$. Считаем стандартное F-отношение: $F_{отношение} = \dfrac{(0,063211 - 0,056448)/2}{0,056448/106} = 6,34$. А теперь посмотрите на приведенную выше табл. 7.2. 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 5% 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.

Наибольшее распространение получило использование распределения Дикки–Фуллера для статистики $\frac{\hat{\gamma}}{s.e.(\hat{\gamma})}$. Мы подробно рассмотрим ее применение в следующих лекциях.

* *
*

СПИСОК ЛИТЕРАТУРЫ

1. *Akaike H.* A New Look at the Statistical Model Identification, IEEE Transactions on Automatic Control, AC-19. 1974. P. 716–723.
2. *Box G. E. P. and Jenkins G. M.* Time Series Analysis, Forecasting and Control, rev. Ed. San Francisco: Holden-Day, 1976.
3. *Dickey D. A. and Fuller W. A.* Distribution of the Estimators for Autoregressive Time-Series with a Unit Root // Journal of the American Statistical Assiciation. 1979. Vol. 74. P. 427–431.
4. *Dickey D. A. and Fuller W. A.* Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root // Econometrica. 1981. Vol. 49. P. 1057–1072.
5. *Judge G. G., Griffits W. E., Hill R. C., Lutkepohl H., Lee Tsoung-Chao.* The Theory and Practice of Econometrics. Second edition. NY: John Willey and Sons, 1985.
6. *Ljung G. M. and Box G. E. P.* On a Measure of Lack of Fit in Time Series Models // Biometrika. 1978. 65. P. 297–303.
7. *MacKinnon J. C.* Critical Values for Cointegration Tests // UC San Diego Discussion Paper. 1990. P. 90–4.
8. *Mills T. C.* The Econometric Modelling of Financial Time Series. Cambridge University Press, 1993.
9. *Mills T. C.* The Econometric Modelling of Financial Time Series. Second edition. Cambridge University Press, 1999.
10. *Nelson C. R. and. Kang H.* Pitfalls in the Use of Time as an Explanatory Variable in Regression // Journal of Business and Economic Statistics. January 1984. Vol. 2. P. 73–82.
11. *Nelson C. R. and Plosser C. I.* Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implication // Journal of Monetary Economics. 1982. Vol. 10. P. 139–162.
12. *Poskitt D. S. and Tremayne A. R.* Determining a Portfolio of Linear Time Series Models // Biometrica. 1987. V. 74. P. 125–37.
13. *Schibata R.* Selection of the Order on an Autoregressive Model by Akaike's Information Criterion // Biometrika. 1976. 63. P. 147–164.
14. *Schwarz G.* Estimating the Dimension of a Model // The Annals of Statistics. 1978. 6. P. 461–464.

ЛЕКЦИОННЫЕ И МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ

Анализ временных рядов

Канторович Г.Г.

В этом номере публикуются очередные три лекции курса «Анализ временных рядов». Они посвящены методологии применения расширенного теста Дикки–Фуллера (ADF-test) для проверки наличия единичного корня, или, другими словами, определения типа нестационарности временного ряда. Рассматривается зависимость распределения так называемого t-отношения от наличия в составе регрессоров свободного члена и/или тренда, а также от лаговой структуры исследуемого ряда. В качестве методики предлагается процедура Доладо и его соавторов. Рассматривается случай кратных единичных корней. Десятая лекция посвящена исследованию наличия единичных корней при возможном наличии структурного скачка в параметрах модели. Обсуждаются результаты Перрона для экзогенного времени скачка и Зивота с Эндрюсом – для эндогенного. Приводятся примеры конкретных экономических рядов. В последующих лекциях предполагается рассмотреть модели взаимосвязи стационарных и нестационарных временных рядов.

Лекция 8

В прошлой лекции мы упоминали распределение Дикки–Фуллера двух типов: типа F-отношения и типа t-отношения $\frac{\hat{\gamma}}{s.e.(\hat{\gamma})}$. 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.

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

Дикки и Фуллер [4] начинали с исследования уравнения: $\Delta X_t = \gamma X_{t-1} + \varepsilon_t$. В этом случае при условии, что выполнена гипотеза H_0, т.е. что $\gamma = 1$, распределение t-отношения (отношения оценки коэффициента γ, полученной МНК, к оценке его среднеквадратичного отклонения) называется DF-распределением. Включив в модель свободный член, получим модель $\Delta X_t = \alpha + \gamma X_{t-1} + \varepsilon_t$. А дополнительное включение линейного тренда дает модель $\Delta X_t = \alpha + \beta t + \gamma X_{t-1} + \varepsilon_t$. Оказалось, что во всех трех случаях распределение t-отношения $\frac{\hat{\gamma}}{s.e.(\hat{\gamma})}$ 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 τ_0, для второго распределения τ_μ, для третьего распределения τ_τ.

Взаимное расположение этих критических значений для одного и того же уровня значимости и одного и того же числа степеней свободы схематически показано на рис. 8.1. 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.

Рис. 8.1.

Прежде чем двигаться дальше зададимся следующим вопросом. Из результатов Нельсона и Канга [11] мы знаем об опасности принятия процесса типа I(1) за процесс типа I(0) и применения к нему обычных процедур метода наименьших квадратов. Если мы будем применять к рядам типа DS обычные формулы регрессионного анализа, то получим кажущиеся зависимости, кажущиеся тренды, т.е. некий статистический феномен вместо экономической взаимосвязи, которую ищем. Может показаться, что достаточно обращаться с процессом типа TS как с процессом типа DS, т.е. брать от него первые разности и дальше оперировать с рядом первых разностей. Ведь при поиске взаимозависимости между процессами y и z можно перейти к рядам Δy и Δz, которые останутся процессами типа I(0), и строить обычную регрессию. При этом может оказаться, что один из рядов (или оба) уже были стационарными и не требовали взятия первой разности. Приводит ли такое *излишнее* взятие разностей к негативным последствиям? Другими сло-

вами, существуют ли негативные последствия, если мы примем ряд типа TS за ряд типа DS? Эта ситуация может встретиться при применении метода Бокса–Дженкинса всякий раз, когда мы переоценили величину параметра *d*.

Рассмотрим опасность излишнего взятия разностей на конкретном примере. Пусть рассматривается процесс MA(1), стационарный процесс, имеющий вид: $x_t = \varepsilon_t + \alpha\varepsilon_{t-1} = (1+\alpha L)\varepsilon_t$. Если, не зная, что он уже стационарный, взять от процесса первую разность, то получим:

$$y_t = \Delta x_t = (1-L)x_t = (1-L)(1+\alpha L)\varepsilon_t = (1-(1-\alpha)L-\alpha L^2)\varepsilon_t.$$

Получившаяся модель теперь относится к типу MA(2), оценивать нужно уже два параметра, а не один, но сейчас более важен другой аспект. Оценим обычным способом коэффициенты модели $y_t = u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2}$. Посмотрим, какова остаточная дисперсия процессов x_t и y_t. В первом случае дисперсия будет равна: $Var(x_t) = (1+\alpha^2)\sigma_\varepsilon^2$. Во втором случае, поскольку взята излишняя разность, один из корней соответствующего характеристического уравнения равен единице. Это означает, что получившийся процесс необратим, для него не существует AR(∞) представления, и ясно, что это само по себе приведет к сложностям в оценивании. Дисперсия этого процесса будет равна $Var(y_t) = (1+\theta_1^2+\theta_2^2)\sigma_u^2$. Или можно записать по-другому: $Var(y_t) = [1+(1-\alpha)^2+\alpha^2]\sigma_\varepsilon^2$. Невооруженным глазом видно, что $Var(y_t) > Var(x_t)$, т.е. взятие излишней разности приводит к увеличению дисперсии, в том числе и остаточной дисперсии после применения метода наименьших квадратов. Мы позже рассмотрим пример, в котором численно убедимся, что это увеличение действительно имеет место. 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.

Таким образом, неверное установление типа процесса ведет к негативным последствиям при ошибке «в обе стороны».

Теперь вернемся к *unit root test*. Специфицируем процесс типа TS в виде: $x_t = \alpha + \beta t + \varepsilon_t$. Попробуем выписать такую модель, в которой различие между типами TS и DS будет выражаться в значениях коэффициентов. Для этого заменим в модели белый шум ε_t на процесс u_t, подчиняющийся марковской схеме первого порядка $u_t = \rho u_{t-1} + \varepsilon_t$. Тогда нулевая гипотеза, заключающаяся в том, что ряд принадлежит типу DS, эквивалентна тому, что $\rho = 1$, а альтернативная – тому, что $\rho < 1$.

Чтобы убедиться в этом, запишем: $x_{t-1} = \alpha + \beta(t-1) + u_{t-1}$. Теперь умножим левую и правую части на ρ, что дает следующее выражение:

$$\rho x_{t-1} = \rho\alpha + \rho\beta(t-1) + \rho u_{t-1}.$$

Вычитая, получаем: $x_t = [\alpha(1-\rho)+\rho\beta]+\beta(1-\rho)t+\rho x_{t-1}+\varepsilon_t$. Это соотношение включает свободный член, линейный тренд, авторегрессионый член и белый шум. При выполнении нулевой гипотезы $\rho = 1$ процесс принимает вид:

$$x_t = \beta + x_{t-1} + \varepsilon_t .$$

Мы получили случайное блуждание с дрейфом, если $\beta \neq 0$. Если же $\rho < 1$, то в модели присутствует тренд и авторегрессионый 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 ε_t стоит стационарный процесс типа ARMA.

В соответствии с полученным соотношением принято рассматривать следующую модель для проверки наличия единичного корня: $x_t = \mu + bt + \rho x_{t-1} + \varepsilon_t$. Если оценить эту модель методом наименьших квадратов, то нужно проверить сформулированную нулевую гипотезу: $H_0 : \rho = 1$. Если бы при нулевой гипотезе ряд остался стационарным, то при предположении о нормальности ε_t надо просто проверить равенство коэффициента b единице, для чего сравнить t-статистику для этого коэффициента с квантилями t-распределения Стьюдента. Чтобы еще ближе свести процедуру к привычной, вычтем из обеих частей x_{t-1}. Тогда получается уравнение $\Delta x_t = \mu + bt + \underbrace{(\rho - 1)}_{\gamma} x_{t-1} + \varepsilon_t$. Получаем, что надо проверить следующую нулевую гипотезу против альтернативной:

$$H_0 : DS \Leftrightarrow \gamma = 0$$
$$H_1 : TS \Leftrightarrow \gamma < 0 \ .$$

Поскольку $\rho > 1$ соответствует нестационарному процессу взрывного типа, который исключен из рассмотрения, используется односторонний тест. При обычной регрессии для проверки этих гипотез применяется t-отношение, т.е. $t = \frac{\hat{\gamma}}{\hat{\sigma}_\gamma}$, 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 гипотезы такая же, только мы должны сравнивать статистику с квантилями не t-распределения, а DF-распределения. И решающее правило тоже чрезвычайно простое. Если $|\tau| > \tau_{кр}$, 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 лежат правее для одного и того же уровня значимости.

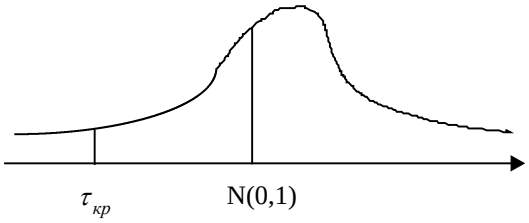

Рис. 8.2.

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 $x_t = \alpha + \beta t + u_t$, 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 $y_t = \alpha + u_t$, 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 $\rho = 1$, 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 α 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 таблицы распределения. Существует несколько систем обозначений, мы будем пользоваться такими: $\tau_0, \tau_\mu, \tau_\tau$.

При практическом применении теста Дикки–Фуллера возникает вопрос, каким образом выбрать вид модели, или, что тоже 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.

Для простоты выкладок рассмотрим процесс $x_t = u_t$, опустив без потери общности и тренд, и свободный член для уменьшения громоздкости выкладок. Пусть u_t подчиняется авторегрессионной схеме второго порядка:

$$u_t = \rho_1 \cdot u_{t-1} + \rho_2 \cdot u_{t-2} + \varepsilon_t .$$

Если мы хотим исключить u_t, то должны выписать $\rho_1 \cdot x_{t-1} = \rho_1 \cdot u_{t-1}$ и $\rho_2 \cdot x_{t-2} = \rho_2 u_{t-2}$ и затем вычесть эти выражения из исходного. В результате получим: $x_t = \rho_1 x_{t-1} + \rho_2 x_{t-2} + \varepsilon_t$. Мы видим, что сделанное предположение о виде авторегрессионной зависимости случайного процесса u_t эквивалентно тому, что процесс x_t также подчиняется схеме AR(2).

Прежде всего, нужно выразить в терминах коэффициентов этого уравнения ρ_1 и ρ_2 нулевую гипотезу о том, что ряд относится к типу DS, т.е. имеет единичный корень. Перепишем уравнение в операторном виде: $(1-\alpha_1-\alpha_2 L^2)x_t = \varepsilon_t$. Что значит, что уравнение имеет единичный корень? Если подставить L=1, получается $1-\alpha_1-\alpha_2 = 0$. Следовательно, нулевая гипотеза принимает вид:

$$\rho_1 + \rho_2 = 1.$$

Далее вычтем из обеих частей x_{t-1}, получаем:

$$\Delta x_t = (\rho_1 - 1)x_{t-1} + \rho_2 x_{t-2} + \varepsilon_t .$$

Добавим и вычтем $\rho_2 x_{t-1}$ в правой части. Получаем:

$$\Delta x_t = (\rho_1 + \rho_2 - 1)x_{t-1} - \rho_2 \Delta x_{t-1} + \varepsilon_t .$$

Теперь нулевая гипотеза о том, что ряд принадлежит типу DS или что у него есть единичный корень, сводится к проверке равенства нулю коэффициента $(\rho_1 + \rho_2 - 1) = 0$. Это соотношение позволяет переформулировать нулевую гипотезу $H_0 : DS \Leftrightarrow \rho_1 + \rho_2 - 1 = 0$. То есть если построить регрессию Δx_t на x_{t-1} и на приращение того же 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 Δx_{t-1}. 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 Δx_{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:

$$\Delta x_t = \gamma x_{t-1} + \sum_{i=1}^{p} w_i \Delta x_{t-i} + \sum_{i=1}^{q} \delta_i \varepsilon_{t-i} .$$

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:

$$\Delta x_t = \alpha + \beta t + \gamma x_{t-1} + \sum_{i=1}^{p} w_i \Delta x_{t-i} + \sum_{i=1}^{q} \delta_i \varepsilon_{t-i} .$$

Нулевая гипотеза о том, что процесс относится к типу DS, по смыслу эквивалентна тому, что $\gamma = 0$. А альтернативная гипотеза, что ряд относится к типу TS, означает, что $\gamma < 0$.

Ключевой вопрос: изменило ли добавление регрессоров распределение t-отношения для коэффициента γ? 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 $\Delta x_t = \alpha + \beta t + \gamma x_{t-1} + \sum_{i=1}^{p} w_i \Delta x_{t-i} + \sum_{i=1}^{q} \delta_i \varepsilon_{t-i}$ присутствуют и свободный член, и тренд, то нулевую гипотезу надо проверять, используя статистику τ_τ, если – только свободный член α, то статистику τ_μ, если нет ни того, ни другого, то надо использовать статистику τ_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 $\tau_0, \tau_\mu, \tau_\tau$ «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 ε_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 длиной реализации ряда, распределение t-отношения для коэффициента γ 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 [16]: число лагов выбирается равным $\left[T^{1/3}\right]$, где [x] означает целую часть числа x. Для квартальных данных было предложено [17] выбирать число лагов по формуле: $\left[4\left(T/100\right)^{1/4}\right]$, и для месячных данных – $\left[12\left(T/100\right)^{1/4}\right]$. Наконец, Диболд и Нерлов показали [6], что на практике хорошо работает приближение $\left[T^{1/4}\right]$.

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 выбор: 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. К сожалению, мне не удалось выяснить, как именно он это выбирает.

В литературе [8] 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) $\rho = 0{,}998$, 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 укажете, или по умолчанию применяет встроенный алгоритм.

ADF-тест допускает возмущение вокруг тренда или в отсутствии тренда в виде общего процесса ARMA(p,q). Вторая возможность, которую стоит рассмотреть, это возможность наличия гетероскедастичности случайной составляющей ε_t. Наличие гетероскедастичности или волатильность более характерны для финансовых рядов. Филлипс и Перрон предложили [14] так называемый *непара-*

метрический тест Филлипса–Перрона (PP-test). Правда, практика его применения показала его чрезвычайно малую мощность, так что в настоящее время он почти не применяется, хотя и включен в состав ряда специализированных компьютерных пакетов, в частности в Econometric Views. Тест Филлипса–Перрона использует уравнение без приращений, т.е. в виде: $x_t = \mu + bt + \rho x_{t-1} + \varepsilon_t$. Но в качестве случайного возмущения допускается любой стационарный, но может быть гетероскедастичный и автокоррелированный процесс.

Для учета более общей структуры случайного возмущения в тесте Филлипса–Перрона рассчитывается следующая статистика:

$$Z(\tau_\mu) = \tau_\mu\left(\hat{\sigma} \Big/ \hat{\sigma}_{\tau l}\right) - \frac{1}{2}\left(\hat{\sigma}_{\tau l}^2 - \hat{\sigma}^2\right)\cdot T \cdot \left\{\hat{\sigma}_{\tau l}^2 \sum_{l=2}^{T}(x_{l-1} - \bar{x}_{-1})^2\right\}^{-1/2}.$$

Здесь, $\hat{\sigma}^2$ – оценка дисперсии остатков от регрессии, $\bar{x}_{-1} = \frac{1}{T-1}\sum_{t=1}^{T-1} x_t$;

$\hat{\sigma}_{\tau l}^2 = \frac{1}{T}\sum_{l=1}^{T} e_l^2 + \frac{2}{T}\sum_{j=1}^{l} w_{jl}\cdot\sum_{l=j+1}^{T} e_l e_{l-j}$. Веса $w_{jl} = 1 - \frac{j}{l+1}$ здесь подобраны так, чтобы обеспечивать положительность оценки дисперсии $\hat{\sigma}_{\tau l}^2$. Параметр l можно выбирать, используя вышеприведенные рекомендации Шверта. Интересно, что статистика *Z* подчиняется тому же распределению Дикки–Фуллера, и мы можем воспользоваться теми же самыми таблицами Мак–Киннона.

Лекция № 9

Продолжим исследование наличия единичных корней или порядка интеграции временного ряда. Следует сказать, что, несмотря на наличие ряда тестов, в частности теста ADF, процедура исследования не так проста. Еще раз аккуратно выпишем, какие гипотезы проходят проверку.

Ряд относится к типу $DS \Rightarrow H_0 : x_t = x_{t-1} + \varepsilon_t$ – случайное блуждание без дрейфа, нестационарный ряд типа I(1);

Ряд относится к типу $TS \Rightarrow H_1 : x_t = \rho x_{t-1} + \varepsilon_t$; $|\rho| < 1$ – стационарный ряд, у него даже нет никакого тренда, это самый простейший вид.

Если $x_t = x_{t-1} + \varepsilon_t$, то общее решение разностного уравнения будет иметь следующий вид: $x_t = x_0 + \sum_{i=1}^{t} \varepsilon_i$.

У нас может быть и более сложная ситуация, когда к ряду типа DS относится ряд вида: $H_0 : x_t = \mu + x_{t-1} + \varepsilon_t$, случайное блуждание с дрейфом. Для него общее решение будет иметь вид: $x_t = x_0 + \mu t + \sum_{i=1}^{t} \varepsilon_i$. Стоит отметить, что процес-

сы $x_t = \rho x_{t-1} + \varepsilon_t$ и $x_t = \mu + x_{t-1} + \varepsilon_t$ ведут себя очень по-разному и не являются реальными альтернативами друг другу. В самом деле, нулевая гипотеза соответствует нестационарному процессу с трендом, а альтернативная – стационарному процессу с нулевым математическим ожиданием и без тренда. Более подходящая альтернативная гипотеза имеет вид: $H_1 : x_t = \alpha + \beta t + \varepsilon_t$, также содержащий линейный тренд. В результате мы получали ADF-тест в следующем виде: $\Delta x_t = \alpha + \beta t + (\rho - 1)x_{t-1} + \sum_{i=1}^{p} \delta_i \Delta x_{t-i} + \varepsilon_t$. Построение такой регрессии – это и есть проверка ADF-теста. При этом статистика, которую мы используем для проверки гипотезы, – это стандартное t-отношение для коэффициента $(\rho - 1)$, т.е. $t = \dfrac{\hat{\rho} - 1}{\hat{\sigma}_\rho} \sim \tau_\tau$. Эта статистика имеет распределение, таблицы которого рассчитаны Мак-Кинноном. Дополнительный регрессор βt меняет предельное распределение, которое первоначально было в тесте DF. Это распределение выражается через стохастические интегралы немного по-другому, и его критические значения получили обозначение τ_τ.

Это означает, что статистика будет разной в зависимости от того, равно β нулю или $\beta \neq 0$. Более того, мы знаем, что если и $\alpha = 0$, то распределение вновь меняется. Если мы строили уравнение при $\beta = 0$, т.е. не включали добавочный трендовый регрессор, критические значения обозначались τ_μ. И, наконец, если одновременно и $\alpha = 0$, и $\beta = 0$, т.е. мы не включаем их в качестве регрессоров в оцениваемую модель, то тогда соответствующие критические значения получили обозначение τ_0. Это просто одна из возможных систем обозначений.

Сложившаяся ситуация отличается от обычной линейной множественной регрессии. Там добавление дополнительного регрессора не меняет распределение используемой t-статистики (лишь меняется число степеней свободы). Теперь это не так. Мы заранее не знаем, какие регрессоры на самом деле включать в модель, а какие нет, и оказываемся в дополнительно неопределенной ситуации. Нужна некоторая методика применения ADF-теста.

Рассмотрим процедуру, предложенную Доладо, Дженкинсоном и Сосвилла-Риверо [7]. Назначение ее все то же, мы хотим разделить ряды типа TS и DS, но не знаем, надо включать тренды в модель или не надо. Напомним, что вопрос заключается в том, как приводить ряд к стационарному виду: с помощью построения регрессии с линейным трендом или с помощью перехода к первым разностям. Процедура состоит из следующих шагов.

1. Оцениваем ADF уравнение вида:

$$\Delta x_t = \alpha + \beta t + (\rho - 1)x_{t-1} + \sum_{i=1}^{p} \delta_i \Delta x_{t-i} + \varepsilon_t,$$

т.е. включаем в него и свободный член, и линейный тренд. При этом нулевая гипотеза имеет вид $H_0 : DS \Rightarrow \rho = 1$. Если $\rho = 1$, то ряд фактически относится к

новому типу, который мы почти не исследовали, в нем наличествуют два типа тренда. Пусть для простоты p=0, тогда получаем процесс вида $x_t = \alpha + \beta \cdot t + x_{t-1} + \varepsilon_t$. Он на самом деле содержит два тренда. Будем называть линейный тренд, соответствующий ряду типа TS, *детерминированным трендом*, а случайное блуждание с дрейфом – *стохастическим трендом*, т.е. случайно меняющимся трендом. Ряд нового типа содержит оба тренда: и стохастический, и детерминированный. Общее решение такого уравнения имеет вид:

$$x_t = x_0 + \sum(\alpha + \beta \cdot t + \varepsilon_t) = x_0 + \alpha \cdot t + \beta \cdot \frac{t(t+1)}{2} + \sum \varepsilon_t .$$

Таким образом, процесс содержит уже квадратичный тренд. Напомним, что $\sum t = \frac{t(t+1)}{2}$, это просто арифметическая прогрессия. Сравнение величины t-отношения для коэффициента при x_{t-1} с критическим значением распределения Дикки–Фуллера может привести к двум возможностям. И выводы зависят от того, попадаем ли мы в критическую область отвержения нулевой гипотезы или не попадаем. Если мы попадаем в критическую область, т.е. t-отношение (с учетом знака) удовлетворяет неравенству $t < \tau_\tau$, тогда наш ряд не имеет единичного корня, т.е. является типа TS. Если же $t > \tau_\tau$, то прежде чем сделать вывод о типе ряда, нужно ответить на вопрос, правильно ли мы включили в наше уравнение линейный тренд βt. Для этого предположим, что нулевая гипотеза выполнена и инкорпорируем эту гипотезу в исходное уравнение, получая модель без регрессора x_{t-1}: $\Delta x_t = \alpha + \beta t + \sum_{i=1}^{p} \delta_i \Delta x_{t-i} + \varepsilon_t$.

2. Оценим уравнение $\Delta x_t = \alpha + \beta t + \sum_{i=1}^{p} \delta_i \Delta x_{t-i} + \varepsilon_t$. Это уравнение заведомо не имеет единичных корней. Поэтому вопрос о том, значим или не значим статистически коэффициент β, наклон линейного тренда, может быть решен стандартной t-статистикой для коэффициента при тренде. Если коэффициент β не значим, т.е. $\beta = 0$, то делаем вывод, что мы зря включили трендовую переменную в уравнение ADF-теста. Если же коэффициент β значим, то наш вывод на первом шаге о том, что ряд принадлежит типу DS, получил подтверждение, и процедура окончена. Если же коэффициент β не значим, то надо перейти к следующему шагу.

3. Поскольку правомерность включения линейного тренда в спецификацию ADF-теста не подтвердилась, переходим к ADF-тесту без тренда, т.е. оцениваем регрессию $\Delta x_t = \alpha + (\rho - 1)x_{t-1} + \sum_{i=1}^{p} \delta_i \Delta x_{t-i} + \varepsilon_t$. Рассчитываем опять t-статистику для коэффициента при регрессоре x_{t-1}, но теперь ее нужно сравнивать с τ_μ, по-

тому что анализируется уже другое уравнение. Если $t < \tau_\mu$, то наш вывод состоит в том, что нет единичного корня и нет линейного тренда. Если $t > \tau_\tau$, то мы начинаем проверять правомерность включения свободного члена α в спецификацию ADF-теста. Строим регрессию $\Delta x_t = \alpha + \sum_{i=1}^{p} \delta_i \Delta x_{t-i} + \varepsilon_t$ и проверяем значимость свободного члена по обычной t-статистике и распределению Стьюдента. Часто длина реализации составляет 70−80, тогда можно проверять гипотезу по нормальному распределению. Если мы считаем, что α не значимо, то не должны его включать в уравнение для ADF теста. В этом случае оцениваем регрессию $\Delta x_t = (\rho - 1) x_{t-1} + \sum_{i=1}^{p} \delta_i \Delta x_{t-i} + \varepsilon_t$. Для проверки нулевой гипотезы о стационарности используем критические значения τ_0.

Проиллюстрируем применение этой процедуры примером ранее рассматриваемого нами ряда индекса деловой активности для Великобритании (UK FTA All Share) [9]. Количество лагов было выбрано равным трем.

1. Методом наименьших квадратов была оценена следующая модель:

$$\underset{t}{\Delta x_t} = \underset{(2,31)}{0{,}124} + \underset{(2,29)}{0{,}00025} \cdot t - \underset{(-2,27)}{0{,}0284} x_{t-1} + \sum_{i=1}^{3} \hat{\delta}_i \cdot \Delta x_{t-i} + e_t \text{ (в скобках − t-отношения),}$$

где t-отношение для коэффициента при x_{t-1} равно −2,27. Мы должны это число сравнить со значением τ_τ на уровне значимости 5%. Оно равно −3,49. Мы видим, что −2,27>−3,49. Поэтому мы не можем отвергнуть нулевую гипотезу, которая заключается в том, что ряд принадлежит к типу DS, т.е. что он имеет единичный корень.

2. На втором шаге оцениваем регрессию вида:

$$\underset{t}{\Delta x_t} = \underset{(-0,27)}{-0{,}0023} + \underset{(1,18)}{0{,}00007} t + \sum_{i=1}^{3} \hat{a}_i \cdot x_{t-i} + e_t,$$

т.е. из числа регрессоров убираем регрессор x_{t-1}. Мы сравниваем t-статистику коэффициента при линейном тренде с таблицами нормального распределения. Видно, что коэффициент не значим, следовательно, тренд был включен напрасно, т.е. его включение не согласуется с нашими реальными данными. Поэтому переходим к третьему шагу.

3. Оцениваем регрессию вида: $\underset{t}{\Delta x_t} = \underset{(0,63)}{0{,}0166} - \underset{(-0,38)}{0{,}00176} x_{t-1} + \sum_{i=1}^{3} \hat{b}_i \Delta x_{t-i} + e_t$.

Сравниваем t-отношение при регрессоре x_{t-1} с τ_μ. Поскольку t-отношение очень маленькое, мы не можем отвергнуть гипотезу о наличии единичного корня. Но, может быть, мы зря включили свободный член в нашу модель.

4. Оцениваем модель $\underset{t}{\Delta x_t} = \underset{(1,78)}{0,0067} + \sum_{i=1}^{3} c_i \Delta x_{t-i} + e_t$. Здесь t-статистика, равная 1,78, при сравнении с критической величиной стандартного нормального распределения оказывается значимой на 5-процентном уровне по одностороннему критерию. Раз коэффициент значим, то количество лагов выбрано правильно, эта модель правильно специфицирована.

Общий вывод применения процедуры Доладо, Дженкинсона и Сосвилла-Риверо заключается в том, что ряд содержит единичный корень, он относится к типу DS. Также по ходу ее применения выяснилось, что ряд не содержит линейного тренда.

Если бы на четвертом шаге мы проверяли нулевую гипотезу по двустороннему критерию, то тогда бы должны были сказать, что свободный член не значим, но, скорее всего, пришли бы к тому же самому выводу о наличии единичного корня. Но спецификация модели не содержала бы свободный член.

Мы видим, что установление типа нестационарности ряда не сводится к однократному применению соответствующего теста. Требуется тщательное исследование правильности спецификации тестовой модели. Описанная методика исследования наличия единичного корня не является единственной. Другие подходы можно найти в монографии Паттерсона [12].

У нас осталось еще два момента, связанных с различением между рядами типа DS и TS. Первый вопрос поставил Перрон. Правильно ли то, что мы сравниваем только такие простые модели, у которых детерминированный тренд на всем временном промежутке один и тот же, в то время как стохастический тренд меняется от точки к точке. В качестве более общей альтернативы Перрон предложил рассмотреть следующую модель. У процесса есть линейный тренд, но он может структурно меняться под влиянием каких-то шоков, т.е. детерминированный тренд претерпевает структурные изменения. Ряд с таким трендом конечно нестационарный, но он разбивается на два временных участка, на каждом из которых ряд стационарен или, по крайней мере, относится к типу TS. Но происходит изменение параметра, изменение структуры модели. Имеет смысл рассмотреть такую сегментированную модель в качестве альтернативы стохастическому тренду. При этом ясно, что в некоторых случаях момент такого структурного скачка известен заранее, в других случаях – нет, и что таких «переломов» может быть, вообще говоря, несколько.

И еще один момент. Мы говорили, что ищем единичный корень, но на самом деле пытались различить процессы типа I(0) и I(1). Если процесс содержит несколько единичных корней, то такая возможность пока не рассматривалась. А ведь такой случай не исключен. Мы уже рассматривали процессы ARIMA(p,d,q) и говорили, что параметр d может быть любым числом. Хотя в экономических задачах параметр d больше чем 2 практически не встречается. Но, по крайне мере, процесс может содержать два единичных корня. В этом случае появляется третья возможность: ряд относится к типу I(2). Классическая процедура проверки статистических гипотез проводится таким образом, что третья возможность фактически «объединяется» с альтернативной гипотезой. Это ведет к увеличению ошибки второго рода и потере эффективности статистического теста. Рассмотрим эту проблему подробнее.

В наиболее общем виде ADF-тест имеет вид:

$$\Delta x_t = \alpha + \beta t + (\rho - 1)x_{t-1} + \sum_{i=1}^{p} \delta_i \Delta x_{t-i} + \varepsilon_t .$$

В качестве гипотез мы фактически рассматривали следующие:

$$H_0 : d = 1;$$
$$H_1 : d = 0 .$$

В классической процедуре проверки гипотез ищется критическое множество при фиксированной величине ошибки 1-ого рода $P\{x \in S | H_0\} = \alpha$. Критическое множество строится, чтобы разделить две гипотезы: основную и альтернативную. Есть два случая. Первый, когда мы точно знаем, что либо d=0, либо d=1. То есть у нас может быть максимально один единичный корень. И вторая ситуация, когда у нас может быть ни одного, один или два единичных корня. Что из общих соображений должно произойти в этом случае? В первом случае, если не верна нулевая гипотеза, мы точно знаем, что d=0. А во втором случае, если не верна нулевая гипотеза, то либо d=0, либо d=2. Поэтому критические значения должны измениться, у нас будет другая структура критического множества. Фактически, когда строился ADF-тест, мы проверяли несколько иную гипотезу, что параметр d не больше единицы. А альтернативная гипотеза, что нет ни одного единичного корня.

Предположим, мы хотим проверить, имеет ли процесс x_t два единичных корня. Естественным путем представляется двукратное применение ADF-теста. Если нулевая гипотеза ADF-теста не отвергается, надо рассмотреть ряд Δx_t и исследовать его на наличие единичного корня. Но возникает вопрос: может быть, процесс содержит не два, а три или четыре единичных корня? У нас есть два конкурирующих разумных подхода. Первый заключается в продолжении выше намеченной процедуры. Если обнаруживается наличие единичного корня, то тогда берем первую разность Δx_t и исследуем, нет ли у преобразованного процесса единичного корня. Если есть, то переходим ко второй разности и вновь проверяем наличие единичного корня у преобразованного ряда. И так далее, пока гипотеза о наличии единичного корня не будет отвергнута. При этом подходе мы последовательно проверяем гипотезы о наличии все большего числа единичных корней.

Второй разумный подход состоит в проверке гипотез в обратной последовательности. Сначала выдвигается предположение о максимально возможном числе единичных корней, например равном трем. В этом случае берем вторую разность процесса и проверяем преобразованный ряд на наличие единичного корня. Если у преобразованного ряда нет единичного корня, это означает, что возможность наличия трех единичных корней у исходного процесса исключена. Если нет трех единичных корней, переходим к проверке на наличие двух единичных корней, для чего проверяем наличие единичного корня у первой разности исходного ряда. Этот подход характеризуется последовательной проверкой гипотез о наличии все меньшего числа единичных корней. На каждом шаге основная и альтернативная гипотезы формулируются следующим образом:

$$H_0 : \max \quad k \quad \textit{единичных корней};$$
$$H_1 : \textit{меньше чем} \quad k \quad \textit{единичных корней}.$$

Диккки и Пантула [5] показали, используя моделирование по методу Монте-Карло, что второй подход эффективнее.

Схема его применения для случая, например, трех возможных единичных корней такова.

Применяем обычный ADF-тест для второй разности исходного процесса, т.е. оцениваем регрессию вида $\Delta^2 x_t = \alpha + \beta t + (\rho - 1)\Delta x_{t-1} + \sum_{i=1}^{p} \delta_i \Delta^2 x_{t-i} + \varepsilon_t$. При этом нулевая и альтернативная гипотезы в терминах исходного процесса выглядят следующим образом.

$$H_0 : 2 \textit{ корня};$$
$$H_1 : 1 \textit{ корень}.$$

Стандартным образом проверяем t-статистику, сравнивая ее с τ_τ, и так далее. Если процедура Доладо привела к выводу, что процесс содержит три единичных корня, то процедура окончена. Если же гипотеза о наличии трех единичных корней отвергается, применяем процедуру Доладо к первой разности исходного процесса, что эквивалентно сравнению гипотез о наличии двух и одного единичного корня. Если гипотеза о наличии двух единичных корней отвергается, то переходим к ADF-тесту для исходного ряда.

Лекция № 10

Вернемся к подходу Перрона, упомянутому в предыдущей лекции. Все альтернативы, которые мы до сих пор рассматривали при различении TS и DS моделей, исходили из того, что параметры, которыми модели характеризуются, в частности свободный член и коэффициент при линейном тренде, сохраняются неизменными все время наших наблюдений. Однако исторические макроэкономические ряды подсказывают, что может наблюдаться и другая ситуация. Тренд или свободный член могут скачком изменить свое численное значение, реагируя на так называемые внешние шоки. Такими внешними шоками являются, например, великая депрессия 1929 г. в США, нефтяной шок 1973 г., а теперь и российский дефолт августа 1998 г. Модели с такими скачкообразными изменениями параметров называют моделями со структурными скачками (*structural breaks*). Поскольку стандартный ADF-тест не допускает возможности таких структурных скачков, существует опасность, что скачки будут трактоваться как нестационарность, и нулевая гипотеза не будет отвергнута.

Нам хотелось бы расширить наше рассмотрение на естественные классы моделей, которые содержат скачкообразные изменения параметров. При этом время структурного скачка может быть известным (экзогенный скачок) или неизвестным (эндогенный скачок). Начнем с простой ситуации, когда изменение происходит в некоторый известный момент. В 1989 г. Перрон [13] сумел распространить схему

исследования типа ряда (TS или DS) на случай структурных изменений. Если модель типа TS (альтернативная в ADF-тесте) включает свободный член и линейный тренд, то внешний шок может вызвать три типа структурных изменений. В некоторый момент времени может произойти скачок уровня, и линия тренда пойдет дальше с тем же наклоном, т.е. с тем же темпом роста. Может произойти скачок наклона линии тренда без изменения уровня ряда. А может произойти ситуация, когда изменился наклон и одновременно произошел сдвиг. При этом следует исследовать уже кусочно-линейные модели. Вообще говоря, эта модель не является линейной, но на каждом из двух интервалов она линейна.

Перрон специфицировал модели, которые соответствуют такому интуитивному пониманию трех указанных случаев для случая ряда типа TS. Для каждого такого случая он выписал также альтернативную модель типа DS с качественно похожим поведением. Напомню, что в тесте Дикки–Фуллера тип DS соответствует нулевой гипотезе, а тип TS соответствует альтернативной гипотезе:

$$H_0 : DS$$
$$H_1 : TS \ .$$

Перрон рассмотрел 3 типа моделей: A, B и C, которые соответствуют трем различным типам структурных изменений. Сведем эти модели в таблицу.

	$H_1 : TS$	$H_0 : DS$
A	$x_t = \mu_1 + \beta \cdot t + (\mu_2 - \mu_1) \cdot (DU)_t + \varepsilon_t$	$x_t = \mu + x_{t-1} + d \cdot (t_B)_t + \varepsilon_t$
B	$x_t = \mu + \beta_1 \cdot t + (\beta_2 - \beta_1)(DT^*)_t + \varepsilon_t$	$x_t = \mu_1 + x_{t-1} + (\mu_2 - \mu_1)(DU)_t + \varepsilon_t$
C	$x_t = \mu_1 + \beta_1 \cdot t + (\mu_2 - \mu_1)(DU)_t + (\beta_2 - \beta_1)(DT^*)_t + \varepsilon_t$	$x_t = \mu_1 + x_{t-1} + d \cdot D(t_B)_t + (\mu_2 - \mu_1) \cdot (DU)_t + \varepsilon_t$

Первая модель (тип A) – это случай, когда происходит изменение свободного члена скачком в некоторый момент времени t_B+1. Следуя Перрону, введем следующие обозначения. $(DU)_t$ – это единый символ, который является обычной dummy-переменной, определенной как: $(DU)_t = \begin{cases} 1, & t > t_B, \\ 0, & t \le t_B \end{cases}$. Эта dummy-переменная несколько непривычно определена, знак «строго больше» означает скачок в момент t_B+1. Причем, видно, что до скачка в модели TS «работает» свободный член уровня μ_1, а после скачка «работает» уровень μ_2. Это совершенно стандартная dummy-переменная.

Для этого же случая (тип A) нулевую модель DS можно написать в следующем виде: $x_t = \mu + x_{t-1} + d \cdot (t_B)_t + \varepsilon_t$, где d – это некий коэффициент, и введена еще одна dummy-переменная $D(t_B)_t = \begin{cases} 1, & t_B + 1 \\ 0, & \textit{в остальных случаях} \end{cases}$. То есть альтернативная гипотеза – это случайное блуждание с дрейфом, как и положено. Дрейф в этой модели равен μ во все моменты времени, кроме момента струк-

турного скачка. Обратите внимание, что $\mu \neq \mu_1$, это разные параметры. Эти модели, обе типа A, описывают одно и то же явление, но по-разному. Одна при детерминированном скачке в случайном члене, а другая при стохастическом. Dummy-переменная обращается в единицу для одного единственного наблюдения. Если описывать «черный вторник» в России, когда курс доллара к рублю скачком возрос, а на следующий день вернулся к практически прежнему значению, то пригодится как раз такая переменная.

Вторая модель (тип B) – это излом наклона линейного тренда. Ее детерминированный вариант нам знаком: $x_t = \mu + \beta_1 \cdot t + (\beta_2 - \beta_1)(DT^*)_t + \varepsilon_t$.

Причем новая dummy-переменная определяется следующим образом:

$$(DT^*)_t = \begin{cases} t - t_B, & t > t_B \\ 0, & t \le t_B \end{cases}.$$

Это обычная смешанная dummy-переменная, тренд с началом отсчета в момент t_B, умноженный на ранее введенную dummy-переменную:

$$(t - t_B)(DU)_t = (DT^*)_t.$$

Просто Перрон ввел для нее специальное обозначение.

Модель типа B для ряда DS означает, что параметр дрейфа изменился и остался измененным, а в модели типа A он изменился на один период и вернулся к старому значению. Для ряда DS модель типа B описывает случайное блуждание с разными дрейфами $x_t = \mu_1 + x_{t-1} + (\mu_2 - \mu_1)(DU)_t + \varepsilon_t$.

Третья модель (тип C) – это самый общий случай, когда происходят оба изменения, и уровня ряда как в модели типа A, и наклона тренда как в модели типа B.

Используя моделирование по методу Монте-Карло, Перрон показал, что если данные сгенерированы одной из моделей из левого столбца, т.е. являются процессом типа TS, то МНК-оценка коэффициента $\hat{\gamma}$ в регрессии $x_t = \mu_1 + \gamma x_{t-1} + \beta t + \varepsilon_t$ все сильнее смещена к единице по мере увеличения величины скачка. Это означает, что ADF-тест будет ошибочно диагностировать наличие единичного корня, подтверждая ранее высказанное опасение. Поскольку смещение не исчезает при увеличении длины ряда, говорят, что ADF-тест несостоятелен при структурном скачке тренда. Для модели типа A, когда присутствует только скачок свободного члена, ADF-тест, вообще говоря, состоятелен (по результатам Перрона), но имеет низкую мощность. Этот результат не является неожиданным, потому что в случае присутствия единичного корня добавка одного регрессора, как мы уже видели, часто меняет выборочную статистику, наверно и здесь так.

Перрон предложил обобщение стратегии тестирования на наличие единичного корня применительно к возможному присутствию структурных скачков. На первом шаге предложенной процедуры исходный ряд «очищается» от тренда со структурным скачком, для чего находятся остатки регрессии исходного ряда на:

а) константу, линейный тренд и переменную $(DU)_t$;

в) константу, линейный тренд и переменную $(DT^*)_t$;

с) константу, линейный тренд, переменную $(DU)_t$ и переменную $(DT^*)_t$, для вышеупомянутых моделей типа A, B и C соответственно. К остаткам от регрессии применяется тест Дикки–Фуллера (не расширенный), но критические точки распределения t-отношения не совпадают со значениями, рассчитанными Фуллером. Поэтому Перрон рассчитал их сам. Оказалось, что эти критические значения больше по абсолютной величине, чем соответствующие значения Дикки–Фуллера, и зависят не от длины реализации T, а от месторасположения момента скачка относительно длины ряда. Попробуйте включить интуитивное понимание. Представьте себе, что имеется длинная серия наблюдений и скачок произошел в самом ее начале, и второй случай, когда скачок произошел в середине. Как вы думаете, когда легче отличить, был скачок или нет? Ответ ясен: во втором случае. Фактически мы сравниваем две модели: на одном временном интервале и на другом. Если интервал маленький, различие как бы не успело набрать силу, нам будет сложно сравнивать. Поэтому критические значения теста Перрона зависят от соотношения $\left(\frac{t_B}{T}\right)$. При любом месте расположения скачка, как уже упоминалось, критические значения статистики Перрона по абсолютной величине больше, чем соответствующие для DF-распределения. И самые большие по модулю критические значения соответствуют $\left(\frac{t_B}{T}\right) = 0{,}5$. Например, для уровня значимости 5% критическое значение, полученное Перроном для модели типа A при $\left(\frac{t_B}{T}\right) = 0{,}5$, составило −3,76, а соответствующее значение статистики Дикки–Фуллера составляет −3,41.

Перрон применил свой подход к историческим макроэкономическим рядам США и выяснил, что допущение возможности структурных скачков изменяет результаты Нельсона и Плоссера, уже не все ряды относятся к типу DS.

Очевидным недостатком изложенного подхода является предположение, что в ответ на внешний шок происходит мгновенное изменение параметра (параметров) модели. Более реалистичным представляется постепенное изменение параметров как реакция на внешний шок. Другими словами, изменение имеет лаговую структуру, может быть, бесконечную. Реакцию первого типа в литературе принято называть аддитивным выбросом (additive outlier, AO), а реакцию второго типа – инновационным выбросом (innovation outlier, IO). В своей работе Перрон предложил для случая инновационного выброса оценить регрессию (для наиболее общего случая – типа C):

$$x_t = \alpha + \theta(DU)_t + \beta t + \delta(DT^*)_t + \gamma x_{t-1} + d \cdot D(t_B)_t + \sum_{i=1}^{k} c_i \Delta x_{t-i} + \varepsilon_t .$$

Нулевой гипотезе наличия единичного корня соответствуют значения параметров $\gamma = 1; \theta = \beta = \delta = 0$, альтернативной гипотезе – $\gamma < 1; \theta, \beta, \delta \neq 0$. Кроме того, при нулевой гипотезе параметр d не должен быть равен 0, в то время как при альтернативной гипотезе он должен быть близок к 0. Для типов B и C число регрессоров соответственно уменьшается. Чрезвычайно удобным для пользователя про-

цедуры Перрона является тот факт, что критические значения, рассчитанные Перроном, являются одними и теми же как для аддитивных, так и для инновационных выбросов.

Одним из обобщений подхода Перрона было распространение его на случай нескольких структурных скачков в разные моменты времени. Раппопорт и Рейхлин [15] рассмотрели случай двух скачков в тренде, а в работах Бая [1] и Бая с Перроном [2] предложенная методика распространена на случай нескольких структурных скачков. Ввиду громоздкости и необходимости набора таблиц новых критических значений большого применения это обобщение пока не нашло.

Слабым местом подхода Перрона является предположение о том, что время структурного скачка известно заранее. Зивот и Эндрюс [18] показали, что момент скачка не может быть предопределен заранее. Они предложили обобщение подхода Перрона на случай эндогенного скачка, т.е. структурного скачка, момент которого оценивается одновременно с оценкой коэффициентов модели. Зивот и Эндрюс полагают, что нет необходимости вводить структурный скачок в нулевую модель, и рассматривают большое по величине изменение уровня просто как осуществление редкого события, соответствующего «хвосту» распределения. Поэтому нулевая гипотеза в их подходе представляет собой случайное блуждание с дрейфом. Альтернативная модель имеет вид:

$$x_t = \alpha + \theta(DU)_t + \beta t + \delta(DT^*)_t(\lambda) + \gamma x_{t-1} + \sum_{i=1}^{k} c_i \Delta x_{t-i} + \varepsilon_t ,$$

сходный с подходом Перрона, но без dummy-переменной $D(t_B)_t$ и с дополнительным параметром λ, равным отношению неизвестного момента скачка к длине ряда. Дальнейшее напоминает поиск на решетке. Строятся альтернативные модели для всевозможных моментов времени скачка от t=2 до t=T−1 и выбирается та из них, для которой t-отношение коэффициента γ принимает наименьшее значение. Не забудьте, эта величина отрицательна.

Поскольку момент скачка оценивается процедурой последовательного перебора, а не известен заранее, критические значения, рассчитанные Перроном, больше не подходят. Зивот и Эндрюс, а также Банерджи с соавторами [3] рассчитали, используя метод Монте-Карло, критические значения для этого подхода. Они лежат еще левее критических значений Перрона. Так, например, 5-процентное критическое значение Перрона для заданного значения параметра λ=0,5 составляет −4,24, а соответствующее значение для оцененного значения параметра равно −5,08. Применив свой подход к все тем же данным Нельсона и Плоссера, Зивот и Эндрюс в четырех из десяти случаях установили в отличие от Перрона наличие единичного корня.

В качестве иллюстрации рассмотрим интересный пример применения этой методики (см. рис. 10.1). Рассмотрим сначала его участок с 1971 по 1988 гг. [9]. На этом участке ряд демонстрирует некоторое подобие наличия структурного скачка. Если применить к этому ряду обычный ADF-тест на наличие единичного корня, то $\tau_\tau = -1{,}84$, а соответствующее критическое значение на 10-процентном уровне значимости равно −3,15, так что мы не можем отвергнуть нулевую гипотезу о наличии единичного корня.

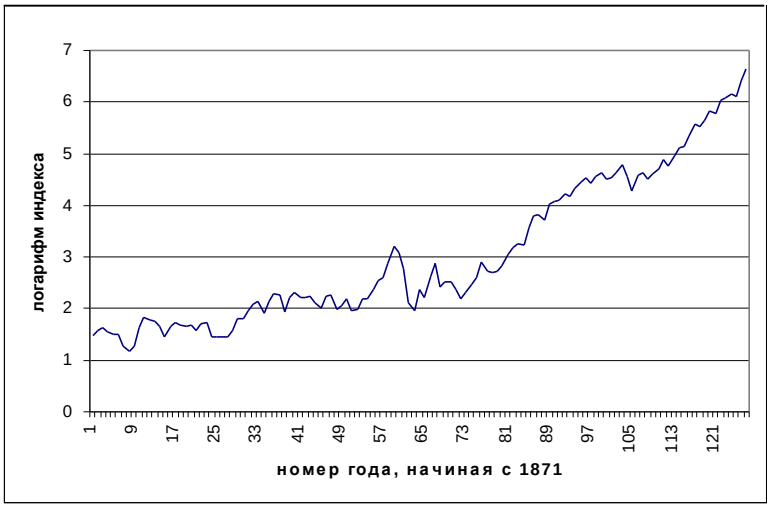

Рис. 10.1. Логарифм среднегодовых значений индекса S&P для США в 1871–1997 гг.[1)]

Следуя Перрону, попробуем включить в рассматриваемые гипотезы наличие структурного скачка. Естественно предположить его в 1929 г. (58-е значение ряда), то, что в Америке называют Great Crash, или великая депрессия. Отношение времени момента структурного скачка к общей длине ряда (118 точек) равно примерно 0,5. Оценка модели типа C дает (в скобках – значения стандартных ошибок):

$$\underset{s.e.}{x_t} = \underset{(0,117)}{0,570} + \underset{(0,0017)}{0,0070}\cdot t - \underset{(0,068)}{0,216}(DU)_t + \underset{(0,003)}{0,011}(DT^*)_t - \underset{(0,185)}{0,177}\,D(t_B)_t + \underset{(0,062)}{0,707}\,x_{t-1} + \underset{(0,092)}{0,156}\,\Delta x_{t-1}.$$

Нулевая гипотеза по-прежнему имеет вид: $H_0 : DS \Rightarrow \gamma = 0$. Уравнение ADF-теста записано не в стандартном виде. Чтобы привести его к стандартному виду, надо вычесть из оценки коэффициента единицу, поэтому получаем: $t = \dfrac{0,707-1}{0,062} = -4,74$.

Пятипроцентное критическое значение из таблиц Перрона для $\lambda = \dfrac{t_B}{T} = 1/2$ равно −4,24. Поэтому нулевую гипотезу о наличии единичного корня нужно отвергнуть на этом уровне значимости.

Поскольку все коэффициенты, кроме коэффициента при dummy-переменной $D(t_B)_t$, статистически значимы, пересчитаем модель, исключив эту переменную. Построенная модель содержит авторегрессию второго порядка. Как мы уже видели раньше, можно представить модель в эквивалентном виде без авторегрессии, но с автокорреляцией второго порядка у случайной составляющей. Такое представление в этом случае является более удобным для интерпретации. В результате, оценивая, получаем:

[1)] Источник данных: файл S&P500 на сайте
http://www.lboro.ac.uk/departments/ec/cup/

$$\underset{s.e.}{x} = \underset{(0,085)}{1,859} + \underset{(0,0047)}{0,0188}\cdot t + \underset{(0,007)}{0,0366}(DT^*)_t - \underset{(0,152)}{0,312}(DU)_t + e_t.$$

Наши остатки стали коррелированными, т.е. $e_t = \underset{(0,093)}{0,932}\, e_{t-1} - \underset{(0,094)}{0,200}\, e_{t-2} + u_t$. Причем $\hat{\sigma}_u = 0,1678$. Что можно заключить по этой модели? До структурного скачка наблюдается значимый линейный тренд, он соответствует приросту индекса на 1,88% в год. Что произошло в момент скачка? С 1930 г. темп роста изменился. Он вырос на 3,66%, и стал 5,55%. Изменение тренда одновременно сопровождалось скачком вниз на 0,312.

К этому же участку ряда была применена методика Зивота и Эндрюса, т.е. дополнительно оценивался момент структурного скачка. В этом случае скачок обнаружился совсем в другой точке, он соответствует 1950 г. И модель после определения времени скачка приняла вид: $\underset{s.e.}{x_t} = \underset{(0,064)}{1,871} + \underset{(0,0024)}{0,0193}t + \underset{(0,0069)}{0,0391}(DT^*)_t + \underset{(0,144)}{0,377}(DU)_t + u_t$.

Здесь u_t тоже подчиняется авторегрессионной схеме второго порядка, мы опускаем это уравнение. Согласно этой модели скачок произошел в 1950 г., и затем темп роста стал больше, чем 1,93%, а точнее возрос на 0,0391. Обратите внимание, что скачок в свободном члене тоже положительный.

Даже на этом примере видна важность учета структурных изменений: если не учитывать структурного скачка, исследуемый ряд относится к типу DS с единичным корнем. Учет структурного скачка позволил описать ряд процессом типа TS. Если искать по формальной процедуре период скачка, то мы нашли другое значение, причем его трудно интерпретировать. Если наложить на график обе модели, то расхождение между ними качественно заметно только между 1929 и 1950 гг. На промежутках до и после этого периода модели хорошо согласуются с рядом и между собой. Если мы занимаемся прогнозированием на ближайшее время, обе эти модели приведут примерно к одному и тому же, если же будем темп роста экстраполировать на длительный период, то, конечно, придем к разным результатам.

Интересно сравнить результаты аналогичного подхода применительно ко всему диапазону с 1871 по 1997 гг. [10]. Стандартный ADF-тест на наличие единичного корня теперь дает величину $\tau_\tau = -1,15$, критическое значение не изменилось, так что по-прежнему мы не можем отвергнуть нулевую гипотезу о наличии единичного корня.

Применив подход Перрона со структурным скачком в 1929 г., получаем следующую модель (в скобках – значения стандартных ошибок):

$$\underset{s.e.}{x_t} = \underset{(0,089)}{0,334} + \underset{(0,0017)}{0,0066}\cdot t - \underset{(0,065)}{0,235}(DU)_t + \underset{(0,003)}{0,012}(DT^*)_t + \underset{(0,181)}{0,184}D(t_B)_t + \underset{(0,058)}{0,731}x_{t-1} + \underset{(0,088)}{0,128}\Delta x_{t-1}.$$

Видно, что коэффициенты модели (за исключением незначимого коэффициента при dummy-переменной $D(t_B)_t$) практически не изменились. Значение t-отношения теперь составило −4,65, и нулевая гипотеза вновь отвергается на 5-процентном уровне значимости.

Пересчитаем модель, исключив переменную $D(t_B)_t$ и перенеся лаговую структуру в остатки. В результате, оценивая, получаем:

$$\underset{s.e.}{x} = \underset{(0,206)}{1,335} + \underset{(0,0053)}{0,0175}\cdot t + \underset{(0,0074)}{0,0430}(DT^*)_t - \underset{(0,156)}{0,346}(DU)_t + e_t.$$

Остатки описываются моделью $e_t = \underset{(0,091)}{0,945}e_{t-1} - \underset{(0,092)}{0,177}e_{t-2} + u_t$. По-прежнему $\hat{\sigma}_u = 0,1678$. По сравнению с аналогичной моделью для 1871–1988 гг. увеличение тренда несколько больше: не на 3,66 %, а на 4,3%. Изменение тренда по-прежнему сопровождается скачком вниз, но теперь на 0,346. По графику можно понять, что причиной этих изменений в параметрах модели является более быстрый рост индекса в 1989–1997 гг. В целом подход Перрона дал сопоставимые результаты для обоих участков.

Совсем по-другому обстоит дело с методикой Зивота и Эндрюса. В этом случае скачок обнаружился в точке, соответствующей 1931 г., что весьма близко к великой депрессии. При этом нулевая модель о наличии единичного корня по-прежнему отвергается. Такое расхождение в оценках времени скачка при добавлении относительно небольшого числа точек ряда ставит интересные вопросы о доверительном интервале этой оценки и о робастности всей процедуры Зивота и Эндрюса, но они выходят за рамки нашего курса.

* *
*

СПИСОК ЛИТЕРАТУРЫ

1. *Bai J.* Estimating Multiple Breaks One at a Time // Econometric Theory. 1997. 13. P. 315–52.

2. *Bai J.* and *Perron P.* Estimating and Testing Linear Models with Multiple Structural Changes // Econometrica. 1998. 66. P. 47–78.

3. *Banerjee A., Lumsdaine R.L.* and *Stock J.H.* Recursive and Sequential Test of the Unit Root and Trend Break Hypothesis: Theory and International Evidence // Journal of Business and Economic Statistics. 1992. 10. P. 271–87.

4. *Dickey D.A.* and *Fuller W.A.* Distribution of the Estimators for Autoregressive Time-Series with a Unit Root // Journal of the American Statistical Assiciation. Vol. 74. 1979. P. 427–431.

5. *Dickey D.A.* and *Pantula S.G.* Determining the Order of Differencing in Autoregressive Processes // Journal of Business and Economic Statistics. 1987. Vol. 5. P. 455–461.

6. *Diebold F.X.* and *Nerlove M.* Unit Roots in Economic Time Series: A Selective Survey // Rhodes G.F. and Fomby T.B.(eds.). Advances in Econometrics. Vol. 8. Greenwich, CT: JAI Press. 1990. P. 3–69.

7. *Dolado J.J., Jenkinson T.* and *Sosvilla-Rivero S.* Cointegration and Unit Roots // Journal of Economic Survey. 1990. Vol. 4. P. 249–73.

8. *Johnston* and *DiNardo J.* Econometric Methods. Fourth Edition. The McGraw-Hill Companies, Inc., 1977.

9. *Mills T.C.* The Econometric Modelling of Financial Time Series. Cambridge University Press, 1993.

10. *Mills T.C.* The Econometric Modelling of Financial Time Series. Second Edition. Cambridge University Press, 1999.

11. *Nelson C.R.* and *Kang H.* Pitfalls in the Use of Time as an Explanatory Variable in Regression // Journal of Business and Economic Statistics. January 1984. Vol. 2. P. 73–82.

12. *Patterson K.* An Introduction to Applied Econometrics: A Time Series Approach. Palgrave, 2000.

13. *Perron P.* The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis // Econometrica. 1989. 57. P. 1361–401.

14. *Phillips P.C.B.* and *Perron P.* Testing for Unit Roots in Time Series Regression // Biometrica. 1988. Vol. 75. P. 335–46.

15. *Rappoport P.* and *Reichlin L.* Segmented Trends and Non-Stationary Time Series // Economic Journal. 1989. 99 (Supplement). P. 168–77.

16. *Said S.T.* and *Dickey D.A.* Testing for Unit Roots in Autoregressive Moving-Average Models with Unknown Order // Biometrica. 1984. Vol. 71. P. 599–607.

17. *Schwert G.W.* Effects of Model Specification on Tests for Unit Roots in Macroeconomic Data // Journal of Monetary Economics. 1987. Vol. 20. P. 73–105.

18. *Zivot E.* and *Andrews D.W.K.* Further Evidence on the Great Crash, the Oil Price Shock? And the Unit Root Hypothesis // Journal of Business and Economic Statistics. 1992. 10. P. 251–70.

ЛЕКЦИОННЫЕ И МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ

Анализ временных рядов[1)]

Канторович Г.Г.

Продолжается публикация курса «Анализ временных рядов». В этом номере рассматривается моделирование сезонности в терминах моделей ARIMA, построение авторегрессионных моделей с распределенными лагами (ADL), модели корреляции ошибками (ECM), различные типы экзогенности переменных, начато рассмотрение многомерных временных рядов.

В следующем выпуске будет продолжено рассмотрение моделей многомерных случайных процессов, коинтеграционные регрессии, тестирование коинтеграции.

Лекция 11

Сезонность

Статистические данные динамики экономических величин часто демонстрируют регулярные или почти регулярные колебания. Это явление носит название сезонности. Так, если рассматривать данные об объеме промышленного производства, то уровень производства в январе имеет большую связь с производством в прошлогоднем январе, чем с производством в предыдущем месяце – декабре. Обычно в России в январе наблюдается сильное падение производства по сравнению с декабрем. В развитых странах в летние месяцы происходит снижение безработицы в связи с сезонными работами. Данные о розничных продажах обычно имеют заметный «всплеск» в декабре, что носит название рождественского эффекта.

Из предыдущих лекций видно, что модели типа ARIMA плохо описывают такое сезонное циклическое поведение данных. Другими словами, неявно предполагается, что сезонность предварительно устранена. В традиционной экономической практике такое «десезонирование» данных получило широкое применение. Одним из распространенных методов является хорошо вам известное из курса эконометрики применение dummy-переменных. Например, для квартальных данных строится регрессия на 4 dummy-переменные, каждая из которых равна 1 в соот-

[1)] Подготовлено при содействии Национального фонда подготовки кадров (НФПК) в рамках программы «Совершенствование преподавания социально-экономических дисциплин в вузах».

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

ветствующем квартале и 0 – в остальных (не забудем, что свободный член в этом случае в регрессию не включается). Остатки этой регрессии и являются данными с устраненной сезонностью. Этот метод предполагает «жесткую» сезонность с неизменной величиной влияния каждого квартала в отдельности. Для месячных данных регрессия строится на 12 dummy-переменных, для полугодовых – на 2 dummy-переменные. Иногда рассматривают временные отрезки длиной в четыре недели, тогда в течение года насчитывается 13 промежутков и, соответственно, 13 dummy-переменных. В дальнейшем будем называть такой подход детерминированной сезонностью.

Предположение о «жесткой» сезонности является слишком ограничительным, поэтому официальная статистика применяет более сложные методы устранения сезонности. Широкое распространение получила программа X-11, разработанная в Американском бюро цензов (US Census Bureau), и ее дальнейшие развития. Разработаны и применяются на практике и другие методы. На рис. 11.1 приведены официальные ежемесячные данные Госкомстата РФ о динамике индекса промышленного производства в России в период с января 1990 г. по апрель 1999 г. до и после устранения сезонности [1]. На графике хорошо видно, что сезонность в этом случае наложена на тренд, т.е. данные демонстрируют одновременно и циклическое поведение, и тенденцию к снижению уровня.

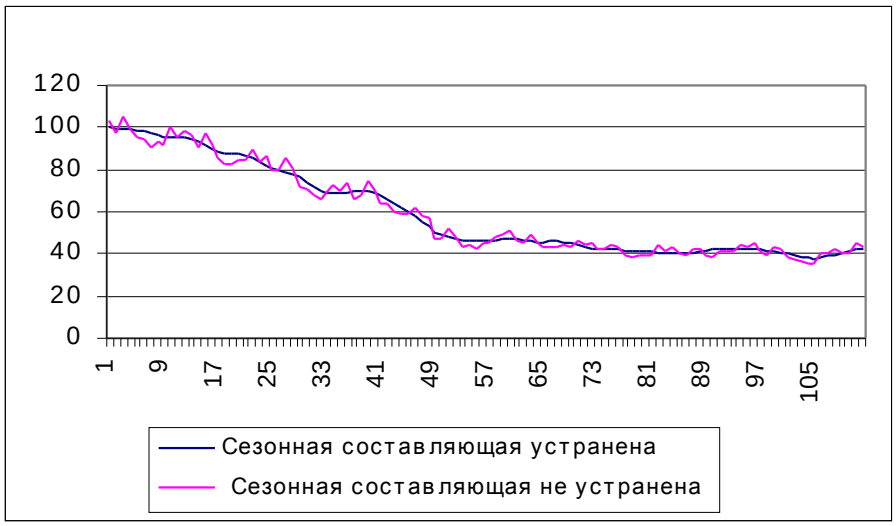

Рис. 11.1. Динамика индекса промышленного производства в России с января 1990 г. по апрель 1999 г., %

Рассмотренные до сих пор модели временной структуры ряда не предназначались для описания именно такого поведения. Мы строили авторегрессионные модели, когда текущее значение ряда было связано с несколькими предыдущими значениями. В принципе, циклическое поведение можно попытаться описать в терминах моделей ARMA, наложив некоторые ограничения на коэффициенты модели. Например, если исследуемые данные являются квартальными, сезонность означает, что данные I квартала текущего года сильнее связаны с данными I квартала предыдущего года, чем с данными предыдущего квартала. Разумно рассмат-

ривать модель, в которой явно присутствует связь только с данными одноименного квартала предыдущего года, т.е. модель вида: $x_t = \varphi x_{t-4} + \varepsilon_t$. А для месячных данных разумной будет модель $x_t = \Psi x_{t-12} + \varepsilon_t$. Используя оператор сдвига, получаем для квартальных данных: $(1-\varphi L^4)x_t = \varepsilon_t$, а для месячных данных: $(1-\varphi L^{12})x_t = \varepsilon_t$. Видно, что сезонность – это специальный случай ARMA. По нашей терминологии это ARMA(4,0) или AR(4). Но у него промежуточные коэффициенты равны нулю.

Что будет, если в модели с квартальными данными $\varphi = 1$? Модель принимает вид $x_t = x_{t-4} + \varepsilon_t$. Очевидно, характеристическое уравнение имеет единичный корень. Модель приводится к виду $(1-L^4)x_t = \varepsilon_t$. Я думаю, вы помните, что $(1-L^4)$ раскладывается на 4 сомножителя, и $L=1$ является корнем. Три остальных корня характеристического уравнения равны соответственно -1 и $\pm i$, где $i = \sqrt{-1}$. Все четыре корня по модулю равны 1. Говорят, что уравнение содержит единичный корень в обычном смысле и, кроме того, три сезонных единичных корня [11]. Аналогично для месячных данных: если $\Psi = 1$, то модель принимает вид $(1-L^{12})x_t = \varepsilon_t$. Разумеется, здесь тоже есть единичный корень в обычном смысле, т.е. и этот ряд относится к типу DS.

Модель, в которой имеется такой четко выраженный сомножитель $(1-\varphi L^4)$ или $(1-\varphi L^{12})$ принято называть *сезонной авторегрессией* (SAR). Аналогично можно рассмотреть сезонную схему для скользящего среднего, т.е. рассмотреть модель типа SMA, которая означает, что в правую часть будет добавлено слагаемое $\beta\varepsilon_{t-4}$. Эта модель представляет собой специальный вид обычного MA(4) или MA(12) с нулевыми промежуточными коэффициентами. Общая модель, описывающая сезонность в подходе Бокса–Дженкинса, носит название SARIMA.

Разработана примерно такая же теория и практика нахождения единичных сезонных корней, как это сделано для обычных единичных корней. Если TS и DS процессы мы называли также процессами с детерминированным и стохастическим трендом соответственно, то для описания сезонности употребляется аналогичная терминология. Если $\varphi = 1$, то присутствует *стохастическая сезонность*. Если $|\varphi| < 1$, то присутствует *детерминированная сезонность*.

Разумеется, в одной модели можно без большого труда сочетать и сезонную составляющую, и обычную ARMA схему. Например, рассмотрим модель типа SARIMA: $(1-L^4)B(L)x_t = \varepsilon_t$, где $B(L) = (1-\beta L)$ (для упрощения записи в модель не включена MA часть). После раскрытия скобок получаем: $(1-L^4-\beta L+\beta L^5)x_t = \varepsilon_t$ или $x_t = \beta x_{t-1} + x_{t-4} - \beta x_{t-5} + \varepsilon_t$. То есть исходная модель приводится к AR(5), но с некоторыми дополнительными ограничениями на коэффициенты. Они выражаются в том, что второй и третий коэффициенты равны нулю, и в том, что коэффициенты первого и пятого слагаемых в сумме равны нулю, т.е. имеют противоположные знаки. Другими словами, это некий частный случай общего AR(5). При

оценивании модели для нас важно, что мы уже знаем эти соотношения между коэффициентами и будем оценивать именно такую модель, а не общую, в которой у нас будут лишние переменные и понизится точность оценки наших коэффициентов. Эту модель принято обозначать SAR(4)×AR(1).

Econometric Views позволяет легко строить такие модели, вы просто указываете, что вам надо добавить в модель SAR или SMA соответствующего порядка. Следует отметить, что при задании порядка сезонности нужно учитывать, как описана структура данных в базе EViews. Если данные описаны как нерегулярные, то порядок сезонности задается числом 4, а если вы их уже описали как квартальные, то можно задать SAR(1).

Получается, что если в данных присутствует сезонность, нельзя обойтись моделью ARMA низкого порядка. Допустим, у вас есть ежемесячные данные за 4 или 5 лет. Наличие сезонности означает, что имеется значимая статистическая зависимость от соответствующего месяца предыдущего года. Вы можете сразу построить модель порядка, скажем, 15, т.е. AR(15), но AR(15) приведет к утрате 15 первых точек для построения модели. Коэффициентов определяется очень много, опасность мультиколлинеарности возрастает. При явном использовании схемы SAR возникает другая ситуация, в модели остается только 4 коэффициента. Конечно, некоторое число наблюдений все равно будет потеряно. Но число коэффициентов гораздо меньше, число степеней свободы существенно больше, значит, точность их оценивания возрастает. Кроме того, уменьшается опасность мультиколлинеарности.

Если учет сезонности в терминах подхода Бокса–Дженкинса находит широкое применение на практике, то исследование единичных сезонных корней не столь популярно. Среди работ по этой тематике, кроме уже отмеченной статьи Хиллеберга и других, стоит отметить монографию Франсеса [5].

Авторегрессионные модели с распределенными лагами

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 (auturegressive 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:

$$Y_t = \theta + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + ... + \alpha_p Y_{t-p} + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + ... + \beta_q X_{t-q} + \varepsilon_t, t = 1,2,...T.$$

Случайное возмущение по-прежнему полагается белым шумом. Используя операторные полиномы, получим $\alpha_p(L)Y_t = \theta + \beta_q(L)X_t + \varepsilon_t$. В отличие от модели

ARIMA мы отказываемся от нормирующего условия $\beta_0 \equiv 1$. Мы будем использовать обозначения ADL(p,q) и ARMAX(p,q), чтобы указать количество лагов независимой и зависимой переменных.

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,∞), 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: $Y_t = \alpha_p(L)^{-1}\theta + \alpha_p(L)^{-1}\beta_q(L)X_t + \alpha_p(L)^{-1}\varepsilon_t$, т.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 представляется в виде суммы динамических откликов на изменения величины X_t. Коэффициент влияния значения X_t на величину Y_{t+k} представляет собой частную производную $\frac{\partial Y_{t+k}}{\partial X}$, которую можно выразить как значение производной от операторного выражения:

$$\frac{d^k}{dL^k}\left(\frac{\beta_q(L)}{\alpha_p(L)}\right) \text{ при } L=0.$$

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:

$$\beta_0, \beta_1 + \alpha_1\beta_0, \alpha_1\beta_1 + \alpha_1{}^2\beta_0 \text{ и т.д.}$$

Долгосрочный отклик определяется как сумма всех промежуточных откликов. Используя стационарность процессов X_t и Y_t, долгосрочный отклик можно найти, просто взяв математические ожидания от обеих частей соотношения. В результате получаем уравнение статического или долгосрочного равновесия.

$$\overline{Y} = \frac{\theta}{\alpha_p(1)} + \frac{\beta_q(1)}{\alpha_p(1)}\overline{X},$$

где через $\overline{Y}$ и $\overline{X}$ 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): $Y_t = \theta + \alpha_1 Y_{t-1} + \beta_0 X_t + \beta_1 X_{t-1} + \varepsilon_t$. Заменив Y_t на $Y_{t-1} + \Delta Y_t$, и X_t на $X_{t-1} + \Delta X_t$, получаем:

$$\Delta Y_t = \theta + \beta_0 \Delta X_t - (1-\alpha_1) Y_{t-1} + (\beta_0 + \beta_1) X_{t-1} + \varepsilon_t.$$

Перегруппировка членов дает:

$$\Delta Y_t = \beta_0 \Delta X_t - (1-\alpha_1)\left[Y_{t-1} - \frac{\theta}{1-\alpha_1} - \frac{\beta_0 + \beta_1}{1-\alpha_1} X_{t-1} \right] + \varepsilon_t.$$

Это представление ADL модели называется *моделью коррекции ошибками* (error correction model), сокращенно – ECM[2)]. 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

$$\left[\overline{Y} - \frac{\theta}{1-\alpha_1} - \frac{\beta_0 + \beta_1}{1-\alpha_1} \overline{X} \right] = 0,$$

поэтому выражение в квадратных скобках положительно, если значение Y_{t-1} превышает равновесное значение, соответствующее X_{t-1}. Таким образом, текущее (краткосрочное) изменение Y представлено в виде суммы двух слагаемых. Первое из них – это мгновенный отклик на текущее (краткосрочное) изменение X, 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 невязке 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 отклоне-

[2)] Впервые введена в [14].

ния, например $Y_{t-2} = Y_{t-1} - \Delta Y_t$, $Y_{t-3} = Y_{t-1} - \Delta Y_t - \Delta Y_{t-2}$. В результате получаем ECM представление:

$$\Delta Y_t = \beta_0 \Delta X_t + \sum_{i=1}^{p-1} \delta_i \Delta Y_{t-i} + \sum_{i=1}^{q-1} \gamma_i \Delta X_{t-i} - \alpha_p(1)\left[Y_{t-1} - \frac{\theta}{\alpha_p(1)} - \frac{\beta_q(1)}{\alpha_p(1)} X_{t-1} \right] + \varepsilon_t .$$

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 также отметить, что коэффициенты δ_i и γ_i, также 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,...,q_k), а общее уравнение принимает вид:

$$\alpha_p(L) Y_t = \theta + \beta_{q_1}(L) X_{1t} + \beta_{q_2}(L) X_{2t} + \ldots + \beta_{q_k}(L) X_{kt} + \varepsilon_t .$$

Преобразование к модели коррекции ошибками проводится в этом случае точно также, 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 [8]. 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, гетероскедастичность 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 численными значениями оценок коэффициентов.

Лекция 12

При оценивании ADL моделей существует еще одна особенность, которую следует иметь в виду. Впрочем, она присутствует при оценивании любой модели со стохастическими регрессорами, но в курсе эконометрики мы оставляли ее за рамками рассмотрения. Рассматривая несколько случайных величин (или процессов) одновременно, мы всегда неявно предполагаем существование их совместного распределения. Для упрощения выкладок и, не теряя общности, ограничимся двумя процессами: Y_t и X_t. Пусть, например, мы строим ADL модель процесса Y_t от одной объясняющей переменной X_t. Совместная плотность распределения $f(Y_t, X_t)$, если она существует, может быть представлена в виде произведения $f(Y_t, X_t) = f(Y_t | X_t) f(X_t)$ условной плотности Y_t при условии X_t и одномерной плотности процесса X_t. Такое разложение называется факторизацией совместного распределения.

В то же время, теоретическая регрессия представляет собой условное математическое ожидание $E(Y_t | X_t) = \int Y_t f(Y_t | X_t) dY_t$, для вычисления которого используется только условное распределение. Фактически для вычисления регрессии не используется часть информации, содержащаяся в совместном распределении. Вопрос о том, в какой степени такой неполный учет информации влияет на оценивание регрессии, представляет очень интересный и важный аспект современного эконометрического исследования.

Рассмотрим два случайных процесса y_t и x_t, определяемые следующей моделью, в которой для упрощения записи математические ожидания положены равными нулю:

$$y_t = \beta x_t + \varepsilon_{1t}$$
$$x_t = \alpha_1 x_{t-1} + \alpha_2 y_{t-1} + \varepsilon_{2t}.$$

Пусть возмущения образуют двумерный нормальный случайный процесс $\vec{\varepsilon}$, каждая компонента которого не имеет автокорреляций (т.е. является белым шумом), но корреляция между одновременными значениями возмущений, вообще говоря, может быть ненулевой. Это означает, что

$$E(\vec{\varepsilon}) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \operatorname{cov}(\vec{\varepsilon}) = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}, \quad \rho = \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}}.$$

Двумерное распределение x_t и y_t полностью определяется свойствами вектора $\vec{\varepsilon}$, но у нас нет нужды выписывать распределение в явном виде. Заметим, что процессы x_t и y_t очевидно автокоррелированы. Первое из соотношений является уравнением регрессии, и вопрос состоит в том, в какой степени можно

оценивать эту зависимость y_t от x_t, не принимая во внимание второе из соотношений. В ранее применявшихся терминах [10] вопрос звучал так: является ли x_t экзогенным (exogenous) для уравнения регрессии? В соответствии же с современными представлениями [4] этот вопрос недостаточно конкретен: следует заранее определить способ использования модели. В эконометрике мы уже встречались с ситуацией, когда выбор, например, наилучшей модели зависел от того, как мы собираемся ее использовать. Мы различали, в основном, два типа использования: а) для целей прогнозирования и б) для изучения механизма модели, когда нас интересовали численные оценки всех или нескольких параметров. В соответствии с [4] будем различать следующие три цели анализа моделей:

- сделать статистические выводы об одном или нескольких параметрах модели (изучить механизм модели);
- предсказать y_t при заданном x_t, (условный прогноз);
- проверить, является ли уравнение регрессии структурно инвариантным при изменениях одномерного распределения x_t.

Слабая экзогенность (weak exogeneity)

Пусть мы рассматриваем регрессию некой величины y_t на несколько регрессоров. Обозначим через $\vec{\varphi}$ вектор параметров совместной плотности величины y_t и регрессоров, через $\vec{\theta}_1$ – вектор параметров, от которых зависит условная плотность вероятности y_t при заданных регрессорах, а через $\vec{\theta}_2$ – вектор параметров совместной плотности вероятности регрессоров.

Наконец, пусть параметры, относительно которых мы хотим сделать статистические выводы, образуют множество, записанное в виде вектора $\vec{\alpha}$. Мы скажем, что набор регрессоров является слабо экзогенным для параметров $\vec{\alpha}$, если статистические выводы об этих параметрах, сделанные на основании условного распределения, эквивалентны выводам по совместному распределению y_t и регрессоров.

Очевидно, что необходимым условием независимости статистических выводов от информации, содержащейся в условном распределении регрессоров, является возможность выразить $\vec{\alpha}$ в виде функции параметров $\vec{\theta}_1$, т.е. $\vec{\alpha} = \vec{\alpha}(\vec{\theta}_1)$. Кроме того, на компоненты векторов параметров $\vec{\theta}_1$ и $\vec{\theta}_2$ не должны быть наложены совместные ограничения ни в виде равенств, ни в виде неравенств. Это означает, что область допустимых значений параметров $\vec{\theta}_1$ не зависит от того, какие значения принимают параметры из множества $\vec{\theta}_2$, и наоборот. Это свойство по-английски выражается термином variation-free. По-русски можно использовать предложенный Э. Б. Ершовым термин «свободно-варьируемые».

Проиллюстрируем свойство слабой экзогенности на примере вышеприведенной модели. Умножив второе из уравнений на $-\frac{\sigma_{12}}{\sigma_{22}}$ и сложив полученные уравнения, получим:

$$y_t = (\beta + \frac{\sigma_{12}}{\sigma_{22}})x_t - \alpha_1 \frac{\sigma_{12}}{\sigma_{22}} x_{t-1} - \alpha_2 \frac{\sigma_{12}}{\sigma_{22}} y_{t-1} - \varepsilon_{1t} - \frac{\sigma_{12}}{\sigma_{22}} \varepsilon_{2t}.$$

Введя для упрощения записи обозначения:

$$\delta_0 = \beta + \frac{\sigma_{12}}{\sigma_{22}}$$

$$\delta_0 = -\alpha_1 \frac{\sigma_{12}}{\sigma_{22}}$$

$$\delta_0 = -\alpha_2 \frac{\sigma_{12}}{\sigma_{22}}$$

$$u_t = \varepsilon_{1t} - \frac{\sigma_{12}}{\sigma_{22}} \varepsilon_{2t},$$

приходим к следующей модели: $y_t = \delta_0 x_t + \delta_1 x_{t-1} + \delta_2 y_{t-1} + u_t$.

Случайный процесс u_t, как линейная комбинация нормальных процессов, сам является нормальным. Очевидно, что $\mathrm{var}(u_t) = \sigma_{11} - \frac{\sigma_{12}}{\sigma_{22}} = \sigma_{11}(1 - p^2)$ и $\mathrm{cov}(u_t, \varepsilon_{2t}) = \mathrm{cov}(\varepsilon_{1t}, \varepsilon_{2t}) - \frac{\sigma_{12}}{\sigma_{22}} \mathrm{cov}(\varepsilon_{2t}, \varepsilon_{2t}) = 0$ по определению параметров σ_{11}, σ_{12}, σ_{22} и ρ. Таким образом, нормальные случайные процессы u_t и ε_{2t} – некоррелированы и, следовательно, независимы. Полученное уравнение для процесса y_t является другой параметризацией уравнения регрессии $y_t = \beta x_t + \varepsilon_{1t}$.

Введенные ранее векторы параметров конкретизируются для данного примера следующим образом:

$$\vec{\phi} = (\beta, \alpha_1, \alpha_2, \sigma_{11}, \sigma_{12}, \sigma_{22})^T$$

$$\vec{\theta}_1 = (\delta_0, \delta_1, \delta_2, \sigma_{11}^2)^T,\ \vec{\theta}_2 = (\alpha_1, \alpha_2, \sigma_{22})^T.$$

Предположим, что вектор $\vec{\alpha}$ содержит только один параметр β, т.е. нас интересует только коэффициент регрессии y_t на x_t. Параметр β выражается через параметры из множеств $\vec{\theta}_1$ и $\vec{\theta}_2$ следующим образом:

$$\beta = \delta_0 + \frac{\delta_1}{\alpha_1} = \delta_0 + \frac{\delta_2}{\alpha_2}.$$

Очевидно, что β невозможно представить в виде функции от параметров, входящих только в множество $\vec{\theta}_1$. Более того, параметры из множеств $\vec{\theta}_1$ и $\vec{\theta}_2$ связаны очевидным соотношением $\alpha_2\delta_1 = \alpha_1\delta_2$, т.е. не являются свободно варьируемыми. Вывод очевиден: переменная (регрессор) x_t не является слабо экзогенной для параметра β.

Несколько изменим модель. Пусть возмущения ε_{1t} и ε_{2t} будут независимыми ($\sigma_{12} = 0$). Тогда

$$\vec{\phi} = (\beta, \alpha_1, \alpha_2, \sigma_{11}, \sigma_{12}, \sigma_{22})^T$$
$$\vec{\theta}_1 = (\beta, \sigma_{11})^T,\ \vec{\theta}_2 = (\alpha_1, \alpha_2, \sigma_{22})^T.$$

Теперь β является элементом множества $\vec{\theta}_1$ и, конечно, является функцией от самого себя. Также очевидно, что параметры множеств $\vec{\theta}_1$ и $\vec{\theta}_2$ являются свободно варьируемыми. Таким образом, в этом случае x_t является слабо экзогенной переменной для параметра β.

Как еще одну модификацию рассмотрим случай, когда σ_{12} вновь не равно нулю, но мы интересуемся другим параметром $-\delta_0$ из преобразованного уравнения. Очевидно, что параметр δ_0 представим как функция параметров из множества $\vec{\theta}_1$, но условие свободной варьируемости не выполнено. Следовательно, процесс x_t не является слабо экзогенным и для параметра δ_0.

Обратите внимание, что если переменная не является слабо экзогенной для параметра, то это еще не означает, что этот параметр нельзя оценить. Например, попытка оценить параметр β методом наименьших квадратов дает несостоятельную оценку, поскольку x_t очевидно коррелирует с ε_{1t}

$$(\operatorname{cov}(x_t, \varepsilon_{2t}) = \operatorname{cov}(\varepsilon_{1t}, \varepsilon_{2t}) = \sigma_{12}).$$

А вот корреляция переменной x_t с возмущением u_t равна нулю, поэтому оценка параметра δ_0 методом наименьших квадратов состоятельна. Но вот точность оценки (ее эффективность) снижена, так как в модель не инкорпорирована отмеченная связь между параметрами множеств $\vec{\theta}_1$ и $\vec{\theta}_2$.

Сильная экзогенность (strong exogeneity)

В соответствии с [4] переменная x_t называется строго экзогенной для параметра β, если она слабо экзогенна для него, и объясняемая переменная y_t не

является причиной по Грэнджеру для переменной x_t. Понятие причинности по Грэнджеру (Granger causality) было введено в работе [6]. 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 [12] 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: $E\left(Z_{t+1}|\Omega_t\right)$, где Ω_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 $E\left(Z_{t+1}|\Omega_t\right)= E\left(Z_{t+1}|\Omega_t \setminus x_s(s \le t)\right)$, то x_t не является причиной по Гренжеру 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 [6] 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: $Z_t = \alpha_0 + \sum_{i=1}^{k} \alpha_i Z_{t-i} + \sum_{i=1}^{k} \beta_i x_{t-i} + \varepsilon_t$. А после этого проверить обычную гипотезу о равенстве нулю группы коэффициентов: $H_0 : \beta_1 = ... = \beta_k = 0$ $H_1 : \beta_1^2 + ... + \beta_k^2 > 0$. Это обычный F-тест. Как обычно, строится полное уравнение и укороченное, и сравниваются остаточные суммы по F-статистике. Если ну-

левая 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 в [16] 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

$$y_t = \sum_{i=-k}^{m} \beta_i x_{t-1} + \varepsilon_t$$

и проверяем гипотезу H_0: $\beta_{-i} = 0$ ($i = 1, \ldots, k$) 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 [3]. 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.

Если группа регрессоров является строго экзогенной для параметра β, то этот параметр можно оценить из уравнения регрессии, используя только информацию об условном распределении. В этом случае также можно прогнозировать процесс y_t, основываясь на прогнозе процесса x_t по его собственным прошлым значениям.

Суперэкзогенность (super exogeneity)

Говорят, что группа регрессоров является суперэкзогенной, если изменение их совместного распределения не меняет условного распределения объясняемой переменной y_t. Это свойство связано с так называемой критикой Лукаса [13]. Она заключается в следующем. Одна из целей эконометрического моделирования состоит в прогнозировании эффекта от изменений экзогенных переменных. Однако при изменении этих переменных экономические агенты видят, что происходят изменения, меняют свое поведение и, тем самым, меняют параметры экономической системы. Поэтому модель с постоянными параметрами, говорил Лукас, не адекватна реальным экономическим системам. Свойство суперэкзогенности выделяет модели экономических систем, к которым критика Лукаса не применима.

В соответствии с уже упоминавшимся подходом Cowles Foundation [10] экзогенность переменных и причинно-следственные отношения между переменны-

ми не тестируются, они специфицируются из содержательных априорных соображений. Современный же подход [4] предполагает тестирование экзогенности. По самому смыслу этого свойства для проверки, влияет или нет учет распределения регрессоров на оценки регрессии, использующие только условное распределение, нужно построить обе модели: учитывающую DGP (Data Generating Process) полностью и уравнение регрессии. Но если мы в состоянии построить модель, учитывающую DGP полностью, то необходимость в регрессии просто пропадает. Рассмотренный выше пример с исследованием слабой экзогенности процесса x_t относительно параметра β показывает, что нарушение этого свойства связано с наличием регрессора y_{t-1} во втором из уравнений и с неравенством нулю ковариации σ_{12}. Следовательно, наличие слабой экзогенности является частным случаем общей модели при ограничениях на значения параметров: $\sigma_{12} = 0, \alpha_2 = 0$. Поэтому для тестирования слабой экзогенности представляется естественным применение теста множителей Лагранжа, при котором оценивается только модель с ограничениями на параметры. Тест, использующий тест множителей Лагранжа для тестирования слабой экзогенности, был разработан Энглом [4].

В условиях нулевой гипотезы оба уравнения для y_t и для x_t могут быть оценены МНК по отдельности. Обозначим остатки соответствующих регрессий через e_y и e_x, причем в обе регрессии включены свободные члены, если только они не должны там отсутствовать из содержательных соображений. Далее строим регрессию e_y на константу, x_t и $e_x : e_{yt} = \mathrm{a} + \mathrm{b}\,x_t + \mathrm{g}\,e_{xt} + \varepsilon_t$.

Нулевая гипотеза принимает вид $H_0 : \gamma = 0$, и при условии справедливости H_0 асимптотическое распределение статистики $(T-1)R^2$ имеет асимптотическое распределение χ_1^2 с одной степенью свободы. Здесь R^2 – множественный коэффициент детерминации последней регрессии, а $(T-1)$ – количество наблюдений при ее оценке.

Асимптотически эквивалентный результат можно получить, используя R^2 регрессии e_x на константу, x_{t-1}, y_{t-1} и e_y. Этот подход требует построения трех регрессий. Можно ограничиться построением только двух, если воспользоваться теоремой Фриша–Вау. В этом варианте теста мы строим регрессию y_t на константу, x_t и e_x, а затем проверяем значимость коэффициента при e_x с помощью статистики Стьюдента. Статистический вывод асимптотически эквивалентен предыдущим модификациям. Последняя версия заключается в регрессии x_t на константу, x_{t-1}, y_{t-1} и e_y и проверке значимости коэффициента при e_y.

Для тестирования сильной экзогенности мы тестируем слабую экзогенность и затем причинность по Грэнджеру, как это описано ранее. Проверка суперэкзогенности проводится в три этапа. Сначала проверяется слабая экзогенность. Затем проверяется стабильность параметров условного распределения из множест-

ва $\vec{\theta}_1$. Для этой цели могут использоваться введение dummy-переменных и проверка их значимости, тест Чау и другие тесты.

Если стабильность параметров условного распределения установлена, то на третьем этапе проверяется стабильность параметров из множества $\vec{\theta}_2$. Проверку имеет смысл проводить, используя dummy-переменные. Если параметры из множества $\vec{\theta}_2$ оказываются стабильными, то имеющиеся данные не дают оснований для вывода о суперэкзогенности. Если же нестабильность параметров из $\vec{\theta}_2$ установлена, то значимые dummy-переменные добавляются в качестве дополнительных регрессоров в исследуемую регрессию. Если введенные dummy-переменные оказываются незначимыми в совокупности, то суперэкзогенность установлена.

На другой идее проверки слабой экзогенности основан тест Ву–Хаусмана [7, 18]. При отсутствии слабой экзогенности регрессор x_t коррелирован со случайным возмущением, поэтому МНК оценки коэффициентов регрессии являются смещенными и несостоятельными. Состоятельные оценки дает метод инструментальных переменных (IV), о котором мы говорили в курсе эконометрики. Напомню, что инструментальной переменной (инструментом) для x_t называется любая переменная z_t, некоррелированная со случайным возмущением, но коррелированная с x_t. Тест Ву–Хаусмана сравнивает две оценки: метода наименьших квадратов и метода инструментальных переменных. При нулевой гипотезе слабой экзогенности обе оценки состоятельны и асимптотически эквивалентны. При альтернативной гипотезе оценки не эквивалентны.

В качестве инструмента выбирается оценка x_t методом наименьших квадратов. Проиллюстрируем применение теста Ву–Хаусмана на рассматриваемом примере.

$$y_t = \beta x_t + \varepsilon_{1t}$$
$$x_t = \alpha_1 x_{t-1} + \alpha_2 y_{t-1} + \varepsilon_{2t}$$

с теми же свойствами случайных возмущений. На первом шаге строим регрессию, соответствующую второму уравнению, и обозначим через $\hat{x}_t$ рассчитанные значения x_t. Далее оцениваем регрессию

$$y_t = \beta x_t + \gamma \hat{x}_t + \varepsilon_t$$

и проверяем значимость коэффициента γ. В общем случае нескольких регрессоров строятся инструменты для каждого из них, все инструменты добавляются в регрессию, и проверяется гипотеза о совместной значимости коэффициентов при инструментальных переменных. Если нулевая гипотеза не отвергается, делается вывод о наличии слабой экзогенности.

Впрочем, на практике достаточно часто просто постулируют слабую экзогенность переменных, исходя из содержательных соображений. Напомним, что стационарность переменных и слабая экзогенность регрессоров позволяют получить состоятельные оценки коэффициентов ADL моделей, и обычные процедуры проверки гипотез являются асимптотически верными.

Лекция 13

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. Как видно из прошлой лекции, 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 $\vec{X}_t = (x_t^1, x_t^2, ..., x_t^k)^T$, каждая компонента которого является временным рядом. Верхним индексом будем обозначать номер компоненты, а нижним по-прежнему – момент времени. Распределение компонент характеризуется семейством совместных плотностей распределения вида: $f_n(x_{t_1}^{i_1}, x_{t_2}^{i_2}, ..., x_{t_n}^{i_n}),\ \ n = 1,2,...$. 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: $f_2(x_t^1, x_t^2) = f_2(x_{t+\tau}^1, x_{t+\tau}^2)$ для любого τ. 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 $f_3(x_t^1, x_{t+h}^1, x_{t+s}^2) = f_3(x_{t+\tau}^1, x_{t+h+\tau}^1, x_{t+s+\tau}^2)$. Можно сказать, что это свойство инвариантности к сдвигу во времени. То есть, если к каждому моменту времени прибавить величину τ, 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 $E(\vec{X}_t) = \begin{pmatrix} \mu_1 \\ ... \\ \mu_k \end{pmatrix}$ 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 τ: $Corr(\tau) = Corr(x_t^i, x_{t+\tau}^i) = \rho_i(\tau)$. Однако теперь можно рассмотреть второй смешанный момент для различных компонент, а также $Corr(x_t^i, x_{t+\tau}^j)$. Такую величину естественно назвать кросс-корреляционной функцией. Если компоненты образуют многомерный стационарный процесс, то кросс-корреляция будет функцией сдвига во времени τ. Обозначим эту функцию $R_{ij}(\tau)$. Довольно очевидно, что $R_{ij}(\tau) = R_{ji}(-\tau)$. При фиксированном значении τ элементы $R_{ij}(\tau)$ образуют матрицу R, зависящую от τ. Значению τ, равному нулю, соответствует корреляционная матрица вектора $\vec{X}_t$.

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. Эта особенность не имеет специфики временных рядов, а относится именно к оценке одновременных уравнений. Поскольку системы одновременных уравнений не входили еще в данный курс эконометрики, кратко рассмотрим возникающие трудности.

Пусть задана простая Кейнсианская макроэкономическая модель национального дохода, состоящая только из двух следующих уравнений:

$$\begin{cases} Y_t = C_t + Z_t \\ C_t = \alpha + \beta Y_t + u_t \end{cases}$$

где Y_t – национальный доход, C_t – потребление, Z_t – инвестиции и государственные расходы, а u_t – белый шум. Первое из уравнений представляет собой условие равновесия национального дохода, а второе – функцию потребления. Коэффициент β – предельная склонность к потреблению, а коэффициент α – автономное потребление. В соответствии с экономическим смыслом считаем эндогенными переменными Y_t и C_t – ВВП и потребление, а Z_t считаем экзогенной переменной. Эти уравнения и их коэффициенты отражают экономический смысл соотношений, в этом случае говорят, что уравнения заданы в *структурной (structural) форме*. Индекс t указывает на то, что переменные являются временными рядами, хотя уравнения и не содержат лаговых составляющих. Пусть собраны данные о потреблении, о расходах и о национальном доходе. Можно ли, от-

влекаясь от проблемы экзогенности, оценить функцию потребления просто из второго соотношения методом наименьших квадратов?

Модель представленного типа называется одновременными уравнениями. В данном случае у нас два уравнения, причем эконометрическое у нас только одно. Первое соотношение – это обычное экономическое тождество, мы ничего не оцениваем в этом уравнении. Мы хотим оценить только коэффициенты α и β, но из-за того, что существует дополнительное экономическое тождество, появляется проблема. Если просто подставить C_t в первое соотношение, то получим $Y_t = \alpha + \beta Y_t + Z_t + u_t$. Разрешив полученное уравнение относительно национального дохода, получаем $Y_t = \frac{\alpha}{1-\beta} + \frac{1}{1-\beta} Z_t + \frac{1}{1-\beta} u_t$. Подстановка Y_t во второе уравнение дает так называемую *приведенную (reduced) форму* модели:

$$\begin{cases} C_t = \frac{\alpha}{1-\beta} + \frac{\beta}{1-\beta} Z_t + \frac{1}{1-\beta} u_t \\ Y_t = \frac{\alpha}{1-\beta} + \frac{1}{1-\beta} Z_t + \frac{1}{1-\beta} u_t \end{cases}.$$

Теперь эндогенные переменные выражены через экзогенные переменные и случайные возмущения. Различие между структурной и приведенной формой носит не только математический, но и содержательный характер. Вообще-то нас интересуют коэффициенты структурных уравнений, они имеют экономический смысл. Например, β – это предельная склонность к потреблению. А коэффициенты в приведенной форме уже не несут экономического смысла.

Приведенная форма явно показывает, что переменная Y_t коррелирована со случайным возмущением u_t. Формально в структурное уравнение, определяющее Y_t, не входит никакая случайная компонента. Тем не менее, из-за того, что это одновременные уравнения, Y_t зависит от u_t. Это означает, что метод наименьших квадратов даст смещенную и несостоятельную оценку при оценивании функции потребления. Он приведет к смещению, вызванному совместными уравнениями. Даже асимптотически метод наименьших квадратов, примененный к структурному уравнению, приводит к плохим результатам. Из-за одновременности уравнений ошибка из любого уравнения как бы проникает во все остальные уравнения.

Можно использовать МНК для оценки коэффициентов приведенной формы, если экзогенная переменная Z_t детерминирована или не коррелирует с возмущением u_t. Для нахождения коэффициентов структурной формы применяются два способа.

1. Косвенный (indirect) метод наименьших квадратов. Методом наименьших квадратов оцениваем коэффициенты приведенной формы, для чего оцениваем регрессии:

$$\begin{cases} Y_t = \pi_{10} + \pi_{11} \cdot Z_t + \varepsilon_t^1 \\ C_t = \pi_{20} + \pi_{21} \cdot Z_t + \varepsilon_t^2 \end{cases}.$$

Каждая из регрессий оценивается отдельно. Для нахождения параметров структурной формы α и β можно попытаться решить систему уравнений:

$$\frac{\alpha}{1-\beta} = \hat{\pi}_{10}, \quad \frac{\alpha}{1-\beta} = \hat{\pi}_{20}, \quad \frac{\beta}{1-\beta} = \hat{\pi}_{21}, \quad \frac{1}{1-\beta} = \hat{\pi}_{11}.$$

Очевидно, что эта система четырех уравнений с двумя неизвестными, вообще говоря, не имеет решения. В этом и состоит основной недостаток косвенного метода наименьших квадратов. Приведенную систему всегда можно оценить. Предположим, система состоит не из двух, а из 10 уравнений с 10 переменными. Выбираем эндогенные и экзогенные переменные и строим регрессию каждой эндогенной переменной на все экзогенные. После этого основной вопрос заключается в том, можно ли от коэффициентов приведенной формы вернуться однозначно к структурным коэффициентам, которые и имеют содержательный смысл. Говорят, что система идентифицируема, если по коэффициентам приведенной формы можно однозначно вернуться к структурным коэффициентам. Иногда соответствующая система уравнений противоречива, иногда решения будут существовать, но не будут единственными. Существуют системы одновременных уравнений, в которых косвенный МНК приводит к однозначному результату. Идентифицируемость зависит в том числе и от того, сколько экзогенных переменных присутствуют в модели и как они входят в структурные уравнения. В этом, на самом деле, одна из причин того, что в нашем примере мы не можем оценить структурные коэффициенты – у нас одна единственная экзогенная переменная, всего два структурных параметра и четыре – в приведенной системе. С проблемой идентифицируемости вы познакомитесь подробней в дальнейшем курсе эконометрики при изучении темы «системы одновременных уравнений». Кажущийся таким простым косвенный метод наименьших квадратов не всегда может дать оценки структурных параметров.

Для дальнейшего изложения важно понять, что, имея систему одновременных уравнений, нельзя оценивать каждое уравнение в отдельности из-за того, что эндогенные переменные входят и в левую, и в правую часть, и из-за того, что равновесие определяется всеми уравнениями вместе. Случайная ошибка из любого уравнения проникает во все остальные, из-за чего оказываются коррелированны регрессор и случайное возмущение.

2. Метод инструментальных переменных. Этот метод известен из курса эконометрики и кратко упоминался в предыдущей лекции. В случае системы одновременных уравнений можно легко предложить набор инструментальных переменных. Они строятся как линейные комбинации экзогенных переменных, подобранные так, чтобы не было коррелированности инструментов и случайных возмущений. Этот метод тоже зависит от того, насколько идентифицируема оцениваемая система уравнений. При отсутствии лаговых переменных двумя относительно широко применяемыми вариантами этого подхода являются так называемые двухшаговый МНК и трехшаговый МНК. Мы не будем на них специально останавливаться. Это отдельная проблема.

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 (как в вышерассмотренном примере), 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.

Векторная авторегрессия

В 1980 г. Симс [15] 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 VARIMA [17], но, как мы увидим дальше, это скорее теоретически, чем практически. 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:

$$\begin{cases} X_t^1 = \alpha_1 + \beta_{11} X_{t-1}^1 + \beta_{12} X_{t-1}^2 + \varepsilon_t^1 \\ X_t^2 = \alpha_2 + \beta_{21} X_{t-1}^1 + \beta_{22} X_{t-1}^2 + \varepsilon_t^2 \end{cases}.$$

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:

$$\vec{X}_t = \vec{a} + A_1 \vec{X}_{t-1} + A_2 \vec{X}_{t-2} + \ldots + A_p \vec{X}_{t-p} + \vec{\varepsilon}_t,$$

где $\vec{\varepsilon}_t$ – векторный белый шум со следующими свойствами: $E(\vec{\varepsilon}_t) = 0$ для всех t, $\operatorname{cov}(\vec{\varepsilon}_t) = E(\vec{\varepsilon}_t \vec{\varepsilon}_s^{\,T}) = \begin{cases} \Omega, s = t \\ 0, s \neq t \end{cases}$. Модель сразу выписывается в приведенной форме. Одновременные компоненты не входят в правую часть. Для рассмотренного двумерного примера получаем:

$$\begin{pmatrix} X_t^1 \\ X_t^2 \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} + \underbrace{\begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix}}_{A_1} \begin{pmatrix} X_{t-1}^1 \\ X_{t-1}^2 \end{pmatrix} + \begin{pmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \end{pmatrix}.$$

Модель VAR(p) напоминает модель ARMA, поэтому многие их свойства схожи. Перепишем модель VAR(p), используя оператор лага,:

$$(E - A_1 L - A_2 L^2 - \ldots - A_p L^p)\vec{x}_t = \vec{\varepsilon}_t.$$

Для упрощения записи мы предполагаем, что нет свободного члена, это не приводит к потери общности, но немного сокращает запись. Для рассматриваемого примера можем записать:

$$\begin{pmatrix} 1-\beta_{11}L & -\beta_{12}L \\ -\beta_{21}L & 1-\beta_{22}L \end{pmatrix} \begin{pmatrix} x_t^1 \\ x_t^2 \end{pmatrix} = \begin{pmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \end{pmatrix}.$$

Также, как и в одномерном случае, возникает вопрос, можно ли выразить вектор $\vec{x}_t$ через текущие и предыдущие значения вектора случайных возмущений $\vec{\varepsilon}_t$. Очевидно, что если существует обратная матрица, то можно записать:

$$\begin{pmatrix} x_t^1 \\ x_t^2 \end{pmatrix} = \frac{1}{\Delta} \begin{pmatrix} 1-\beta_{22}L & \beta_{12}L \\ \beta_{21}L & 1-\beta_{11}L \end{pmatrix} \begin{pmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \end{pmatrix}.$$

Если обозначить через I единичную матрицу, то определитель матрицы системы $\Pi = I - A_1$ равен:

$$\Delta = (1-\beta_{11}L)(1-\beta_{22}L) - \beta_{12}\beta_{21}L^2 = 1-(\beta_{11}+\beta_{22})L + (\beta_{11}\beta_{22} - \beta_{12}\beta_{21})L^2 = (1-\mu_1 L)(1-\mu_2 L),$$

где μ_1 и μ_2 – собственные числа матрицы Π, возможно комплексные.

Если хотя бы одно из собственных чисел равно 0, то матрица системы уравнений вырождена, и $\vec{x}_t$ не выражается через текущие и предыдущие значения векторного 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 $\vec{x}_t$ 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 $\vec{x}_t$ будет стационарной в широком смысле.

Для изучения условий стационарности и свойств моделей VAR(1) удобно использовать сведение матриц A_1 и Π к простейшей форме. Заметим, что матрицы Π и A_1 очевидно имеют одни 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 Π 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 λ_1 и λ_2 соответственно. Обозначим через Λ диагональную матрицу diag$\{\lambda_1, \lambda_2\}$. Тогда очевидно, что $C^{-1}A_1C = \Lambda$ и $A_1 = C\Lambda C^{-1}$. Определим новый вектор переменных соотношением $\vec{y}_t = C^{-1}\vec{x}_t$, отсюда следует также, что $\vec{x}_t = C\vec{y}_t$. Умножая уравнение $\vec{x}_t = A_1\vec{x}_{t-1} + \vec{\varepsilon}_t$ слева на матрицу C^{-1}, получим $\vec{y}_t = \Lambda\vec{y}_{t-1} + \vec{\eta}_t$, где через $\vec{\eta}_t$ обозначен «новый» векторный белый шум.

В новых переменных система уравнений распадается на два отдельных уравнения:

$$y_t^1 = \lambda_1 y_{t-1}^1 + \eta_t^1$$
$$y_t^2 = \lambda_2 y_{t-1}^2 + \eta_t^2,$$

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 $\vec{y}_t$ 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 $\vec{x}_t$ 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 ограничимся только кратким рассмотрением этого понятия без доказательств. Итак, доказано, что существует невырожденная матрица P, такая что $P^{-1}A_1P = J$ и $A_1 = PJP^{-1}$, где J – жорданова матрица: $J = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$. Определив вектор $\vec{y}_t = P^{-1}\vec{x}_t$ ($\vec{x}_t = P\vec{y}_t$), сведем исходную модель к виду $\vec{y}_t = J\vec{y}_{t-1} + \vec{\eta}_t$, где сейчас векторный белый шум $\vec{\eta}_t = P^{-1}\vec{\varepsilon}_t$. В этом случае система уравнений:

$$y_t^1 = \lambda_1 y_{t-1}^1 + y_{t-1}^2 + \eta_t^1$$
$$y_t^2 = \lambda_2 y_{t-1}^2 + \eta_t^2$$

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 Π) должны лежать строго внутри единичного круга.

Этот результат легко обобщить на случай VAR(1) модели произвольной размерности. Если все собственные числа $\lambda_1, \lambda_2, ..., \lambda_k$ матрицы A_1 различны, то модель распадается на отдельные уравнения $y_t^i = \lambda_i y_{t-1}^i + \eta_t^i \quad (i = 1,2,..,k)$. Каждое из них стационарно тогда и только тогда, когда $|\lambda_i| < 1$. Если же среди собственных чисел $\lambda_1, \lambda_2, ..., \lambda_k$ есть группа равных по величине, то, обозначив их без потери общности номерами 1,2,...,*k*, приводим модель к тому же виду: $\vec{y}_t = J\vec{y}_{t-1} + \vec{\eta}_t$. Матрица J содержит так называемую жорданову клетку S размера r в левом верхнем углу:

$$S = \begin{pmatrix} \lambda & 1 & . & 0 & 0 \\ 0 & \lambda & . & 0 & 0 \\ . & . & . & . & . \\ 0 & 0 & . & \lambda & 1 \\ 0 & 0 & . & 0 & \lambda \end{pmatrix}$$

На главной диагонали жордановой клетки стоят собственные числа, первая над-диагональ содержит 1, а остальные элементы равны 0. Отдельные уравнения системы принимают вид:

$$\begin{aligned} & y_t^1 = \lambda y_{t-1}^1 + y_{t-1}^2 + \eta_t^1 \\ & . \\ & . \\ & . \\ & y_t^{r-1} = \lambda y_{t-1}^{r-1} + y_{t-1}^r + \eta_t^{r-1} \\ & y_t^r = \lambda y_{t-1}^r + \eta_t^r \\ & y_t^{r+1} = \lambda_{r+1} y_{t-1}^{r+1} + \eta_t^{r+1} \\ & . \\ & . \\ & . \\ & y_t^k = \lambda_k y_{t-1}^k + \eta_t^k . \end{aligned}$$

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 $\vec{\varepsilon}_t$. По аналогии со скалярным случаем будем искать общее решение однородной системы $\vec{x}_t - A_1\vec{x}_{t-1} - ... - A_p\vec{x}_{t-p} = 0$ в виде $\vec{x}_t = \vec{C}(\lambda)^t$. Подставив это выражение в уравнение, получим:

$$(\lambda^p I + \lambda^{p-1}A_1 + ... + A_p)\vec{C} = 0.$$

Однородная система линейных уравнений имеет ненулевые решения при λ равных собственным числам матрицы $\lambda^p I + \lambda^{p-1}A_1 + ... + A_p$, т.е. корням характеристического уравнения $\det(\lambda^p I + \lambda^{p-1}A_1 + ... + A_p) = 0$.

Соответствующими нетривиальными решениями системы

$$(\lambda^p I + \lambda^{p-1}A_1 + ... + A_p)\vec{C} = 0$$

являются собственные векторы ее матрицы. Тогда слагаемое, соответствующее собственному числу λ_i, стремится к нулю при неограниченном увеличении *t*. Если среди $p \times k$ собственных чисел есть одинаковые, то нужно заменить в выражении $\vec{x}_t = \vec{C}(\lambda)^t$ постоянный вектор $\vec{C}$ зависящим от времени вектором вида $\vec{C}_0 + \vec{C}_1 t + ... + \vec{C}_{r-1}t^{r-1}$. Во всех случаях условие стационарности выражается компактно: все корни характеристического уравнения должны лежать строго внутри единичного круга.

Если условие стационарности выполнено, и случайное возмущение $\vec{\varepsilon}_t$ 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_t^i на предыдущие значения всех компонент, включая предыдущие значения самой x_t^i. 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 и легко реализуемо в специализированных пакетах. Обычно в компьютерных программах, например в том же Econometric views, задается список переменных, число лагов, и этого достаточно для построения модели.

Одним из больших препятствий при практическом применении VAR моделей является очень большое число подлежащих оцениванию параметров. Рассмотрим простой случай. Как вы знаете, стандартная кейнсианская модель состоит из 7 уравнений. Переменные модели являются макроэкономическими показателями. По каким данным вы будете оценивать такую модель? Ну, не меньше, чем по квартальным. Сколько лагов включать в модель? Конечно, существуют

статистические методы оценки числа лагов, но попробуем использовать содержательные знания. Знаменитая Сент-Луисская модель [2] демонстрирует, что в уравнение темпа роста номинального ВВП входят темпы роста денежного агрегата M_1 и темпы роста правительственных расходов на поддержание занятости с запаздыванием вплоть до четырех кварталов. Давайте возьмем модель с 5 лагами, пусть используются квартальные данные. Сколько параметров нам надо оценить, если мы строим VAR модель? 5 лагов – это значит 5 матриц размерности 7 на 7. В каждой 49 параметров, плюс столбец свободных членов, плюс 7 дисперсий случайных составляющих. Итого мы получаем, что надо оценить $49 \cdot 5 + 7 + 7 = 259$ коэффициентов. В каждом из уравнений нужно оценить 37 параметров. Для этого необходимы данные хотя бы лет за 15. Вот эта простая арифметика называется перепараметризацией VAR, т.е. VAR вводит слишком много параметров в модель. Это свойство препятствует широкому применению VAR моделей.

В то же время оказалось, что VAR представление тесно связано с исследованием стационарности и возможностью построения эконометрических моделей с нестационарными регрессорами. Это и будет предметом нашего рассмотрения в последующих лекциях.

* *
*

СПИСОК ЛИТЕРАТУРЫ

1. Индексы интенсивности промышленного производства / Центр экономической конъюнктуры при Правительстве РФ, ежемесячное издание. Также официальный сайт Центра http://www.cea.gov.ru

2. *Carlson K.M.* Does the St. Louis Equation Now Believe in Fiscal Policy // Review., Federal Bank of St. Louis. 1978. V. 60. № 2. P. 17.

3. *Chambarlain G.* The General Equivalence of Granger and Sims Causality // Econometrica. 1982. V. 50. P. 569–582.

4. *Engle R.F.* Wald, Likelihood Ratio and Lagrange Multiplier Tests in Econometrics // Handbook of Econometrics, eds. Z. Griliches and M.D. Intriligator. Chapter 13. North-Holland, 1984.

5. *Franses P.H.* Periodicity and Stochastic Trends in Economic Time Series. Oxford Univercity Press, 1996.

6. *Granger C.V.J.* Investigating Causal Relations by Econometric Methods and Cross-Spectral Methods // Econometrica. 1969. V. 37. P. 424–438.

7. *Hausman A.* Specification Tests in Econometrics // Econometrica. V. 46. P. 1251–1271.

8. *Hendry D.F.* and *Ericsson N.R.* Modeling the Demand for Narrow Money in the United Kingdom and the United States // European Economic Review. 1991. V. 35. P. 833–866.

9. *Hendry D.F.* and *Richard J.-F.* Exogeneity // Econometrica. 1983. V. 51. P. 277–304.

10. *Hood W. C. and Coopmans T. S.* Studies in Econometric Method. Wiley, 1953.

11. *Hylleberg S.*, *Engle R.F.*, *Granger C.W.J.*, Yoo *B.S.* Seasonal Integration and Cointegration // Journal of Econometrics. 1990. № 44. P. 215–238.

12. *Leamer E.E.* Vector Autoregressions for Cansal Inference // K.L. Brunner and A. Meltrer (eds.) Understanding Monetary Regims (Supplement to the Journal of Monetary Economics). 1985. P. 255–304.

13. *Lucas R.E.* Econometric Policy Evaluation: A Critique // K.L. Brunner (eds.) The Phillips Curve and Labor Markets (Supplement to the Journal of Monetary Economics). 1976. P. 19–46.

14. *Sargan J.D.* Wages and Prices in the United Kingdom: A Study in Econometric Methodology // P.E. Hart, G. Mills, K. Whiteker (eds.) Econometric Analysis for National Economic Planning, Butterworth. London, 1964; reprinted in D.F. Hendry and K.F. Wallis (eds.) Econometrics and Quantitative Economics. Oxford: Basil Blackwell, 1984.

15. *Sims C.A.* Macroeconomics and Reality // Econometrica. 1980. V. 48. P. 1–48.

16. *Sims C.A.* Money, Income and Causality // American Economic Review. 1972. V. 62. P. 450–552.

17. *Tsiao G.C.* and *Box G.E.P.* Modelling Multiple Time Series with Applications // Journal of the American Statistical Association. 1981. V. 76. P. 802–816.

18. *Wu D.* Alternative Tests of Independence between Stochastic Regressors and Disturbances // Econometrica. 1973. V. 41. P. 733–750.

ЛЕКЦИОННЫЕ И МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ

Анализ временных рядов[1)]

Канторович Г.Г.

Заканчивается публикация курса «Анализ временных рядов». В этом номере рассматривается построение экономических моделей с нестационарными регрессорами, понятие коинтеграции и ее связь с VAR-представлением многомерного процесса, тестирование наличия коинтеграционной зависимости, оценивание параметров модели при наличии коинтеграции. Отдельно рассматриваются модели с кластеризованной волатильностью, нашедшие широкое применение при анализе финансовых временных рядов.

Лекция 14

Коинтеграция

Рассмотренные до сих пор методы оценки ADL-уравнений, связывающих несколько экономических переменных, применимы только для стационарных рядов. Более точно, переменные могут быть и нестационарными, но типа TSP, тогда в качестве дополнительного регрессора в правую часть модели добавляется дополнительная объясняющая переменная – линейный тренд, и основным способом оценивания остается метод 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 показывает, что большинство макроэкономических переменных относятся именно к типу DSP [15]. Фактически это означало бы отсутствие долгосрочных зависимостей в макроэко-

[1)] Подготовлено при содействии Национального фонда подготовки кадров (НФПК) в рамках программы «Совершенствование преподавания социально-экономических дисциплин в вузах».

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

номике. 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 Грэнжером и Энглом [6, 4] 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, $Z_t = \alpha x_t + \beta y_t \sim I(1)$, вообще говоря, тоже является рядом порядка интеграции 1. Однако может оказаться, что существуют такие коэффициенты (α, β), 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 (α, β) 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 (α, β), такой что $\alpha x_t + \beta y_t \sim I(d-b)$, 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 распространенном 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 $\vec{w}_t$ имеет порядок интеграции d: $\vec{w}_t \sim I(d)$, 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 $\vec{w}_t$ называются коинтегрированными: $(\vec{\alpha}^T \vec{w}_t) \sim I(d-b)$, а вектор $\vec{\alpha}$ 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. Если попробовать изобразить это свойство графически, то мы получим следующую символическую картинку:

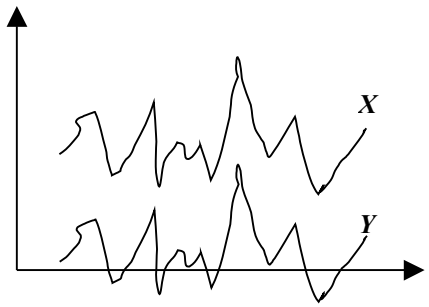

Рис. 1.

График демонстрирует, что существует некоторое соотношение между двумя экономическими величинами (линейная комбинация $\alpha x_t + \beta y_t$), которое является стационарным процессом. Соотношение можно переписать в виде $y_t = \gamma x_t + u_t$, где u_t – стационарный процесс. Если взять математическое ожидание от обеих 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. Коинтеграция совместима с понятием долгосрочного равновесия. Хотя каждый из нестационарных процессов «блуждает» случайным образом, наличие коинтеграции «заставляет» 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 коинтеграцию 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 коинтегрированы), 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 коинтегрирующего соотношения, то просто бессмысленно называть равновесием то значение, к которому процесс практически никогда не вернется.

Следовательно, коинтегрирующее 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 коинтегрирующим соотношением. Посмотрим, как описывается краткосрочное поведение. Пусть коинтегрирующее соотношение представляется в нормированном виде следующим образом: $y_t = \gamma x_t + \varepsilon_t$, где $\varepsilon_t \sim WN$ – белый шум. Поскольку $x_t, y_t \sim I(1)$ являются процессами $I(1)$, то процессы их разностей Δx_t и Δy_t являются процессами типа $I(0)$, т.е. стационарными. Содержательно Δx_t и Δy_t описывают собой краткосрочные изменения исходных переменных. Поэтому уже знакомая нам модель коррекции ошибками (ECM) $\Delta y_t = \alpha \cdot \Delta x_t + \beta(y_{t-1} - \gamma x_{t-1}) + \nu_t$ связывает между собой стационарную величину Δy_t, стационарную Δx_t и стационарное коинтегрирующее 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. Выяснилось, 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 коинтегрированных переменных.

Теперь убедимся, что коинтегрированные процессы существуют. Рассмотрим, следуя Энглу и Грэнжеру [4, p. 263], следующий пример. Пусть случайные процессы X_t и Y_t определены следующим образом. $X_t + \beta Y_t = u_t$, $X_t + \alpha Y_t = \nu_t$. Пусть также $u_t = u_{t-1} + \varepsilon_{1t}$ и $\nu_t = \rho\nu_{t-1} + \varepsilon_{2t}$, где ε_{1t} и ε_{2t} – некоррелированные белые шумы. При условии, что $|\rho| < 1$, ν_t является стационарным процессом, а u_t – случайным блужданием без дрейфа. Другими словами: $u_t \sim I(1)$, $\nu_t \sim I(0)$. Совместное распределение X_t и Y_t полностью определяется приведенными уравнениями. Отметим, что имеет смысл рассматривать модель только при $\alpha \neq \beta$, так как иначе система противоречива.

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:

$$X_t = \frac{\alpha}{\alpha - \beta} u_t - \frac{\beta}{\alpha - \beta} v_t \sim I(1)$$
$$Y_t = -\frac{1}{\alpha - \beta} u_t + \frac{1}{\alpha - \beta} v_t \sim I(1) \cdot$$

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 $X_t + \alpha Y_t$, 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 предыдущей главы, рассмотрим модель

$$\begin{pmatrix} X_t^1 \\ X_t^2 \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} + \underbrace{\begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix}}_{A_1} \begin{pmatrix} X_{t-1}^1 \\ X_{t-1}^2 \end{pmatrix} + \begin{pmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \end{pmatrix}.$$

В предыдущей главе 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 в предыдущей лекции, определим вектор новых переменных соотношением $\vec{y}_t = C^{-1}\vec{x}_t$, где C – матрица, столбцы которой являются собственными векторами, соответствующими собственным числам λ_1 и λ_2 соответственно.

В новых переменных система уравнений распадается на два отдельных уравнения:

$$y_t^1 = \lambda_1 y_{t-1}^1 + \eta_t^1$$
$$y_t^2 = \lambda_2 y_{t-1}^2 + \eta_t^2,$$

где через $\vec{\eta}_t$ 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 $\vec{y}_t$ не приводится к стационарному процессу взятием последовательных разностей. Но тогда этим свойством обладают и компоненты процесса $\vec{x}_t$, так как его компоненты яв-

ляются линейными комбинациями компонент вектора $\vec{y}_t$ (напомним, что $\vec{x}_t = C\vec{y}_t$). Поэтому в дальнейшем мы не будем рассматривать случай «взрывной» нестационарности. Если же оба корня по модулю меньше единицы, то компоненты процесса $\vec{x}_t$ 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 $\vec{x}_t$, как линейные комбинации процессов нулевого и первого порядков интеграции, будут первого порядка интеграции. Но нельзя забывать, что справедливо и обратное: компоненты вектора $\vec{y}_t$ являются линейными комбинациями компонент вектора $\vec{x}_t$. 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.

В предыдущей лекции мы ввели в рассмотрение матрицу $\Pi = I - A_1$ и отметили, что собственные векторы матриц Π 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 Π, а это, в свою очередь, эквивалентно равенству единице ранга матрицы Π. Используя диагональное представление матрицы Π, получаем:

$$\Pi = C\begin{pmatrix} 0 & 0 \\ 0 & (1-\lambda_2) \end{pmatrix} C^{-1} = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & (1-\lambda_2) \end{pmatrix} \begin{pmatrix} \tilde{c}_{11} & \tilde{c}_{12} \\ \tilde{c}_{21} & \tilde{c}_{22} \end{pmatrix} = \begin{pmatrix} (1-\lambda_2)c_{12} \\ (1-\lambda_2)c_{22} \end{pmatrix} \begin{pmatrix} \tilde{c}_{21} & \tilde{c}_{22} \end{pmatrix}.$$

Здесь через c_{ij} обозначены элементы матрицы C, а через $\tilde{c}_{ij}$ – элементы матрицы C^{-1}. Обратим внимание, что матрица Π разложена в произведение (факторизована) 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 Π.

Пусть теперь $\lambda_1 = \lambda_2$. 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: $J = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$. Определив вектор $\vec{y}_t = P^{-1}\vec{x}_t$ ($\vec{x}_t = P\vec{y}_t$), получаем систему соотношений:

$$y_t^1 = \lambda_1 y_{t-1}^1 + y_{t-1}^2 + \eta_t^1$$
$$y_t^2 = \lambda_2 y_{t-1}^2 + \eta_t^2.$$

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 $\vec{x}_t = P\vec{y}_t$, 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: $x_1, x_2 \sim CI(2,1)$. Оба собственных числа матрицы Π равны в этом случае нулю, а ее ранг равен 1. В этом случае матрица $\Pi = P\begin{pmatrix} 0 & -1 \\ 0 & 0 \end{pmatrix}P^{-1} = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix}\begin{pmatrix} 0 & -1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} \tilde{p}_{11} & \tilde{p}_{12} \\ \tilde{p}_{21} & \tilde{p}_{22} \end{pmatrix} = \begin{pmatrix} -p_{11} \\ -p_{21} \end{pmatrix}\begin{pmatrix} \tilde{p}_{21} & \tilde{p}_{22} \end{pmatrix}$ также факторизуется в произведение прямоугольных матриц, и вторая из них дает 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 Π приводятся к диагональной форме, и система уравнений для вектора $\vec{y}_t$ принимает вид: $y_t^1 = y_{t-1}^1 + \eta_t^1$, $y_t^2 = y_{t-1}^2 + \eta_t^2$. Обе компоненты относятся к типу $I(1)$, поэтому обе компоненты вектора x_t^1 также относятся к типу $I(1)$, а коинтегрирующей комбинации нет. Заметим, что матрица Π в этом случае состоит из одних нулей, и ее ранг равен 0.

Теперь рассмотрим случай модели VAR(1) для трехмерного случайного вектора, матрица A_1 имеет 3 собственных числа, и набор возможностей несколько расширяется. Начнем со случая, когда $|\lambda_1| < 1$, $|\lambda_2| < 1$, $\lambda_3 = 1$ и $\lambda_2 \neq \lambda_1$. Тогда матрица C из собственных векторов существует, и система уравнений для компонент вектора $\vec{y}_t$ распадается на отдельные уравнения:

$$y_t^1 = \lambda_1 y_{t-1}^1 + \eta_t^1$$
$$y_t^2 = \lambda_2 y_{t-1}^2 + \eta_t^2$$
$$y_t^3 = y_{t-1}^3 + \eta_t^3.$$

Очевидно, что третья компонента будет типа $I(1)$, а две остальные – типа $I(0)$. Поэтому все компоненты вектора $\vec{x}_t$ будут типа $I(1)$, в то же время две их линейные комбинации, соответствующие переменным y_t^1 и y_t^2, будут стационарными. Это означает, что переменные x_t^1, x_t^2, x_t^3 коинтегрированы, и существует два линейно независимых коинтегрирующих вектора. Эти два вектора определяют линейное подпространство, каждый элемент которого будет коинтегрирующим вектором. Эта ситуация препятствует непосредственной интерпретации коинтегрирующих векторов как долгосрочных зависимостей между переменными, ведь теперь таких соотношений бесконечно много.

Ранг матрицы Π равен 2 и вновь совпадает с числом линейно независимых коинтегрирующих векторов. Ничего принципиально не меняет случай, когда $\lambda_1 = \lambda_2$. В этом случае y_t^3 по-прежнему процесс типа $I(1)$, y_t^2 – типа $I(0)$, а y_t^1 определяется моделью ADL(1, 1) и также является стационарным, т.е. типа $I(0)$. Все компоненты вектора $\vec{x}_t$ являются процессами типа $I(1)$, а переменные y_t^2 и y_t^3 являются стационарными линейно независимыми комбинациями компонент вектора $\vec{x}_t$, т.е. существует два линейно независимых коинтегрирующих вектора.

Матрица Π легко приводится либо к диагональному виду:

$\begin{pmatrix} 1-\lambda_1 & 0 & 0 \\ 0 & 1-\lambda_2 & 0 \\ 0 & 0 & 0 \end{pmatrix}$, либо к виду $\begin{pmatrix} 1-\lambda & -1 & 0 \\ 0 & 1-\lambda & 0 \\ 0 & 0 & 0 \end{pmatrix}$, что дает соответственно

$$\Pi = C\begin{pmatrix} 1-\lambda_1 & 0 & 0 \\ 0 & 1-\lambda_2 & 0 \\ 0 & 0 & 0 \end{pmatrix}C^{-1} = \begin{pmatrix} (1-\lambda_1)\begin{pmatrix} 1 \end{pmatrix} & (1-\lambda_2)\begin{pmatrix} 2 \end{pmatrix} \end{pmatrix}\begin{pmatrix} \begin{pmatrix} 1 \end{pmatrix} \\ \begin{pmatrix} 2 \end{pmatrix} \end{pmatrix},$$

либо $\Pi = P\begin{pmatrix} 1-\lambda & -1 & 0 \\ 0 & 1-\lambda & 0 \\ 0 & 0 & 0 \end{pmatrix}P^{-1} = \begin{pmatrix} (1-\lambda)\begin{pmatrix} 1 \end{pmatrix} & (1-\lambda)\begin{pmatrix} 2 \end{pmatrix} - \begin{pmatrix} 1 \end{pmatrix} \end{pmatrix}\begin{pmatrix} \begin{pmatrix} 1 \end{pmatrix} \\ \begin{pmatrix} 2 \end{pmatrix} \end{pmatrix}.$

Через $\begin{pmatrix} 1 \end{pmatrix}$ и $\begin{pmatrix} 2 \end{pmatrix}$ обозначены первый и второй столбцы матрицы C или P, а через $\begin{pmatrix} 1 \end{pmatrix}$ и $\begin{pmatrix} 2 \end{pmatrix}$ – соответствующие строки матрицы C^{-1} или P^{-1}. В обоих случаях матрица Π разложена в произведение прямоугольных матриц размерностью $(n \times r)$ и $(r \times n)$ соответственно, где n – размерность вектора $\vec{x}_t$, а r – ранг матрицы Π. При этом строки второй из матриц являются коинтегрирующими векторами. В литературе почему-то принято обозначать эти матрицы строчными греческими буквами α и β^T, и факторизация записывается в виде $\Pi = \alpha\beta^T$.

Теперь рассмотрим случай, когда $\lambda_1 = \lambda_2 = 1$, а $|\lambda_3| < 1$. Если при этом кратному собственному числу $\lambda_1 = 1$ соответствует только один собственный вектор, то система уравнений для вектора $\vec{y}_t = P^{-1}\vec{x}_t$ принимает вид:

$$y_t^1 = y_{t-1}^1 + y_{t-1}^2 + \varepsilon_t^1$$

$$y_t^2 = y_{t-1}^2 + \varepsilon_t^2$$

$$y_t^3 = \lambda_3 y_{t-1}^3 + \varepsilon_t^3.$$

Очевидно, что $y_t^3 \sim I(0)$, $y_t^2 \sim I(1)$, $y_t^1 \sim I(2)$. Все компоненты вектора $\vec{x}_t$ относятся тогда к типу $I(2)$, и существуют две коинтегрирующие линейные комбинации. Одна из них, соответствующая y_t^2, понижает порядок интеграции до 1, а вторая, соответствующая y_t^3, понижает порядок до 0. Ранг матрицы Π равен 2, ее факторизация в виде $\Pi = \alpha\beta^T$ опять имеет место.

Если же кратному единичному собственному значению соответствует два линейно независимых собственных вектора, то VAR-модель для вектора $\vec{y}_t$ распадается на уравнения для отдельных компонент.

$$y_t^1 = y_{t-1}^1 + \varepsilon_t^1$$
$$y_t^2 = y_{t-1}^2 + \varepsilon_t^2$$
$$y_t^3 = \lambda_3 y_{t-1}^3 + \varepsilon_t^3.$$

Тогда $y_t^3 \sim I(0)$, y_t^2, $y_t^1 \sim I(1)$. Компоненты вектора $\vec{x}_t$ будут процессами типа $I(1)$, и существует лишь одна коинтегрирующая комбинация, соответствующая y_t^3. Ранг матрицы Π равен 1, и она может быть представлена в виде произведения вектора-столбца на вектор-строку, которая и является коинтегрирующим вектором.

Осталось рассмотреть случай, когда $\lambda_1 = \lambda_2 = \lambda_3 = 1$. Этому трехкратному собственному числу может соответствовать один, два или три линейно независимых собственных вектора. Если такой вектор один, то для вектора $\vec{y}_t$ получается следующая система:

$$y_t^1 = y_{t-1}^1 + y_{t-1}^2 + \eta_t^1$$
$$y_t^2 = y_{t-1}^2 + y_{t-1}^3 + \eta_t^2$$
$$y_t^3 = y_{t-1}^3 + \eta_t^3.$$

Компонента $y_{t-1}^3 \sim I(1)$, компонента $y_{t-1}^2 \sim I(2)$, и компонента $y_{t-1}^1 \sim I(3)$. Все компоненты вектора $\vec{x}_t$ будут типа $I(3)$, и существует две коинтегрирующие комбинации, имеющие порядок интеграции 2 (соответствует y_t^2) и 1 (соответствует y_t^3). Ранг матрицы Π вновь равен числу линейно независимых коинтегрирующих векторов, т.е. двум.

Если собственному числу соответствует два собственных вектора, то для вектора $\vec{y}_t$ получается следующая система:

$$y_t^1 = y_{t-1}^1 + y_{t-1}^2 + \eta_t^1$$
$$y_t^2 = y_{t-1}^2 + \eta_t^2$$
$$y_t^3 = y_{t-1}^3 + \eta_t^3.$$

Компоненты $y_{t-1}^2, y_{t-1}^3 \sim I(1)$ и компонента $y_{t-1}^1 \sim I(2)$. Все компоненты вектора $\vec{x}_t$ будут типа $I(2)$, и существует только одна коинтегрирующая комби-

нация порядка интеграции 1 (соответствует y_t^2). Ранг матрицы П вновь равен числу линейно независимых коинтегрирующих векторов, т.е. единице.

Наконец, если собственному числу соответствуют три линейно независимых собственных вектора, то система уравнений для вектора $\vec{y}_t$ распадается на 3 отдельных уравнения:

$$y_t^1 = y_{t-1}^1 + \eta_t^1$$
$$y_t^2 = y_{t-1}^2 + \eta_t^2$$
$$y_t^3 = y_{t-1}^3 + \eta_t^3.$$

Все компоненты вектора $\vec{y}_t$ являются интегрированными процессами порядка 1. Все компоненты вектора $\vec{x}_t$ также будут типа $I(1)$, и не существует ни одной коинтегрирующей комбинации. Ранг матрицы П, состоящей теперь из одних нулей, вновь равен числу линейно независимых коинтегрирующих векторов, т.е. нулю.

Оказывается, результат рассмотрения удобнее и элегантнее выразить в терминах ранга матрицы П, чем в количестве единичных корней матрицы A_1. Мы фактически установили, что если двумерный или трехмерный случайный процесс имеет представление в виде модели VAR(1) и среди его характеристических корней ни один по модулю не превышает единицы, 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 П 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. И, наконец, рассмотрим общий случай модели VAR(p): $\vec{y}_t = A_1\vec{y}_{t-1} + ... + A_p\vec{y}_{t-p} + \vec{\varepsilon}_t$. Это уравнение можно переписать в другом виде после таких же преобразований, как и при выводе уравнения ADF-теста:

$$\Delta\vec{y}_t = -\Pi\vec{y}_{t-1} + B_1\Delta\vec{y}_{t-1} + ... + B_{p-1}\Delta\vec{y}_{t-p+1} + \vec{\varepsilon}_t.$$

Здесь матрица $\Pi = I - A_1 - A_2 - ... - A_p$, а матрицы $B_1,.., B_{p-1}$ выражаются через матрицы $A_1,..., A_p$ следующим образом: $B_j = -\sum_{i=j+1}^{p} A_i$. Доказательство того, что наличие коинтеграции связано с вырожденностью матрицы П, или, более точно, что количество коинтегрирующих векторов в точности равно рангу матрицы П, требует использования блочных матриц и останется за рамками нашего курса. Доказательство можно найти, например, для $p = 2$ в учебнике Джонстона и Ди Нардо [12]. 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 наи-

более сложной технически. Для исходного уравнения, как мы уже упоминали, характеристическое уравнение имеет вид: $\det(\lambda^p I - \lambda^{p-1} A_1 - ... - A_p) = 0$. Следовательно, существование единичного корня у исходного уравнения эквивалентно существованию нулевого корня у вышеприведенной матрицы Π. С помощью рассуждений, аналогичных проведенным выше, мы приходим к следующему заключению. Ранг матрицы $\Pi = I - A_1 - A_2 - ... - A_p$ равен числу линейно независимых коинтегрирующих векторов. Этот показатель принято называть *коинтегрирующим рангом*.

Итак, если между компонентами вектора $\vec{y}_t$ существуют коинтегрирующие соотношения, то их количество равно количеству нулевых собственных чисел матрицы Π, причем не все коинтегрирующие соотношения дают стационарные линейные комбинации.

Лекция 15

Коинтеграционная регрессия и тестирование коинтеграции

До сих пор мы рассматривали коинтеграцию как некоторое сугубо теоретическое свойство. Посмотрим, что меняет наличие этого свойства для оценивания регрессий с нестационарными переменными. Вернемся к рассмотренному в предыдущей лекции примеру Грэнжера и Энгла: случайные процессы X_t и Y_t определены следующими соотношениями: $X_t + \beta Y_t = u_t$, $X_t + \alpha Y_t = \nu_t$, где $u_t = u_{t-1} + \varepsilon_{1t}$, и $\nu_t = \rho \nu_{t-1} + \varepsilon_{2t}$ (ε_{1t} и ε_{2t} – некоррелированные белые шумы). Мы уже установили, что X_t и Y_t коинтегрированы. Пусть у нас есть выборка объема T из обеих переменных. 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.

Давайте посмотрим, почему это так. Оценка коэффициента наклона в регрессии $X_t = \alpha + \gamma Y_t + \varepsilon_t$ будет выражена известным выражением $\hat{\gamma} = \dfrac{\sum x_t y_t}{\sum y_t^2}$. Поскольку математические ожидания обеих переменных равны нулю, можно заменить в этом выражении строчные буквы прописными. Используем приведенную форму нашей системы:

$$X_t = \frac{\alpha}{\alpha - \beta} u_t - \frac{\beta}{\alpha - \beta} \nu_t$$

$$Y_t = -\frac{1}{\alpha - \beta} u_t + \frac{1}{\alpha - \beta} \nu_t .$$

Используя теорему Слуцкого и тот факт, что дисперсия стационарного процесса v_t постоянна, а дисперсия процесса случайного блуждания u_t стремится к бесконечности, получим:

$$p\lim \hat{\gamma} = \frac{p\lim\left(\frac{1}{T}\sum x_t y_t\right)}{p\lim\left(\frac{1}{T}\sum y_t^2\right)} = \frac{Cov\left(\frac{\alpha}{\alpha-\beta}u_t - \frac{\beta}{\alpha-\beta}v_t, -\frac{1}{\alpha-\beta}u_t + \frac{1}{\alpha-\beta}v\right)}{\mathrm{var}\left(-\frac{1}{\alpha-\beta}u_t + \frac{1}{\alpha-\beta}v\right)} =$$

$$= \frac{-\alpha\ \mathrm{var}(u_t) - \beta\ \mathrm{var}(v_t)}{\mathrm{var}(u_t) + \mathrm{var}(v_t)} \xrightarrow[T\to\infty]{} -\alpha$$

Следовательно, оценка МНК коэффициента наклона является состоятельной оценкой нормированного коинтегрирующего соотношения. Более того, можно показать, что скорость сходимости оценки выше, чем в случае оценки МНК для стационарных регрессий [22]. Поскольку дисперсия случайного блуждания растет пропорционально T, скорость сходимости при наличии коинтеграции пропорциональна $\frac{1}{T}$, в то время как для стационарных регрессоров скорость сходимости пропорциональна только $\frac{1}{\sqrt{T}}$. Это свойство оценок МНК при наличии коинтеграции называют *суперсостоятельностью*.

Однако регрессия X_t на Y_t использует не всю информацию, содержащуюся в исходной модели. Умножим второе из уравнений приведенной формы на α и сложим с первым уравнением. Получим $X_t + \alpha Y_t = \upsilon_t = \rho\upsilon_{t-1}$. Стандартным образом получаем:

$$X_t + \alpha Y_t - \rho(X_{t-1} + \alpha Y_{t-1}) = \upsilon_t - \rho\upsilon_{t-1} = \varepsilon_{2t},$$
$$\text{или } \Delta X_t = -\alpha\Delta Y_t + (\rho - 1)(X_{t-1} + \alpha Y_{t-1}) + \varepsilon_{2t}.$$

Полученное выражение представляет собой уже знакомую нам модель коррекции ошибками, в котором коинтегрирующее соотношение играет роль долгосрочного соотношения между переменными. Эта модель соответствует модели ADL(1,1) с нестационарными регрессорами, а именно: $X_t = \rho X_{t-1} - \alpha Y_t - \rho\alpha Y_{t-1} + \varepsilon_{2t}$. Понятно, что ECM и ADL(1,1) являются различными параметризациями одной и той же модели. При этом, если оценивать модель в виде ECM методом наименьших квадратов, то как объясняемая переменная ΔX_t, так и регрессоры ΔY_t и $(X_{t-1} + \alpha Y_{t-1})$ являются стационарными. Поэтому оценки коэффициентов ECM-представления заведомо состоятельны. Поскольку оценки ADL-представления однозначно пересчитываются через оценки ECM-представления, они также состоятельны.

Общий результат, полученный Симсом, Стоком и Уотсоном [21], состоит в следующем. Если при некоторой параметризации модели параметр может быть выражен как коэффициент при переменной типа $I(0)$, обычные тестовые статистики для этого коэффициента асимптотически справедливы. Более того, если

при некоторой параметризации модели подмножество параметров является коэффициентами при переменных типа $I(0)$, обычные тестовые статистики для этого подмножества коэффициентов асимптотически справедливы. Доказательство этого чрезвычайно важного результата выходит за рамки нашего курса. Обратите внимание, что для применимости обычных тестовых процедур и таблиц распределений достаточно принципиальной возможности подходящей параметризации, вовсе не требуется оценивать коэффициенты именно этой параметризации. Например, это означает, что оценки методом наименьших квадратов ECM и ADL представлений асимптотически эквивалентны, если переменные коинтегрированы. Впрочем, в конечных выборках эквивалентность не гарантирована, и существует обширная литература, где обсуждаются преимущества той или иной параметризации в конечных выборках.

Общий вывод таков: одни и те же методы оценивания и диагностические процедуры применимы как для стационарных, так и для нестационарных регрессоров типа $I(1)$, если последние коинтегрированы. Поэтому ключевым вопросом является тестирование наличия коинтеграции нестационарных процессов. Первая тестовая *двухшаговая* процедура была предложена Энглом и Грэнжером [4]. Рассмотрим ее на примере регрессии между двумя рядами типа $I(1)$.

1 шаг. Убеждаемся, что оба регрессора относятся к типу $I(1)$. Строим обычную регрессию переменной X_t на переменную Y_t методом наименьших квадратов.

2 шаг. Проверяем остатки регрессии на стационарность. Если остатки являются стационарным временным рядом, то процессы X_t и Y_t коинтегрированы. Если же остатки относятся к типу $I(1)$, то коинтеграция отсутствует. Если процессы коинтегрированы, то регрессия, построенная на первом шаге, называется *коинтегрирующей (cointegrating regression).*

Для проверки порядка интеграции остатков авторы предложили применять тест Дикки–Фуллера, но критические значения таблиц этого теста в данном случае не годятся. Вспомните, что таблицы получены в результате моделирования Монте–Карло для сгенерированного случайного блуждания. В этом же случае исследуемый процесс получен в результате процедуры метода наименьших квадратов, т.е. является «оцененным». В частности, сумма ошибок равна 0. Таблицы критических значений для процедуры Энгла–Грэнжера были приведены в работе Мак-Киннона [14]. Они больше по абсолютной величине, чем критические значения статистики Дикки–Фуллера τ_μ для того же уровня значимости. Другими словами, критические значения теста на наличие коинтеграции лежат левее критических значений теста Дикки–Фуллера. Например, на уровне значимости 5% для $T = 100$, отсутствия лагов и тренда в модели критическое значение для теста Дикки–Фуллера равно $-1{,}945$, а для теста на наличие коинтеграции соответствующее значение равно $-3{,}396$. Поскольку нулевая гипотеза заключается в наличии единичного корня у ряда остатков, т.е. в *отсутствии коинтеграции*, критические значения смещены в сторону гипотезы о наличии коинтеграции. В вышеупомянутой работе Мак-Киннон предложил аппроксимирующую формулу для расчета критических чисел, которая и используется в большинстве специализированных компьютерных программ. Эта аппроксимирующая функция носит название поверхности отклика и представляется следующим образом:

$$C(\alpha, T) = k_0 + \frac{k_1}{T} + \frac{k_2}{T^2},$$

где $C(\alpha,T)$ – односторонняя α-процентная точка для выборки размера T, k_0 – асимптотическое значение критической величины, а два вторых слагаемых выражают поправки на конечность выборки. Мак-Киннон привел таблицу значений параметров поверхности отклика k_0, k_1, k_2 в зависимости от числа наблюдений T и количества регрессоров в тестовом уравнении без учета константы. Включение и невключение линейного тренда в исходное уравнение регрессии дает два набора таблиц критических значений.

Линейный тренд также может быть добавлен в коинтегрирующее соотношение, что требует дополнительных таблиц. Необходимость включения линейного тренда в коинтегрирующее соотношение может возникнуть в следующем случае. Исследование нестационарности ADF-тестом может показать наличие детерминированного тренда и единичного корня одновременно у одной из исследуемых переменных. Если вторая переменная (или не все из переменных в общем случае) не содержит детерминированный тренд, следует включить его в коинтегрирующее соотношение, в противном случае невозможно обеспечить «баланс» слагаемыми различного типа в коинтегрирующем уравнении. Если более чем одна переменная содержит наряду с единичным корнем детерминированный тренд, то эти детермированные тренды могут, вообще говоря, компенсировать друг друга. Но чтобы допустить возможность отсутствия такой компенсации и тестировать ее, следует все же включить детерминированный тренд в коинтегрирующее соотношение. Ситуацию, когда две переменные обе содержат детерминированный тренд, и эти тренды компенсируют друг друга, иногда называют *детерминированной коинтеграцией*.

Отметим также, что существуют ситуации, когда экономическая теория предписывает определенный коэффициент в долгосрочном соотношении (равновесии). Например, в классической теории потребления по Кейнсу предполагается единичная долгосрочная эластичность потребления по располагаемому доходу. Для данных, выраженных в логарифмах, это означает, что нам заранее известен коинтегрирующий вектор (1 −1). В этом случае на первом шаге процедуры Энгла–Грэнжера нет нужды оценивать регрессию, и остатки получаются просто вычитанием переменных. Теперь проверка гипотезы о наличии коинтеграции может быть проведена непосредственно по таблицам теста Дикки–Фуллера, так как остатки более не являются «оценкой».

Если тест Энгла–Грэнжера показал наличие коинтеграции, то возможно описать в модели не только долгосрочное равновесие, но и динамику движения к этому долгосрочному равновесию. Для этого разумно добавить в исходную регрессию дополнительный регрессор – коинтегрирующее соотношение в предыдущий момент времени. Это означает, что мы строим Error correction model. Обратите внимание, насколько полезней этот подход подхода Бокса–Дженкенса, из которого следует, что если процессы оказываются нестационарными типа $I(1)$, то во избежание кажущейся регрессии мы строим регрессию первой разности ΔX_t на первую разность ΔY_t. При переходе к разностям всякая информация о долгосрочной взаимосвязи рядов теряется. Понятие коинтеграции помогает резко улучшить си-

туацию. Теперь кроме составляющих, описывающих краткосрочное поведение, полученная модель связывает и долгосрочное, и краткосрочное поведение.

Мы уже упоминали, что наличие коинтеграции позволяет оценивать модель и в виде ADL-представления. Это представление также отражает и краткосрочное, и долгосрочное поведение, но его параметры уже не допускают непосредственной интерпретации. Но для того, чтобы быть уверенным, что оценивание ADL-модели приведет к состоятельным оценкам ее параметров, нужно предварительно установить наличие коинтеграции переменных. Оба метода оценки асимптотически эквивалентны, но в конечных выборках дают несколко разные оценки. Сток [22] показал, что в конечных выборках процедура Энгла–Грэнжера дает несколько большее смещение оценок коэффициентов.

Тест Йохансена

Основные из применяемых в настоящее время тестов наличия коинтеграции [24] основаны на связи количества коинтегрирующих векторов с рангом матрицы Π в VAR-представлении многомерного временного ряда, а следовательно, с количеством нулевых собственных чисел этой матрицы. Мы рассмотрим лишь один из тестов, предложенный Йохансеном в серии работ [9, 11, 12]. 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

$$\vec{Y}_t = \{\vec{\mu} + t\vec{\beta}\} + A_1\vec{Y}_{t-1} + ... + A_p\vec{Y}_{t-p} + \vec{\varepsilon}_t .$$

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 гомоскедастичны и не коррелированны. 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 [10]. 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 [7]:

$$\ln l = const + \frac{T}{2}\ln\left|\hat{\Omega}^{-1}\right|,$$

где $\hat{\Omega}$ – оценка ковариационной матрицы остатков VAR-модели. Отношение правдоподобий поэтому принимает вид:

$$LR = -2(\ln l_q - \ln l_p) = T(\ln\left|\hat{\Omega}_q^{-1}\right| - \ln\left|\hat{\Omega}_p^{-1}\right|).$$

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 $\Pi = I - A_1 - A_2 - ... - A_p$, соответствующая ECM-представлению модели $\Delta\vec{y}_t = \{\vec{\mu} + t\vec{\beta}\} + -\Pi\vec{y}_{t-1} + B_1\Delta\vec{y}_{t-1} + ... + B_{p-1}\Delta\vec{y}_{t-p+1} + \vec{\varepsilon}_t$. Нулевой гипотезой является то, что ранг матрицы Π не превышает некоторого числа $r < k$. В качестве альтернативной гипотезы используется $H_1 : rank\,\Pi = k$, или $H_1 : rank\,\Pi = r+1$. Статистика для проверки нулевой гипотезы против первой из приведенных альтернативных имеет вид $-T\sum_{i=r+1}^{k}\ln(1-\tilde{\lambda}_i)$ и называется *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: $-T\ln(1-\tilde{\lambda}_{r+1})$ и называется *max statistic*. Название связано с тем, что статистика пропорциональна логарифму максимального из собственных чисел матрицы, считающихся нулевыми. В приведенных выражениях через $\tilde{\lambda}_i$ обозначена оценка максимального правдоподобия i-го корня полученного Йохансеном уравнения. При этом предполагается, что корни упорядочены в порядке убывания.

В пакете Eviwes применяется *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 обеих 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 достаточны для отказа от такой спецификации теста Йохансена (первой 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. Раз коинтеграция существует, то матрица $\Pi = \alpha\beta^T$ может быть факторизована, и ECM представление в этом случае может быть записано в виде: $\Delta\vec{y}_t = -\alpha\begin{pmatrix}\beta^T\\ \vec{\gamma}^T\end{pmatrix}\begin{pmatrix}\vec{y}_{t-1}\\ 1\end{pmatrix} + B_1\Delta\vec{y}_{t-1} + ... + B_{p-1}\Delta\vec{y}_{t-p+1} + \vec{\varepsilon}_t$. Через $\begin{pmatrix}\beta^T\\ \vec{\gamma}^T\end{pmatrix}$ здесь обозначена блочная матрица, нижний блок которой является вектор-строкой, составленной из входящих в коинтегрирующие соотношения констант. Вектор $\begin{pmatrix}\vec{y}_{t-1}\\ 1\end{pmatrix}$, в свою очередь, содержит в нижнем блоке число 1. Другими словами, в ECM-модели константы входят только в «долгосрочную часть» и отсутствуют в «краткосрочной». При переходе от ECM-представления к VAR-модели мы должны «смириться» с присутствием свободного члена и в VAR-модели.

Кроме того, само VAR-представление может включать свободный член, даже если долгосрочное равновесие не подразумевает константы. Ведь априори не всегда есть основания отрицать ненулевое математическое ожидание всех компонент многомерного случайного процесса. В этом случае переход к ECM-представлению даст вектор констант в «краткосрочной части» модели, независимо от наличия/отсутствия констант в «долгосрочной части». Но если компоненты вектора данных интегрированы, то наличие константы в уравнении краткосрочного поведения означает наличие дополнительного детерминированного тренда в данных. Ситуация полностью аналогична случайному блужданию с дрейфом. Таким образом, наличие вектора свободных членов в VAR-модели может соответствовать двум качественно очень разным случаям:

• если вектор свободных членов VAR-модели полностью переходит в коинтеграционные соотношения ECM-представления, то порожденные такой моделью данные не демонстрируют наличия детерминированного тренда;

• если вектор свободных членов VAR-модели переходит не только в коинтеграционные соотношения ECM-представления, но и в дополнительный вектор свободных членов «краткосрочной части», то порожденные такой моделью данные демонстрируют наличие детерминированного тренда.

Поэтому, если визуальная инспекция данных и предварительные оценки свидетельствуют об отсутствии в данных линейного тренда, а в долгосрочном соотношении тренд должен присутствовать, мы выбираем вторую из пяти вышеуказанных возможностей. Если же анализ свидетельствует в пользу наличия детерминированного тренда в данных, то мы предпочитаем третью возможность.

Аналогично рассматривается случай включения линейного тренда в коинтегрирующее соотношение. Если линейный тренд из VAR-модели при переходе в ECM-представление входит только в «долгосрочную часть», то порождаемые данные должны демонстрировать детерминированный линейный тренд, и для тестирования коинтеграции следует использовать четвертую из возможностей. Если же тренд переходит и/или в «краткосрочную часть», данные должны демонстрировать уже квадратичный детерминированный тренд. Для такого случая предназначена пятая возможность применения теста Йохансена. К сожалению, в литературе и в описаниях компьютерных пакетов эти пять опций описаны в терминах отсутствия/наличия константы и тренда в VAR-модели и коинтегрирующем уравнении, хотя по сути речь идет не о VAR, а о «краткосрочной части» ECM-представления. Наше рассмотрение показывает, что наиболее употребительными, подходящими к экономическим данным являются вторая и третья опции теста Йохансена.

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 β из разложения $\Pi = \alpha\beta^T$. Возможность факторизации матрицы явно указывает на наличие линейных соотношений между ее элементами. В свою очередь, это означает, что учет этих ограничений при оценивании даст более эффективные оценки коэффициентов модели. Такой учет может быть осуществлен путем оценки модели в форме ECM с прямым включением в нее 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 интерпритировать как экономические равновесия. Можно сослаться, например, на опыт

построения функции спроса на деньги Йохансеном и Джузелиусом [12] 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 дает 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.

Лекция 16

Модели с условной гетероскедастичностью

Моделирование цен активов на современных финансовых рынках часто приводит к тому, что временные ряды неплохо могут быть аппроксимированы моделью случайного блуждания. Этот эмпирический факт не противоречит теории эффективных рынков, но в то же время теория эффективных рынков допускает более широкий класс моделей. В этих моделях приращения процесса не обязательно должны быть независимы и иметь одинаковое распределение. Существенно, чтобы процесс был таким, что допускал бы так называемую «честную игру» (fair play), чтобы невозможно было достичь гарантированного выигрыша, основываясь на прошлой информации. Такое расширение класса процессов случайного блуждания достигается введением понятия *мартингала* (*martingale*). В английском языке слово martingale имеет два значения: часть конской сбруи, препятствующая лошади задирать голову слишком высоко, и система назначения ставок в азартной игре, при которой ставка удваивается при каждом проигрыше. Как мы увидим, свойства процесса-мартингала действительно позволяют провести некоторые аналогии с этими житейскими понятиями.

Формально мартингалом называется стохастический процесс y_t, обладающий следующими свойствами:

- математическое ожидание процесса конечно для любого момента времени t;
- Условное математическое ожидание процесса на основе всей информации, известной к моменту s, равно значению процесса в момент s: $E(y_t|\Omega_s) = y_s$ для всех $s \le t$, где Ω_s – совокупность всей информации к моменту s. Это свойство носит название мартингального.

При анализе одномерного дискретного временного ряда вся информация к некоторому моменту времени s представляет собой просто реализацию процесса на отрезке [1, s], и мартингальное свойство может быть переписано в виде $E(y_t|y_1 y_2,..., y_s) = y_s$ для всех $s \le t$. Это означает, что условное математическое ожидание приращения процесса равно нулю, и прогноз приращения мартингала, обеспечивающий минимальную среднеквадратическую ошибку, равен нулю. В соответствии с этим определением случайное блуждание является частным случаем мартингала. При анализе реальных финансовых временных рядов часто встречаются и иные процессы, для которых мартингальное равенство может быть

заменено неравенством $E(y_t|y_1, y_2, ..., y_s) \geq y_s$ для всех $s \leq t$. Такой процесс называется *субмартингалом*. Если же неравенство выполняется с противоположным знаком, процесс называется *супермартингалом*.

Исходя из мартингального свойства, мартингал может быть записан в эквивалентной форме, очень напоминающей случайное блуждание $y_t = y_{t-1} + \eta_t$, где η_t называется мартингальным приращением, или мартингальной разностью. Свойства процесса η_t должны быть такими, чтобы гарантировать выполнение мартингального равенства. При этом процесс η_t не обязательно должен быть стационарным, например, может существовать зависимость между условными моментами процесса, условная дисперсия мартингальной разности η_t может зависеть от прошлого, а процесс y_t останется мартингалом.

Практика работы на финансовых рынках выявила еще некоторые черты реальных процессов, которые не очень хорошо описывались уже разработанными моделями. В 1982 г. Энглом [3] была предложена модель, получившая название модели с условной гетероскедастичностью (autoregressive conditional heteroskedastisity, обычно используемая аббревиатура ARCH, или ARCH-модель). Энгл предложил эту модель, чтобы описать хорошо известную из практики черту поведения многих финансовых рядов: кластеризацию волатильности, когда периоды большой изменчивости показателя сменяются периодом малой изменчивости. В то же время статистическая проверка постоянства дисперсии (как меры изменчивости) не отвергает гипотезы о гомоскедастичности.

Начнем с простой авторегрессионной модели

$$y_t = \lambda y_{t-1} + \varepsilon_t, \ \varepsilon_t \sim SWN\ (0, \sigma^2).$$

Через $SWN\ (0, \sigma^2)$ обозначен так называемый сильный белый шум, для которого свойство некоррелированности заменено на свойство независимости. В этой модели условное математическое ожидание $E(y_t|y_{t-1}) = \lambda y_{t-1}$ представляет собой теоретическую регрессию и, естественно, зависит от времени *t*. Условная дисперсия $\text{var}(y_t|y_{t-1}) = \sigma^2$ от времени не зависит. Соответственно безусловное математическое ожидание равно нулю, а безусловная дисперсия равна $\frac{\sigma^2}{1-\lambda^2}$. Энгл предложил рассматривать зависимость условной дисперсии от прошлого, он специфицировал эту зависимость в простейшем случае в линейном виде:

$$h_t = \text{var}(\varepsilon_t) = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2.$$

Поскольку дисперсия случайной ошибки должна быть положительной, параметр α_0 также должен быть положительным, а параметр α_1 – неотрицательным. Процесс, определяемый соотношением

$$u_t = \varepsilon_t h_t^{1/2} = \varepsilon_t \left(\alpha_0 + \alpha_1 u_{t-1}^2\right)^{1/2}, \ \varepsilon_t \sim SWN\ (0, \sigma^2),$$

называется процессом ARCH(1). Параметр в скобках указывает наибольшую длину лага в уравнении для дисперсии. Рассмотрим свойства этого процесса.

Для сокращения записи обозначим условное математическое ожидание произвольного процесса y_t $E(y_t|y_1, y_2,..., y_{t-1})$ через $E_{t-1}(y_t)$. Вычисление условного математического ожидания процесса ARCH(1) дает

$$E_{t-1}(u_t) = \left(\alpha_0 + \alpha_1 u_{t-1}^2\right)^{1/2} E_{t-1}(\varepsilon_t) = 0 .$$

Применяя формулу итерированного случайного ожидания, получим

$$E(u_t) = E_0(E_1(...(E_{t-2}(E_{t-1}(u_t)))...) = 0 .$$

Следовательно, безусловное математическое ожидание процесса рано нулю.

Перейдем к расчету условной дисперсии процесса ARCH(1).

$$E_{t-1}(u_t^2) = \left(\alpha_0 + \alpha_1 u_{t-1}^2\right)E_{t-1}(\varepsilon_t^2) = \sigma^2\left(\alpha_0 + \alpha_1 u_{t-1}^2\right).$$

Далее $E_{t-2}(E_{t-1}(u_t^2)) = E_{t-2}(\sigma^2(\alpha_0 + \alpha_1 u_{t-1}^2)) =$

$$= \sigma^2(\alpha_0 + \alpha_1 E_{t-2}(u_{t-1}^2)) = \sigma^2(\alpha_0 + \alpha_1(\alpha_0 + \alpha_1 u_{t-2}^2)) = \sigma^2(\alpha_0 + \alpha_1\alpha_0 + \alpha_1^2 u_{t-2}^2).$$

Для нахождения безусловной дисперсии процесса вычисляем итерированные ожидания.

$$E(u_t^2) = E_0(E_1(...(E_{t-2}(E_{t-1}(u_t^2)))...) = \sigma^2\alpha_0\left(1 + \alpha_1 + \alpha_1^2 + ... + \alpha_1^{t-1}\right) + \sigma^2\alpha_1^t u_0^2 .$$

Если уравнение AR(1) для дисперсии стационарно, то $|\alpha_1| < 1$, и перейдя к пределу при $t \to \infty$, получим $\mathrm{var}(u_t) = \dfrac{\alpha_0}{1-\alpha_1}\sigma^2$. Полученный результат показывает, что безусловная дисперсия процесса ARCH(1) не зависит от времени, т.е. сам процесс гомоскедастичен.

Условная автоковариация первого порядка равна нулю, так как $E_{t-1}(u_t u_{t-1}) = u_{t-1}E_{t-1}(u_t) = 0$. Очевидно, что автоковариации более высокого порядка также равны нулю. Мы видим, что процесс ARCH(1) является стационарным белым шумом. Попутно обратим внимание, что хотя значения процесса ARCH(1) некоррелированы, существует зависимость дисперсии возмущения от дисперсии предыдущего возмущения. Это означает, что некоторые тесты могут воспринять условную гетероскедастичность как свидетельство автокорреляции остатков. В частности, таким «недостатком» обладает популярный тест Дарбина–Уотсона. Поэтому наличие условной гетероскедастичности делает этот тест неприменимым.

Уже обозначение ARCH(1) подразумевает обобщение на процесс ARCH(p). Этот процесс определяется уравнением типа AR(p) для дисперсии $h_t = \mathrm{var}(\varepsilon_t) = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + ... + \alpha_p\varepsilon_{t-p}^2$, что дает для процесса ARCH(p) следующее выражение:

$$u_t = \varepsilon_t h_t^{1/2} = \varepsilon_t\left(\alpha_0 + \alpha_1 u_{t-1}^2 + ... + \alpha_p u_{t-p}^2\right)^{1/2},\ \varepsilon_t \sim SWN\ (0, \sigma^2) .$$

Очевидно, что как условное, так и безусловное математическое ожидание процесса ARCH(p) равно нулю. Условная дисперсия процесса фактически задана его определением: $E_{t-1}(u_t^2) = \sigma^2\left(\alpha_0 + \alpha_1 u_{t-1}^2 + ... + \alpha_p u_{t-p}^2\right)$, а безусловная дис-

персия в результате применения формулы итерированного ожидания с последующим переходом к пределу будет сложной дробно-рациональной функцией от параметров $\alpha_0, \alpha_1, ..., \alpha_p$. Самое важное, что безусловная дисперсия не будет зависеть от времени, т.е. процесс по-прежнему будет гомоскедастичным. Так же сохранится свойство некоррелированности, поскольку это свойство не зависит от конкретного уравнения процесса, а лишь от равенства нулю условного математического ожидания. Таким образом, процесс ARCH(p) является случайным блужданием и мартингалом. Установленные свойства процесса ARCH(p) означают, что он является слабо стационарным.

Тестирование наличия условной гетероскедастичности в остатках модели или для исходного процесса не требует разработки новых специальных тестов. Нулевая гипотеза об отсутствии условной гетероскедастичности эквивалентна равенству нулю всех коэффициентов кроме коэффициента α_0 в уравнении дисперсии. Поэтому проверка состоит из следующих шагов.

1. Строим необходимую модель и получаем ряд остатков e_t. Если исследованию на условную гетероскедастичность подвергается одномерный ряд исходных данных, то этот шаг не нужен.

2. Методом наименьших квадратов оцениваем модель

$$e_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + ... + \alpha_p e_{t-p}^2 + \nu_t.$$

3. Проверяем гипотезу об адекватности регрессии, построенной на втором шаге: $H_0 : \alpha_1 = ... = \alpha_p = 0$ против естественной альтернативы $H_1 : \alpha_1^2 + ... + \alpha_p^2 > 0$.

Если нулевая гипотеза отвергается, то не только устанавливается присутствие условной гетероскедастичности, но и определяется длина наибольшего лага, входящего в уравнение (параметр p).

Оценивание моделей, в которых остатки проявляют условную гетероскедастичность, можно вести в том же ключе, что и для моделей с гетероскедастичными остатками, например после использования тестов Парка или Глейзера. Оцененное на втором шаге уравнение используется для расчета весов с последующим применением взвешенного метода наименьших квадратов или для преобразования переменных, что означает применение обобщенного метода наименьших квадратов с оцененной ковариационной матрицей. Так же как при использовании вышеупомянутых тестов Парка и Глейзера, существенной практической трудностью является возможность получения отрицательных или нулевых значений подкоренного выражения при расчете весов по обычной формуле

$$w_t = \frac{1}{\sqrt{\hat{\alpha}_0 + \hat{\alpha}_1 e_{t-1}^2 + ... + \hat{\alpha}_p e_{t-p}^2}}.$$

Чтобы избежать этого, иногда вводят ограничения на параметры, например задавая характер убывания коэффициентов в уравнении

$$e_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + ... + \alpha_p e_{t-p}^2 + \nu_t.$$

Боллерслев [1] предложил обобщение ARCH-модели, позволяющее достичь большей гибкости. Он предложил добавить в уравнение для дисперсии зависимость от предыдущих значений дисперсии:

$$h_t = \mathrm{var}(\varepsilon_t) = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + ... + \alpha_p \varepsilon_{t-p}^2 + \gamma_1 h_{t-1} + ... + \gamma_q h_{t-q}.$$

Эта модель получила название обобщенная (generalized) ARCH, GARCH(p,q) – обобщенная авторегрессионная модель с условной гетероскедастичностью. В ней условная дисперсия подчиняется модели ADL(p,q) и зависит от предыдущих значений условной дисперсии и квадратов ошибок. Используя операторные полиномы, модель GARCH(p,q) может быть переписана в виде:

$$h_t = \mathrm{var}(\varepsilon_t) = \alpha_0 + \alpha_p(L)\varepsilon_t^2 + \gamma_q(L)h_{t-1}.$$

Аналогично моделям типа ARMA можно перейти от модели GARCH(p,q) к ARCH-модели с бесконечным числом лагов:

$$h_t = \mathrm{var}(\varepsilon_t) = \frac{\alpha_0}{1-\gamma_q(L)} + \frac{\alpha_p(L)}{1-\gamma_q(L)}\varepsilon_t^2.$$

Такой переход возможен, если корни характеристического полинома $1-\gamma_q(L)$ лежат вне единичного круга (в данном случае удобней использовать характеристическое уравнение в такой форме). Если у полиномов $\alpha_p(L), \gamma_q(L)$ кроме того нет общих корней, то положительность дисперсии будет гарантирована, если все коэффициенты полинома бесконечной степени $\frac{\alpha_p(L)}{1-\gamma_q(L)}$ неотрицательны. Необходимые и достаточные условия этого были получены Нельсоном и Као [19]. Для процесса GARCH(1,1), заданного уравнением $h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \gamma_1 h_{t-1}$ и весьма популярного в настоящее время при моделировании финансовых рядов, эти условия означают неотрицательность всех трех параметров.

Установить свойства моментов процесса GARCH(p,q) и ограничения на коэффициенты модели, при которых процесс GARCH(p,q) стационарен, значительно сложнее, чем для ARCH-процессов. Общие условия стационарности для процесса GARCH(p,q) в терминах лаговых операторов были выведены Бужеролем и Пикаром [2]. Для процесса GARCH(1,1), заданного уравнением $h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \gamma_1 h_{t-1}$, Нельсон [18] показал, что как u_t, так и h_t будут строго стационарны тогда и только тогда, когда $E(\ln(\beta_1 + \alpha_1 \frac{u_t^2}{h_t})) < 0$.

Для оценивания процессов GARCH(p,q) применяется метод максимального правдоподобия в предположении о гауссовости сильного белого шума ε_t. При увеличении порядка процесса GARCH трудности оценивания параметров быстро нарастают, так что широкое применение нашел, пожалуй, только процесс GARCH(1,1).

Было предложено много модификаций моделей GARCH, подробное рассмотрение которых выходит за рамки нашего курса. Следуя Миллсу [15], рассмотрим на примере модели GARCH(1,1) лишь основные идеи, положенные в основу таких модификаций.

Одна из наиболее ранних модификаций была предложена Тейлором [23] и развита Свертом [20]. В этой модели та же схема зависимости, что и в GARCH(1,1)

применена к условному стандартному отклонению процесса, а не к условной дисперсии. Эта модель имеет вид $\sigma_t = \sqrt{h_t} = \alpha_0 + \alpha_1 |u_{t-1}| + \gamma_1 \sigma_{t-1}$. В ней усредненная условная дисперсия получается возведением в квадрат усредненного среднеквадратичного отклонения, а не усреднением квадрата отклонения. Благодаря неравенству Йенсена и выпуклости вниз квадратичной функции влияние больших отклонений в такой модели будет оказывать меньшее влияние, чем в модели GARCH.

При анализе многих финансовых проблем была замечена асимметрия реакции на отклонения, например инфляция «охотнее» увеличивается, чем уменьшается. На финансовых рынках волатильность более быстро реагирует на падение рынка, чем на его рост. В 1991 г. Нельсон [17] предложил модель, получившую название экспоненциальная (exponential) GARCH (EGARCH) для отражения такой асимметрии. Модель имеет вид: $\ln h_t = \alpha_0 + \alpha_1 f\left(\frac{u_{t-1}}{\sqrt{h_t}}\right) + \gamma_1 \ln h_{t-1}$, где

$$f\left(\frac{u_{t-1}}{\sqrt{h_{t-1}}}\right) = \theta_1 \frac{u_{t-1}}{\sqrt{h_{t-1}}} + \left(\left|\frac{u_{t-1}}{\sqrt{h_{t-1}}}\right| - E\left(\left|\frac{u_{t-1}}{\sqrt{h_{t-1}}}\right|\right)\right).$$

Функция $f(.)$ отражает реакцию на «новости» и связана с коррекцией текущего уровня волатильности $\ln h_t$ уровнем отклонения u_{t-1}. Очевидно, что реакция асимметрична. При выполнении условия $u_{t-1} > 0$ получаем $\frac{\partial f}{\partial u_{t-1}} = \theta_1 + 1$, а при $u_{t-1} < 0$ получаем $\frac{\partial f}{\partial u_{t-1}} = \theta_1 - 1$. Можно показать, что $f(u_{t-1})$ является сильным белым шумом с нулевым математическим ожиданием и постоянной дисперсией. Следовательно, процесс $\ln h_t$ подчиняется модели ARMA(1,1) и будет стационарным при выполнении условия $\gamma_1 < 1$.

Модели GARCH(1,1), EGARCH и модель Шверта могут быть представлены как частные случаи единой модели, предложенной Хиггинсом и Бера [8]. Эта общая модель носит название NARCH (Non-linear ARCH) и может быть представлена в следующем виде: $\sigma_t^\beta = \left(\sqrt{h_t}\right)^\beta = \alpha_0 + \alpha_1 f^\beta(u_{t-1}) + \gamma_1 \sigma_{t-1}^\beta$. Еще одним обобщением является пороговый ARCH-процесс: $\sigma_t^\beta = \left(\sqrt{h_t}\right)^\beta = \alpha_0 + \alpha_1 g^{(\beta)}(u_{t-1}) + \gamma_1 \sigma_{t-1}^\beta$, где $g^{(\beta)}(u_{t-1}) = \theta I(u_{t-1} > 0)|u_{t-1}|^\beta + \theta I(u_{t-1} \le 0)|u_{t-1}|^\beta$. Здесь через $I(.)$ обозначена функция-индикатор, равная 1 при выполнении условия, указанного в скобках, и равная нулю – в остальных случаях. При $\beta = 1$ получаем TARCH-модель (Threshold ARCH) [25], а при $\beta = 2$ получаем модель GJR, предложенную Глостеном, Джаганнатаном и Ранклем [5]. Количество модификаций модели GARCH к настоящему времени весьма велико, более полный обзор их можно найти в ранее цитированной монографии Т. Миллса [15].

* * *

СПИСОК ЛИТЕРАТУРЫ

1. *Bollerslev T.* Generalized Autoregressive Conditional Heteroscedastisity // Journal of Econometrics. 1986. Vol. 31. P. 307–327.

2. *Bougerol P., Picard N.* Stationarity of GARCH Processes and of Some Nonnegative Time Series // Journal of Econometrics. 1992. Vol. 52. P. 115–128.

3. *Engle R.F.* Autoregressive Conditional Heteroskedastisity with Estimates of the Variance of U.K. Inflation // Econometrica. 1982. Vol. 50. P. 987–1007.

4. *Engle R.F., Granger C.W.J.* Cointegration and Error Correction: Representation, Estimation and Testing // Econometrica. 1987. Vol. 55. № 2. P. 251–276.

5. *Glosten L.R., Jagannathan R., Runkle D.* Relationship Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks // Journal of Finance. 1993. Vol. 48. P. 1779–1801.

6. *Granger C.W.J.* Some Properties of Time Series Data and Their Use in Econometric Model Specification // Journal of Econometrics. 1981. Vol. 16. № 1. P. 121–130.

7. *Hamilton J.D.* Time Series Analysis. Princeton University Press. 1994. P. 296.

8. *Higgins M.L., Bera A.K.* A Joint Test for ARCH and Bilinearity in the Regression Model // Econometrics Reviews. Vol. 7. P. 171–181.

9. *Johansen J.* Estimation and Hypothesis Testing of Cointegrating Vectors in Gaussian Vector Autoregressive Models // Econometrica. 1991. Vol. 59. P. 1551–1580.

10. *Johansen J.* Likelihood-based Inference in Cointegrating Vector Autoregressive Models. Oxford: Oxford University Press, 1995.

11. *Johansen J.* Statistical Ananlysis of Cointegrating Vectors // Journal of Economic Dynamics and Control. 1998. Vol. 12. P. 231–254.

12. *Johansen J., Juselius K.* Maximum Likelihood Estimation and Inference on Cointegration – With Application to the Demand for Money // Oxford Bulletin of Economics and Statistics. 1992. Vol. 54. P. 461–471.

13. *Johnston J., DiNardo J.* Econometric Methods. Fourth Edition. The McGrow-Hill Componies. Inc., 1997. P. 320.

14. *MacKinnon J.G.* Critical Values for Cointegration Tests, Chapter 13 // Long-Run Economic Relationships / R.F. Engle, C.W.J. Granger (eds.) Oxford University Press, 1991.

15. *Mills T.* The Econometric Modelling of Financial Time Series. Second Edition. Cambridge University Press, 1999.

16. *Nelson C.R., Plosser C.I.* Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications // Journal of Monetary Economics. 1982. Vol. 10. P. 139–162.

17. *Nelson D.B.* Conditional Heteroskedastsity in Asset Returns // Econometrica. Vol. 59. P. 347–370.

18. *Nelson D.B.* Stationarity and Persistence in the GARCH(1,1) Model // Econometric Theory. 1990. Vol. 6. P. 318–34.

19. *Nelson D.B., Cao C.Q.* Inequality Constraints in Univariate GARCH Models // Journal of Business and Economic Statistics. 1992. Vol. 10. P. 229–235.

20. *Schwert G.W.* Why does Stock Market Volatility Change Over Time? // Journal of Finance. 1989. Vol. 44. P. 1115–1153.

21. *Sims C.A., Stock J.H., Watson M.W.* Inference in Linear Time Series Models with Some Unit Roots // Econometrica. 1990. Vol. 58. P. 113–144.

22. *Stock J. H.* Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors // Econometrica. 1987. Vol. 55. P. 1035–1056.

23. *Teylor S.J.* Modelling Financial Time Series. New York: Wiley, 1986.

24. *Watson M.W.* Vector Avtoregression and Cointegration // Handbook of Econometrics. 1994. Vol. 4. Amsterdam: North-Holland. P. 2844–2915.

25. *Zakoian J.M.* Threshold Heteroskedastic Models //Journal of Economic Dynamics and Control, 1994. Vol. 18. P. 931–955.