

## 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)  $X_t = \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} + \varepsilon_t$  можно применить обычный метод наименьших квадратов (МНК). Поскольку регрессоры относятся к предыдущим моментам времени, а  $\varepsilon_t$  – белый шум, то корреляция регрессоров со случайным возмущением  $\varepsilon_t$  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  $\varepsilon_t$  – 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  $\theta$ . 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  $\bar{X} = \frac{1}{T} \sum_{t=1}^T X_t$  и  $\hat{\mu} = \frac{\hat{\theta}}{1 - \hat{\alpha}_1 - \dots - \hat{\alpha}_p}$ , в которой использованы

оценки МНК. Для гауссова процесса  $\varepsilon_t$  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  $\sum e_t^2$  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  $\varepsilon_t$  выражаем ковариационную матрицу ошибок (ее элементы – значения автокорреляционной функции) через параметры  $\beta_1, \beta_2, \dots, \beta_q$  обратимой модели MA(q). Функция правдоподобия для нормально распределенного вектора  $\bar{X}$  имеет вид:

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

Здесь через  $\Sigma_X$  обозначена ковариационная матрица процесса  $\bar{X}$ . Элементы этой матрицы выражаются, как мы видели, через параметры модели  $\beta_1, \beta_2, \dots, \beta_q$ , 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  $\bar{X}$  в предположении о его стационарности и эргодичности. Затем выберем некоторые значения  $(\beta_1, \beta_2)$ , например (0,5; 0,5). Исходя из уравнения процесса,  $X_1$  выражается через предыдущие ошибки, но они не известны. Поэтому будем считать, что  $X_1 = \bar{X} + e_1$ . Это дает нам  $e_1 = X_1 - \bar{X}$ . Затем полагаем  $e_2 = X_2 - \bar{X} - \beta_1 e_1$ ,  $e_3 = X_3 - \bar{X} - \beta_1 e_2 - \beta_2 e_1$  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  $(\beta_1, \beta_2)$ , мы можем рассматривать ее как функцию этих переменных. Далее какой-либо численной процедурой, например поиском на сетке, можно перебирать комбинации  $\beta_1$  и  $\beta_2$  или численно искать минимум этой функции:  $\min_{\beta_1, \beta_2} \sum e_i^2$ . Коэффициенты  $(\beta_1^*, \beta_2^*)$ ,

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

В общем случае модели МА(q) мы фиксируем начальные значения  $\beta_1^0, \dots, \beta_q^0$  и определяем остатки через наблюдения:  $e_1 = X_1 - \bar{X}$ ,  $e_2 = X_2 - \bar{X} - \beta_1^0 e_1$  и так далее. Процедура получения остатков достаточно проста, все остатки с отрицательными и нулевыми индексами в используемых выражениях заменяем нулями. Затем находим  $\min_{\beta} \sum e_i^2$ . 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  $X_t = \frac{(1 + \beta_1 L + \beta_2 L^2) \varepsilon_t}{1 - \alpha_1 L - \alpha_2 L^2}$ . Введем вспомогательный слу-

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

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

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  $\beta_1, \beta_2$ , можно оценить коэффициенты  $\alpha_1$  и  $\alpha_2$  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  $X_1, X_2, \dots, X_T$ , 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  $\varepsilon_t$ , 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:

$$BIC(p, q) = \ln \hat{\sigma}^2 + \ln T \frac{(p+q)}{T}. \text{ Разумеется, } \hat{\sigma}^2 = \frac{RSS}{T-p-q}.$$

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  $\hat{\sigma}^2$  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  $H_0: \rho_1 = \rho_2 = \dots = \rho_m$  против альтернативной

гипотезы  $H_1: \sum_{i=1}^m \rho_i^2 > 0$ . Бокс и Пирс показали, что при увеличении длины выбор-

ки статистика  $Q^* = T \sum_{k=1}^m r_k^2$  имеет асимптотическое распределение  $\chi_m^2$ . Этот тест

было предложено применять для проверки наличия автокорреляции остатков моде-  
ли 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  $\chi_m^2$ , 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 знакомимся с этим тестом еще в курсе эконометрики, сформулируем его применительно к анализу автокорреляции остатков.

Пусть рассматривается модель множественной регрессии

$$Y_t = \beta_0 + \beta_1 X_{1t} + \dots + \beta_k X_{kt} + \varepsilon_t,$$

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

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

$$\begin{aligned} H_0 : \gamma_1 = \dots = \gamma_p &= 0 \\ H_1 : \gamma_1^2 + \dots + \gamma_p^2 &> 0. \end{aligned}$$

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

$$Y_t = \beta_0 + \beta_1 X_{1t} + \dots + \beta_k X_{kt} + \varepsilon_t.$$

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  $e_{t-1}, e_{t-2}, \dots, e_{t-p}$ ).

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  $\chi^2$  с  $p$  степенями свободы.

Для проверки гипотезы о равенстве нулю группы переменных мы привыкли использовать F-статистику, проверяющую, фактически, статистическую значимость уменьшения остаточной суммы квадратов от включения дополнительных переменных. Разумеется, F-статистика применима и в этом случае. Но, 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  $\sigma$  – 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  $\chi^2$  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  $\rho_k$ . 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  $\rho_k$  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  $\chi_k^2$ , 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):  $X_t = 0,176 + 0,856 X_{t-1} + e_t$ ;  $\hat{\sigma} = 0,870$ . 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 ожида-

ние. Его оценка равна  $\hat{\mu} = \frac{0,176}{1-0,856} = 1,227$ , 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:

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

(0,101) (0,054) (0,084)

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

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

(0,104) (0,051) 0,095

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 г. [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  $\rho_k$  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,  $p \leq 3, q \leq 3$ . 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  $p \leq 3$  и  $q \leq 3$ . Здесь их совсем немного. Поэтому составим сводную таблицу для значений критериев АИС и ВИС в зависимости от параметров модели.

Таблица 6.4.

$p$	Критерий АИС			
	$q$			
	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	<b>3,683</b>	3,693	3,698	3,707

Продолжение таблицы

p	Критерий ВИС			
	q			
	0	1	2	3
0	3,701	<b>3,696</b>	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 АИС, 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):

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

s.e. (0,62) (0,057) (0,057) (0,057)

Для модели ARMA(0,1) было получено:

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

(0,35) (0,056)

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:

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

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 моде-

лей следующую величину:  $R = \exp\left\{-\frac{1}{2}T[AIC(p_1, q_1) - AIC(p, q)]\right\}$ . 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 АИС отбирает в нашем случае 6 моделей:

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

которые разумно рассматривать.

А вот критерий ВИС выбрал всего 3 модели:  $\{(0,1);(1,0);(0,0)\}$ . Любопытно, что по критерию ВИС 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 ВИС 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  $\varepsilon$ ;
- отсутствие точных значений коэффициентов модели (у нас есть только их оценки, оцененные по имеющейся выборке).

При прогнозировании по модели множественной регрессии мы оцениваем значение зависимой переменной при заданных значениях независимых переменных – регрессоров. Это приводит к тому, что характеристики прогноза, как случайной величины, являются, по сути, условными характеристиками при условии

имеющейся выборки независимых переменных. Иная ситуация в моделях типа ARIMA. Значение переменной прогнозируется для некоторого будущего момента времени, при этом лаговые значения этой переменной, служащие регрессорами модели, можно рассматривать фиксированными на выборочных значениях, или случайными. Первая возможность приводит к условному прогнозу, как и для модели множественной регрессии, а вторая – к безусловному прогнозу. Таким образом, при прогнозировании по модели типа ARIMA можно рассматривать как условный, так и безусловный прогнозы. Из курса теории вероятностей известно, что условная дисперсия случайной величины не превышает ее безусловную дисперсию, поэтому точность условного прогноза всегда выше.

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

Показано, что если мы хотим достичь минимума среднеквадратической ошибки (MSE), т.е. не требовать несмещенности, то надо взять условное математическое ожидание:  $E\{X_{T+h} | X_1, \dots, X_T\}$ . Такое условное математическое ожидание будет гарантировать получение MSE, или иногда ее обозначают MMSE, т.е. Minimum mean square error.

Начнем с модели MA(q):  $X_t = \theta + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \dots + \beta_q \varepsilon_{t-q}$ . 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.  $\theta$ . Условным прогнозом для момента времени  $T+1$  будет условное математическое ожидание

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

Среди случайных величин  $\varepsilon$ , 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  $\varepsilon_{T+1}$ , не равны нулю. Рассмотрим сначала  $E\{\varepsilon_T | X_1, \dots, X_T\}$ , потому что схема одна и та же. Это математическое ожидание – остаток между наблюдением и расчетом, прогнозом по модели, т.е.  $e_T = X_T - \tilde{X}_T$ . 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  $\varepsilon$  заменяются нулями, а предыдущие – заменяются реально наблюдаемыми остатками. Таким образом, для модели MA(q) прогноз зависит от того, какие ошибки были на предыдущих шагах. Начиная с шага  $(q+1)$  условный прогноз представляет собой просто математическое ожидание  $\theta$ , т.е. условный прогноз совпадает с безусловным.

Рассмотрим условную дисперсию ошибки прогноза на 1 шаг:

$$\text{var}(X_{T+1} - \tilde{X}_{T+1} | X_1, \dots, X_T) = E\{(\theta + \varepsilon_{T+1} + \beta_1 \varepsilon_T + \dots + \beta_q \varepsilon_{T-q+1} - \theta - \beta_1 \varepsilon_T + \dots + \beta_q \varepsilon_{T-q+1})^2 | X_1, \dots, X_T\} = \sigma_\varepsilon^2.$$

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

$$X_t = \theta + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \dots + \alpha_p X_{t-p} + \varepsilon_t.$$

Для прогноза на 1 шаг вперед можно записать:

$$\tilde{X}_{T+1} = E\{X_{T+1} | X_1, \dots, X_T\} = E\{\theta + \alpha_1 X_T + \dots + \alpha_p X_{T-p} + \varepsilon_{T+1} | X_1, \dots, X_T\} = \theta + \alpha_1 X_T + \dots + \alpha_p X_{T-p}.$$

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

$$E\{X_{T+2} | X_1, \dots, X_T\} = E\{\theta + \alpha_1 X_{T+1} + \dots + \alpha_p X_{T-p+2} + \varepsilon_{T+2} | X_1, \dots, X_T\}.$$

Разумеется, математическое ожидание от случайной ошибки  $\varepsilon$  опять даст 0, условные математические ожидания от  $X_1, X_2, \dots, X_T$  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  $\frac{\theta}{1 - \alpha_1 - \dots - \alpha_p}$ , 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  $\tilde{X}_{T+1} = \alpha_1 X_T + \alpha_2 X_{T-1}$ , а  $X_{T+1} = \alpha_1 X_T + \alpha_2 X_{T-1} + \varepsilon_{T+1}$ . Очевидно, что дисперсия ошибки прогноза на 1 шаг равна  $\sigma_\varepsilon^2$ . Для прогноза на 2 шага соответственно получаем:

$$\begin{aligned} \tilde{X}_{T+2} &= \alpha_1 \tilde{X}_{T+1} + \alpha_2 X_T = \alpha_1(\alpha_1 X_T + \alpha_2 X_{T-1}) + \alpha_2 X_T, \\ X_{T+2} &= \alpha_1 X_{T+1} + \alpha_2 X_T + \varepsilon_{T+2} = \alpha_1(\alpha_1 X_T + \alpha_2 X_{T-1} + \varepsilon_{T+1}) + \alpha_2 X_T + \varepsilon_{T+2}. \end{aligned}$$

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

$$\begin{aligned} \tilde{X}_{T+3} &= \alpha_1 \tilde{X}_{T+2} + \alpha_2 \tilde{X}_{T+1} = \alpha_1(\alpha_1(\alpha_1 X_T + \alpha_2 X_{T-1}) + \alpha_2 X_T) + \alpha_2(\alpha_1 X_T + \alpha_2 X_{T-1}), \\ X_{T+3} &= \alpha_1 X_{T+2} + \alpha_2 X_{T+1} + \varepsilon_{T+3} = \\ &= \alpha_1(\alpha_1(\alpha_1 X_T + \alpha_2 X_{T-1} + \varepsilon_{T+1}) + \alpha_2 X_T + \varepsilon_{T+2}) + \alpha_2(\alpha_1 X_T + \alpha_2 X_{T-1} + \varepsilon_{T+1}) + \varepsilon_{T+3}. \end{aligned}$$

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 МА представлению  $X_t = \theta + \varepsilon_t + \psi_1\varepsilon_{t-1} + \dots + \psi_q\varepsilon_{t-q} + \dots$ , 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  $[1, T]$  – наблюдаемые 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 МА 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  $X_t = (1 - \alpha L)^{-1}(1 + \beta L)\varepsilon_t = (1 + \alpha L + \alpha^2 L^2 + \dots)(1 + \beta L)\varepsilon_t = \varepsilon_t + \eta_1 \varepsilon_{t-1} + \eta_2 \varepsilon_{t-2} + \dots$ , где  $\eta_1 = \alpha + \beta$ ,  $\eta_2 = \alpha(\alpha + \beta)$ , ...,  $\eta_k = \alpha^{k-1}(\alpha + \beta)$ . Начиная со второго коэффициенты убывают в геометрической прогрессии. Теперь легко посчитать дисперсию ошибки прогноза на  $h$  шагов:  $\sigma_\varepsilon^2 \sum_{i=0}^{h-1} \eta_i^2 = \sigma_\varepsilon^2 (1 + \sum_{i=0}^{h-1} \alpha^{2i} (\alpha + \beta)^2) = \sigma_\varepsilon^2 (1 + \frac{(\alpha + \beta)(1 - \alpha^{2h})}{1 - \alpha^2})$ . Срав-

нив 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 общепринятых приемов является разбиение имеющейся реализации на 2 части. По первым  $n$  наблюдениям выбирается и оценивается модель, а по последним ( $T-n$ ) наблюдениям проводится сравнение наблюдаемых 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  $\varepsilon_t$  – 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  $\varepsilon_t$  – стационарный процесс, то его первая разность также стационарный процесс, хотя если  $\varepsilon_t$  – 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  $\beta$ . 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 tun*: 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 tun*: 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
<ul style="list-style-type: none"> <li>• Не стационарен из-за непостоянного тренда.</li> <li>• 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.</li> </ul>	<ul style="list-style-type: none"> <li>• Не стационарен из-за непостоянной дисперсии.</li> <li>• Так как в явном решении стоит сумма всех предыдущих <math>\varepsilon</math>, то шоки помнятся все время. Это процесс с бесконечной памятью. Экономически это не очень понятно, шоки не должны сказываться постоянно.</li> </ul>

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  $R^2 \rightarrow 1$  при  $T \rightarrow \infty$ , т.е. для процесса DSP с дрейфом  $R^2 \rightarrow 1$  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:

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

где  $v_t$  и  $\varepsilon_t$  – 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 мы 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 неправилен. 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 Пlossера [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 Саргент 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–Пlossера, 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 спреда 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.

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

$H_1$ : ряд является TS  $\Rightarrow |\rho| < 1$  (тогда  $\varepsilon$  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 ин-

тегралы 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 заданном 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
$\infty$	6,25	3,00

На рис. 7.1 схематически показаны графики плотностей распределения Дикки-Фуллера 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).

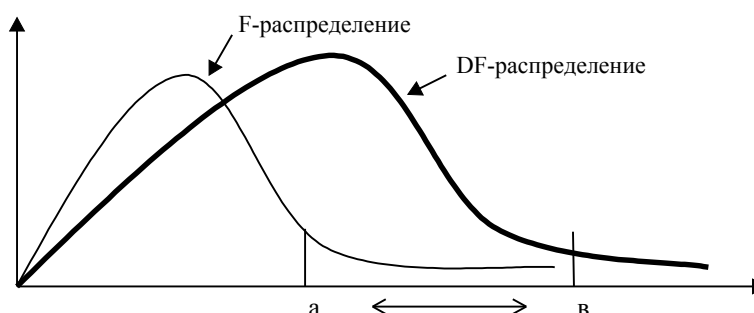


Рис. 7.1.

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  $\Delta Y$ . Поэтому были оценены 2 следующие модели, соответствующие нулевой и альтернативной гипотезе (в скобках – стандартные ошибки оценок коэффициентов).

$$\text{Первая модель: } Y_t - Y_{t-1} = \Delta Y_t = 0,52 + 0,0012 t - 0,119 Y_{t-1} + 0,498 (Y_{t-1} - Y_{t-2});$$

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

**Вторая модель:** Пусть  $\beta_1 = 0; \alpha_1 = 1$ . МНК дает следующую модель:  $Y_t - Y_{t-1} = 0,0054 + 0,447(Y_{t-1} - Y_{t-2})$ . Соответственно  $RSS = 0,063211$  ( $RSS_R$ ). Считаем

стандартное  $F$ -отношение:  $F_{\text{отношение}} = \frac{(0,063211 - 0,056448) / 2}{0,056448 / 106} = 6,34$ . А теперь по-

смотрите 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  $F_{\text{критическое}}$  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 Association. 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., Griffiths 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 // Biometrika. 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.