

В.П.Носко

Эконометрика

Введение в регрессионный анализ временных рядов

Москва

2002

Оглавление

Оглавление

Введение

Глава 1. Особенности регрессионного анализа для стохастических объясняющих переменных

Глава 2. Стационарные ряды. Модели ARMA

2.1. Общие понятия.

2.2. Процесс белого шума

2.3. Процесс авторегрессии

2.4. Процесс скользящего среднего

2.5. Смешанный процесс авторегрессии – скользящего среднего (процесс авторегрессии с остатками в виде скользящего среднего)

2.6. Модели ARMA, учитывающие наличие сезонности

Глава 3. Подбор стационарной модели ARMA для ряда наблюдений

3.1. Идентификация стационарной модели ARMA

3.2. Оценивание коэффициентов модели

3.3. Диагностика оцененной модели

Глава 4. Регрессионный анализ для стационарных объясняющих переменных

4.1. Асимптотическая обоснованность стандартных процедур

4.2. Динамические модели

4.3. Векторная авторегрессия

4.4. Некоторые частные случаи динамических моделей

Глава 5. Нестационарные временные ряды

5.1. Нестационарные ARMA модели

5.2. Проблема определения принадлежности временного ряда классу TS рядов или классу DS рядов

5.3. Различение TS и DS рядов в классе моделей ARMA. Гипотеза единичного корня.

Глава 6. Процедуры для различения TS и DS рядов

6.1. Предварительные замечания

6.2. Критерии Дики – Фуллера

6.3. Расширенные критерии Дики - Фуллера

6.4. Краткий обзор критериев Дики – Фуллера

6.5. Некоторые другие сочетания DGP и SM

6.6. Ряды с квадратичным трендом.

6.7. Многовариантная процедура проверки гипотезы единичного корня

6.8. Обзор некоторых других процедур

6.8.1. Критерий Филлипса – Перрона

6.8.2. Критерий Лейбурна

- 6.8.3. Критерий Шмидта – Филлипса.
- 6.8.4. Критерий DF-GLS
- 6.8.5. Критерий Квятковского – Филлипса – Шмидта – Шина (KPSS)
- 6.8.6. Процедура Кохрейна (отношение дисперсий)
- 6.9. Некоторые проблемы, возникающие при различении TS и DS гипотез
 - 6.9.1. Коррекция сезонности
 - 6.9.2. Протяженность ряда и мощность критерия
 - 6.9.3. Проблема согласованности статистических выводов при различении TS и DS гипотез
 - 6.9.4. Наличие нескольких единичных корней
- 6.10. Критерий Перрона и его обобщение
 - 6.10.1. Критерий Перрона
 - 6.10.2. Обобщенная процедура Перрона
- Глава 7. Регрессионный анализ для нестационарных объясняющих переменных
 - 7.1. Проблема ложной регрессии
 - 7.2. Коинтегрированные временные ряды. Модели коррекции ошибок
 - 7.3. Проверка нескольких рядов на коинтегрированность. Критерии Дики – Фуллера
 - 7.4. Оценивание коинтегрированных систем временных рядов
- Глава 8. Процедура Йохансена
 - 8.1. Оценивание ранга коинтеграции
 - 8.2. Оценивание модели коррекции ошибок
- Заключение
- Список литературы
- Указатель

Введение

В начальных курсах эконометрики, в том числе и в ранее изданном автором учебном пособии “Эконометрика для начинающих” [Носко (2000)], первоочередное внимание уделяется статистическим выводам в рамках **классической нормальной линейной модели наблюдений**

$$y_t = \theta_1 x_{t1} + \theta_2 x_{t2} + \dots + \theta_p x_{tp} + \varepsilon_t, \quad t = 1, 2, \dots, n,$$

в которой предполагается, что значения объясняющих переменных $x_{t1}, x_{t2}, \dots, x_{tp}, t = 1, 2, \dots, n$, **фиксированы**, а случайные составляющие $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ (“ошибки”) являются независимыми случайными величинами, имеющими одинаковое нормальное распределение с нулевым математическим ожиданием и конечной дисперсией (такие предположения об ошибках мы называем “**стандартными**”). Далее обычно анализируются последствия различного типа нарушений таких предположений об ошибках и рассматриваются методы коррекции статистических выводов о коэффициентах модели при наличии соответствующих нарушений стандартных предположений.

В упомянутом пособии обсуждались методы коррекции статистических выводов

- при наличии гетероскедастичности (неоднородности дисперсий ошибок);
- при наличии автокоррелированности ошибок;
- при наличии сезонности.

При этом сохранялось предположение о фиксированности значений объясняющих переменных.

Последнее предположение в совокупности со стандартными предположениями об ошибках удобно с чисто математической точки зрения: при таких предположениях оценки параметров, получаемые методом наименьших квадратов, имеют нормальное распределение, что, в свою очередь, дает возможность

- строить доверительные интервалы для коэффициентов линейной модели, используя квантили t -распределения Стьюдента;
- проверять гипотезы о значениях отдельных коэффициентов, используя квантили t -распределения Стьюдента;
- проверять гипотезы о выполнении тех или иных линейных ограничений на коэффициенты модели, используя квантили F -распределения Фишера;
- строить интервальные прогнозы для “будущих” значений объясняемой переменной, соответствующих заданным будущим значениям объясняющих переменных.

Вместе с тем, используемое в классической модели предположение о фиксированности значений объясняющих переменных в n наблюдениях означает, что мы можем повторить наблюдения значений объясняемой переменной при том же наборе значений объясняющих переменных $x_{t1}, x_{t2}, \dots, x_{tp}, t = 1, 2, \dots, n$; при этом мы получим другую реализацию (другой набор значений) случайных составляющих $\varepsilon_t, t = 1, 2, \dots, n$, что приведет к значениям объясняющей переменной, отличным от значений y_1, y_2, \dots, y_n , наблюдавшихся ранее.

С точки зрения моделирования реальных экономических явлений, предположение о фиксированности значений объясняющих переменных можно считать реалистическим

лишь в отдельных ситуациях, связанных с проведением контролируемого эксперимента. Например, можно представить себе гипотетический эксперимент, в котором в течение определенного периода времени значения располагаемого месячного дохода y_t домохозяйств остаются неизменными и наблюдаются ежемесячные расходы этих домохозяйств на личное потребление. Величина этих расходов для каждого отдельного домохозяйства будет изменяться от месяца к месяцу за счет факторов, не известных исследователю в рамках проводимого эксперимента, а потому эта изменчивость, с точки зрения исследователя, носит “случайный” характер, что формально описывается включением в модель наблюдений случайной составляющей – “ошибки”.

Между тем в реальных ситуациях по большей части нет возможности сохранять неизменными значения объясняющих переменных. Более того, и сами наблюдаемые значения объясняющих переменных (как и “ошибки”) часто интерпретируются как реализации некоторых случайных величин. В таких ситуациях становится проблематичным использование техники статистических выводов, разработанной для классической нормальной линейной модели.

В наиболее выраженной степени сказанное относится к моделям, в которых наблюдения развернуты во времени, т.е. производятся в последовательные моменты времени (годы, кварталы, месяцы, дни и т.п.). При этом значения отдельной объясняющей переменной в последовательные моменты времени образуют **временной ряд**, и если говорить опять о линейной модели наблюдений, то теперь уже такая модель связывает временные ряды, и это самым существенным образом сказывается на свойствах оценок коэффициентов линейной модели. Упомянем в связи с этим простую по форме модель

$$y_t = \beta y_{t-1} + \varepsilon_t, \quad t = 1, 2, \dots, n, \quad y_0 = 0,$$

в которой $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ – независимые случайные величины, имеющие одинаковое нормальное распределение с нулевым математическим ожиданием и дисперсией σ^2 .

Это линейная модель наблюдений, в которой в качестве значения объясняющей переменной x_t в момент времени t выступает запаздывающее на одну единицу времени значение объясняемой переменной, т.е. $x_t = y_{t-1}$.

Предполагая, что процесс порождения данных описывается такой моделью (с не известными нам значениями параметров β и σ), мы можем получить для β оценку наименьших квадратов, которая вычисляется по формуле

$$\hat{\beta} = \frac{\sum_{t=2}^n y_t y_{t-1}}{\sum_{t=2}^n y_{t-1}^2}.$$

В отличие от ситуации, когда значения объясняющей переменной фиксированы, теперь наблюдаемые значения объясняющей переменной сами являются реализациями случайных величин и выражаются через значения случайных величин ε_t :

$$x_1 = 0,$$

$$x_2 = y_1 = \varepsilon_1,$$

$$x_3 = y_2 = \beta y_1 + \varepsilon_2 = \beta \varepsilon_1 + \varepsilon_2,$$

$$x_4 = y_3 = \beta y_2 + \varepsilon_3 = \beta(\beta \varepsilon_1 + \varepsilon_2) + \varepsilon_3 = \beta^2 \varepsilon_1 + \beta \varepsilon_2 + \varepsilon_3,$$

...

$$x_n = y_{n-1} = \beta^{n-2} \varepsilon_1 + \beta^{n-3} \varepsilon_2 + \dots + \varepsilon_{n-1}.$$

Это приводит к тому, что на сей раз распределение оценки наименьших квадратов для параметра β не является нормальным, так что t - и F -статистики, используемые для статистических выводов в классической нормальной линейной модели, уже не имеют t - и F -распределений, соответственно.

Более того, если в действительности в процессе порождения данных $\beta = 1$, так что мы имеем дело с популярной моделью *случайного блуждания* (выходящего из нуля),

$$y_t = y_{t-1} + \varepsilon_t, \quad t = 1, 2, \dots, n, \quad y_0 = 0,$$

то тогда распределение (центрированной и нормированной) оценки наименьших квадратов для β не сближается с нормальным даже при неограниченном возрастании количества наблюдений. Иначе говоря, в такой модели оценка наименьших квадратов для β даже не является асимптотически нормальной. Впервые на этот факт было указано в работе [White (1958)], и это открытие привело впоследствии к полному пересмотру методологии эконометрического анализа статистических данных, представляемых в виде временных рядов.

На первый план вышло фундаментальное различие между TS (*trend stationary*) рядами (стационарными или стационарными относительно детерминированного тренда) и DS (*difference stationary*) рядами, приводящимися к стационарным только в результате однократного или многократного дифференцирования. Это различие определяется наличием у DS рядов так называемого *стохастического тренда*, собственно и приводящего к неприменимости стандартной асимптотической теории при работе с такими рядами. В создании и развитии асимптотической теории, учитывающей возможное наличие у рассматриваемых переменных стохастического тренда, приняло участие большое количество авторов, среди которых непременно следует упомянуть Dickey D.A., Fuller W.A., Granger C.W.J., Hansen B.E., Johansen S., Juselius K., Perron P., Phillips P.S.B., Sims C.A., Stock J.H., Watson M.W.J. Прекрасный обзор полученных в этом направлении результатов содержится в работе [Maddala G.S., Kim In-Moo (1998)]; см. также [Hatanaka M. (1996)].

В предлагаемом учебном пособии мы даем краткое введение в современные методы эконометрического анализа статистических данных, представленных в виде временных рядов, которые учитывают возможное наличие у рассматриваемых переменных стохастического тренда. Основные акценты, как и в работе [Носко (2000)], смещены в сторону разъяснения базовых понятий и основных процедур статистического анализа данных с привлечением смоделированных и реальных экономических данных. Вместе с тем, от читателя требуется несколько большая осведомленность в отношении вероятностно-статистических методов исследования. Предполагается, что читатель имеет представление о совместной функции распределения, многомерном нормальном распределении, методе максимального правдоподобия, свойстве состоятельности оценок, характеристиках статистических критериев (ошибки первого и второго рода, мощность), а также владеет методами регрессионного анализа в рамках начального курса эконометрики. Кроме того он должен иметь некоторое представление о комплексных числах и комплексных корнях полиномов.

Пособие написано на основании курса лекций, прочитанных автором в Институте экономики переходного периода.

Глава 1. Особенности регрессионного анализа для стохастических объясняющих переменных

Как уже было указано во Введении, в начальных курсах эконометрики первоочередное внимание уделяется статистическим выводам в рамках *классической нормальной линейной модели наблюдений*

$$y_t = \theta_1 x_{t1} + \theta_2 x_{t2} + \dots + \theta_p x_{tp} + \varepsilon_t, \quad t = 1, 2, \dots, n,$$

в которой предполагается, что значения объясняющих переменных $x_{t1}, x_{t2}, \dots, x_{tp}$, $t = 1, 2, \dots, n$, фиксированы, а случайные составляющие $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ (“ошибки”) являются независимыми случайными величинами, имеющими одинаковое нормальное распределение с нулевым математическим ожиданием и конечной дисперсией (такие предположения об ошибках мы называем “*стандартными*”). Далее анализируются последствия различного типа нарушений таких предположений об ошибках и рассматриваются методы коррекции статистических выводов о коэффициентах модели при наличии соответствующих нарушений стандартных предположений.

Предположение о фиксированности значений объясняющих переменных в совокупности со стандартными предположениями об ошибках удобно с чисто математической точки зрения: при таких предположениях оценки параметров, получаемые методом наименьших квадратов, имеют нормальное распределение, что, в свою очередь, дает возможность

- строить доверительные интервалы для коэффициентов линейной модели, используя квантили t -распределения Стьюдента;
- проверять гипотезы о значениях отдельных коэффициентов, используя квантили t -распределения Стьюдента;
- проверять гипотезы о выполнении тех или иных линейных ограничений на коэффициенты модели используя квантили F -распределения Фишера;
- строить интервальные прогнозы для “будущих” значений объясняемой переменной, соответствующих заданным будущим значениям объясняющих переменных.

Вместе с тем, используемое в классической модели предположение о фиксированности значений объясняющих переменных в n наблюдениях означает, что мы можем повторить наблюдения значений объясняемой переменной при том же наборе значений объясняющих переменных $x_{t1}, x_{t2}, \dots, x_{tp}$, $t = 1, 2, \dots, n$; при этом мы получим другую реализацию (другой набор значений) случайных составляющих ε_t , $t = 1, 2, \dots, n$, что приведет к значениям объясняемой переменной, отличающимся от значений y_1, y_2, \dots, y_n , наблюдавшихся ранее.

С точки зрения моделирования реальных экономических явлений, предположение о фиксированности значений объясняющих переменных можно считать реалистическим лишь в отдельных ситуациях, связанных с проведением контролируемого эксперимента. Между тем в реальных ситуациях по большей части нет возможности сохранять неизменными значения объясняющих переменных. Более того, и сами наблюдаемые значения объясняющих переменных (как и “ошибки”) часто интерпретируются как реализации некоторых случайных величин. В таких ситуациях становится проблематичным использование техники статистических выводов, разработанной для классической нормальной линейной модели.

Поясним последнее, обратившись к матрично-векторной форме записи классической линейной модели с p объясняющими переменными:

$$y = X\theta + \varepsilon,$$

в которой

$y = (y_1, \dots, y_n)^T$ - вектор значений объясняемой переменной в n наблюдениях,

$X = (n \times p)$ -матрица значений объясняющих переменных в n наблюдениях, $n > p$,

$\theta = (\theta_1, \theta_2, \dots, \theta_p)^T$ - вектор коэффициентов,

$\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^T$ - вектор случайных ошибок (возмущений) в n наблюдениях.

Если матрица X имеет полный ранг p , то матрица $X^T X$ является невырожденной, для нее существует обратная матрица $(X^T X)^{-1}$, и оценка наименьших квадратов для вектора θ неизвестных коэффициентов имеет вид

$$\hat{\theta} = (X^T X)^{-1} X^T y.$$

Математическое ожидание вектора оценок коэффициентов равно

$$E(\hat{\theta}) = E((X^T X)^{-1} X^T (X\theta + \varepsilon)) = E((X^T X)^{-1} X^T X\theta) + E((X^T X)^{-1} X^T \varepsilon) = \theta + E((X^T X)^{-1} X^T \varepsilon).$$

Если матрица X фиксирована, то $E((X^T X)^{-1} X^T \varepsilon) = (X^T X)^{-1} X^T E(\varepsilon) = 0$, так что

$$E(\hat{\theta}) = \theta,$$

т.е. $\hat{\theta}$ – несмещенная оценка для θ .

Если же мы имеем дело со **стохастическими (случайными, недетерминированными)** объясняющими переменными, то в общем случае $E((X^T X)^{-1} X^T \varepsilon) \neq 0$, так что

$$E(\hat{\theta}) \neq \theta,$$

и $\hat{\theta}$ – смещенная оценка для θ , и, кроме того, эта оценка уже не имеет нормального распределения.

Если объясняющие переменные стохастические, то в некоторых случаях все же остается возможным использовать стандартную технику статистических выводов, предназначенную для классической нормальной линейной модели, по крайней мере, в асимптотическом плане (при большом количестве наблюдений).

В этом отношении наиболее благоприятной является

Ситуация А

- случайная величина ε_s не зависит (статистически) от $x_{t1}, x_{t2}, \dots, x_{tp}$ при всех t и s ,
- $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ являются независимыми случайными величинами, имеющими одинаковое нормальное распределение с нулевым математическим ожиданием и конечной дисперсией $\sigma^2 > 0$. (Далее мы кратко будем обозначать это как $\varepsilon_t \sim i.i.d. N(0, \sigma^2)$. Здесь **i.i.d.** – **independent, identically distributed**.)

При выполнении таких условий имеем

$$E((X^T X)^{-1} X^T \varepsilon) = E((X^T X)^{-1} X^T) \cdot E(\varepsilon) = 0,$$

так что оценка наименьших квадратов для θ является несмещенной. Распределение статистик критериев (“тестовых статистик”) можно найти с помощью двухшаговой процедуры. На первом шаге находим условное распределение при фиксированном значении X ; при этом значения объясняющих переменных рассматриваются как детерминированные (как в классической модели). На втором шаге мы получаем безусловное распределение соответствующей статистики, умножая условное распределение на плотность X и интегрируя по всем возможным значениям X .

Если применить такую процедуру для получения безусловного распределения оценки наименьших квадратов $\hat{\theta}$, то на первом шаге находим:

$$\hat{\theta} | X \sim N(\theta, \sigma^2 (X^T X)^{-1}).$$

Интегрирование на втором этапе приводит к распределению, являющемуся смесью нормальных распределений $N(\theta, \sigma^2 (X^T X)^{-1})$ по X . Это распределение, в отличие от классического случая, не является нормальным.

В то же время, для оценки j -го коэффициента имеем:

$$\hat{\theta}_j | X \sim N(\theta_j, \sigma^2 (X^T X)^{-1}_{jj}),$$

где $(X^T X)^{-1}_{jj}$ – j -й диагональный элемент матрицы $(X^T X)^{-1}$,

так что

$$\frac{\hat{\theta}_j - \theta_j}{\sigma \sqrt{(X^T X)^{-1}_{jj}}} | X \sim N(0, 1).$$

Условным распределением отношения S^2/σ^2 , где $S^2 = RSS/(n-p)$, RSS – остаточная сумма квадратов, является распределение хи-квадрат с $(n-p)$ степенями свободы,

$$S^2/\sigma^2 | X \sim \chi^2(n-p).$$

Заметим теперь, что t -статистика для проверки гипотезы $H_0: \theta_j = \theta_j^*$ определяется соотношением

$$t = \frac{\hat{\theta}_j - \theta_j^*}{S \sqrt{(X^T X)^{-1}_{jj}}} = \frac{(\hat{\theta}_j - \theta_j^*) / \sigma \sqrt{(X^T X)^{-1}_{jj}}}{\sqrt{S^2/\sigma^2}}.$$

Из предыдущего вытекает, что если гипотеза H_0 верна, то условное распределение этой t -статистики имеет t -распределение Стьюдента с $(n-p)$ степенями свободы,

$$t | X \sim t(n-p).$$

Это условное распределение одно и то же для всех X . Поэтому вне зависимости от того, какое именно распределение имеет X , безусловным распределением t -статистики для $H_0: \theta_j = \theta_j^*$ при выполнении этой гипотезы будет все то же распределение $t(n-p)$.

Аналогичное рассмотрение показывает возможность использования стандартных F -критериев для проверки линейных гипотез о значениях коэффициентов.

Те же самые выводы остаются в силе при замене предположений ситуации А следующим предположением.

Ситуация А'

- $\varepsilon \mid X \sim N(0, \sigma^2 I_n)$, где I_n – единичная матрица (размера $n \times n$).

Для краткости мы будем далее обозначать

$x_t = (x_{t1}, x_{t2}, \dots, x_{tp})^T$ – вектор значений p объясняющих переменных в t -м наблюдении.

Ситуация В

- случайная величина ε_s не зависит (статистически) от $x_{t1}, x_{t2}, \dots, x_{tp}$ при всех t и s ;
- распределение случайной величины ε_t не является нормальным, но $\varepsilon_t \sim i.i.d.$, $E(\varepsilon_t) = 0$, $D(\varepsilon_t) = \sigma^2 > 0$ и $E(\varepsilon_t^4) = \mu_4 < \infty$;
- $E(x_t x_t^T) = Q_t$ – положительно определенная матрица, $(1/n)(Q_1 + \dots + Q_n) \rightarrow Q$ при $n \rightarrow \infty$, где Q – положительно определенная матрица;
- $E(x_{it} x_{jt} x_{kt} x_{st}) < \infty$ для всех i, j, k, s ;
- $(1/n)(x_1 x_1^T + \dots + x_n x_n^T) \rightarrow Q$ по вероятности.

В силу первого предположения, оценка наименьших квадратов вектора коэффициентов θ остается несмещенной, как и в ситуации А. Однако при конечном количестве наблюдений n из-за негауссовости (ненормальности) распределения ε_t , распределения статистики S^2 , а также t - и F -статистик, будут отличаться от стандартных, получаемых в предположении гауссовости. Чтобы продолжать пользоваться обычной техникой регрессионного анализа, мы должны здесь сослаться на следующие асимптотические результаты, строгий вывод которых можно найти, например, в книге [Hamilton (1994)].

Пусть $\hat{\theta}(n)$ – оценка наименьших квадратов вектора θ по n наблюдениям, X_n – матрица значений объясняющих переменных для n наблюдений, а S_n^2 , t_n , F_n – статистики S^2 , t , F , вычисляемые по n наблюдениям. Если выполнены предположения, перечисленные при описании ситуации В, то при $n \rightarrow \infty$

$$\sqrt{n} (\hat{\theta}(n) - \theta) \rightarrow N(0, \sigma^2 Q^{-1}),$$

$$\sqrt{n} (S_n^2 - \sigma^2) \rightarrow N(0, \mu_4 - \sigma^4),$$

$$t_n \rightarrow N(0, 1),$$

$$qF_n \rightarrow \chi^2(q), \text{ где } q \text{ – количество линейных ограничений на компоненты вектора } \theta.$$

Здесь везде имеются в виду сходимости по распределению, т.е. распределения случайных величин, стоящих слева от стрелки, неограниченно сближаются при $n \rightarrow \infty$ с распределениями, стоящими справа от стрелки. При этом имеют место приближенные соотношения

$$\hat{\theta}(n) \approx N(\theta, \sigma^2 Q^{-1}/n), \text{ или } \hat{\theta}(n) \approx N(\theta, \sigma^2 (X_n^T X_n)^{-1}),$$

(последнее аналогично точному соотношению в гауссовской модели),

$$S_n^2 \approx N(\sigma^2, (\mu_4 - \sigma^4)/n),$$

$$t_n \approx N(0, 1),$$

$$qF_n \approx \chi^2(q).$$

Если в ситуации В при имеющемся количестве наблюдений n использовать не асимптотические распределения, а распределение Стьюдента $t(n-p)$ для t -статистики (вместо $N(0, 1)$) и распределение Фишера $F(q, n-p)$ для F -статистики (вместо $\chi^2(q)$ для qF_n), то это приводит к более широким доверительным интервалам (по сравнению с интервалами, построенными по асимптотическим распределениям). Многие исследователи предпочитают поступать именно таким образом, учитывая это обстоятельство и то, что при конечных n распределения Стьюдента и Фишера могут давать лучшую аппроксимацию истинных распределений статистик t_n и F_n .

Ситуация С

В рассмотренных выше ситуациях, как и в классической модели, предполагалось, что $\varepsilon_t | X \sim i.i.d.$ Теперь мы откажемся от этого предположения и предположим, что

- условное распределение случайного вектора ε относительно матрицы X является n -мерным нормальным распределением $N(0, \sigma^2 V)$;
- V – известная положительно определенная матрица размера $n \times n$.

Поскольку V – ковариационная матрица, она к тому же и симметрична; таковой же будет и обратная к ней матрица V^{-1} . Но тогда существует такая невырожденная $(n \times n)$ -матрица P , что $V^{-1} = P^T P$. Используя матрицу P , преобразуем вектор ε к вектору

$$\varepsilon^* = P \varepsilon.$$

При этом $E(\varepsilon^*) = 0$ и условная (относительно X) ковариационная матрица вектора ε^*

$$\begin{aligned} \text{Cov}(\varepsilon^* | X) &= E(\varepsilon^* \varepsilon^{*T} | X) = E(P \varepsilon (P \varepsilon)^T | X) \\ &= P E(\varepsilon \varepsilon^T | X) P^T = P \sigma^2 V P^T. \end{aligned}$$

Но $V = (V^{-1})^{-1} = (P^T P)^{-1}$, так что

$$\text{Cov}(\varepsilon^* | X) = P \sigma^2 V P^T = \sigma^2 P (P^T P)^{-1} P^T = \sigma^2 I_n.$$

Преобразуя с помощью матрицы P обе части основного уравнения

$$y = X\theta + \varepsilon,$$

получаем:

$$Py = PX\theta + P\varepsilon,$$

или

$$y^* = X^* \theta + \varepsilon^*,$$

где

$$y^* = Py, \quad X^* = PX, \quad \varepsilon^* = P\varepsilon.$$

В преобразованном уравнении

$$\varepsilon^* | X \sim N(0, \sigma^2 I_n),$$

так что преобразованная модель удовлетворяет условиям, характеризующим ситуацию А'. Это означает, что все результаты, полученные в ситуации А, применимы к модели $y^* = X^* \theta + \varepsilon^*$.

В частности, оценка наименьших квадратов

$$\theta^* = (X^{*T} X^*)^{-1} X^{*T} y^* = (X^T P^T P X)^{-1} X^T P^T P y = (X^T V^{-1} X)^{-1} X^T V^{-1} y$$

является несмещенной, т.е. $E(\theta^*) = \theta$, ее условное распределение (относительно X) нормально и имеет ковариационную матрицу

$$\text{Cov}(\theta^* | X) = \sigma^2 (X^{*T} X^*)^{-1} = \sigma^2 (X^T V^{-1} X)^{-1}.$$

Эта оценка известна как **обобщенная оценка наименьших квадратов (GLS – generalized least squares)**.

В рамках модели $y^* = X^* \theta + \varepsilon^*$ можно использовать обычные статистические процедуры, основанные на t - и F -статистиках.

Если ковариационная матрица V не известна априори, то обычно ограничиваются моделями, в которых она параметризована, так что $V = V(\beta)$, где β – векторный параметр, который приходится оценивать по имеющимся наблюдениям. При этом достаточно часто можно использовать стандартные выводы в асимптотическом плане, заменяя в выражении для GLS оценки $\theta^* = (X^T V^{-1} X)^{-1} X^T V^{-1} y$ неизвестную матрицу $V = V(\beta_0)$ ($V(\beta_0)$ – истинная ковариационная матрица вектора ε) матрицей $V(\hat{\beta}_n)$, где $\hat{\beta}_n$ – любая состоятельная оценка для β_0 . Более того, такую состоятельную оценку часто можно получить простым анализом остатков при оценивании обычной процедурой наименьших квадратов.

Рассмотренные ситуации не охватывают, однако, наиболее интересных для нас моделей стационарных и нестационарных временных рядов. Мы перейдем теперь к обсуждению основных понятий и фактов, касающихся стационарных и нестационарных временных рядов, и рассмотрению процедур регрессионного анализа временных рядов.

Глава 2. Стационарные ряды. Модели ARMA

2.1. Общие понятия.

Под **временным рядом** (*time series*) понимается последовательность наблюдений значений некоторой переменной, произведенных через равные промежутки времени. Если принять длину такого промежутка за единицу времени (год, квартал, день и т.п.), то можно считать, что последовательные наблюдения x_1, \dots, x_n произведены в моменты $t = 1, \dots, n$.

Основная отличительная особенность статистического анализа временных рядов состоит в том, что последовательность наблюдений x_1, \dots, x_n рассматривается как реализация последовательности, вообще говоря, статистически зависимых случайных величин X_1, \dots, X_n , имеющих некоторое совместное распределение с функцией распределения

$$F(v_1, v_2, \dots, v_n) = P\{X_1 < v_1, X_2 < v_2, \dots, X_n < v_n\}.$$

Мы будем рассматривать в основном временные ряды, у которых совместное распределение случайных величин X_1, \dots, X_n имеет совместную плотность распределения $p(x_1, x_2, \dots, x_n)$.

Чтобы сделать задачу статистического анализа временных рядов доступной для практического решения, приходится так или иначе ограничивать класс рассматриваемых моделей временных рядов, вводя те или иные предположения относительно структуры ряда и структуры его вероятностных характеристик. Одно из таких ограничений предполагает стационарность временного ряда.

Ряд x_t , $t = 1, \dots, n$, называется **строго стационарным** (или **стационарным в узком смысле**), если для любого m ($m < n$) совместное распределение вероятностей случайных величин X_{t_1}, \dots, X_{t_m} такое же, как и для $X_{t_1+\tau}, \dots, X_{t_m+\tau}$, при любых t_1, \dots, t_m и τ , таких, что $1 \leq t_1, \dots, t_m \leq n$ и $1 \leq t_1+\tau, \dots, t_m+\tau \leq n$.

Другими словами, свойства строго стационарного временного ряда не изменяются при изменении начала отсчета времени. В частности, при $m = 1$ из предположения о строгой стационарности временного ряда x_t следует, что закон распределения вероятностей случайной величины X_t не зависит от t , а значит, не зависят от t и все его основные числовые характеристики (если, конечно, они существуют), в том числе: математическое ожидание $E(X_t) = \mu$ и дисперсия $D(X_t) = \sigma^2$.

Значение μ определяет постоянный уровень, относительно которого колеблется анализируемый временной ряд x_t , а постоянная σ характеризует размах этих колебаний.

Как мы уже говорили, одно из главных отличий последовательности наблюдений, образующих временной ряд, заключается в том, что члены временного ряда являются, вообще говоря, статистически взаимозависимыми. Степень тесноты статистической связи между случайными величинами X_t и $X_{t+\tau}$ может быть измерена парным коэффициентом корреляции

$$\text{Corr}(X_t, X_{t+\tau}) = \frac{\text{Cov}(X_t, X_{t+\tau})}{\sqrt{D(X_t)}\sqrt{D(X_{t+\tau})}},$$

где

$$\text{Cov}(X_t, X_{t+\tau}) = E[(X_t - E(X_t))(X_{t+\tau} - E(X_{t+\tau}))].$$

Если ряд x_t стационарный, то значение $Cov(X_t, X_{t+\tau})$ не зависит от t и является функцией только от τ ; мы будем использовать для него обозначение $\gamma(\tau)$:

$$\gamma(\tau) = Cov(X_t, X_{t+\tau}).$$

В частности,

$$D(X_t) = Cov(X_t, X_t) \equiv \gamma(0).$$

Соответственно, для стационарного ряда и значение коэффициента корреляции $Corr(X_t, X_{t+\tau})$ зависит только от τ ; мы будем использовать для него обозначение $\rho(\tau)$, так что

$$\rho(\tau) = Corr(X_t, X_{t+\tau}) = \gamma(\tau) / \gamma(0).$$

В частности, $\rho(0) = 1$.

Практическая проверка строгой стационарности ряда x_t на основании наблюдения значений x_1, x_2, \dots, x_n в общем случае затруднительна. В связи с этим под стационарным рядом на практике часто подразумевают временной ряд x_t , у которого

- $E(X_t) \equiv \mu$,
- $D(X_t) \equiv \sigma^2$,
- $Cov(X_t, X_{t+\tau}) = \gamma(\tau)$ для любых t и τ .

Ряд, для которого выполнены указанные три условия, называют **стационарным в широком смысле (слабо стационарным, стационарным второго порядка или ковариационно стационарным)**.

Если ряд является стационарным в широком смысле, то он не обязательно является строго стационарным. В то же время, и строго стационарный ряд может не быть стационарным в широком смысле просто потому, что у него могут не существовать математическое ожидание и/или дисперсия. (В отношении последнего примером может служить случайная выборка из распределения Коши.) Кроме того, возможны ситуации, когда указанные три условия выполняются, но, например, $E(X_t^3)$ зависит от t .

Ряд x_t , $t = 1, \dots, n$, называется **гауссовским**, если совместное распределение случайных величин X_1, \dots, X_n является n -мерным нормальным распределением. Для гауссовского ряда понятия стационарности в узком и в широком смысле совпадают.

В дальнейшем, говоря о стационарности некоторого ряда x_t , мы (если не оговаривается противное) будем иметь в виду, что этот ряд стационарен в широком смысле (так что у него существуют математическое ожидание и дисперсия).

Итак, пусть x_t – стационарный ряд с $E(X_t) \equiv \mu$, $D(X_t) \equiv \sigma^2$ и $\rho(\tau) = Corr(X_t, X_{t+\tau})$. Поскольку в данном случае коэффициент $\rho(\tau)$ измеряет корреляцию между членами одного и того же временного ряда, его принято называть **коэффициентом автокорреляции** (или просто **автокорреляцией**). По той же причине о ковариациях $\gamma(\tau) = Cov(X_t, X_{t+\tau})$ говорят как об **автоковариациях**. При анализе изменения величины $\rho(\tau)$ в зависимости от значения τ принято говорить об **автокорреляционной функции** $\rho(\tau)$. Автокорреляционная функция безразмерна, т.е. не зависит от масштаба измерения анализируемого временного ряда. Ее значения могут изменяться в пределах от -1 до $+1$; при этом $\rho(0) = 1$. Кроме того, из стационарности ряда x_t следует, что $\rho(\tau) = \rho(-\tau)$, так что при анализе поведения автокорреляционных функций обычно ограничиваются рассмотрением только неотрицательных значений τ .

График зависимости $\rho(\tau)$ от τ часто называют **коррелограммой**. Он может использоваться для характеристики некоторых свойств механизма, порождающего временной ряд. Для дальнейшего заметим, что если x_t – стационарный временной ряд и

c – некоторая постоянная, то временные ряды x_t и $(x_t + c)$ имеют одинаковые коррелограммы.

Если предположить, что временной ряд описывается моделью стационарного гауссовского процесса, то полное описание совместного распределения случайных величин X_1, \dots, X_n требует задания $n+1$ параметров: $\mu, \gamma(0), \gamma(1), \dots, \gamma(n-1)$ (или $\mu, \gamma(0), \rho(1), \dots, \rho(n-1)$). Это намного меньше, чем без требования стационарности, но все же больше, чем количество наблюдений. В связи с этим, даже для стационарных гауссовских временных рядов приходится производить дальнейшее упрощение модели с тем, чтобы ограничить количество параметров, подлежащих оцениванию по имеющимся наблюдениям. Мы переходим теперь к рассмотрению некоторых простых по структуре временных рядов, которые, в то же время, полезны для описания эволюции во времени многих реальных экономических показателей.

2.2. Процесс белого шума

Процессом белого шума (“белым шумом”, “чисто случайным временным рядом”) называют стационарный временной ряд x_t , для которого

$$E(X_t) \equiv 0, D(X_t) \equiv \sigma^2 > 0$$

и

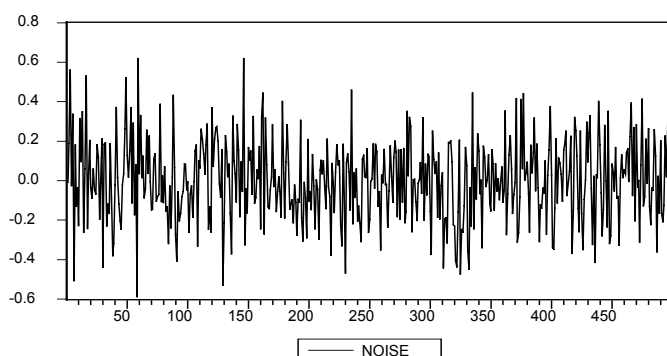
$$\rho(\tau) = 0 \text{ при } \tau \neq 0.$$

Последнее означает, что при $t \neq s$ случайные величины X_t и X_s , соответствующие наблюдениям процесса белого шума в моменты t и s , некоррелированы.

В случае, когда X_t имеет нормальное распределение, случайные величины X_1, \dots, X_n взаимно независимы и имеют одинаковое нормальное распределение $N(0, \sigma^2)$, образуя случайную выборку из этого распределения, т.е. $X_t \sim i.i.d. N(0, \sigma^2)$. Такой ряд называют *гауссовским белым шумом*.

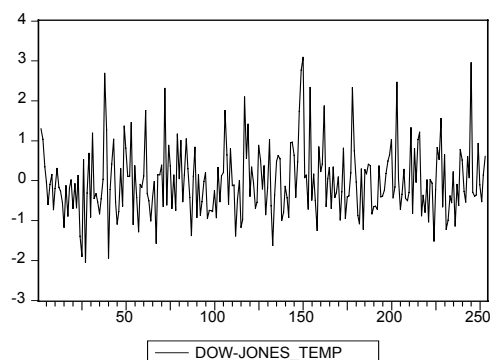
В то же время, в общем случае, даже если некоторые случайные величины X_1, \dots, X_n взаимно независимы и имеют одинаковое распределение, то это еще не означает, что они образуют процесс белого шума, т.к. случайная величина X_t может просто не иметь математического ожидания и/или дисперсии (в качестве примера мы опять можем указать на распределение Коши).

Временной ряд, соответствующий процессу белого шума, ведет себя крайне нерегулярным образом из-за некоррелированности при $t \neq s$ случайных величин X_t и X_s . Это иллюстрирует приводимый ниже график смоделированной реализации гауссовского процесса белого шума (NOISE) с $D(X_t) \equiv 0.04$.



В связи с этим процесс белого шума не годится для непосредственного моделирования эволюции большинства временных рядов, встречающихся в экономике. В то же время, как мы увидим ниже, такой процесс является базой для построения более реалистичных моделей временных рядов, порождающих “более гладкие” траектории ряда. В связи с частым использованием процесса белого шума в дальнейшем изложении, мы будем отличать этот процесс от других моделей временных рядов, используя для него обозначение ε_t .

В качестве примера ряда, траектория которого похожа на реализацию процесса белого шума, можно указать, например, на ряд, образованный значениями темпов изменения (прироста) индекса Доу-Джонса в течение 1984 года (дневные данные). График этого ряда имеет вид



Заметим, однако, что здесь наблюдается некоторая асимметрия распределения вероятностей значений x_t (скошенность этого распределения в сторону положительных значений), что исключает описание модели этого ряда как гауссовского белого шума.

2.3. Процесс авторегрессии

Одной из широко используемых моделей временных рядов является **процесс авторегрессии** (**модель авторегрессии**). В своей простейшей форме модель авторегрессии описывает механизм порождения ряда следующим образом:

$$X_t = a X_{t-1} + \varepsilon_t, \quad t = 1, \dots, n,$$

где ε_t – процесс белого шума, имеющий нулевое математическое ожидание и дисперсию σ_ε^2 , X_0 – некоторая случайная величина, а $a \neq 0$ – некоторый постоянный коэффициент.

При этом

$$E(X_t) = a E(X_{t-1}),$$

так что рассматриваемый процесс может быть стационарным только если $E(X_t) = 0$ для всех $t = 0, 1, \dots, n$.

Далее,

$$X_t = a X_{t-1} + \varepsilon_t = a (a X_{t-2} + \varepsilon_{t-1}) + \varepsilon_t = a^2 X_{t-2} + a \varepsilon_{t-1} + \varepsilon_t = \dots =$$

$$= a^t X_0 + a^{t-1} \varepsilon_1 + a^{t-2} \varepsilon_2 + \dots + \varepsilon_t,$$

$$X_{t-1} = a X_{t-2} + \varepsilon_{t-1} = a^{t-1} X_0 + a^{t-2} \varepsilon_1 + a^{t-3} \varepsilon_2 + \dots + \varepsilon_{t-1},$$

$$X_{t-2} = a X_{t-3} + \varepsilon_{t-2} = a^{t-2} X_0 + a^{t-3} \varepsilon_1 + a^{t-4} \varepsilon_2 + \dots + \varepsilon_{t-2},$$

...

$$X_1 = a X_0 + \varepsilon_1.$$

Если случайная величина X_0 не коррелирована со случайными величинами $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$, то отсюда следует, что

$$\text{Cov}(X_0, \varepsilon_1) = 0, \text{Cov}(X_1, \varepsilon_2) = 0, \dots, \text{Cov}(X_{t-2}, \varepsilon_{t-1}) = 0, \text{Cov}(X_{t-1}, \varepsilon_t) = 0$$

и

$$D(X_t) = D(aX_{t-1} + \varepsilon_t) = a^2 D(X_{t-1}) + D(\varepsilon_t), \quad t = 1, \dots, n.$$

Предполагая, наконец, что

$$D(X_0) = D(X_t) = \sigma_X^2 \quad \text{для всех } t = 1, \dots, n,$$

находим:

$$\sigma_X^2 = a^2 \sigma_X^2 + \sigma_\varepsilon^2.$$

Последнее может выполняться только при выполнении условия $a^2 < 1$, т.е. $|a| < 1$.

При этом получаем выражение для σ_X^2

$$\sigma_X^2 = \sigma_\varepsilon^2 / (1 - a^2).$$

Что касается автоковариаций и автокорреляций, то

$$\begin{aligned} \text{Cov}(X_t, X_{t+\tau}) &= \text{Cov}(a^t X_0 + a^{t-1} \varepsilon_1 + a^{t-2} \varepsilon_2 + \dots + \varepsilon_t, \\ &\quad a^{t+\tau} X_0 + a^{t+\tau-1} \varepsilon_1 + a^{t+\tau-2} \varepsilon_2 + \dots + \varepsilon_{t+\tau}) = \\ &= a^{2t+\tau} D(X_0) + a^\tau (1 + a^2 + \dots + a^{2(t-1)}) \sigma_\varepsilon^2 = \\ &= a^\tau [a^{2t} \sigma_\varepsilon^2 / (1 - a^2) + (1 - a^{2t}) \sigma_\varepsilon^2 / (1 - a^2)] = [a^\tau / (1 - a^2)] \sigma_\varepsilon^2, \end{aligned}$$

и

$$\text{Corr}(X_t, X_{t+\tau}) = a^\tau,$$

т.е. при сделанных предположениях автоковариации и автокорреляции зависят только от того, насколько разнесены по времени соответствующие наблюдения.

Таким образом, механизм порождения последовательных наблюдений, заданный соотношениями

$$X_t = a X_{t-1} + \varepsilon_t, \quad t = 1, \dots, n,$$

порождает стационарный временной ряд, если

- $|a| < 1$;
- случайная величина X_0 не коррелирована со случайными величинами $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$;
- $E(X_0) = 0$;
- $D(X_0) = \sigma_\varepsilon^2 / (1 - a^2)$.

При этом $\text{Corr}(X_t, X_{t+\tau}) = \rho(\tau) = a^\tau$.

Рассмотренная модель порождает (при указанных условиях) стационарный ряд, имеющий нулевое математическое ожидание. Однако ее можно легко распространить и на временные ряды y_t с ненулевым математическим ожиданием $E(Y_t) = \mu$, полагая, что указанная модель относится к центрированному ряду $X_t = Y_t - \mu$:

$$Y_t - \mu = a (Y_{t-1} - \mu) + \varepsilon_t, \quad t = 1, \dots, n,$$

так что

$$Y_t = a Y_{t-1} + \delta + \varepsilon_t, \quad t = 1, \dots, n,$$

где

$$\delta = \mu (1 - a).$$

Поэтому без ограничения общности можно обойтись в текущем рассмотрении моделями авторегрессии, порождающими стационарный процесс с нулевым средним.

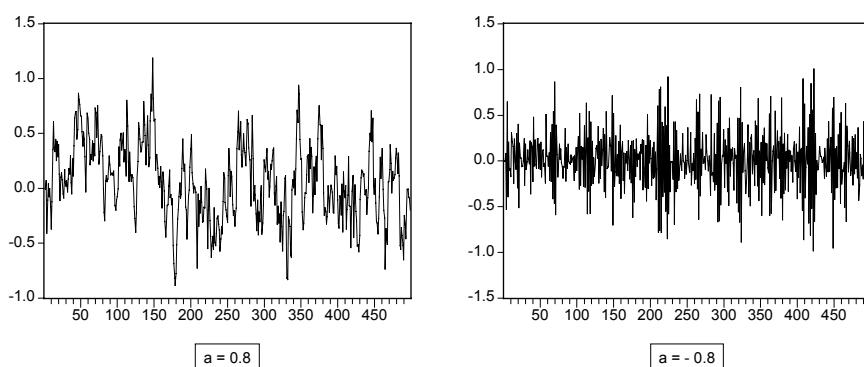
Продолжая рассмотрение для ранее определенного процесса X_t (с нулевым математическим ожиданием), заметим, что для него

$$\gamma(1) = E(X_t \cdot X_{t-1}) = E[(a X_{t-1} + \varepsilon_t) \cdot X_{t-1}] = a\gamma(0),$$

так что

$$\rho(1) = \gamma(1)/\gamma(0) = a,$$

и при значениях $a > 0$, близких к 1, между соседними наблюдениями имеется сильная положительная корреляция, что обеспечивает более гладкий характер поведения траекторий ряда по сравнению с процессом белого шума. При $a < 0$ процесс авторегрессии, напротив, имеет менее гладкие реализации, поскольку в этом случае проявляется тенденция чередования знаков последовательных наблюдений. Следующие два графика демонстрируют поведение смоделированных реализаций временных рядов, порожденных моделями авторегрессии $X_t = a X_{t-1} + \varepsilon_t$ с $\sigma_\varepsilon^2 = 0.2$ при $a = 0.8$ (первый график) и $a = -0.8$ (второй график).



Теперь мы должны обратить внимание на следующее важное обстоятельство. В практических ситуациях “стартовое” значение $X_0 = x_0$, на основе которого в соответствии с соотношением $X_t = a X_{t-1} + \varepsilon_t$ строятся последующие значения ряда X_t , может относиться к концу предыдущего периода, на котором просто в силу других экономических условий эволюция соответствующего экономического показателя следует иной модели, например, модели $X_t = a X_{t-1} + \varepsilon_t$ с другими значениями a и σ_ε^2 . Более того, статистические данные о поведении ряда до момента $t = 0$ могут отсутствовать вовсе, так что значение x_0 является просто некоторой наблюдаемой числовой величиной. В обоих случаях ряд X_t уже не будет стационарным даже при $|a| < 1$. Рассмотрим подробнее характеристики и поведение ряда в таких ситуациях.

Если не конкретизировать модель, в соответствии с которой порождались наблюдения до момента $t = 1$, то значение x_0 можно рассматривать как фиксированное. При этом

$$X_t = a^t x_0 + a^{t-1} \varepsilon_1 + a^{t-2} \varepsilon_2 + \dots + \varepsilon_t,$$

$$E(X_t) = a^t x_0 + a^{t-1} E(\varepsilon_1) + a^{t-2} E(\varepsilon_2) + \dots + E(\varepsilon_t) = a^t x_0,$$

$$\begin{aligned} D(X_t) &= (a^{2(t-1)} + a^{2(t-2)} + \dots + 1) \sigma_\varepsilon^2 = [(1 - a^{2t}) / (1 - a^2)] \sigma_\varepsilon^2 \\ &= \sigma_\varepsilon^2 / (1 - a^2) - [a^{2t} / (1 - a^2)] \sigma_\varepsilon^2, \end{aligned}$$

$$\begin{aligned} \text{Cov}(X_t, X_{t+\tau}) &= \text{Cov}(X_t - a^t x_0, X_{t+\tau} - a^{t+\tau} x_0) = a^t (1 + a^2 + \dots + a^{2(t-1)}) \sigma_\varepsilon^2 \\ &= a^t (1 - a^{2t}) \sigma_\varepsilon^2 / (1 - a^2), \end{aligned}$$

так что и математическое ожидание и дисперсия случайной величины X_t , а также ковариации $\text{Cov}(X_t, X_{t+\tau})$ зависят от t .

В то же время, если $|a| < 1$, то при $t \rightarrow \infty$ получаем

$$E(X_t) \rightarrow 0, D(X_t) \rightarrow \sigma_\varepsilon^2 / (1 - a^2), \text{Cov}(X_t, X_{t+\tau}) \rightarrow a^t [\sigma_\varepsilon^2 / (1 - a^2)],$$

т.е. при $t \rightarrow \infty$ значения математического ожидания и дисперсии случайной величины X_t , а также автоковариации $Cov(X_t, X_{t+\tau})$ стабилизируются, приближаясь к своим предельным значениям.

С этой точки зрения, условие $|a| < 1$ можно трактовать как условие стабильности ряда, порождаемого моделью $X_t = aX_{t-1} + \varepsilon_t$ при фиксированном значении $X_0 = x_0$. Рассмотрим в этой ситуации наряду с только что исследованным рядом X_t ,

$$X_t = a^t x_0 + \sum_{k=0}^{t-1} a^k \varepsilon_{t-k}, \quad |a| < 1,$$

ряд, порождаемый моделью

$$\tilde{X}_t = \sum_{k=0}^{\infty} a^k \varepsilon_{t-k}.$$

Имеем:

$$\tilde{X}_t - X_t = -a^t x_0 + \sum_{k=t}^{\infty} a^k \varepsilon_{t-k};$$

при $t \rightarrow \infty$

$$a^t x_0 \rightarrow 0 \text{ и } E \left| \sum_{k=t}^{\infty} a^k \varepsilon_{t-k} \right|^2 = \sigma_{\varepsilon}^2 \sum_{k=t}^{\infty} a^{2k} \rightarrow 0.$$

Таким образом, ряд \tilde{X}_t является предельным для X_t ; ряд X_t “выходит на режим” \tilde{X}_t при $t \rightarrow \infty$. При этом выход ряда X_t на режим \tilde{X}_t происходит тем быстрее, чем ближе X_0 и a к нулю.

Для ряда \tilde{X}_t

$$E(\tilde{X}_t) = E \left(\sum_{k=0}^{\infty} a^k \varepsilon_{t-k} \right) = \sum_{k=0}^{\infty} a^k E(\varepsilon_{t-k}) = 0,$$

$$D(\tilde{X}_t) = D \left(\sum_{k=0}^{\infty} a^k \varepsilon_{t-k} \right) = \sum_{k=0}^{\infty} a^{2k} D(\varepsilon_{t-k}) = \sigma_{\varepsilon}^2 \sum_{k=0}^{\infty} a^{2k} = \frac{\sigma_{\varepsilon}^2}{1-a^2},$$

$$Cov(\tilde{X}_t, \tilde{X}_{t+\tau}) = E \left[\left(\sum_{k=0}^{\infty} a^k \varepsilon_{t-k} \right) \left(\sum_{k=0}^{\infty} a^k \varepsilon_{t+\tau-k} \right) \right] = a^{\tau} \left(\sum_{k=0}^{\infty} a^{2k} E(\varepsilon_{t-k}^2) \right) = a^{\tau} \frac{\sigma_{\varepsilon}^2}{1-a^2},$$

так что \tilde{X}_t – стационарный ряд (в широком смысле). Кроме того,

$$\tilde{X}_{t-1} = \frac{1}{a} \sum_{k=1}^{\infty} a^k \varepsilon_{t-k},$$

так что

$$a\tilde{X}_{t-1} + \varepsilon_t = \sum_{k=0}^{\infty} a^k \varepsilon_{t-k} = \tilde{X}_t,$$

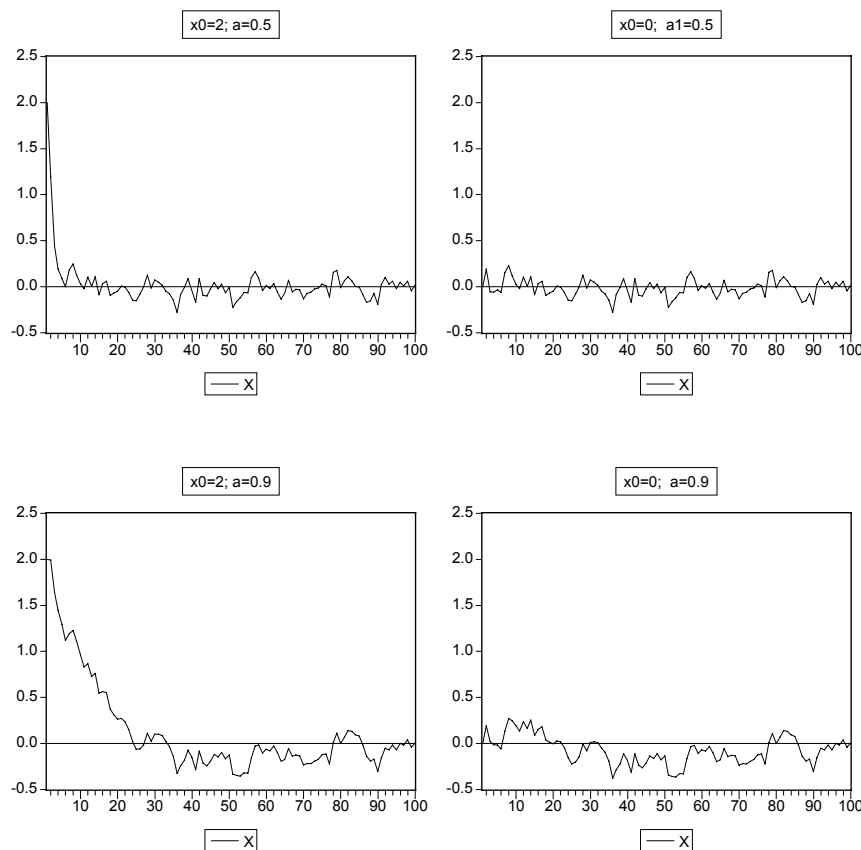
т.е. \tilde{X}_t удовлетворяет соотношению

$$\tilde{X}_t = a\tilde{X}_{t-1} + \varepsilon_t.$$

Поскольку ε_t не входит в правую часть выражений для $\tilde{X}_{t-1}, \tilde{X}_{t-2}, \dots$, то случайная величина ε_t не коррелирована с $\tilde{X}_{t-1}, \tilde{X}_{t-2}, \dots$, т.е. ε_t является инновацией

(обновлением). В итоге получаем, что \tilde{X}_t – стационарный процесс авторегрессии первого порядка, и фактически именно этот процесс имеется в виду, когда говорят о стационарном процессе AR(1).

Проиллюстрируем сказанное выше с помощью смоделированных реализаций ряда x_t , порожденных моделью $X_t = a X_{t-1} + \varepsilon_t$ с $\sigma_\varepsilon = 0.2$ и различными значениями коэффициента a и стартового значения x_0 .



Рассмотренную только что модель $X_t = a X_{t-1} + \varepsilon_t$ называют процессом авторегрессии первого порядка. **Процесс авторегрессии порядка p** (в кратком обозначении – **AR(p)**) определяется соотношениями

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} + \dots + a_p X_{t-p} + \varepsilon_t, \quad a_p \neq 0,$$

где ε_t – процесс белого шума с $D(\varepsilon_t) = \sigma_\varepsilon^2$. Для простоты мы будем теперь сразу полагать, что $Cov(X_{t-s}, \varepsilon_t) = 0$ для всех $s > 0$; при этом говорят, что случайные величины ε_t образуют **инновационную (обновляющую) последовательность**, а случайная величина ε_t называется **инновацией** для наблюдения в момент t . Такая терминология объясняется тем, что наблюдаемое значение ряда в момент t получается здесь как линейная комбинация p предшествующих значений этого ряда плюс не коррелированная с этими предшествующими значениями случайная составляющая ε_t , отражающая обновленную информацию, скажем, о состоянии экономики, на момент t , влияющую на наблюдаемое значение X_t .

При рассмотрении процессов авторегрессии и некоторых других моделей удобно использовать **оператор запаздывания L (lag operator)**, который воздействует на временной ряд и определяется соотношением

$$L X_t = X_{t-1};$$

в некоторых руководствах его называют оператором обратного сдвига и используют для него обозначение B (*backshift operator*).

Если оператор запаздывания применяется k раз, что обозначается как L^k , то это дает в результате

$$L^k X_t = X_{t-k}.$$

Выражение

$$a_1 X_{t-1} + a_2 X_{t-2} + \dots + a_p X_{t-p}$$

можно записать теперь в виде

$$(a_1 L + a_2 L^2 + \dots + a_p L^p) X_t,$$

а соотношение, определяющее процесс авторегрессии p -го порядка, в виде

$$a(L) X_t = \varepsilon_t,$$

где

$$a(L) = 1 - (a_1 L + a_2 L^2 + \dots + a_p L^p).$$

Для того, чтобы такой процесс был стационарным, все корни алгебраического уравнения

$$a(z) = 0$$

(вещественные и комплексные) должны лежать вне единичного круга $|z| \leq 1$. (В частности, для процесса AR(1) имеем $a(z) = 1 - a z$, уравнение $a(z) = 0$ имеет корень $z = 1/a$, и условие стационарности $|z| > 1$ равносильно уже знакомому нам условию $|a| < 1$.) При этом решение уравнения $a(L) X_t = \varepsilon_t$ можно представить в виде

$$X_t = \frac{1}{a(L)} \varepsilon_t = \sum_{j=0}^{\infty} b_j \varepsilon_{t-j}, \quad \text{где} \quad \sum_{j=0}^{\infty} |b_j| < \infty,$$

откуда, в частности, следует, что

$$E(X_t) = E\left(\sum_{j=0}^{\infty} b_j \varepsilon_{t-j}\right) = \sum_{j=0}^{\infty} b_j E(\varepsilon_{t-j}) = 0.$$

Стационарный процесс AR(p) с ненулевым математическим ожиданием μ удовлетворяет соотношению

$$a(L) (X_t - \mu) = \varepsilon_t,$$

или

$$a(L) X_t = \delta + \varepsilon_t,$$

где

$$\delta = a(L)\mu = \mu(1 - a_1 - a_2 - \dots - a_p) = \mu a(1).$$

При этом решение уравнения $a(L) (X_t - \mu) = \varepsilon_t$ имеет вид

$$X_t = \mu + \frac{1}{a(L)} \varepsilon_t.$$

Таким образом, если стационарный процесс AR(p) задан в виде $a(L) X_t = \delta + \varepsilon_t$, то следует помнить о том, что в этом случае математическое ожидание этого процесса равно не δ , а

$$\mu = \frac{\delta}{(1 - a_1 - a_2 - \dots - a_p)}.$$

Для процесса AR(1) имеем $a(L) = 1 - aL$, так что (вне зависимости от того, равно μ нулю или нет)

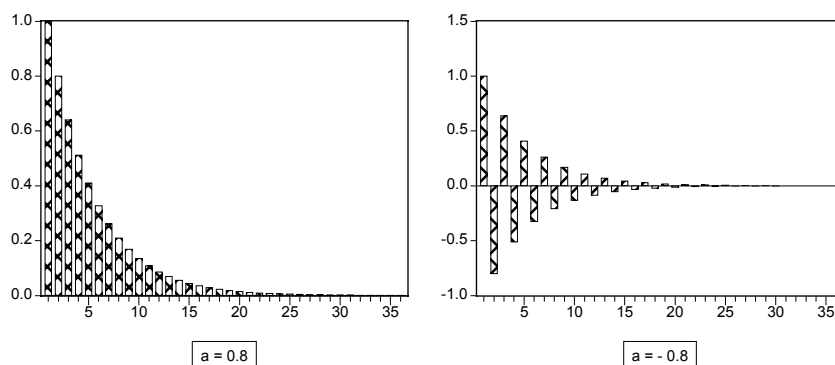
$$X_t - \mu = (1 / (1 - aL)) \varepsilon_t = (1 + aL + a^2L^2 + \dots) \varepsilon_t = \varepsilon_t + a\varepsilon_{t-1} + a^2\varepsilon_{t-2} + \dots$$

Из последнего выражения сразу видно, что

$$\rho(k) = \text{Corr}(X_t, X_{t+k}) = a^k, \quad k = 0, 1, 2, \dots$$

При $0 < a < 1$ коррелограмма (график функции $\rho(k)$ для $k = 0, 1, 2, \dots$) отражает показательное убывание корреляций с возрастанием интервала между наблюдениями; при $-1 < a < 0$ коррелограмма имеет характер затухающей косинусоиды.

Сравним поведение коррелограмм стационарного процесса AR(1) при $a = \pm 0.8$:



Коррелограмма процесса AR(p) при $p > 1$ имеет более сложную форму, зависящую от расположения (на комплексной плоскости) корней уравнения $a(z) = 0$. Однако для больших значений k автокорреляция $\rho(k)$ хорошо аппроксимируется значением $A\theta^k$, где $\theta = 1/z_{\min}$ и z_{\min} – наименьший по абсолютной величине корень уравнения $a(z) = 0$, если этот корень является вещественным и положительным, или заключена в интервале $\pm A |\theta^k|$ в противном случае. Здесь $A > 0$ – некоторая постоянная, определяемая коэффициентами a_1, \dots, a_p .

Если умножить на X_{t-k} ($k > 0$) обе части соотношения, определяющего процесс AR(p), и после этого взять от обеих частей математическое ожидание, то получим соотношение

$$\gamma(k) = a_1 \gamma(k-1) + a_2 \gamma(k-2) + \dots + a_p \gamma(k-p), \quad k > 0.$$

Разделив обе части последнего на $\gamma(0)$, приходим к системе уравнений Юла–Уокера

$$\rho(k) = a_1 \rho(k-1) + a_2 \rho(k-2) + \dots + a_p \rho(k-p), \quad k > 0.$$

Эта система позволяет последовательно находить значения автокорреляций и дает возможность, используя первые p уравнений, выразить коэффициенты a_j через значения первых p автокорреляций, что можно непосредственно использовать при подборе модели авторегрессии к реальным статистическим данным (см. разделы 3.1 и 3.2).

Пример. Рассмотрим процесс авторегрессии AR(2)

$$X_t = 4.375 + 0.25X_{t-1} - 0.125X_{t-2} + \varepsilon_t.$$

Уравнение $a(z) = 0$ принимает в этом случае вид

$$1 - 0.25z + 0.125z^2 = 0, \quad \text{или} \quad z^2 - 2z + 8 = 0,$$

и имеет корни $z_{1,2} = 1 \pm i\sqrt{7}$. Оба корня по абсолютной величине больше единицы, так что процесс стационарный. Математическое ожидание этого процесса равно

$$\mu = \delta / (1 - a_1 - a_2) = 4.375 / (1 - 0.25 + 0.125) = 5,$$

так что траектории этого процесса флуктуируют вокруг уровня 5.

Для построения коррелограммы воспользуемся уравнениями Юла – Уокера. У нас $p = 2$, так что

$$\rho(k) = 0.25 \rho(k-1) - 0.125 \rho(k-2), \quad k > 0.$$

По определению, $\rho(0) = 1$. Для $\rho(1)$ имеем

$$\rho(1) = 0.25 \rho(0) - 0.125 \rho(-1) = 0.25 - 0.125 \rho(1),$$

откуда находим:

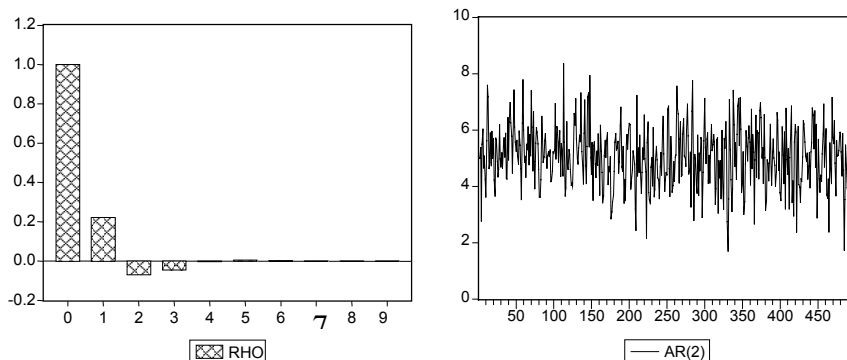
$$\rho(1) = 0.25 / (1 + 0.125) = 2/9 = 0.222.$$

Далее последовательно находим:

$$\rho(2) = 0.25 \rho(1) - 0.125 \rho(0) = 0.25 * 0.222 - 0.125 = -0.069,$$

$$\rho(3) = -0.045, \quad \rho(4) = -0.003, \quad \rho(5) = 0.005 \text{ и т.д.}$$

Корреляции даже между соседними наблюдениями очень малы, и можно ожидать, что поведение траекторий такого ряда не очень существенно отличается от поведения реализаций процесса белого шума. Теоретическая коррелограмма рассматриваемого процесса и смоделированная реализация этого процесса приведены ниже.



2.4. Процесс скользящего среднего

Еще одной простой моделью порождения временного ряда является процесс *скользящего среднего порядка q* (MA(q)). Согласно этой модели,

$$X_t = \varepsilon_t + b_1 \varepsilon_{t-1} + b_2 \varepsilon_{t-2} + \dots + b_q \varepsilon_{t-q}, \quad b_q \neq 0,$$

где ε_t – процесс белого шума.

Такой процесс имеет нулевое математическое ожидание. Модель можно обобщить до процесса, имеющего ненулевое математическое ожидание μ , полагая

$$X_t - \mu = \varepsilon_t + b_1 \varepsilon_{t-1} + b_2 \varepsilon_{t-2} + \dots + b_q \varepsilon_{t-q},$$

т.е.

$$X_t = \mu + \varepsilon_t + b_1 \varepsilon_{t-1} + b_2 \varepsilon_{t-2} + \dots + b_q \varepsilon_{t-q}.$$

Для процесса скользящего среднего порядка q используется обозначение MA(q) (скользящее среднее – *moving average*).

При $q = 0$ и $\mu = 0$ получаем процесс белого шума. Если $q = 1$, то

$$X_t - \mu = \varepsilon_t + b \varepsilon_{t-1}$$

– скользящее среднее первого порядка. В последнем случае

$$D(X_t) = (1 + b^2) \sigma_\varepsilon^2, \quad E[(X_t - \mu)(X_{t-1} - \mu)] = b \sigma_\varepsilon^2, \quad E[(X_t - \mu)(X_{t-k} - \mu)] = 0, \quad k > 1,$$

так что процесс X_t является стационарным с $E(X_t) = 0$, $D(X_t) = (1 + b^2)\sigma_\varepsilon^2$,

$$\gamma(k) = \begin{cases} (1 + b^2)\sigma_\varepsilon^2, & k = 0, \\ b\sigma_\varepsilon^2, & k = 1, \\ 0, & k > 1. \end{cases}$$

Автокорреляции этого процесса равны

$$\rho(k) = \begin{cases} 1, & k = 0, \\ b/(1 + b^2), & k = 1, \\ 0, & k > 1, \end{cases}$$

т.е. коррелограмма процесса имеет весьма специфический вид. Коррелированными оказываются только соседние наблюдения. Корреляция между ними положительна, если $b > 0$, и отрицательна при $b < 0$. Соответственно, процесс $MA(1)$ с $b > 0$ имеет более гладкие, по сравнению с белым шумом, реализации, а процесс $MA(1)$ с $b < 0$ имеет менее гладкие, по сравнению с белым шумом, реализации. Заметим, что для любого процесса $MA(1)$

$$|\rho(1)| \leq 0.5,$$

т.е. корреляционная связь между соседними наблюдениями невелика, тогда как у процесса $AR(1)$ такая связь может быть сколь угодно сильной (при значениях $|a|$, близких к 1).

Модель $MA(q)$ кратко можно записать в виде

$$X_t - \mu = b(L) \varepsilon_t,$$

где

$$b(L) = 1 + b_1L + \dots + b_qL^q.$$

Для нее

$$\gamma(k) = E[(X_t - \mu)(X_{t-k} - \mu)] = \begin{cases} \left(\sum_{j=0}^{q-k} b_j b_{j+k} \right) \sigma_\varepsilon^2, & 0 \leq k \leq q, \\ 0, & k > q, \end{cases}$$

так что $MA(q)$ является стационарным процессом с нулевым математическим ожиданием, дисперсией

$$\sigma_X^2 = (1 + b_1^2 + \dots + b_q^2) \sigma_\varepsilon^2$$

и автокорреляциями

$$\rho_k = \begin{cases} \left(\sum_{j=0}^{q-k} b_j b_{j+k} \right) / \left(\sum_{j=0}^q b_j^2 \right), & k = 0, 1, \dots, q, \\ 0, & k = q+1, q+2, \dots \end{cases}$$

Здесь статистическая связь между наблюдениями сохраняется в течение q единиц времени (т.е. “длительность памяти” процесса равна q).

Подобного рода временные ряды соответствуют ситуации, когда некоторый экономический показатель находится в равновесии, но отклоняется от положения

равновесия в силу последовательно возникающих непредсказуемых событий, причем система такова, что влияние таких событий отмечается на протяжении некоторого периода времени.

Если влияние прошлых событий ослабевает с течением времени показательным образом, так что $b_j = a^j$, $0 < a < 1$, то искусственное предположение о том, что ряд ε_t начинается в “бесконечном прошлом”, приводит к модели бесконечного скользящего среднего $MA(\infty)$

$$X_t = \sum_{j=0}^{\infty} a^j \varepsilon_{t-j} = \sum_{j=0}^{\infty} b_j \varepsilon_{t-j}, \quad \text{где} \quad \sum_{j=0}^{\infty} |b_j| < \infty.$$

Ранее мы видели, что такое же представление допускает стационарный процесс авторегрессии первого порядка $AR(1)$

$$X_t = a X_{t-1} + \varepsilon_t, \quad |a| < 1,$$

т.е. в рассматриваемом случае процесс $MA(\infty)$ эквивалентен процессу $AR(1)$.

Вообще, всякий стационарный процесс $AR(p)$ можно записать в форме процесса $MA(\infty)$:

$$X_t = \mu + \frac{1}{a(L)} \varepsilon_t = \mu + \sum_{j=0}^{\infty} b_j \varepsilon_{t-j} = \mu + b(L) \varepsilon_t,$$

где

$$b(L) = \sum_{j=0}^{\infty} b_j L^j = \frac{1}{a(L)} \quad \text{и} \quad \sum_{j=0}^{\infty} |b_j| < \infty.$$

Примеры.

а) Рассмотрим процесс $MA(1)$ с $b = 0.8$ и $E(X_t) = 6$, т.е.

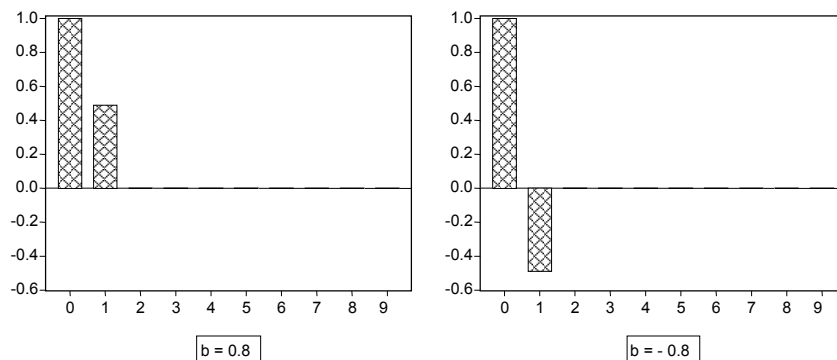
$$X_t = 6 + \varepsilon_t + 0.8 \varepsilon_{t-1}.$$

Для него $\rho(1) = 0.8/(1 + 0.8^2) = 0.488$.

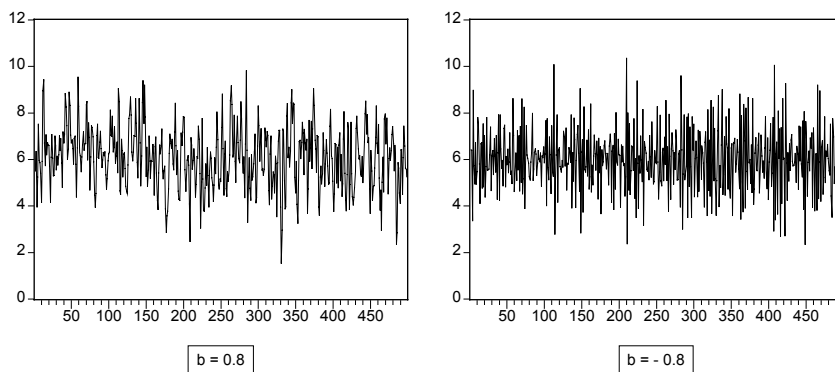
б) Для процесса $MA(1)$ с $b = -0.8$ и $E(X_t) = 6$ имеем

$$\rho(1) = -0.8/(1 + 0.8^2) = -0.488.$$

Коррелограммы этих двух процессов имеют вид



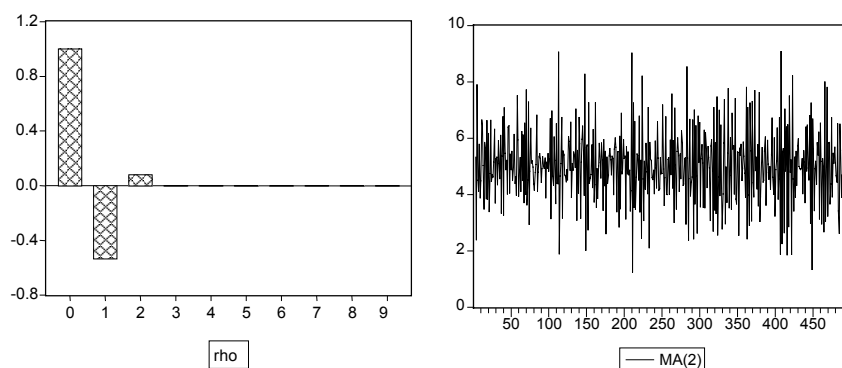
Смоделированные реализации этих двух процессов с $\sigma_\varepsilon^2 = 1$:



с) Для МА(2) процесса $X_t = 5 + \varepsilon_t - 0.75\varepsilon_{t-1} + 0.125\varepsilon_{t-2}$ имеем:

$$\begin{aligned}\rho(1) &= (b_0b_1 + b_1b_2)/(b_0^2 + b_1^2 + b_2^2) \\ &= (-0.75 - 0.75 \cdot 0.125)/(1 + 0.75^2 + 0.125^2) = -0.535, \\ \rho(2) &= 0.125/1.578 = 0.079.\end{aligned}$$

Ниже приводятся коррелограмма и смоделированная реализация этого процесса.



2.5. Смешанный процесс авторегрессии – скользящего среднего (процесс авторегрессии с остатками в виде скользящего среднего)

Процесс X_t с нулевым математическим ожиданием, принадлежащий такому классу процессов, характеризуется порядками p и q его AR и MA составляющих и обозначается как процесс **ARMA(p, q)** (*autoregressive moving average, mixed autoregressive moving average*). Более точно, процесс X_t с нулевым математическим ожиданием принадлежит классу ARMA(p, q), если

$$X_t = \sum_{j=1}^p a_j X_{t-j} + \sum_{j=0}^q b_j \varepsilon_{t-j}, \quad a_p \neq 0, b_q \neq 0,$$

где ε_t – процесс белого шума и $b_0 = 1$. В операторной форме последнее соотношение имеет вид

$$a(L) X_t = b(L) \varepsilon_t,$$

где $a(L)$ и $b(L)$ имеют тот же вид, что и в определенных ранее моделях AR(p) и MA(q). Если процесс имеет постоянное математическое ожидание μ , то он является процессом типа ARMA(p, q), если

$$X_t - \mu = \sum_{j=1}^p a_j (X_{t-j} - \mu) + \sum_{j=0}^q b_j \varepsilon_{t-j}.$$

Отметим следующие свойства процесса ARMA(p, q) с $E(X_t) = \mu$.

- Процесс стационарен, если все корни уравнения $a(z) = 0$ лежат вне единичного круга $|z| \leq 1$.

- Если процесс стационарен, то существует эквивалентный ему процесс $MA(\infty)$

$$X_t - \mu = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}, \quad c_0 = 1, \quad \sum_{j=0}^{\infty} |c_j| < \infty,$$

или

$$X_t - \mu = c(L) \varepsilon_t,$$

где

$$c(z) = \sum_{j=0}^{\infty} c_j z^j = \frac{b(z)}{a(z)}.$$

- Если все корни уравнения $b(z) = 0$ лежат вне единичного круга $|z| \leq 1$ (условие обратимости), то существует эквивалентное представление процесса X_t в виде процесса авторегрессии бесконечного порядка $AR(\infty)$

$$X_t - \mu = \sum_{j=1}^{\infty} d_j (X_{t-j} - \mu) + \varepsilon_t,$$

или

$$d(L)(X_t - \mu) = \varepsilon_t,$$

где

$$d(z) = 1 - \sum_{j=1}^{\infty} d_j z^j = \frac{a(z)}{b(z)}.$$

Отсюда вытекает, что стационарный процесс $ARMA(p, q)$ всегда можно аппроксимировать процессом скользящего среднего достаточно высокого порядка, а при выполнении условия обратимости его можно также аппроксимировать процессом авторегрессии достаточно высокого порядка.

Специфику формы коррелограммы процесса $ARMA(p, q)$ в общем случае указать труднее, чем для моделей $AR(p)$ и $MA(q)$. Отметим только, что для значений $k > q$ коррелограмма процесса $a(L) X_t = b(L) \varepsilon_t$ выглядит так же, как и коррелограмма процесса авторегрессии $a(L) X_t = \varepsilon_t$. Так, для процесса $ARMA(1, 1)$

$$\rho(k) = a_1 \rho(k-1) \text{ для } k = 2, 3, \dots,$$

как и у процесса $X_t = a_1 X_{t-1} + \varepsilon_t$. При этом, однако, $\rho(1) \neq a_1$.

Предпосылкой для обоснования использования моделей $ARMA$ является следующий факт. Если $ARMA(p_1, q_1)$ ряд X_t и $ARMA(p_2, q_2)$ ряд Y_t статистически независимы между собой, и $Z_t = X_t + Y_t$, то типичным является положение, когда Z_t является $ARMA(p, q)$ рядом, у которого

$$p = p_1 + p_2,$$

$$q = p_1 + q_2, \text{ если } p_1 + q_2 > p_2 + q_1,$$

$$q = p_2 + q_1, \text{ если } p_2 + q_1 > p_1 + q_2.$$

Возможны также ситуации, когда значения p и q оказываются меньше указанных значений. (Такие ситуации возникают в случаях, когда многочлены $a_X(z)$ и $a_Y(z)$, соответствующие авторегрессионным частям процессов X_t и Y_t , имеют общие корни.)

В частном случае, когда оба ряда имеют тип $AR(1)$, но с различными параметрами, их сумма имеет тип $ARMA(2, 1)$.

В экономике многие временные ряды являются агрегированными. Из указанного выше факта вытекает, что если каждая из компонент отвечает простой модели AR , то при независимости этих компонент их сумма будет $ARMA$ процессом. Такого же рода

процесс мы получим, если часть компонент имеет тип AR, а остальные компоненты имеют тип MA. Единственным исключением является случай, когда все компоненты являются MA процессами – в этом случае в результате получаем MA процесс.

Предположим, наконец, что “истинный” экономический ряд отвечает $AR(p)$ модели, но значения этого ряда измеряются со случайными ошибками, образующими процесс белого шума (т.е. $MA(0)$). Тогда наблюдаемый ряд имеет тип $ARMA(p, p)$.

Замечание

Ранее мы уже говорили о том, что если $ARMA(p, q)$ процесс X_t удовлетворяет условию обратимости, то его можно представить в виде стационарного процесса $AR(\infty)$. Последний, в свою очередь, можно аппроксимировать стационарным процессом $AR(p)$, быть может, достаточно высокого порядка.

Таким образом, в практических задачах можно было бы и вовсе обойтись без использования моделей ARMA, ограничиваясь либо AR либо MA моделями. При этом, однако, количество коэффициентов, подлежащих оцениванию, может оказаться слишком большим (что снижает точность оценивания) и даже превосходить количество имеющихся наблюдений. В этом смысле модели ARMA могут быть “*более экономными*”.

2.6. Модели ARMA, учитывающие наличие сезонности

Если наблюдаемый временной ряд обладает выраженной сезонностью, то модель ARMA, соответствующая этому ряду, должна содержать составляющие, обеспечивающие проявление сезонности в порождаемой этой моделью последовательности наблюдений.

Для квартальных данных чисто сезонными являются стационарные модели *сезонной авторегрессии первого порядка (SAR(1))*

$$X_t = a_4 X_{t-4} + \varepsilon_t, \quad |a_4| < 1$$

и *сезонного скользящего среднего первого порядка (SMA(1))*

$$X_t = \varepsilon_t + b_4 \varepsilon_{t-4}.$$

В первой модели

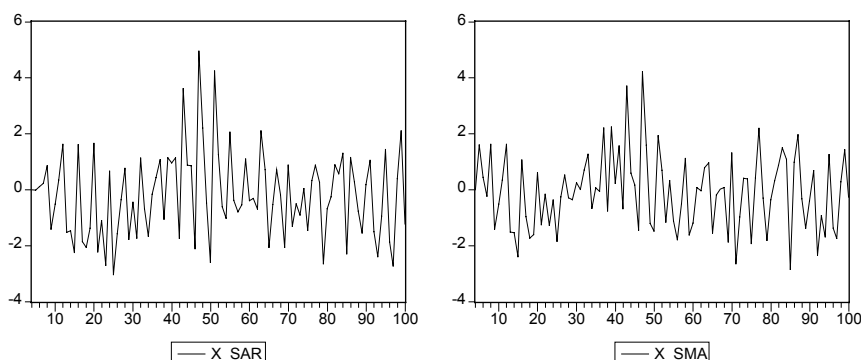
$$\rho(k) = a_4^{k/4} \quad \text{для } k = 4m, \quad m = 0, 1, 2, \dots,$$

$$\rho(k) = 0 \quad \text{для остальных } k > 0.$$

Во второй модели

$$\rho(0) = 1, \quad \rho(4) = b_4, \quad \rho(k) = 0 \quad \text{для остальных } k > 0.$$

Ниже приведены смоделированные реализации модели SAR(1) с $a_4 = 0.8$ и модели SMA(1) с $b_4 = 0.8$.



Комбинации несезонных и сезонных изменений реализуются, например, в моделях

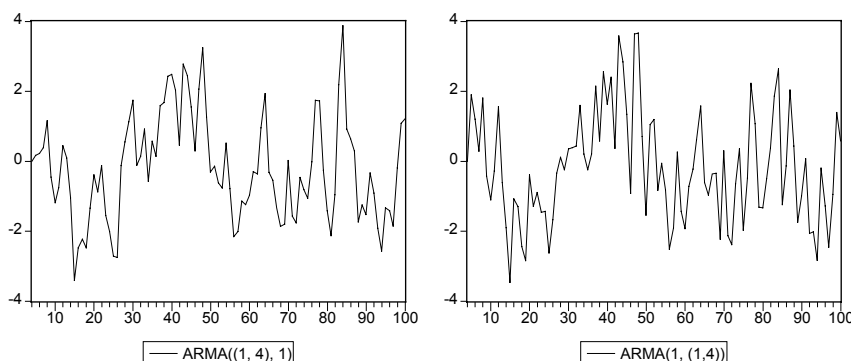
$$\text{ARMA}((1, 4), 1)$$

$$X_t = a_1 X_{t-1} + a_4 X_{t-4} + \varepsilon_t + b_1 \varepsilon_{t-1}$$

$$\text{и } \text{ARMA}(1, (1,4))$$

$$X_t = a_1 X_{t-1} + \varepsilon_t + b_1 \varepsilon_{t-1} + b_4 \varepsilon_{t-4}.$$

Следующие два графика показывают поведение смоделированных реализаций таких рядов при $a_1 = 2/3$, $a_4 = -1/48$, $b_4 = 1/5$ у первого ряда и при $a_1 = 0.4$, $b_1 = 0.3$, $b_4 = 0.8$ у второго ряда.



Заметим, что для первой модели уравнение $a(z) = 0$ принимает вид $1 - 2/3z + 1/48z^4 = 0$, т.е. $z^4 - 32z + 48 = 0$; корни этого уравнения $z_1 = 2$, $z_2 = 2$, $z_3 = -2 + i\sqrt{8}$, $z_4 = -2 - i\sqrt{8}$ лежат вне единичного круга, что обеспечивает стационарность рассматриваемого процесса. Во второй модели $a(z) = 0$ принимает вид $1 - 0.4z = 0$; корень этого уравнения $z = 2.5 > 1$, так что и эта модель стационарна.

Кроме рассмотренных примеров аддитивных сезонных моделей, употребляются также и мультипликативные спецификации, например,

$$(1 - a_1 L)X_t = (1 + b_1 L)(1 + b_4 L^4) \varepsilon_t,$$

$$(1 - a_1 L)(1 - a_4 L^4) X_t = (1 + b_1 L) \varepsilon_t.$$

Первая дает

$$X_t = a_1 X_{t-1} + \varepsilon_t + b_1 \varepsilon_{t-1} + b_4 \varepsilon_{t-4} + b_1 b_4 \varepsilon_{t-5},$$

а вторая

$$X_t = a_1 X_{t-1} + a_4 X_{t-4} - a_1 a_4 X_{t-5} + \varepsilon_t + b_1 \varepsilon_{t-1}.$$

В первой модели допускается взаимодействие составляющих скользящего среднего на лагах 1 и 4 (т.е. значений ε_{t-1} и ε_{t-4}), а во второй – взаимодействие авторегрессионных составляющих на лагах 1 и 4 (т.е. значений X_{t-1} и X_{t-4}). Конечно, эти две модели являются частными случаями аддитивных моделей

$$X_t = a_1 X_{t-1} + \varepsilon_t + b_1 \varepsilon_{t-1} + b_4 \varepsilon_{t-4} + b_5 \varepsilon_{t-5},$$

$$X_t = a_1 X_{t-1} + a_4 X_{t-4} + a_5 X_{t-5} + \varepsilon_t + b_1 \varepsilon_{t-1}$$

с $b_5 = b_1 b_4$, $a_5 = -a_1 a_4$. При приближенном выполнении последних соотношений (по крайней мере, если гипотезы о наличии таких соотношений не отвергаются), естественно перейти от оценивания аддитивной модели к оцениванию мультипликативной модели, опять следуя **принципу “экономности”** модели (**“parsimony model”**). Впрочем, каких-либо теоретических оснований, ведущих к предпочтению одной формы сезонности перед другой (мультипликативной или аддитивной), не существует.

Более подробно с сезонными ARMA моделями можно ознакомиться, например, в книге [Enders (1995)].

Глава 3. Подбор стационарной модели ARMA для ряда наблюдений

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

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

На этапе идентификации производится выбор некоторой частной модели из всего класса ARMA, т.е. выбор значений p и q . Используемые при этом процедуры являются не вполне точными, что может при последующем анализе привести к выводу о непригодности идентифицированной модели и необходимости замены ее альтернативной моделью. На этом же этапе делаются предварительные грубые оценки коэффициентов $a_1, a_2, \dots, a_p, b_1, b_2, \dots, b_q$ идентифицированной модели.

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

На третьем этапе применяются различные диагностические процедуры проверки адекватности выбранной модели имеющимся данным (*misspecification tests*). Неадекватности, обнаруженные в процессе такой проверки, могут указать на необходимую корректировку модели, после чего производится новый цикл подбора, и т.д. до тех пор, пока не будет получена удовлетворительная модель.

Разумеется, если мы имеем дело с ситуацией, когда уже имеется достаточно отработанная и разумно интерпретируемая модель эволюции того или иного показателя, можно обойтись и без этапа идентификации.

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

3.1. Идентификация стационарной модели ARMA

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

О поведении автокорреляционных функций для различных моделей ARMA мы уже говорили. Однако по поведению только автокорреляционной функции трудно идентифицировать даже порядок чистого (без MA составляющей) процесса авторегрессии. Решению этого вопроса помогает рассмотрение поведения **частной**

автокорреляционной функции (PACF) стационарного процесса X_t . Ее значение $\rho_{part}(k)$ на лаге k определяется как значение коэффициента корреляции между случайными величинами X_t и X_{t+k} , очищенными от влияния случайных величин $X_{t+1}, \dots, X_{t+k-1}$.

Это соответствует тому, что $\rho_{part}(k)$ является коэффициентом при X_{t-k} в линейной комбинации случайных величин X_{t-1}, \dots, X_{t-k} , наилучшим образом приближающей случайную величину X_t . Исходя из последнего, можно показать (см., например, [Hamilton (1994)]), что $\rho_{part}(k)$ определяется как решение относительно a_k системы первых k уравнений Юла – Уокера

$$\rho(s) = a_1 \rho(s-1) + a_2 \rho(s-2) + \dots + a_k \rho(s-k), \quad s = 1, 2, \dots, k,$$

которую в этом случае удобнее записать в виде

$$\rho(s-1) a_1 + \rho(s-2) a_2 + \dots + \rho(s-k) a_k = \rho(s), \quad s = 1, 2, \dots, k,$$

подчеркивая, что неизвестными здесь являются a_1, a_2, \dots, a_k , а $\rho(1-k), \dots, \rho(k-1)$ – известные коэффициенты. Исходя из этого и применяя известное из алгебры правило Крамера решения системы k линейных уравнений с k неизвестными, находим, что вычисление PACF можно производить по формулам

$$\rho_{part}(0) = 1,$$

$$\rho_{part}(1) = \rho(1),$$

$$\rho_{part}(2) = \frac{\begin{vmatrix} 1 & \rho(1) \\ \rho(1) & \rho(2) \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{vmatrix}} = \frac{\rho(2) - \rho^2(1)}{1 - \rho^2(1)},$$

$$\rho_{part}(3) = \frac{\begin{vmatrix} 1 & \rho(1) & \rho(2) \\ \rho(1) & 1 & \rho(2) \\ \rho(2) & \rho(1) & \rho(3) \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) & \rho(2) \\ \rho(1) & 1 & \rho(1) \\ \rho(2) & \rho(1) & 1 \end{vmatrix}},$$

...

$$\rho_{part}(k) = \frac{\begin{vmatrix} 1 & \rho(1) & \rho(2) & \dots & \rho(1) \\ \rho(1) & 1 & \rho(1) & \dots & \rho(2) \\ \rho(2) & \rho(1) & 1 & \dots & \rho(3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho(k-1) & \rho(k-2) & \rho(k-3) & \dots & \rho(k) \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) & \rho(2) & \dots & \rho(k-1) \\ \rho(1) & 1 & \rho(1) & \dots & \rho(k-2) \\ \rho(2) & \rho(1) & 1 & \dots & \rho(k-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho(k-1) & \rho(k-2) & \rho(k-3) & \dots & 1 \end{vmatrix}}.$$

Здесь определитель в числителе выражения для $\rho_{part}(k)$ отличается от определителя в знаменателе этого выражения только заменой последнего столбца столбцом, состоящим из значений $\rho(1), \rho(2), \dots, \rho(k)$.

Замечательным является тот факт, что если X_t – процесс типа $AR(p)$, то тогда

$$\rho_{part}(p) \neq 0,$$

$$\rho_{part}(k) = 0 \text{ для } k > p.$$

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

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

В то же время, вместо не известных нам истинных последовательностей автокорреляций $\rho(k)$ и частных автокорреляций $\rho_{part}(k)$ мы можем довольствоваться только их состоятельными оценками – **выборочной ACF**, образованной **выборочными автокорреляциями**

$$r(k) = \frac{\frac{1}{T-k} \sum_{t=1}^{T-k} (x_t - \hat{\mu})(x_{t+k} - \hat{\mu})}{\frac{1}{T} \sum_{t=1}^T (x_t - \hat{\mu})^2} = \frac{\hat{\gamma}(k)}{\hat{\gamma}(0)}, \quad k = 1, \dots, T-1,$$

где

$$\hat{\mu} = \bar{x} = \frac{1}{T} \sum_{t=1}^T x_t \text{ – оценка для } \mu = E(X_t),$$

$$\hat{\gamma}(k) = \frac{1}{T-k} \sum_{t=1}^{T-k} (x_t - \hat{\mu})(x_{t+k} - \hat{\mu}) \text{ – оценка для } \gamma(k),$$

и **выборочной PACF**, образованной **выборочными частными автокорреляциями** $r_{part}(k)$. Получить последние можно, заменяя входящие в выражения для $\rho_{part}(k)$ автокорреляции $\rho(s)$ их оценками $r(s)$. Однако проще поступить иначе, исходя из того, что $\rho_{part}(k)$ является коэффициентом при X_{t-k} в линейной комбинации случайных величин X_{t-1}, \dots, X_{t-k} , наилучшим образом приближающей случайную величину X_t . Именно, можно просто оценить методом наименьших квадратов коэффициенты в модели

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} + \dots + a_k X_{t-k} + u_t$$

(в которой составляющая u_t получается как разность

$$u_t = X_t - (a_1 X_{t-1} + a_2 X_{t-2} + \dots + a_k X_{t-k}),$$

так что на нее не накладываются какие-либо предварительные ограничения). Полученная в результате оценка коэффициента a_k и есть $r_{part}(k)$.

Если X_t является стационарным процессом типа $ARMA(p, q)$ и $E(X_t^4) < \infty$, то указанные оценки $\hat{\mu}$, $\hat{\gamma}(k)$, $r(k)$ и $r_{part}(k)$ являются состоятельными оценками для μ , $\gamma(k)$, $\rho(k)$ и $\rho_{part}(k)$, соответственно. (См. [Hamilton (1994), p.199].) Поскольку $r(k)$ и $r_{part}(k)$ всего лишь оценки для $\rho(k)$ и $\rho_{part}(k)$, то их наблюдаемые значения могут значительно отличаться от $\rho(k)$ и $\rho_{part}(k)$. В частности, если при некоторых $k = k_1$ и $k = k_2$ в модели, порождающей наблюдения, $\rho(k_1) = 0$ и $\rho_{part}(k_2) = 0$, то, как правило, $r(k_1) \neq 0$ и $r_{part}(k_2) \neq 0$, что вносит дополнительную неопределенность в задачу идентификации. Более того, характер изменения теоретической автокорреляционной функции вовсе не обязательно будет воспроизводиться в ее выборочном аналоге – выборочной автокорреляционной функции.

Тем не менее, во многих случаях поведение теоретических ACF и PACF в какой-то мере отражается и на поведении их выборочных аналогов. Поэтому представление о поведении теоретических ACF и PACF может помочь в решении задачи идентификации соответствующих моделей в рамках общего класса моделей ARMA. В этой связи мы суммируем в следующей таблице свойства ACF и PACF для некоторых популярных моделей стационарных временных рядов.

Модел	ACF	PACF
Белый шум, MA(0)	$\rho(k) = 0$ для $k \neq 0$	$P_{part}(k) = 0$ для $k \neq 0$
AR(1), $a_1 > 0$	Экспоненциальное убывание $\rho(k) = a_1^k$	$\rho_{part}(1) = a_1$ $\rho_{part}(k) = 0, k \geq 2$
AR(1), $a_1 < 0$	Осциллирующее убывание $\rho(k) = a_1^k$	$\rho_{part}(1) = a_1$ $\rho_{part}(k) = 0, k \geq 2$
AR(p)	Убывание к нулю с возможной осцилляцией	Зануление при $k \geq p$
MA(1), $b_1 > 0$	Положительный пик при $k = 1$; зануление при $k > 1$	Осциллирующее убывание; $\rho_{part}(1) > 0$
MA(1), $b_1 < 0$	Отрицательный пик при $k = 1$; зануление при $k > 1$	Убывание по абсолютной величине; $\rho_{part}(k) < 0$ при $k \geq 1$
MA(q)	Зануление при $k \geq p$	
ARM A(1, 1) $a_1 > 0$	Экспоненциальное убывание с лага 1; знак $\rho(1)$ совпадает со знаком $(a_1 + b_1)$	Осциллирующее убывание с лага 1; $\rho_{part}(1) = \rho(1)$
ARM A(1, 1) $a_1 < 0$	Осциллирующее убывание с лага 1; знак $\rho(1)$ совпадает со знаком $(a_1 + b_1)$	Экспоненциальное убывание с лага 1; $\rho_{part}(1) = \rho(1)$; знак $\rho_{part}(k)$ совпадает со знаком $\rho(1), k > 1$
ARM A(p, q)	Осциллирующее или прямое убывание,	Осциллирующее или прямое убывание,

	начинающееся с лага q	начинающееся с лага p
SAR(1)	Затухание на лагах, кратных периоду сезонности; зануление на остальных лагах	Пик на лаге, кратном периоду сезонности; зануление на остальных лагах
SMA(1)	Пик на лаге, кратном периоду сезонности; зануление на остальных лагах	Затухание на лагах, кратных периоду сезонности; зануление на остальных лагах

Имея в виду возможность идентификации моделей $AR(p)$ и $MA(q)$ по графикам функций $r(k)$ и $r_{part}(k)$, желательно иметь статистические критерии для проверки гипотез о равенстве нулю тех или иных значений $\rho(k)$ и $\rho_{part}(k)$ на основании наблюдаемых значений $r(k)$ и $r_{part}(k)$. Вопрос этот весьма сложный, и мы ограничимся только двумя приближенными рецептами, которые предполагают гауссовость инноваций (т.е., что ε_t – гауссовский белый шум).

Если X_t – процесс типа $MA(q)$, то при больших n

$$D(r(k)) \approx \frac{1}{T} \left(1 + 2 \sum_{j=1}^q \rho^2(j) \right) \quad \text{для } k > q,$$

так что чем длиннее ряд наблюдений, тем надежнее выявляются нулевые значения $\rho(k)$, $k > q$. При этом,

$$\lim_{T \rightarrow \infty} E(r(k)) = \rho(k).$$

Более того, при больших T и $k > q$ распределение случайной величины $r(k)$ близко к нормальному распределению. Отсюда вытекает, что естественный приближенный критерий проверки гипотезы H_0 : “ X_t – процесс типа $MA(q)$ ” состоит в том, чтобы отвергать эту гипотезу, если

$$|r(k)| > \frac{2}{\sqrt{T}} \sqrt{1 + \sum_{j=1}^q r^2(j)}$$

для $k > q$. Уровень значимости такого критерия приближенно равен 0.05.

В частности, если $q = 0$, то $X_t \sim MA(0)$ – белый шум, и гипотеза H_0 : “ X_t – белый шум” отвергается указанным приближенным критерием при

$$|r(k)| > \frac{2}{\sqrt{T}}, \quad k > 0.$$

(2) Если X_t – процесс типа $AR(p)$, то при больших T и $k > p$

распределение $r_{part}(k)$ можно аппроксимировать нормальным распределением $r_{part}(k) \approx N(0, T^{-1})$ (так что $D(r_{part}(k)) \approx T^{-1}$).

Следовательно, если гипотезу H_0 : $X_t \sim AR(p)$ отвергать при

$$|r_{part}(k)| > \frac{2}{\sqrt{T}}, \quad k > p,$$

то получим критерий, уровень значимости которого приближенно равен 0.05.

Имея в виду два указанных приближенных критерия, в процедурах анализа временных рядов обычно предусмотрена распечатка графиков выборочных ACF и PACF, на которые нанесены границы полосы $\pm 2/\sqrt{T}$. В этих границах с вероятностью, близкой к 0.95, должно заключаться значение $r(k)$, если X_t – белый шум, и значение $r_{part}(k)$, если $X_t \sim \text{AR}(p)$. Здесь следует сделать одно важное предупреждение. Именно, оба построенных критерия имеют уровень значимости, близкий к 0.05, только когда мы проверяем гипотезу H_0 при некотором фиксированном k .

Что, однако, обычно происходит на практике? Рассмотрим это на примере смоделированного белого шума, график которого уже приводился ранее. Всего там было получено $T = 499$ “наблюдений” x_1, x_2, \dots, x_{499} . В следующей таблице приведены значения выборочных автокорреляционной и частной автокорреляционной функций для значений (“лагов”) $k = 1, 2, \dots, 36$.

	CF	ACF		CF	ACF		CF	ACF
	0.019	0.019	3	.102	.126	5	0.053	0.031
	0.013	0.014	4	0.071	0.051	6	0.015	0.018
	0.083	0.083	5	0.044	0.036	7	0.064	0.035
	.038	.035	6	.017	.034	8	.032	.042
	0.047	0.049	7	0.083	0.115	9	0.057	0.075
	.017	.009	8	.035	.028	0	0.053	0.044
	0.024	0.019	9	0.049	0.085	1	.011	0.006
	.062	.053	0	.069	.032	2	.034	.021
	.061	.069	1	.041	.022	3	.029	.034
0	.074	.073	2	0.014	0.057	4	0.042	0.057
1	.079	.099	3	0.035	0.018	5	.013	.064
2	.021	.034	4	.034	.012	6	.046	.055

Рассматривая значения ACF, мы замечаем, что из полосы $\pm 2/\sqrt{T} = \pm 0.0895$ выбивается значение $r(13) = 0.102$. Означает ли это, что мы должны отвергнуть гипотезу H_0 : X_t – белый шум? Рассматривая значения PACF, мы также обнаруживаем значения, выходящие за пределы этой полосы, что приводит к тому же вопросу.

Поскольку количество наблюдений у нас весьма велико ($T = 499$), можно воспользоваться утверждением об асимптотической независимости $r_{part}(k)$, $k = 1, 2, \dots$. Пусть B_k – событие, состоящее в том, что $r_{part}(k)$ выходит за пределы полосы $\pm 2/\sqrt{T}$. Вероятность этого события приблизительно равна 0.05. Тогда вероятность выхода за пределы полосы ровно двух (из 36) $r_{part}(k)$, $k = 1, 2, \dots, 36$, приблизительно равна

$$P_2 = C_{36}^2 (0.05)^2 (1 - 0.05)^{36-2} = (36 \cdot 35 / 2) \cdot 0.05^2 \cdot 0.95^{34},$$

и

$$\lg(P_2) = \lg(630) + 2\lg(0.05) + 34\lg(0.95) = -0.560.$$

Отсюда находим: $P_2 = 0.275$, так что вероятность двух выходов из полосы графика выборочной PACF при рассмотрении 36 лагов вовсе не мала.

Что касается вероятности единственного выхода из полосы выборочной ACF, то здесь мы можем воспользоваться утверждением об асимптотической независимости $r(k)$, $k = 1, 2, \dots$ при условии, что X_t – белый шум (в случае $MA(q)$ процесса с $q \geq 1$ это не так). При этом вероятность наличия единственного выхода выборочной ACF из все той же полосы приближенно равна

$$P_1 = C_{36}^{-1} (0.05)(1 - 0.05)^{35},$$

так что $\lg(P_1) = -0.780$, откуда находим: $P_1 = 0.166$.

Рассмотренный пример показывает, что к интерпретации графиков выборочных ACF и PACF следует подходить достаточно осторожно. Сюда же относится и то обстоятельство, что выражение, используемое при вычислении значений $r(k)$ в пакете EVIEWS, отличается от приведенного выше: в формуле для $\hat{r}(k)$ деление производится не на $T - k$, а на T . Последнее приводит к тому, что так вычисляемая оценка для $\rho(k)$ имеет смещение в направлении нуля.

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

Существует несколько вариантов Q-статистик. Одна из таких статистик (**статистика Бокса – Пирса**) была предложена Боксом и Пирсом [Box, Pierce (1970)] и имеет вид

$$Q = T \sum_{k=1}^M r^2(k).$$

Вспомним уже упоминавшиеся ранее результаты об асимптотической независимости $r(1), r(2), \dots, r(M)$ в случае, когда X_t – белый шум, и заметим, что при больших T в этом случае $\sqrt{T} \cdot r(k) \approx N(0, 1)$, так что $T r^2(k) \approx [N(0, 1)]^2 = \chi^2(1)$. (Заметим, что в этой ситуации не требуется гауссовость X_t – см. [Хеннан (1974)].) Отсюда вытекает, что при больших T приближенно имеем

$$Q \sim \chi^2(M).$$

Против гипотезы H_0 говорят скорее большие значения этой статистики. Поэтому если выбрать уровень значимости равным 0.05, то естественно отвергать эту гипотезу при выполнении неравенства

$$Q > \chi^2_{0.95}(M).$$

В распечатках коррелограмм приводятся P-значения статистики Q для последовательных значений $M = 1, 2, \dots$. При конкретном значении M гипотеза H_0 отвергается, когда соответствующее P-значение меньше 0.05.

Впрочем, исследования показали, что статистика Бокса – Пирса плохо приближается распределением $\chi^2(M)$ при умеренных значениях T . Вместо нее в таких случаях предпочтительнее использовать **статистику Лjung – Бокса** [Ljung, Box (1979)]

$$Q = T(T+2) \sum_{k=1}^M \frac{r^2(k)}{(T-k)},$$

которая (при $T \rightarrow \infty$) также имеет асимптотическое распределение $\chi^2(M)$, но ближе к этому распределению при умеренных значениях T , чем статистика Бокса – Пирса. В пакете EVIEWS (*Econometric Views*) значения статистики Лjungа – Бокса распечатываются вместе с приближенными P -значениями, соответствующими распределениям $\chi^2(M)$.

Практическое использование Q -статистик наталкивается на определенные трудности. Посмотрим на таблицу P -значений (Prob) Q -статистики Лjungа – Бокса для только что рассмотренного примера с реализацией процесса белого шума.

	Prob		Prob		Prob
1	.670	13	.064	25	.061
	.873	4	.045	6	.077
	.292	5	.049	7	.063
	.348	6	.066	8	.072
	.349	7	.037	9	.065
	.455	8	.044	0	.061
	.539	9	.044	1	.076
	.438	0	.033	2	.084
	.360	1	.037	3	.096
0	.243	2	.049	4	.099
1	.146	3	.056	5	.119
2	.187	4	.064	6	.119

P -значения, соответствующие $M = 14, 15, 17 - 22$, меньше 0.05, так что формально при использовании статистики Лjungа – Бокса с любым из этих значений M гипотеза H_0 : “ X_t – белый шум” должна отвергаться, тогда как при остальных значениях M соответствующие P -значения больше, чем 0.05, и гипотеза H_0 при таких значениях M не отвергается.

Какого-либо определенного рецепта, указывающего, как поступать в подобных ситуациях, на какое значение M следует ориентироваться, до сих пор не существует. Среди многочисленных исследований в этом направлении можно отметить работы [Kwan (1996a,b)].

Уже из рассмотренного примера ясно, что на этапе выбора подходящей модели среди всего множества ARMA моделей используемые процедуры являются не вполне точными и часто приводят к довольно неопределенным выводам. В итоге этого этапа возможно оставление для дальнейшего исследования не одной, а нескольких потенциальных моделей. Более определенные выводы при выборе модели на первом этапе можно получить, применяя информационные критерии отбора моделей.

Использование информационных критериев.

Если заранее ограничиваться рассмотрением только AR моделей, т.е. полагать, что процесс X_t следует модели $AR(k)$

$$\sum_{j=1}^k a_{kj} (X_{t-j} - \mu) = \varepsilon_t, \quad a_{k0} = 1,$$

с неизвестным истинным порядком k , то для определения k в таких ситуациях долгое время использовался **информационный критерий Акаике** [Akaike (1973)]. Согласно этому критерию, среди альтернативных значений k выбирается значение, которое минимизирует величину

$$AIC(k) = \ln \hat{\sigma}_k^2 + (2k/T)$$

где T – количество наблюдений, а $\hat{\sigma}_k^2$ – оценка дисперсии инноваций ε_t в AR модели k -го порядка. Для вычисления $\hat{\sigma}_k^2$ производится подбор модели k -го порядка с использованием уравнений Юла – Уокера

$$\rho_s = \sum_{j=1}^k a_{kj} \rho_{s-j},$$

полученные оценки коэффициентов $\hat{a}_{kj}, j = 1, \dots, k$, подставляются вместо a_{kj} в уравнение модели, μ заменяется на \bar{x} , так что получаются оценки для ε_t ,

$$\hat{\varepsilon}_t = \sum_{j=0}^k \hat{a}_{kj} (x_{t-j} - \bar{x}),$$

после чего $\hat{\sigma}_k^2$ определяется как

$$\hat{\sigma}_k^2 = \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_t^2.$$

Впоследствии было выяснено, что оценка Акаике несостоятельна и асимптотически переоценивает (завышает) истинное значение k_0 с ненулевой вероятностью. В связи с этим, были предложены состоятельные критерии, основанные на минимизации суммы

$$\ln \hat{\sigma}_k^2 + k c_T,$$

где $c_T = O(T^{-1} \ln T)$ (т.е. c_T при $T \rightarrow \infty$ имеет тот же порядок малости, что и $T^{-1} \ln T$).

Одним из таких критериев является часто используемый в настоящее время **информационный критерий Шварца** – SIC [Schwarz (1978)],

$$SIC = \ln \hat{\sigma}_k^2 + k \frac{\ln T}{T}.$$

Несколько позднее был предложен **критерий Хеннана – Куинна** [Hannan, Quinn (1979)], в котором $c_T = 2ck T^{-1} \ln \ln T$, $c > 1$,

$$HQ = \ln \hat{\sigma}_k^2 + k \frac{2 \ln \ln T}{T},$$

обладающий более быстрой сходимостью к истинному значению k_0 при $T \rightarrow \infty$. Однако при небольших значениях T этот критерий недооценивает порядок авторегрессии.

Пример

Рассмотрим модель процесса AR(2)

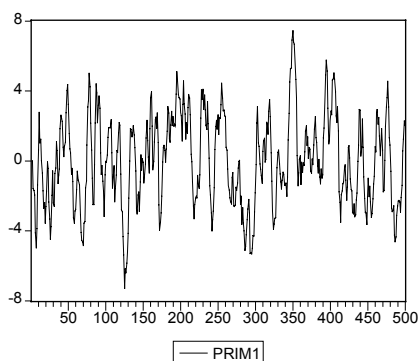
$$X_t = 1.2 X_{t-1} - 0.36 X_{t-2} + \varepsilon_t.$$

Уравнение $a(z) = 0$ принимает в этом случае вид

$$1 - 1.2z + 0.36z^2 = 0$$

и имеет двойной корень $z = 5/3 > 1$, так что процесс, порождаемый такой моделью стационарен.

Смоделированная реализация этого процесса для $t = 1, 2, \dots, 500$ имеет следующий вид.



Построенная по этой реализации выборочная коррелограмма имеет вид:

ACF	PACF	L	C	PAC	Q-Stat	Prob
		0	0	0	4	
. *****	. *****	1	.899	.899	06.25	.000
	*** .	2	.732	0.396	75.97	.000
. *****		3	.561	0.005	34.98	.000
. ****	. .	4	.409	0.027	19.59	.000
. ***		5	.277	0.048	58.40	.000
. **	. .	6	.167	0.015	72.59	.000
. *	. .	7	.095	.071	77.15	.000
.	. *	8	.045	0.045	78.19	.000
.	. .	9				

			.014	.011	78.29	.000
. .	. .		-	с	9	
		0	0.001	.020	78.29	.000
. .	. .		с	с	9	
		1	.003	.055	78.30	.000
. .	. .		с	с	9	
		2	.019	.001	78.49	.000

Здесь из полосы $\pm 2/\sqrt{T} = \pm 0.089$ выходят только значения выборочной PACF, соответствующие лагам $k = 1, 2$. В соответствии с приближенным критерием, упомянутым ранее, это приводит к неотвержению гипотезы $H_0: X_t \sim \text{AR}(2)$.

Для подтверждения этой гипотезы сравним значения информационных критериев Акаике и Шварца, получаемые при оценивании AR моделей 4-го, 3-го, 2-го и 1-го порядков, допускающих ненулевое математическое ожидание соответствующих AR процессов.

	p = 1	p = 2	p = 3	p = 4
IC	3.0 83264	2. 91800	2.9 19244	2.9 24441
IC	3.1 00148	2. 94336	2.9 53116	2.9 66846

Оба критерия выбирают модель AR(2).

Если мы не ограничиваем себя моделями AR и допускаем, что модель, порождающая данные, имеет вид $\text{ARMA}(p_0, q_0)$ (с неизвестными p_0, q_0)

$$a(L)X_t = b(L)\varepsilon_t,$$

то в этом случае имеется несколько процедур оценивания пары (p_0, q_0) , одну из которых мы сейчас рассмотрим (см. [Kavalieris (1991)]).

На первом шаге этой процедуры уже известными нам методами производится подбор модели авторегрессии $\text{AR}(k)$

$$\sum_{j=0}^k a_{kj}X_{t-j} = \varepsilon_t, \quad a_{k0} = 1,$$

вычисляются оценки коэффициентов $\hat{a}_{kj}, j = 1, \dots, k$, и на их основе получают оценки инноваций

$$\hat{\varepsilon}_k(t) = \sum_{j=0}^k \hat{a}_{kj}x_{t-j}, \quad \hat{a}_{k0} = 1.$$

Порядок k авторегрессионной модели на этом шаге должен быть достаточно высоким. Его можно выбрать, опираясь на сравнение значений критерия Акаике для оцененных моделей авторегрессии различных порядков. (Вспомним, что критерий Акаике склонен завышать порядок модели, а это в данном случае нас как раз и устраивает.)

На втором шаге берутся регрессии X_t на $X_{t-j}, j = 1, \dots, p$, и регрессии X_t на $\hat{\varepsilon}_k(t-j), j = 1, \dots, q$. По первым из них получаем начальные оценки наименьших квадратов

для параметров a_j , т.е. \hat{a}_{kj} , $j = 1, \dots, p$, а по вторым – оценки \hat{b}_j для b_j , $j = 1, \dots, q$. Соответственно, оценками полиномов $a(z)$, $b(z)$ служат

$$a(z) = \sum_{j=0}^p \hat{a}_j z^j, \quad b(z) = \sum_{j=0}^q \hat{b}_j z^j,$$

и с помощью оцененных полиномов получаем оценку для инноваций

$$\hat{\varepsilon}_t = \hat{b}^{-1}(L) \hat{a}(L) x_t,$$

на основании которой строим уточненную оценку для дисперсии инноваций

$$\tilde{\sigma}_{p,q}^2 = \frac{1}{T} \sum_{t=1}^T \tilde{\varepsilon}_t^2.$$

При этом предполагается, что сами инновации, зная точно коэффициенты ARMA модели, можно найти по формуле $\varepsilon_t = b^{-1}(L) a(L) x_t$ что соответствует обратимости этой модели.

В качестве оценок для p_0 , q_0 берется пара значений (\tilde{p}, \tilde{q}) , при которой минимизируется величина

$$\text{SIC}(\tilde{\sigma}_{p,q}^2) = \ln \tilde{\sigma}_{p,q}^2 + (p+q) \frac{\ln T}{T}.$$

Существенно, что $\text{SIC}(\tilde{\sigma}_{p,q}^2)$ – возрастающая функция от p и q , когда $p \geq p_0$, $q \geq q_0$, что ведет к состоятельности оценок (\tilde{p}, \tilde{q}) .

3.2. Оценивание коэффициентов модели

После того как произведена идентификация (стационарной) модели ARMA, т.е. на основании имеющихся наблюдений принято решение о значениях p , q в модели ARMA(p , q), порождающей данные, переходят к этапу оценивания коэффициентов модели. На этом этапе обычно используется метод максимального правдоподобия, который в конечном счете сводится к методу наименьших квадратов. За исключением некоторых наиболее простых случаев (например, модели AR(1)), эта задача решается итерационными методами, требующими задания некоторых “начальных” (“стартовых”) значений параметров, которые затем последовательно уточняются.

В качестве таких начальных значений можно использовать предварительные оценки, полученные на первом этапе. Такие начальные значения можно найти, приравнявая неизвестные “истинные” значения автокорреляций $\rho(k)$ значениям $r(k)$ выборочной автокорреляционной функции и используя функциональную связь между значениями $\rho(k)$ и значениями коэффициентов модели. Например, если оценивается модель AR(p), то коэффициенты a_1, \dots, a_p определяются из системы первых p уравнений Юла – Уокера

$$\rho(k) = \sum_{j=1}^p a_j \rho(k-j), \quad k=1, \dots, p,$$

в которые вместо неизвестных значений $\rho(1), \dots, \rho(p)$ автокорреляций подставляются наблюдаемые (вычисляемые по реализации ряда) значения $r(1), \dots, r(p)$ выборочных автокорреляций.

При оценивании моделей с $MA(q)$ составляющей ($q > 0$) существенным оказывается условие обратимости, сформулированное в разд. 2.5. Покажем это на примере $MA(1)$ модели

$$X_t - \mu = \varepsilon_t + b\varepsilon_{t-1}, \quad t = 1, \dots, T.$$

Имея наблюдаемые значения x_1, x_2, \dots, x_T , мы последовательно выражаем $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_T$ через эти значения и (ненаблюдаемое) значение ε_0 :

$$\varepsilon_1 = X_1 - \mu - b\varepsilon_0,$$

$$\varepsilon_2 = X_2 - \mu - b\varepsilon_1 = X_2 - \mu - b(X_1 - \mu - b\varepsilon_0) = (X_2 - \mu) - b(X_1 - \mu) + b^2\varepsilon_0,$$

...

$$\varepsilon_T = X_T - \mu - b\varepsilon_{T-1} = (X_T - \mu) - b(X_{T-1} - \mu) + b^2(X_{T-2} - \mu) - \dots + (-1)^{T-1}b^{T-1}(X_1 - \mu) + (-1)^Tb^T\varepsilon_0.$$

Максимизация (по b) условной функции правдоподобия, соответствующей наблюдаемым значениям x_1, x_2, \dots, x_T при фиксированном значении ε_0 , равносильна минимизации суммы квадратов

$$Q(b) = \varepsilon_1^2 + \varepsilon_2^2 + \dots + \varepsilon_T^2,$$

которая является нелинейной функцией от b . Для поиска минимума этой суммы квадратов приходится использовать численные итерационные методы оптимизации, которые, в свою очередь, требуют задания начального (“стартового”) значения параметра b . Как мы уже говорили, такое стартовое значение может быть получено на этапе идентификации модели. Однако полученное в итоге итераций “оптимальное” значение b зависит от неизвестного нам значения ε_0 , что затрудняет интерпретацию результатов. Задача интерпретации облегчается, если выполнено условие обратимости $|b| < 1$, и при этом значение $|b|$ существенно меньше 1.

Действительно, при выполнении этого условия можно просто положить $\varepsilon_0 = 0$. Эффект от такой замены истинного значения ε_0 на нулевое быстро убывает, так что сумма квадратов, получаемая в предположении $\varepsilon_0 = 0$, может служить хорошей аппроксимацией для суммы, получаемой при истинном значении ε_0 , при достаточно большом количестве наблюдений. Те же аргументы пригодны и для модели $MA(q)$ с $q > 1$: в этом случае можно положить $\varepsilon_0 = \varepsilon_{-1} = \dots = \varepsilon_{-q+1} = 0$. Для получения более точной аппроксимации, в пакетах статистических программ (в том числе и в EVIEWS) предусмотрена процедура (**backcasting**), в которой процесс итераций включает в себя также и оценивание значений $\varepsilon_0, \varepsilon_{-1}, \dots, \varepsilon_{-q+1}$ путем построения для них “**обратного прогноза**”.

Если в результате оценивания получена модель, в которой условие обратимости не выполняется, рекомендуется повторить процедуру оценивания с использованием другого набора начальных значений.

Более подробное изложение процедур оценивания стационарных ARMA моделей методом максимального правдоподобия можно найти, например, в книге [Hamilton (1994)]. Там же можно прочитать о том, каким образом вычисляются приближения для стандартных ошибок оценок коэффициентов этих моделей, которые можно использовать при большом количестве наблюдений обычным образом.

В заключение необходимо только сделать одно важное замечание.

Пусть мы имеем стационарную $AR(p)$ модель

$$a(L)X_t = \delta + \varepsilon_t.$$

Мы уже говорили о том, что в этом случае математическое ожидание μ процесса X_t связано с константой δ соотношением

$$\mu = \frac{\delta}{(1 - a_1 - a_2 - \dots - a_p)}.$$

При этом можно сначала оценить коэффициенты a_1, \dots, a_p и δ , применяя обычный метод наименьших квадратов к модели

$$X_t = \delta + a_1 X_{t-1} + a_2 X_{t-2} + \dots + a_p X_{t-p} + \varepsilon_t,$$

а затем, используя полученные оценки $\hat{a}_1, \dots, \hat{a}_p$ и $\hat{\delta}$, получить оценку для μ в виде

$$\hat{\mu} = \frac{\hat{\delta}}{(1 - \hat{a}_1 - \hat{a}_2 - \dots - \hat{a}_p)}.$$

Однако можно поступить и иначе, как это предусмотрено, например, в пакете EVIEWS (*Econometric Views*), используемом нами в последующих примерах. Именно, мы можем записать ту же модель в виде

$$X_t = \mu(1 - a_1 - a_2 - \dots - a_p) + a_1 X_{t-1} + a_2 X_{t-2} + \dots + a_p X_{t-p} + \varepsilon_t$$

и одновременно оценивать и a_1, \dots, a_p и μ . Такая процедура теоретически более эффективна. Однако в такой форме модель оказывается нелинейной по параметрам, и это обстоятельство, как и при оценивании МА моделей, требует применения нелинейного метода наименьших квадратов (NLLS – *nonlinear least squares*) и численных итерационных методов оптимизации.

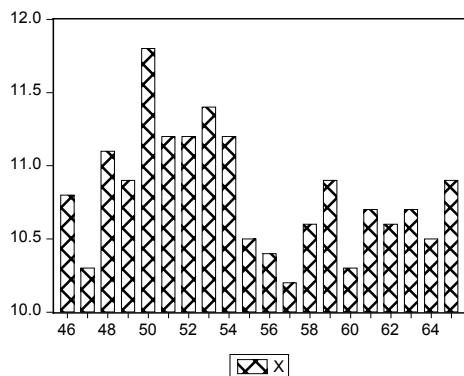
Пример

Рассмотрим данные о годовом потреблении рыбных продуктов в США (на душу населения, в фунтах).

946	0.8	956	0.4
947	0.3	957	0.2
948	1.1	958	0.6
949	0.9	959	0.9
950	1.8	960	0.3
951	1.2	961	0.7
952	1.2	962	0.6
953	11.4	963	0.7
954	1.2	964	0.5

955	0.5	965	0.9
-----	-----	-----	-----

График этого ряда:



Коррелограмма, построенная по этим данным, имеет вид

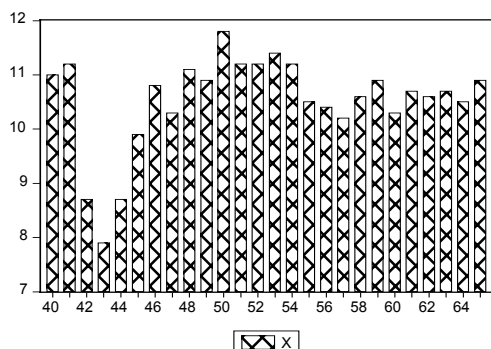
on	Autocorrelati	Partial	C	PAC	Q-Stat	Prob
	. ***	. ***	.429	.429	.2694	.039
	. ***	. **	.366	.222	.5350	.023
	. .	** .	.059	0.204	.6255	.054
016	0.034	.6324	.106
	* .	* .	0.156	0.129	.3498	.138
	** .	** .	0.255	0.195	0.393	.109
	*** .	* .	0.321	0.123	3.879	.053
	. * .	. *	0.133	.175	4.523	.069

Ориентируясь на указанные ранее приближенные критерии, прежде всего найдем значение $2/\sqrt{T} = 2/\sqrt{20} = 0.447$. Из полосы ± 0.447 не выходит ни одна из выборочных автокорреляций и частных автокорреляций. Поэтому с точки зрения этих критериев, мы не должны отвергать гипотезу о том, что наблюдаемый ряд порождается моделью $MA(0)$

$$X_0 = \mu + \varepsilon_t.$$

С другой стороны, если ориентироваться на критерий Лjung – Бокса, то пик ACF на лаге 1 является статистически значимым. Это означает, что в качестве потенциальных моделей порождения данных можно предварительно рассматривать модели $AR(1)$ и $MA(1)$. Таким образом, мы сталкиваемся здесь с конфликтной ситуацией: статистические выводы, получаемые при использовании разных критериев, не соответствуют друг другу. Подобная ситуация не является чем-то исключительным и достаточно часто встречается при идентификации модели, порождающей наблюдаемый ряд, тем более, что используемые критерии – асимптотические, тогда как обычно в распоряжении исследователя имеется не слишком большое количество наблюдений.

Последнее обстоятельство связано в значительной степени с тем, что для многих экономических рядов периоды, на которых порождающая ряд модель может считаться стационарной, обычно непродолжительны из-за изменения общей экономической обстановки, в которой эволюционирует рассматриваемый ряд. Это соображение можно легко проиллюстрировать на примере того же самого ряда данных о потреблении рыбных продуктов в США, если привлечь дополнительно статистические данные за период с 1940 по 1945 годы (годы Второй мировой войны). Эволюция ряда на расширенном периоде с 1940 по 1965 годы представлена следующим графиком:



Провал траектории ряда в 1942 – 1944 г.г. не позволяет трактовать этот ряд как стационарный на всем периоде с 1940 по 1965 годы. Поэтому мы продолжаем далее рассматривать значения ряда только на послевоенном периоде – с 1946 по 1965 г.г.

Если остановиться на модели AR(1), то для нее, как мы знаем, $\rho(1) = a_1$. Поэтому приравняв неизвестное значение $\rho(1)$ значению $r(1) = 0.429$, мы получаем предварительную оценку для неизвестного значения a_1 . В то же время, производя непосредственное оценивание модели AR(1) с ненулевым математическим ожиданием нелинейным методом наименьших квадратов, получаем следующие результаты.

Dependent Variable: X

Method: Least Squares

Sample(adjusted): 1947 1965

Included observations: 19 after adjusting endpoints

Convergence achieved after 3 iterations

Variable	Coef	Std. Error	t-Statistic	Prob
C	10.8	0.159	67.90	0.000
AR(1)	0.43	0.219	1.963	0.062
R-squared	0.18			
Adjusted R-squared	0.13			
S.E. of regression	0.39			
Sum squared resid	2.65			
Mean dependent var				10.81053
S.D. dependent var				0.425434
Akaike info criterion				1.080645
Schwarz criterion				1.180060

В данной ситуации уточненная оценка практически совпадает с предварительной оценкой для a_1 .

Если же остановиться на модели MA(1), то в такой модели $\rho(1) = b_1/(1 + b_1^2)$. Приравнивание неизвестного значения $\rho(1)$ значению $r(1) = 0.429$ приводит к уравнению $b_1/(1 + b_1^2) = 0.429$. Корни последнего уравнения равны 0.567 и 1.704.

Первый корень соответствует обратимой MA(1) модели; второй корень соответствует необратимой MA(1) модели. Уточненное оценивание MA(1) модели с использованием обратного прогноза (backcasting) дает следующие результаты.

Dependent Variable: X
Method: Least Squares
Sample: 1946 1965
Included observations: 20
Convergence achieved after 13 iterations
Backcast: 1945

Backcast: 1949						
Variable		Coef		Std.	t-	Pro
			Error		Statistic	b.
C		10.8	0.113	95.39		0.0
	1379	355	726	000		
MA(1)		0.28	0.228	1.230		0.2
	0610	102	195	345		
R-squared		0.11	Mean dependent		10.	
	7961	var			81000	
Adjusted R-squared	R-	0.06	S.D.	dependent	0.4	
	8959	var			14094	
S.E. of regression	of	0.39	Akaike	info	1.0	
	9561	criterion			97739	
Sum squared resid	squared	2.87	Schwarz	criterion	1.1	
	3684				97313	
Log likelihood		-	F-statistic		2.4	
	8.977395				07257	
Durbin-Watson		1.78	Prob(F-statistic)		0.1	
stat	9895				38178	

В этом случае уточненное значение коэффициента b_1 существенно отличается от предварительной оценки этого коэффициента. Отметим также большое отличие P -значений для t - и F -статистик в отношении значимости коэффициента b_1 . Это можно объяснить тем, что величина стандартной ошибки вычисляется согласно асимптотической процедуре, тогда как в нашем распоряжении имеется всего лишь 20 наблюдений.

Если не производить обратного прогнозирования значения инновации для 1945 года, то результаты получаются близкими:

Dependent Variable: X
Method: Least Squares
Sample: 1946 1965
Included observations: 20
Convergence achieved after 24 iterations
Backcast: OFF

Variable	Coef	Std.	t-	Pro
	ficient	Error	Statistic	b.
C	10.8	0.112	96.06	0.0
	1515	582	431	000
MA(1)	0.27	0.229	1.195	0.2
	4024	231	405	474
R-squared	0.11	Mean dependent		10.
	6800	var		81000
Adjusted R-	0.06	S.D. dependent		0.4

squared	7734	var	14094
S.E. of regression	0.39	Akaike criterion	1.0
Sum squared resid	9824	Schwarz criterion	1.1
Log likelihood	2.87	F-statistic	2.3
Durbin-Watson stat	7464	Prob(F-statistic)	0.1
	8.990543		80443
	1.77		40261
	8286		

3.3. Диагностика оцененной модели

После выбора типа и оценивания коэффициентов модели производится **диагностика** оцененной модели, т.е. выяснение того, насколько хорошо модель соответствует данным наблюдений (адекватна данным наблюдений) – это является третьим этапом процедуры подбора модели.

Для целей диагностики можно использовать целый ряд различных статистических процедур, которые направлены в основном на проверку гипотезы H_0 о том, что в модели, порождающей наблюдения, последовательность ε_t действительно образует процесс белого шума.

Пусть мы остановили свой выбор на этапе идентификации на модели ARMA(p, q)

$$a(L) X_t = b(L) \varepsilon_t,$$

т.е.

$$X_t = a_1 X_{t-1} + \dots + a_p X_{t-p} + \varepsilon_t + b_1 \varepsilon_{t-1} + \dots + b_q \varepsilon_{t-q},$$

и на втором этапе оценили ее как

$$\hat{a}(L) X_t = \hat{b}(L) \varepsilon_t,$$

где

$$\hat{a}(L) = 1 - \hat{a}_1 L - \dots - \hat{a}_p L^p,$$

$$\hat{b}(L) = 1 + \hat{b}_1 L + \dots + \hat{b}_q L^q.$$

Если МА составляющая модели ARMA(p, q) обратима, то тогда

$$\varepsilon_t = \frac{a(L)}{b(L)} X_t,$$

и оценки для ε_t теоретически можно получить заменой $a(L)$ и $b(L)$ на $\hat{a}(L)$ и $\hat{b}(L)$, соответственно:

$$\hat{\varepsilon}_t = \frac{\hat{a}(L)}{\hat{b}(L)} X_t.$$

На практике, конечно, мы можем использовать эту формулу лишь частично, поскольку бесконечный ряд в правой части приходится обрывать из-за наличия только конечного количества наблюдений.

При большом количестве наблюдений поведение $\hat{\varepsilon}_t$ должно имитировать поведение самих ε_t . Следовательно, если ошибки ε_t образуют процесс белого шума, то остатки должны имитировать процесс белого шума.

Основываясь на этом соображении, Бартлетт [Bartlett (1946)] и Бокс и Пирс [Box, Pierce (1970)] предложили исследовать статистическую значимость выборочных автокорреляций для ряда инноваций ε_t

$$r_\varepsilon(k) = \frac{\sum_{t=1}^{T-k} \hat{\varepsilon}_t \hat{\varepsilon}_{t+k}}{\sum_{t=1}^T \hat{\varepsilon}_t^2}$$

и суммы их квадратов

$$Q_{BP} = T \sum_{k=1}^M r_\varepsilon^2(k) \quad (\text{Q-статистика Бокса – Пирса}).$$

Если модель правильно специфицирована, то Q_{BP} имеет распределение, которое близко к распределению $\chi^2(M-p-q)$, при условии, что T и M велики, а отношение (M/T) мало. Гипотеза адекватности подобранной модели отвергается, если

$$Q_{BP} > \chi^2_{0.95}(M-p-q).$$

Однако впоследствии было замечено, что при конечных T распределение статистики Q_{BP} может существенно отличаться от распределения $\chi^2(M-p-q)$. Используя результаты Лjungа и Бокса [Ljung, Box (1978)], можно показать, что

$$E(Q_{BP}) \approx (M-p-q) - \frac{M(M+5)}{2T+2}.$$

Следовательно, если отношение $M(M+5)/(2T+2)$ существенно, то использование $\chi^2(M-p-q)$ приближения не является оправданным.

Лjung и Бокс предложили два способа преодоления проблемы смещения. Первый – прямой метод – состоит в использовании приближения

$$Q_{BP} \approx \chi^2(E(Q_{BP})),$$

где для $E(Q_{BP})$ используется указанное выше выражение (модифицированный критерий Бокса – Кокса).

Второй способ учитывает более точное выражение для $D(r_\varepsilon(k))$ – вместо $(1/T)$ берется $(T-k)/(T^2+2T)$. Это приводит к **Q-статистике Лjungа – Бокса**

$$Q_{LB} = T(T+2) \sum_{k=1}^M \frac{r_\varepsilon^2}{(T-k)},$$

которая имеет то же асимптотическое распределение $\chi^2(M-p-q)$, что и Q_{BP} , но зато при конечных T распределение статистики Q_{LB} гораздо ближе к $\chi^2(M-p-q)$, чем распределение статистики Q_{BP} . При этом качество приближения ухудшается, если значения параметров находятся вблизи границы стационарности или обратимости модели; особенно это заметно при малых M . Заметим, что хотя первоначально вывод асимптотического распределения статистики Лjungа – Бокса производился в предположении, что ε_t – гауссовский белый шум, в дальнейшем было установлено, что этот критерий достаточно устойчив к отклонениям распределения ε_t от нормального. Важно только, чтобы была конечной дисперсия $D(\varepsilon_t)$. (Последнее условие часто нарушается для рядов, описывающих эволюцию быстро изменяющихся финансовых показателей – цен на акции, биржевых индексов, обменных курсов. Для таких рядов распределение ε_t обычно имеет “тяжелые” хвосты, т.е. достаточно часто наблюдаются большие по абсолютной величине значения ε_t . И это требует привлечения для описания таких рядов более сложных моделей.)

В пакете EVIEWS в распечатке результатов оценивания моделей ARMA рядом с коррелограммой ряда остатков приводятся P -значения для наблюдаемых значений Q -статистики Лjungа – Бокса.

Для только что рассмотренных моделей AR(1) и MA(1) коррелограммы рядов остатков имеют следующий вид.

Для модели AR(1):

CF	A	P	L			
	ACF	C	PAC	Q-Stat	Prob	
		-	-	0		
* .	* .	0.096	0.096	.2033		
		C	C	1		
. **	. **	.271	.265	.9334	.164	
		-	-	2		
* .	* .	0.116	0.078	.2687	.322	
		C	-	2		
. *	. .	.076	0.008	.4232	.489	
		-	-	2		
* .	. .	0.099	0.049	.7024	.609	
		-	-	3		
* .	* .	0.125	0.175	.1852	.671	
		-	-	5		
** .	** .	0.257	0.260	.3801	.496	
		C	C	5		
. 019	.051	.3928	.612	
		C	C	5		
. .	*	.047	.189	.4826	.705	
		-	-	6		
* .	** .	0	0.178	0.259	.8945	.648
		C	-	6		
. .	. .	1	.040	0.050	.9748	.728
		-	-	8		
* .	* .	2	0.187	0.132	.9579	.626

Для модели MA(1):

CF	A	P	L			
	ACF	C	PAC	Q-Stat	Prob	
. *		C	C	0		
	. *	.100	.100	.2306		
		C	C	3		
. ***	. ***	.365	.358	.4838	.062	
		-	-	3		
. .	* .	0.050	0.127	.5491	.170	
		C	-	3		
. .	* .	.058	0.068	.6417	.303	

		-	-	4	
*	.	*	.	0.139 0.088 .2051	.379
		-	-	5	
*	.	*	.	0.169 0.182 .1071	.403
		-	-	7	
*	.	**	.	0.277 0.199 .6991	.261
		-	с	7	
*	.	.	*	0.066 .094 .8603	.345
		-	с	7	
.		.	*	0.021 .159 .8772	.446
		-	-	9	
*	.	**	.	0 0.187 0.302 .4191	.400
		с	-	9	
.		.		1 .002 0.042 .4192	.493
		-	-	1	
*	.	*	.	2 0.187 0.104 1.338	.415

В обоих случаях все P -значения для статистики Q_{LB} больше 0.05, так что гипотеза о том, что в специфицированных моделях составляющие ε_t образуют процесс белого шума, не отвергается. Заметим также, что в обоих случаях не отвергается гипотеза нормальности распределения ε_t : P -значения критерия Jarque – Bera равны, соответственно, 0.480 и 0.608. Об оправданности применения последнего критерия при анализе временных рядов будет сказано ниже.

Проверка предположения о нормальности

Многие статистические процедуры, используемые при анализе временных рядов, опираются на предположение гауссовости (нормальности) анализируемого ряда. Последнее означает, что для любого набора t_1, \dots, t_n случайные величины X_{t_1}, \dots, X_{t_n} имеют совместное нормальное распределение.

Имея в распоряжении одну единственную реализацию временного ряда, просто невозможно проверить справедливость такого утверждения. В то же время, еще возможно проверить гипотезу о нормальности одномерного (маргинального) распределения стационарного временного ряда. Соответствующая процедура была предложена в работе Ломницкого [Lomnicki (1961)]. Пусть

$$m_k = \frac{1}{T} \sum_{t=1}^T (X_t - \bar{X})^k,$$

$$G_1 = \frac{m_3}{m_2^{3/2}}, \quad G_2 = \frac{m_4}{m_2^2} - 3.$$

В указанной работе было доказано, что если X_t – стационарный гауссовский временной ряд, то при больших T статистики G_1 и G_2 имеют приближенно нормальные распределения с нулевыми математическими ожиданиями и дисперсиями

$$D(G_1) = \frac{6}{T} \sum_{k=-\infty}^{\infty} \rho^3(k), \quad D(G_2) = \frac{24}{T} \sum_{k=-\infty}^{\infty} \rho^4(k).$$

Оценить эти дисперсии можно, заменив бесконечные суммы степеней автокорреляций $\rho(k)$ конечными суммами степеней выборочных автокорреляций $r(k)$. Используя такие оценки $\hat{D}(G_1)$, $\hat{D}(G_2)$, получаем статистики

$$G_1^* = \frac{G_1}{\sqrt{\hat{D}(G_1)}}, \quad G_2^* = \frac{G_2}{\sqrt{\hat{D}(G_2)}},$$

которые при гипотезе нормальности имеют распределения, аппроксимируемые стандартным нормальным распределением. Поскольку последние статистики еще и асимптотически независимы, то при $T \rightarrow \infty$

$$(G_1^*)^2 + (G_2^*)^2 \approx \chi^2(2).$$

Моделирование показывает, однако, что при умеренных значениях T распределение статистики G_2^* плохо приближается нормальным распределением. Более того, процедура проверки здесь весьма общая (структура временного ряда не специфицируется). Поэтому критерий нормальности, основанный на статистике $(G_1^*)^2 + (G_2^*)^2$, имеет довольно низкую мощность при применении его к моделям AR и MA, т.е. слишком часто не отвергает гипотезу нормальности ряда X_t , когда она не верна.

Более подходящей является в этом отношении аналогичная процедура, применяемая не к самому ряду X_t , а к остаткам, полученным при оценивании специфицированной модели ряда X_t . В моделях AR и MA остатки состоятельно оценивают инновации ε_t , которые, в предположении нормальности, являются независимыми, одинаково распределенными случайными величинами, имеющими распределение $N(0, \sigma^2)$. Поэтому при применении метода Ломницкого для проверки предположения о нормальности инноваций, мы получаем:

$$D(G_1) = \frac{6}{T}, \quad D(G_2) = \frac{24}{T},$$

(т.к. для ряда инноваций $\rho(k) = 0$ при $k \neq 0$),

$$G_1^* = \sqrt{T} \frac{G_1}{\sqrt{6}}, \quad G_2^* = \sqrt{T} \frac{G_2}{\sqrt{24}},$$

так что

$$(G_1^*)^2 + (G_2^*)^2 = T \left(\frac{G_1^2}{6} + \frac{G_2^2}{24} \right),$$

а это есть статистика, используемая для проверки нормальности в популярном критерии Jarque – Bera [Jarque, Bera (1980)].

Таким образом, критерий Jarque – Bera можно использовать не только в рамках классической модели регрессии (с фиксированными значениями объясняющих переменных), но и для проверки нормальности инноваций в моделях временных рядов, помня, конечно, о том, что это всего лишь асимптотический критерий. Для улучшения приближения статистики критерия распределением хи-квадрат, в пакете EVIEWS в статистике критерия вместо множителя T используется множитель $(T - K)$, где K – количество коэффициентов, оцениваемых при построении модели исследуемого ряда.

Правда, здесь мы не заметили еще одного ”подводного камня”. Мы предполагали неявно, что остатки берутся как результат оценивания правильно идентифицированной модели. Как будет влиять на свойства критерия неправильное определение порядка модели?

При больших T критерий Шварца достаточно надежно определяет порядок (p, q) модели ARMA, так что проверка нормальности инноваций по модели, выбранной критерием Шварца, асимптотически равносильна проверке нормальности инноваций по правильно идентифицированной модели.

На третьем шаге производят также проверку выбранной модели на “оптимальность”, имея в виду, что “более сложные” модели не должны существенно отличаться от подобранной модели. Точнее говоря, при увеличении порядка модели оценки коэффициентов при добавленных составляющих должны быть статистически незначимыми, а оценки коэффициентов при сохраняемых составляющих должны изменяться не очень существенно.

Пример

Обращаясь опять к результатам оценивания MA(1) и AR(1) моделей для данных о потреблении рыбных продуктов в США, замечаем, что гипотеза $H_0 : a_1 = 0$ в AR(1) модели и гипотеза $H_0 : b_1 = 0$ в MA(1) модели не отвергаются. Это означает, что обе эти модели могут быть редуцированы к модели MA(0)

$$X_t = \mu + \varepsilon_t.$$

Оценивая последнюю, получаем:

Variable		Coef	Std. Error	t-Statistic	Prob.
C		10.8	0.092	116.7	0.0
	1000	594	460	000	
R-squared		0.00		Mean dependent	10.
	0000	var			81000
Adjusted R-squared	R-	0.00		S.D. dependent	0.4
	0000	var			14094
S.E. of regression	of	0.41		Akaike info	1.1
	4094	criterion			23258
Sum squared resid	squared	3.25		Schwarz criterion	1.1
	8000				73045
Log likelihood	-			Durbin-Watson	1.1
	10.23258	stat			38735

Но оцененная коррелограмма для этой модели была уже приведена выше, в самом начале рассмотрения данного примера, и именно она дала повод рассматривать в качестве возможных кандидатур модели AR(1) и MA(1). При этом, решая вопрос о статистической значимости $\rho(1)$ и $\rho_{part}(1)$, мы опирались на асимптотические результаты, хотя имели в распоряжении лишь небольшое количество наблюдений, и это может быть причиной несогласованности полученных выводов.

Впрочем, мы можем воспользоваться и точным критерием, основанным на статистике Дарбина – Уотсона. Поскольку в последней модели нет никаких объясняющих переменных кроме константы, можно получить таблицы непосредственно для критических значений этой статистики, а не для границ, между которыми заключены эти критические значения. Соответствующие критические значения приведены в работе [Sargan, Bhargava (1983)] . В частности, для уровня значимости 0.05 и $T = 21$ критическое значение равно 1.069. Ориентируясь на него, мы не отвергаем гипотезу о том, что наблюдаемые данные порождены процессом MA(0).

Сравним оцененные модели MA(0), MA(1) и AR(1) по критериям Акаике и Шварца.

	M A(0)	M A(1)	A R(1)
IC	1 .123	1 .098	1 .081
IC	1 .173	1 .197	1 .180

Предпочтительной по критерию Акаике является модель AR(1), тогда как с точки зрения критерия Шварца более предпочтительна модель MA(0). Такое положение в практическом анализе временных рядов возникает достаточно часто: если критерии Акаике и Шварца выбирают разные модели, то критерий Акаике выбирает модель более высокого порядка.

Пример

Обратимся теперь к приведенной в разд. 3.1 реализации процесса авторегрессии второго порядка $X_t = 1.2 X_{t-1} - 0.36 X_{t-2} + \varepsilon_t$. Используя выборочную коррелограмму, построенную по этой реализации, мы (правильно) идентифицировали порядок этого процесса. Среди AR моделей порядков 4, 3, 2 и 1 оба критерия AIC и SIC также выбрали модель второго порядка. Оценивание модели с ненулевым математическим ожиданием нелинейным методом наименьших квадратов приводит к следующим результатам.

Dependent Variable: X
Sample(adjusted): 3 500
Included observations: 498 after adjusting endpoints
Convergence achieved after 3 iterations

Variable	Coef	Std. Error	t-Statistic	Prob.
C	0.00	0.330	0.003	0.9
AR(1)	1.25	0.041	30.45	0.0
AR(2)	-	0.041	-	0.0
	0.397095	0.290	9.617188	0.000

Коррелограмма ряда остатков имеет вид

F	AC	CF	PA	C	PAC	Q-Stat	Prob
				-	-	0	
. .	. .			0.003	0.003	.0042	
				-	-	0	
. .	. .			0.005	0.005	.0165	
				C	C	0	
.000	.000	.0165	.898
				C	C	0	
.036	.036	.6866	.709

				C	C	1	
.		.		.037	.037	.3675	.713
				-	-	5	
*		*		0.086	0.085	.0736	.280
				C	C	5	
.		.		.005	.005	.0882	.405
				-	-	5	
.		.		0.004	0.006	.0977	.531
				-	-	5	
.		.		0.002	0.004	.0993	.648
				-	-	6	
.		.		0	0.054	0.050	.5887 .582
				-	-	6	
.		.		1	0.014	0.008	.6897 .669
				C	C	6	
.		.		2	.019	.011	.8676 .738

Все P -значения для статистики Q_{LB} намного больше 0.05, так что гипотеза о том, что в специфицированной модели составляющие ε_t образуют процесс белого шума, не отвергается. Не отвергается также и гипотеза нормальности ε_t (P -значение в критерии Jarque – Bera равно 0.616). Вместе с тем, оценка математического ожидания процесса X_t статистически незначима, что позволяет не отвергать гипотезу о нулевом математическом ожидании AR(2) процесса. Оценивая модель с нулевым математическим ожиданием, получаем

Dependent Variable: X

Sample(adjusted): 3 500

Included observations: 498 after adjusting endpoints

Convergence achieved after 2 iterations

Variable	Coef	Std.	t-	Pro
	Error		Statistic	b.
AR(1)	1.25	0.041	30.48	0.0
	6581	215	807	000
AR(2)	-	0.041	-	0.0
	0.397096	248	9.627056	000
S.E. of regression	1.03	Akaike info	2.9	
	6707	criterion	1398	
Sum squared resid	533.	Schwarz criterion	2.9	
	0816		3089	
Inverted AR Roots	.63	.63+.05i		
	-.05i			

Исследуем последнюю модель на оптимальность в указанном выше смысле. С этой целью приведем коэффициенты оцененных AR(2), AR(3) и AR(4) моделей и P -значения для тех коэффициентов двух последних моделей, которые являются “лишними” с точки зрения “оптимальной” модели AR(2).

Модель	Коэффициенты при переменных			
	X_{t-1}	X_{t-2}	X_{t-3}	X_{t-4}
AR(2)	1.26	- 0.40		

AR(3)	1.25	– 0.39	$P = 0.87$	
AR(4)	1.25	– 0.40	$P = 0.72$	$P = 0.56$

Эта таблица показывает, что в моделях с неоправданно высоким порядком “лишние” коэффициенты оказались статистически незначимыми, а коэффициенты при переменных, включенных в “оптимальную” модель, практически не изменяются при изменении порядка модели. Именно это и характеризует подобранную модель AR(2) как оптимальную.

Интересно, наконец, обратить внимание на еще одно обстоятельство. Как мы уже отмечали ранее, в теоретической модели AR(2), по которой строилась исследуемая нами реализация, уравнение $a(z) = 0$, т.е. $1 - 1.2z + 0.36z^2 = 0$, имеет двойной корень $z = 5/3 \approx 1.67$. Этот корень больше единицы, что обеспечивает стационарность процесса, порождаемого такой моделью. В то же время, для оптимальной модели, полученной нами в результате подбора, соответствующее уравнение имеет корни, обратные величинам, указанным в последней строке распечатки результатов оценивания этой модели. Указанные в этой строке величины равны $0.63 \pm 0.05i$, так что сами корни равны $z = 1.58 \pm 0.125i$. Хотя эти корни, конечно, отличаются от (двойного) корня уравнения $a(z) = 0$ в теоретической модели, тем не менее оба они больше единицы по абсолютной величине, а значит, подобранная нами AR(2) модель также является стационарной.

Глава 4. Регрессионный анализ для стационарных объясняющих переменных

Прежде, чем переходить к изложению материала этой главы, заметим, что в этой главе мы не будем различать в обозначениях случайные величины и их наблюдаемые значения – в обоих случаях будут использоваться строчные буквы.

4.1. Асимптотическая обоснованность стандартных процедур

В главе 1 мы уже отмечали, что рассмотренные там случаи, в которых можно использовать стандартные процедуры регрессионного анализа несмотря на то, что объясняющие переменные являются стохастическими (ситуации А, А', В, С), не охватывают наиболее интересных для нас моделей стационарных и нестационарных временных рядов. Это замечание относится и к широко используемым на практике моделям авторегрессии, в том числе и стационарным.

Рассмотрим модель авторегрессии $AR(p)$

$$y_t = \alpha + a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_p y_{t-p} + \varepsilon_t,$$

где ε_t – инновации, образующие процесс белого шума с $D(\varepsilon_t) = \sigma_\varepsilon^2$. Эту модель можно представить в виде линейной модели регрессии

$$y_t = x_t^T \theta + \varepsilon_t,$$

где

$$x_t = (1, y_{t-1}, y_{t-2}, \dots, y_{t-p})^T, \quad \theta = (\alpha, a_1, a_2, \dots, a_p)^T.$$

Но мы не можем использовать для нее результаты, полученные в ситуациях А и В. Хотя ε_s и x_t статистически независимы при $s \geq t$, они оказываются зависимыми уже при $s = t - 1$, поскольку ε_{t-1} участвует в формировании случайной величины y_{t-1} , входящей в состав x_t . Это нарушает условие, входящее в определения ситуаций А и В.

Мы не можем также использовать и результаты, полученные в ситуациях А' и С. Там требовалось, чтобы условное распределение вектора ε при фиксированной матрице X имело вид $N(0, \sigma^2 V)$ с положительно определенной (невырожденной) матрицей V (в ситуации А' это единичная матрица). Однако при фиксированных значениях $x_{t+1} = (1, y_t, y_{t-1}, \dots, y_{t-p+1})^T$ и x_t значение ε_t известно с полной определенностью.

Тем не менее, если $AR(p)$ модель стационарна, то положение вполне благополучно:

Ситуация D

- процесс y_t порождается моделью

$$y_t = \alpha + a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_p y_{t-p} + \varepsilon_t \quad (\varepsilon_t \text{ — инновации});$$

- все корни полинома $1 - a_1 z - a_2 z^2 - \dots - a_p z^p = 0$ лежат за пределами единичного круга;
- $\varepsilon_t \sim i.i.d.$, $E(\varepsilon_t) = 0$, $D(\varepsilon_t) = \sigma^2 > 0$, $E(\varepsilon_t^4) = \mu_4 < \infty$.

При выполнении перечисленных условий, для оценки наименьших квадратов $\hat{\theta}_n$ вектора коэффициентов $\theta = (\alpha, a_1, a_2, \dots, a_p)$, полученной по n наблюдениям, выполняется соотношение

$$n^{1/2} (\hat{\theta}_n - \theta) \rightarrow N(0, \sigma^2 Q^{-1}),$$

где Q – положительно определенная матрица, элементы которой выражаются в явной форме через математическое ожидание и автокорреляции процесса y_t . Ковариационная матрица $\sigma^2 Q^{-1}$ асимптотического распределения может быть оценена состоятельно посредством $S_n^{-2} (X_n^T X_n / n)^{-1}$, и это означает, что асимптотически обоснованны статистические процедуры, трактующие распределение $\hat{\theta}_n$ как $N(\theta, S_n^{-2} (X_n^T X_n)^{-1})$. (Здесь X_n – матрица значений объясняющих переменных в n наблюдениях.)

Иными словами, и в рассматриваемой ситуации можно пользоваться стандартными методами регрессионного анализа, имея в виду их асимптотическую обоснованность.

Если перейти к процессам, стационарным относительно детерминированного тренда, то следует отметить возникающую здесь особенность, связанную со сходимостью распределения оценок наименьших квадратов к асимптотическому распределению. Мы поясним эту особенность на следующем примере.

Пусть ряд y_t порождается простой моделью временного тренда

$$y_t = \alpha + \beta t + \varepsilon_t,$$

где $\varepsilon_t \sim i.i.d.$, $E(\varepsilon_t) = 0$, $D(\varepsilon_t) = \sigma^2 > 0$, $E(\varepsilon_t^4) = \mu_4 < \infty$. Если и здесь записать модель в стандартной форме

$$y_t = x_t^T \theta + \varepsilon_t, \quad x_t = (1, t)^T, \quad \theta = (\alpha, \beta),$$

то чтобы получить невырожденное асимптотическое распределение оценки наименьших квадратов $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\beta}_n)$, приходится использовать различные нормирующие множители: $(\hat{\alpha}_n - \alpha_n)$ умножается на $T^{1/2}$, а $(\hat{\beta}_n - \beta_n)$ умножается на $T^{3/2}$. Однако это различие компенсируется тем, что аналогичным образом ведут себя и стандартные ошибки для $\hat{\alpha}_n$ и $\hat{\beta}_n$. Как результат, обычные t -статистики имеют асимптотическое $N(0, 1)$ распределение. Иными словами, можно использовать стандартную технику регрессионного анализа, имея в виду ее асимптотическую обоснованность.

Те же самые принципы можно использовать и для исследования процесса авторегрессии произвольного порядка, стационарного относительно детерминированного временного тренда.

Ситуация Е

- процесс y_t порождается моделью

$$y_t = \alpha + \beta t + a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_p y_{t-p} + \varepsilon_t \quad (\varepsilon_t - \text{инновации});$$

- все корни полинома $1 - a_1 z - a_2 z^2 - \dots - a_p z^p = 0$ лежат за пределами единичного круга;
- $\varepsilon_t \sim i.i.d.$, $E(\varepsilon_t) = 0$, $D(\varepsilon_t) = \sigma^2 > 0$, $E(\varepsilon_t^4) = \mu_4 < \infty$.

При выполнении этих предположений обычные t -статистики и статистики qF (где q – количество линейных ограничений на коэффициенты, а F – обычная F -статистика критерия для проверки выполнения этих ограничений) имеют асимптотические $N(0, 1)$ и $\chi^2(q)$ распределения. Можно использовать стандартную технику регрессионного анализа, имея в виду ее асимптотическую обоснованность.

Если не ограничиваться процессами авторегрессии, но оставаться в классе стационарных моделей, то и в этом случае все еще можно надеяться на возможность использования стандартной техники регрессионного анализа, опять имея в виду ее асимптотическую обоснованность.

Рассмотрим линейную модель

$$y = X\theta + \varepsilon, \quad X = X_n,$$

или, в эквивалентной форме,

$$y_t = x_t^T \theta + \varepsilon_t, \quad t = 1, 2, \dots, n,$$

где

$$x_t = (x_{t1}, x_{t2}, \dots, x_{tp})^T -$$

вектор значений p объясняющих переменных в t -м наблюдении, и пусть $\hat{\theta}_n$ – оценка наименьших квадратов вектора коэффициентов $\theta = (\theta_1, \theta_2, \dots, \theta_p)^T$, полученная по n наблюдениям. Известно (см., например, [Green (1997)]), что следующие три условия обеспечивают состоятельность и асимптотическую нормальность $\hat{\theta}_n$ при $n \rightarrow \infty$:

- $\text{plim} \left(\frac{1}{n} \sum_{t=1}^n x_t \varepsilon_t \right) = 0$

- $\text{plim} \left(\frac{1}{n} \sum_{t=1}^n x_t x_t^T \right) = Q$

[в эквивалентной форме: $\text{plim}(n^{-1} X_n^T X) = Q$], где Q – положительно определенная матрица;

- $\left(\frac{1}{\sqrt{n}} \sum_{t=1}^n x_t \varepsilon_t \right) \rightarrow N(0, \sigma^2 Q)$

[в эквивалентной форме: $(n^{-1/2} X_n^T \varepsilon) \rightarrow N(0, \sigma^2 Q)$].

(Здесь plim – предел по вероятности; стрелка в последнем условии обозначает сходимость по распределению.) Если эти условия выполнены, то при $n \rightarrow \infty$, как и в ситуации D,

$$\sqrt{n} (\hat{\theta}_n - \theta) \rightarrow N(0, \sigma^2 Q^{-1}).$$

В работе [Mann, Wald (1943)] авторы показали следующее (**теорема Манна-Вальда**). Если

$$\bullet \quad \text{plim} \left(\frac{1}{n} \sum_{t=1}^n x_t x_t^T \right) = Q$$

[в эквивалентной форме: $\text{plim} (n^{-1} X_n^T X_n) = Q$], где Q – положительно определенная матрица,

- $\varepsilon_t \sim i.i.d.$, $E(\varepsilon_t) = 0$, $D(\varepsilon_t) = \sigma^2 > 0$, $E|\varepsilon_t|^m < \infty$ для всех $m = 1, 2, \dots$,
- $E(x_t \varepsilon_t) = 0$, $t = 1, 2, \dots, n$,

то тогда выполнены также первое и третье условия из предыдущей тройки условий, обеспечивающих состоятельность и асимптотическую нормальность $\hat{\theta}_n$ при $n \rightarrow \infty$.

Заметим, что условие $E(x_t \varepsilon_t) = 0$, $t = 1, 2, \dots, n$, в сочетании с $E(\varepsilon_t) = 0$, означает, что

$$\text{Cov}(x_{tk}, \varepsilon_t) = 0 \text{ для } k = 1, 2, \dots, p,$$

т.е. означает некоррелированность значений объясняющих переменных с ε_t в совпадающие моменты времени. Условие $E|\varepsilon_t|^m < \infty$ для всех $m = 1, 2, \dots$, выполняется, в частности, для нормального распределения ε_t .

Цитированные результаты можно объединить теперь вместе.

Ситуация F

Пусть в линейной модели

$$y_t = x_t^T \theta + \varepsilon_t, \quad t = 1, 2, \dots, n,$$

где

$$x_t = (x_{t1}, x_{t2}, \dots, x_{tK})^T -$$

вектор значений K объясняющих переменных в t -м наблюдении, $\hat{\theta}_n$ – оценка наименьших квадратов вектора коэффициентов $\theta = (\theta_1, \theta_2, \dots, \theta_K)$, полученная по n наблюдениям. Пусть для этой модели выполнены следующие условия:

$$\bullet \quad \text{plim}_{n \rightarrow \infty} \left(\frac{1}{n} \sum_{t=1}^n x_t x_t^T \right) = Q \quad \left(\text{т.е. } \text{plim}_{n \rightarrow \infty} \frac{1}{n} X_n^T X_n = Q \right),$$

где Q – положительно определенная матрица,

- $\varepsilon_t \sim i.i.d.$, $E(\varepsilon_t) = 0$, $D(\varepsilon_t) = \sigma^2 > 0$, $E|\varepsilon_t|^m < \infty$ для всех $m = 1, 2, \dots$,
- $\text{Cov}(x_{tk}, \varepsilon_t) = 0$ для $k = 1, 2, \dots, K$.

Тогда при $n \rightarrow \infty$

$$\sqrt{n} (\hat{\theta}_n - \theta) \rightarrow N(0, \sigma^2 Q^{-1}).$$

Предположим теперь, что x_t – **стационарный векторный (K -мерный) ряд**, так что

$$E(x_t) = \mu = \text{const}, \quad \text{Cov}(x_t) = Q, \quad \text{Cov}(x_{tk}, x_{t+s,l}) = \gamma_{kl}(s)$$

при всех t, s для каждой пары $k, l = 1, 2, \dots, K$. (Здесь $\gamma_{kl}(s)$ – **кросс-корреляция** значений k -ой и l -ой компонент векторного ряда x_t , разнесенных на s единиц времени. Если рассматривать s как аргумент, а $\gamma_{kl}(s)$ как функцию от s , то $\gamma_{kl}(s)$ – **кросс-корреляционная функция** k -ой и l -ой компонент векторного ряда x_t .) Тогда первое из трех условий, перечисленных в ситуации F, обеспечивает возможность оценивания

неизвестной ковариационной матрицы $Cov(x_t) = Q$ простым усреднением доступных наблюдению матриц $x_t x_t^T$ по достаточно длинному интервалу $t = 1, 2, \dots, n$.

В рамки ситуации F помещается достаточно распространенный класс **ARX моделей** :

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_p y_{t-p} + z_t^T \beta + \varepsilon_t,$$

где

$$z_t = (z_{t1}, z_{t2}, \dots, z_{tM})^T, \beta = (\beta_1, \beta_2, \dots, \beta_M)^T.$$

Подобная модель вписывается в ситуацию F, если положить

$$x_t = (y_{t-1}, y_{t-2}, \dots, y_{t-p}, z_1, z_2, \dots, z_M)^T,$$

$$\theta = (a_1, a_2, \dots, a_p, \beta_1, \beta_2, \dots, \beta_M)^T.$$

Пусть для этой модели выполнены следующие условия:

- z_t – стационарный векторный (M -мерный) ряд;
- $\text{plim}_{n \rightarrow \infty} \left(\frac{1}{n} \sum_{t=1}^n z_t z_t^T \right) = Q_Z$,

где Q_Z – положительно определенная матрица;

- $\varepsilon_t \sim i.i.d.$, $E(\varepsilon_t) = 0$, $D(\varepsilon_t) = \sigma^2 > 0$, $E|\varepsilon_t|^m < \infty$ для всех $m = 1, 2, \dots$;
- $Cov(z_{tm}, \varepsilon_t) = 0$ для $m = 1, 2, \dots, M$;
- $Cov(y_{t-j}, \varepsilon_t) = 0$ для $j = 1, 2, \dots, p$;
- все корни уравнения $a(z) = 1 - a_1 z - a_2 z^2 - \dots - a_p z^p = 0$ лежат вне единичного круга.

Тогда (см. [Green (1993)]) выполнено и первое условие ситуации F, и при $n \rightarrow \infty$ $n^{1/2} (\hat{\theta}_n - \theta) \rightarrow N(0, \sigma^2 Q^{-1})$.

Последнее из перечисленных условий, касающееся корней уравнения $a(z) = 0$, обеспечивает **стабильность** модели ARX. Последнее означает, что по мере продвижения в будущее (т.е. с ростом t) устанавливается определенная **“долговременная” (long-run)** связь между переменными $y_t, z_{t1}, z_{t2}, \dots, z_{tM}$, по отношению к которой происходят достаточно быстрые осцилляции.

4.2. Динамические модели

Среди различных ARX моделей, в эконометрических исследованиях широкое применение нашли **динамические модели (модели с авторегрессионно распределенными запаздываниями – ADL)**

$$\begin{aligned} y_t = & \alpha_0 + a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_p y_{t-p} + \\ & + (\beta_{10} x_{1,t} + \beta_{11} x_{1,t-1} + \dots + \beta_{1r} x_{1,t-r}) + \\ & + \dots + \\ & + (\beta_{s0} x_{s,t} + \beta_{s1} x_{s,t-1} + \dots + \beta_{sr} x_{s,t-r}) + \varepsilon_t. \end{aligned}$$

Для такой модели используют обозначение **ADL($p, r; s$)**, где p – глубина запаздываний по переменной y_t , r – глубина запаздываний по переменным $x_{1,t}, x_{2,t}, \dots, x_{s,t}$, не являющимся запаздываниями переменной y_t , s – количество таких переменных. При такой форме записи допускается, что некоторые из коэффициентов

β_{ij} равны нулю, так что глубина запаздываний может быть различной для различных переменных $x_{i,t}$.

Модель $ADL(p, r; s)$ можно представить в компактном виде

$$a(L) y_t = \mu + b_1(L) x_{1,t} + \dots + b_s(L) x_{s,t} + \varepsilon_t,$$

где

$$a(L) = 1 - a_1 L - a_2 L^2 - \dots - a_p L^p,$$

$$b_i(L) = \beta_{i0} + \beta_{i1} L + \dots + \beta_{ir} L^r, \quad i = 1, \dots, s.$$

Если выполнено условие стабильности, то y_t представляется в виде

$$y_t = \frac{1}{a(L)} \mu + \frac{1}{a(L)} b_1(L) x_{1,t} + \dots + \frac{1}{a(L)} b_s(L) x_{s,t} + \frac{1}{a(L)} \varepsilon_t,$$

или

$$y_t = \frac{1}{a(L)} \mu + c_1(L) x_{1,t} + \dots + c_s(L) x_{s,t} + \frac{1}{a(L)} \varepsilon_t,$$

где

$$c_i(L) = \frac{b_i(L)}{a(L)}.$$

Долговременную связь между переменными можно найти, полагая в выражении для y_t

$$L = 1, \quad \varepsilon_t \equiv 0.$$

При этом получаем

$$y_t = \frac{1}{a(1)} \mu + c_1(1) x_{1,t} + \dots + c_s(1) x_{s,t};$$

строго говоря, в последнем выражении указание на момент t следует исключить:

$$y = \frac{1}{a(1)} \mu + c_1(1) x_1 + \dots + c_s(1) x_s.$$

Коэффициенты $c_1(1), \dots, c_s(1)$ в последнем соотношении называются **долгосрочными мультипликаторами (long-run multipliers)**. Поясним это название на примере модели $ADL(1, 1; 1)$, которую запишем в виде

$$(1 - \alpha_1 L) y_t = \mu + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t.$$

При $|\alpha_1| < 1$ получаем равносильное представление

$$y_t = \frac{1}{(1 - \alpha_1 L)} \mu + \frac{1}{(1 - \alpha_1 L)} (\beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t),$$

т.е.

$$y_t = (1 + \alpha_1 + \alpha_1^2 + \dots) \mu + (1 + \alpha_1 L + \alpha_1^2 L^2 + \dots) (\beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t),$$

из которого последовательно находим:

$$\partial y_t / \partial x_t = \beta_0,$$

$$\partial y_{t+1} / \partial x_t = \partial y_t / \partial x_{t-1} = \beta_1 + \alpha_1 \beta_0,$$

$$\partial y_{t+2} / \partial x_t = \partial y_t / \partial x_{t-2} = \alpha_1 \beta_1 + \alpha_1^2 \beta_0,$$

$$\partial y_{t+3} / \partial x_t = \partial y_t / \partial x_{t-3} = \alpha_1^2 \beta_1 + \alpha_1^3 \beta_0,$$

и т.д. Правые части дают значения **импульсных мультипликаторов**, показывающих влияние единовременного (импульсного) изменения значения x_t на текущее и последующие значения переменной y_t . Просуммировав полученные выражения, получаем:

$$\begin{aligned} \partial y_t / \partial x_t + \partial y_t / \partial x_{t-1} + \partial y_t / \partial x_{t-2} + \partial y_t / \partial x_{t-3} + \dots = \\ = \beta_0 (1 + \alpha_1 + \alpha_1^2 + \dots) + \beta_1 (1 + \alpha_1 + \alpha_1^2 + \dots) = \\ = (1 - \alpha_1 L)^{-1} (\beta_0 + \beta_1). \end{aligned}$$

Правая часть этого соотношения, как легко заметить, представляет собой долгосрчный мультипликатор рассматриваемой ADL(1, 1; 1). В соответствии с левой частью, этот мультипликатор показывает изменение значения y_t при изменении на единицу текущего и всех предыдущих значений переменной x_t .

Прежде, чем перейти к рассмотрению примера оценивания конкретной ADL модели, следует заметить следующее.

При выполнении условий, обеспечивающих возможность использования стандартной техники регрессионного анализа (имеется в виду ее асимптотическая обоснованность – см. разд. 4.1, ситуация F):

- Обычная t -статистика имеет асимптотическое $N(0,1)$ распределение.
- Если F – обычная F -статистика для проверки гипотезы о выполнении q линейных ограничений на коэффициенты модели, то статистика qF имеет асимптотическое χ^2 распределение с q степенями свободы.
- При умеренном количестве наблюдений параллельно с асимптотическими распределениями для t и qF можно для контроля использовать и точные (стандартные) распределения (распределение Стьюдента для t -статистики, распределение Фишера для F -статистики). Согласованность получаемых при этом результатов подкрепляет уверенность в правильности соответствующих статистических выводов.
- При наличии в правой части запаздывающих значений объясняемой переменной проверку гипотезы об отсутствии автокоррелированности у ряда ε_t следует производить, используя **критерий Бройша – Годффри**. Критерий Дарбина – Уотсона не годится для этой цели, поскольку в данном случае значения

статистики Дарбина – Уотсона d смещены в направлении значения $d = 2$, так что использование таблиц Дарбина – Уотсона приводит к неоправданно частому неопровержению указанной гипотезы (“презумпция некоррелированности ε_t ”).

Пример

Рассмотрим модель $ADL(3, 2; 1)$

$$(1 - 0.5L - 0.1L^2 - 0.05L^3)y_t = 0.7 + (0.2 + 0.1L + 0.05L^2)x_t + \varepsilon_t.$$

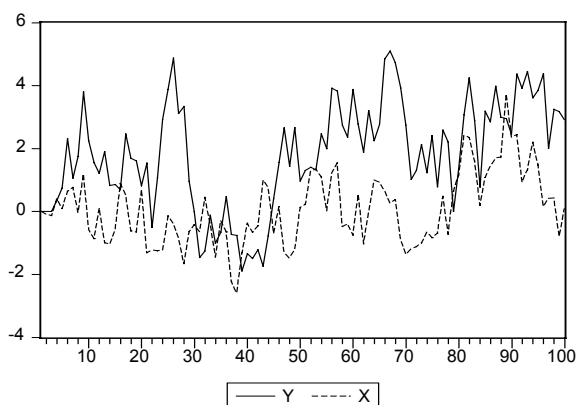
Для нахождения долговременной связи между переменными y и x полагаем $L = 1$ и $\varepsilon_t \equiv 0$:

$$(1 - 0.5 - 0.1 - 0.05)y = 0.7 + (0.2 + 0.1 + 0.05)x,$$

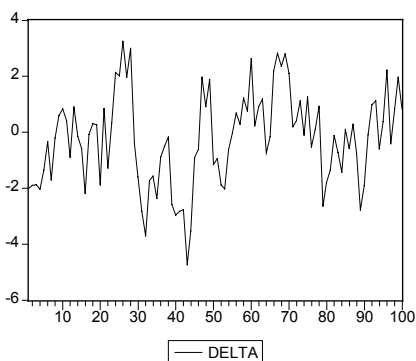
$$\text{т.е. } 0.35y = 0.7 + 0.35x, \text{ или}$$

$$y = 2 + x.$$

На приводимом ниже графике представлены смоделированная реализация ряда $x_t = 0.7x_{t-1} + \varepsilon_{xt}$, $\varepsilon_{xt} \sim i.i.d. N(0, 1)$, и соответствующая ей реализация ряда y_t , порождаемого указанной моделью $ADL(3, 2; 1)$, где $\varepsilon_t \sim i.i.d. N(0, 1)$, причем ряд ε_t порождается независимо от ряда ε_{xt} . В качестве начальных значений при моделировании были взяты: $x_1 = 0, y_1 = y_2 = y_3 = 0$.



Имея в распоряжении только эти две реализации, мы не знаем, с какой моделью имеем дело. Начнем с оценивания статической модели $y_t = \mu + \beta x_t + \varepsilon_t$ методом наименьших квадратов; в результате получаем оцененную модель $y_t = 1.789 + 0.577x_t + e_t$, где e_t – ряд остатков. График ряда остатков имеет вид:



Здесь обнаруживается явная автокоррелированность ряда остатков, которая подтверждается построенной для него коррелограммой

ACF PAC
F C PAC Q-Stat Prob

			C	4		
*****	*****	.696	.696	9.981	.000	
			C	7		
****	*	.536	.099	9.868	.000	
				9		
***		.364	0.081	3.801	.000	
				9		
**		.227	0.056	9.279	.000	
				1		
*		.130	0.015	01.10	.000	
				1		
		.057	0.020	01.46	.000	

и критерием Бройша – Годфри с запаздыванием на один шаг, который дает P -значение 0.0000. Это означает, что мы имеем дело не со статической, а с динамической моделью. Поэтому следует прежде всего рассмотреть характер поведения обоих рядов и произвести их идентификацию.

Для ряда x_t коррелограмма имеет вид

ACF	PACF	AC	PAC	Q-Stat	Prob
. *****	. *****	0.686	0.686	48.468	0.000
. ***	* .	0.429	-0.079	67.594	0.000
. *	* .	0.193	-0.132	71.527	0.000
. .	* .	0.024	-0.066	71.591	0.000
* .	.	-0.058	0.003	71.958	0.000
* .	* .	-0.140	-0.107	74.090	0.000

По этой коррелограмме ряд x_t идентифицируется как AR(1).

Для ряда y_t коррелограмма имеет вид

ACF	PACF	AC	PAC	Q-Stat	Prob
. *****	. *****	0.767	0.767	60.58	0.000

. *	. *	0.629	0.100	101.75	0.000

. .	. .	0.494	-0.042	127.37	0.000

. .	. .	0.399	0.019	144.32	0.000

. .	. .	0.318	-0.003	155.21	0.000
**					
. .	. .	0.257	0.004	162.38	0.000
**					

так что и этот ряд идентифицируется как AR(1).

Такой предварительный анализ ограничивает рассмотрение моделью ADL с глубиной запаздываний, равной единице:

$$y_t = \mu + a_1 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t.$$

Оценивая такую модель ADL(1, 1; 1), получаем:

Dependent Variable: Y

Variable Coefficient Std. Error t-Statistic Prob.

C	0.558588	0.157276	3.55163	0.0006
Y(-1)	0.695204	0.066095	10.51828	0.0000
X	0.208971	0.126135	1.65673	0.1009
X(-1)	0.161690	0.132352	1.22166	0.2249

Анализ остатков не выявляет автокоррелированности (P -значение критерия Бройша–Годфри при AR(1) альтернативе равно 0.164), не выявляет значимых отклонений от нормальности распределения ε_t (P -значение критерия Jarque – Bera = 0.267), не обнаруживает гетероскедастичности (P -значение критерия Уайта = 0.159), так что можно, опираясь на приведенные выше факты, использовать асимптотическую теорию статистических выводов и на ее основе использовать результаты, получаемые при применении t - и F -критериев.

При проверке гипотезы $H_0: \beta_0 = \beta_1 = 0$ получаем при использовании F -распределения P -значение 0.0032; при использовании асимптотического распределения $\chi^2(2)$ получаем P -значение 0.0022. В обоих случаях эта гипотеза отвергается. Исключение из правой части модели запаздывающей переменной x_{t-1} , коэффициент при которой статистически незначим и имеет большее P -значение, чем коэффициент при x_t , дает:

Dependent Variable: Y

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.517868	0.154098	3.360648	0.0011
Y(-1)	0.719738	0.063131	11.40064	0.0000
X	0.310343	0.095241	3.258511	0.0015
R-squared	0.637523	Mean dependent var	1.844751	
Adjusted R-squared	0.629971	S.D. dependent var	1.709520	
S.E. of regression	1.039901	Akaike info criterion	2.945962	
Sum squared resid	103.8138	Schwarz criterion	3.024602	
Log likelihood	-142.8251	F-statistic	84.42207	
Durbin-Watson stat	2.256404	Prob(F-statistic)	0.000000	

Здесь все коэффициенты имеют высокую значимость, а остатки вполне удовлетворительны.

Если из предыдущей модели исключить не x_{t-1} , а x_t , то это приводит к оцененной модели

Dependent Variable: Y

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.567821	0.158600	3.580215	0.0005
Y(-1)	0.692523	0.066673	10.38690	0.0000
X(-1)	0.305939	0.100582	3.041698	0.0030
R-squared	0.632818	Mean dependent var	1.844751	
Adjusted R-squared	0.625169	S.D. dependent var	1.709520	
S.E. of regression	1.046627	Akaike info criterion	2.958857	
Sum squared resid	105.1611	Schwarz criterion	3.037497	
Log likelihood	-143.4634	F-statistic	82.72550	
Durbin-Watson stat	2.221594	Prob(F-statistic)	0.000000	

По критерию Шварца чуть более предпочтительной выглядит модель с исключенной x_{t-1} , так что на ней можно и остановиться. Посмотрим, к какому долговременному соотношению приводит такая модель.

Итак, мы останавливаемся на оцененной модели

$$(1 - 0.720L)y_t = 0.518 + 0.310x_t + e_t.$$

Полагая $L = 1$ и $e_t \equiv 0$, получаем: $0.28y_t = 0.518 + 0.310x_t$, так что долговременное соотношение оценивается как

$$y = 1.839 + 1.107x.$$

Это соотношение, конечно, несколько отличается от теоретического. Посмотрим, однако, что дало бы оценивание динамической модели ADL(3, 2; 1). Оценивая такую модель, получаем:

Dependent Variable: Y

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.539617	0.174057	3.100239	0.0026
Y(-1)	0.590293	0.105583	5.590795	0.0000
Y(-2)	0.153936	0.120517	1.277303	0.2048
Y(-3)	-0.031297	0.099814	-0.313555	0.7546
X	0.205570	0.129483	1.587626	0.1159
X(-1)	0.191959	0.153608	1.249666	0.2147
X(-2)	-0.024779	0.138389	-0.179053	0.8583
R-squared	0.643818	Mean dependent var	1.882787	
Adjusted R-squared	0.620072	S.D. dependent var	1.706160	
S.E. of regression	1.051648	Akaike info criterion	3.008022	
Sum squared resid	99.53667	Schwarz criterion	3.193826	
Log likelihood	-138.8891	F-statistic	27.11327	
Durbin-Watson stat	1.999213	Prob(F-statistic)	0.000000	

Если найти долговременное соотношение между y и x на основе такого оцененного уравнения по той же схеме, что и прежде, то получаем:

$$y = 1.882 + 1.300x,$$

и это соотношение отнюдь не ближе к теоретическому, чем то, которое мы получили по редуцированному уравнению. Впрочем, и по критерию Шварца полная оцененная модель хуже редуцированной.

4.3. Векторная авторегрессия

Популярной моделью связи между временными рядами является **векторная авторегрессия (VAR – vector autoregression)**.

В своей простейшей форме такая модель связывает два ряда y_{1t} и y_{2t} следующим образом:

$$y_{1t} = \mu_1 + \pi_{11.1}y_{1,t-1} + \pi_{12.1}y_{2,t-1} + \varepsilon_{1t},$$

$$y_{2t} = \mu_2 + \pi_{21.1}y_{1,t-1} + \pi_{22.1}y_{2,t-1} + \varepsilon_{2t},$$

т.е., в отличие от простого процесса авторегрессии, значение y_{1t} связывается не только с запаздыванием $y_{1,t-1}$, но и с запаздыванием $y_{2,t-1}$ второй переменной y_{2t} . Случайные величины ε_{1t} и ε_{2t} являются **инновациями**:

- $Cov(\varepsilon_{jt}, \varepsilon_{ls}) = 0$ для $t \neq s$ при любых $j, l = 1, 2$;
- $Cov(\varepsilon_{jt}, y_{l,t-r}) = 0$ для $r \geq 1$ при любых $j, l = 1, 2$.

В то же время, для совпадающих моментов времени случайные величины ε_{1t} и ε_{2t} могут быть коррелированными.

Модель векторной авторегрессии для двух рядов допускает включение в правые части уравнений для y_{1t} и y_{2t} и большого количества запаздываний этих переменных. Наибольший порядок запаздываний, включаемых в правую часть, называется **порядком** векторной авторегрессии. Если этот порядок равен p , то для такой модели используют обозначение **VAR(p)**.

В общем случае рассматривается k временных рядов $y_{1t}, y_{2t}, \dots, y_{kt}$. Модель векторной авторегрессии порядка p предполагает, что связь между этими рядами имеет вид

$$\begin{aligned} y_{1t} = & \mu_1 + \pi_{11.1} y_{1,t-1} + \pi_{11.2} y_{1,t-2} + \dots + \pi_{11.p} y_{1,t-p} + \\ & + \pi_{12.1} y_{2,t-1} + \pi_{12.2} y_{2,t-2} + \dots + \pi_{12.p} y_{2,t-p} + \\ & + \dots + \\ & + \pi_{1k.1} y_{k,t-1} + \pi_{1k.2} y_{k,t-2} + \dots + \pi_{1k.p} y_{k,t-p} + \varepsilon_{1t}, \end{aligned}$$

$$\begin{aligned} y_{2t} = & \mu_2 + \pi_{21.1} y_{1,t-1} + \pi_{21.2} y_{1,t-2} + \dots + \pi_{21.p} y_{1,t-p} + \\ & + \pi_{22.1} y_{2,t-1} + \pi_{22.2} y_{2,t-2} + \dots + \pi_{22.p} y_{2,t-p} + \\ & + \dots + \\ & + \pi_{2k.1} y_{k,t-1} + \pi_{2k.2} y_{k,t-2} + \dots + \pi_{2k.p} y_{k,t-p} + \varepsilon_{2t}, \end{aligned}$$

...

$$\begin{aligned} y_{kt} = & \mu_k + \pi_{k1.1} y_{1,t-1} + \pi_{k1.2} y_{1,t-2} + \dots + \pi_{k1.p} y_{1,t-p} + \\ & + \pi_{k2.1} y_{2,t-1} + \pi_{k2.2} y_{2,t-2} + \dots + \pi_{k2.p} y_{2,t-p} + \\ & + \dots + \\ & + \pi_{kk.1} y_{k,t-1} + \pi_{kk.2} y_{k,t-2} + \dots + \pi_{kk.p} y_{k,t-p} + \varepsilon_{kt}, \end{aligned}$$

где $\pi_{ij,r}$ - коэффициент при $y_{j,t-r}$ в уравнении для y_{it} .

Здесь $\varepsilon_{1t}, \varepsilon_{2t}, \dots, \varepsilon_{kt}$ - случайные величины, для которых

- $Cov(\varepsilon_{jt}, \varepsilon_{ls}) = 0$ для $t \neq s$ при любых $j, l = 1, \dots, k$;
- $Cov(\varepsilon_{jt}, y_{l,t-r}) = 0$ для $r \geq 1$ при любых $j, l = 1, \dots, k$;
- $Cov(\varepsilon_{jt}, \varepsilon_{lt})$ могут отличаться от нуля.

Случайные величины $\varepsilon_{1t}, \varepsilon_{2t}, \dots, \varepsilon_{kt}$ образуют случайный вектор $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t}, \dots, \varepsilon_{kt})^T$, компоненты которого некоррелированы по времени и не коррелированы с запаздывающими значениями переменных $y_{1t}, y_{2t}, \dots, y_{kt}$. Этот вектор называют **вектором инноваций (обновлений) относительно информационного множества**

$$Y_{t-1} = (y_{1,t-1}, y_{1,t-2}, \dots, y_{1,t-p}, \dots, y_{k,t-1}, y_{k,t-2}, \dots, y_{k,t-p}).$$

Пример

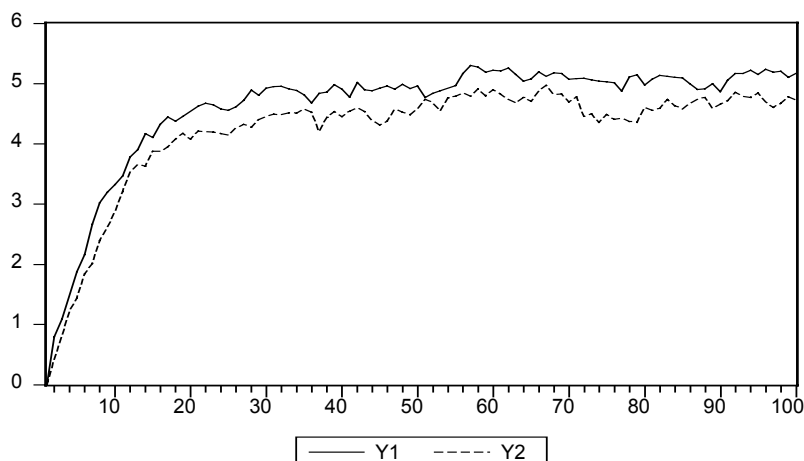
Рассмотрим следующую модель VAR(1) для двух рядов ($k = 2, p = 1$):

$$y_{1t} = 0.6 + 0.7 y_{1,t-1} + 0.2 y_{2,t-1} + \varepsilon_{1t},$$

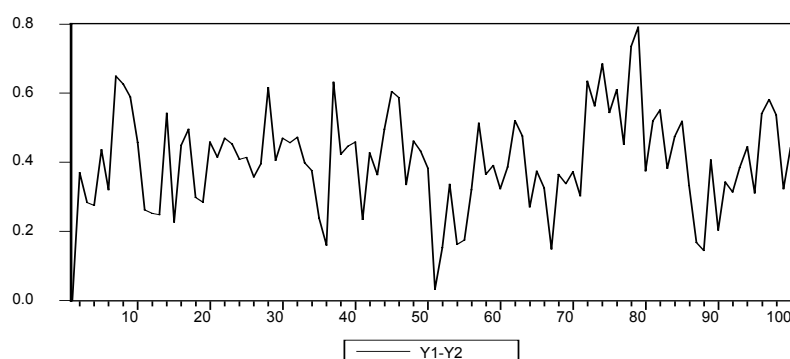
$$y_{2t} = 0.4 + 0.2 y_{1,t-1} + 0.7 y_{2,t-1} + \varepsilon_{2t}.$$

Приводимый ниже график иллюстрирует поведение смоделированной пары y_{1t}, y_{2t} порождаемой этой моделью для $t = 2, 3, \dots, 100$. В качестве начальных значений

были взяты $y_{11} = y_{21} = 0$; ε_{1t} и ε_{2t} моделировались как независимые случайные величины, имеющие одинаковое нормальное распределение $N(0, 0.1^2)$.



Следующий график представляет поведение разности $(y_{1t} - y_{2t})$.



Мы видим, что с течением времени поведение рядов стабилизируется: они осциллируют вокруг установившихся уровней. Второй график показывает, что установившийся уровень для ряда y_{1t} превышает установившийся уровень для ряда y_{2t} приблизительно на 0.4 (среднее арифметическое разности $y_{1t} - y_{2t}$ равно 0.403). Такой характер поведения пары y_{1t}, y_{2t} указывает на стабильность данной модели VAR.

Предсказать стабильный характер поведения реализаций рядов, связанных VAR моделью, можно, анализируя коэффициенты модели. Для этого удобно записать VAR(p) модель для k рядов в более компактной форме

$$y_t = \mu + \Pi_1 y_{t-1} + \Pi_2 y_{t-2} + \dots + \Pi_p y_{t-p} + \varepsilon_t.$$

Здесь

$$y_t = (y_{1t}, y_{2t}, \dots, y_{kt})^T, \mu = (\mu_1, \mu_2, \dots, \mu_k)^T, \varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t}, \dots, \varepsilon_{kt})^T,$$

$\Pi_r = (\pi_{ij,r})$ — матрица размера $k \times k$ коэффициентов при $y_{1,t-r}, y_{2,t-r}, \dots, y_{k,t-r}$ в k уравнениях.

Последнее представление можно записать как

$$y_t - \Pi_1 y_{t-1} - \Pi_2 y_{t-2} - \dots - \Pi_p y_{t-p} = \mu + \varepsilon_t,$$

$$(I_k - \Pi_1 L - \Pi_2 L^2 - \dots - \Pi_p L^p) y_t = \mu + \varepsilon_t,$$

или

$$A(L) y_t = \mu + \varepsilon_t,$$

где

$$A(L) = I_k - \Pi_1 L - \Pi_2 L^2 - \dots - \Pi_p L^p.$$

Условие стабильности такой VAR модели состоит в следующем:

- Все k корней уравнения

$$\det(I_k - z \Pi_1 - z^2 \Pi_2 - \dots - z^p \Pi_p) = 0 \quad (\text{т.е. } \det A(z) = 0)$$

лежат за пределами единичного круга на комплексной плоскости (т.е. модули всех k корней больше единицы).

Если это условие выполняется, то при продвижении вперед по оси времени система постепенно “забывает” о том, при каких начальных значениях y_1, y_2, \dots, y_p она начала реализовываться. Стабильное состояние системы находится путем приравнивания $L = 1$ и $\varepsilon_t = 0$. При этом получаем

$$A(1) y_t = \mu,$$

так что стабильное состояние определяется как

$$y_t = A^{-1}(1) \mu.$$

Пример

Продолжим рассмотрение модели VAR(1) для двух рядов

$$y_{1t} = 0.6 + 0.7 y_{1,t-1} + 0.2 y_{2,t-1} + \varepsilon_{1t},$$

$$y_{2t} = 0.4 + 0.2 y_{1,t-1} + 0.7 y_{2,t-1} + \varepsilon_{2t}.$$

В компактной форме эта система имеет вид

$$y_t = \mu + \Pi_1 y_{t-1} + \varepsilon_t,$$

где

$$y_t = \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix}, \quad \mu = \begin{pmatrix} 0.6 \\ 0.4 \end{pmatrix}, \quad \Pi_1 = \begin{pmatrix} 0.7 & 0.2 \\ 0.2 & 0.7 \end{pmatrix}, \quad \varepsilon_t = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix},$$

или

$$A(L) y_t = \mu + \varepsilon_t,$$

где

$$A(L) = I_2 - \Pi_1 L = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 0.7L & 0.2L \\ 0.2L & 0.7L \end{pmatrix} = \begin{pmatrix} 1-0.7L & -0.2L \\ -0.2L & 1-0.7L \end{pmatrix},$$

так что

$$A(1) = \begin{pmatrix} 0.3 & -0.2 \\ -0.2 & 0.3 \end{pmatrix}, \quad A^{-1}(1) = \begin{pmatrix} 6 & 4 \\ 4 & 6 \end{pmatrix}.$$

Уравнение $\det A(z) = 0$ принимает здесь вид

$$\det A(z) = \begin{vmatrix} 1-0.7z & -0.2z \\ -0.2z & 1-0.7z \end{vmatrix} = 0,$$

т.е. $(1 - 0.7z)^2 - (0.2z)^2 = 0$, или $(1 - 0.9z)(1 - 0.5z) = 0$. Оба корня $z = 1/0.9$ и $z = 1/0.5$ больше 1, т.е. условие стабильности выполняется.

Долгосрочное (стабильное) поведение системы находим по формуле

$$y_t = A^{-1}(1) \mu = \begin{pmatrix} 6 & 4 \\ 4 & 6 \end{pmatrix} \begin{pmatrix} 0.6 \\ 0.4 \end{pmatrix} = \begin{pmatrix} 5.2 \\ 4.8 \end{pmatrix}.$$

Таким образом, стабильное состояние системы определяется здесь как

$$y_{1t} = 5.2, \quad y_{2t} = 4.8,$$

так что стабильное состояние разности $y_{1t} - y_{2t}$ есть $y_1 - y_2 = 0.4$.

Соответственно, с течением времени, независимо от начальных условий, ряд y_{1t} начинает осциллировать вокруг уровня 5.2, а ряд y_{2t} начинает осциллировать вокруг уровня 4.8; разность $(y_{1t} - y_{2t})$ осциллирует вокруг уровня 0.4. Именно такое поведение смоделированных реализаций рассматриваемой VAR(1) мы и наблюдали ранее.

Векторные авторегрессии, определенные так, как было указано выше, называют также **замкнутыми VAR**, отличая тем самым эти модели от **открытых VAR**, в правые части которых наряду с запаздывающими значениями переменных, находящихся в левых частях уравнений (**эндогенные переменные**), входят и некоторые другие переменные и их запаздывания (**экзогенные переменные**). Проводя различие между эндогенными и экзогенными переменными, по-существу предполагают, что значения экзогенных переменных формируются вне рассматриваемой системы, а значения эндогенных переменных порождаются в рамках этой системы. Фактически, система в этом случае рассматривается как условная по отношению к экзогенным переменным. Заметим, что в замкнутой VAR экзогенные переменные отсутствуют.

Открытую VAR можно представить в виде

$$A(L)y_t = \mu + B(L)x_t + \varepsilon_t,$$

где

$$A(L) = I - \Pi_1 L - \Pi_2 L^2 - \dots - \Pi_p L^p$$

и $B(L)$ – матричные полиномы.

Если все решения уравнения $\det A(z) = 0$ лежат за пределами единичного круга на комплексной плоскости, что необходимо для обеспечения стабильности системы, то тогда справедливо также представление

$$y_t = A^{-1}(L)\mu + C(L)x_t + A^{-1}(L)\varepsilon_t,$$

где $C(L) = A^{-1}(L)B(L)$ – **передаточная функция (transfer function)**. Функция $C(L)$ – матричная функция; она устанавливает влияние единичных изменений в экзогенных переменных на эндогенные переменные.

Долговременную (долгосрочную, стабильную, long-run) связь между экзогенными и эндогенными переменными можно найти, полагая в последнем представлении $L = 1$ и $\varepsilon_t \equiv 0$. При этом получаем:

$$y_t = A^{-1}(1)\mu + C(1)x_t.$$

Матрица $C(1)$ называется **матрицей долгосрочных мультипликаторов**. Ее (i, j) -й элемент $c_{ij}(1)$ представляет влияние единичного изменения x_{jt} на y_{it} в долговременном плане (см. интерпретацию долгосрочных мультипликаторов в разд. 4.2).

Пример

На базе рассмотренной выше замкнутой модели VAR(1) для двух рядов построим открытую VAR

$$y_{1t} = 0.6 + 0.7 y_{1, t-1} + 0.2 y_{2, t-1} + 0.1 x_{1, t-1} + 0.2 x_{2, t} + \varepsilon_{1t},$$

$$y_{2t} = 0.4 + 0.2 y_{1, t-1} + 0.7 y_{2, t-1} + 0.2 x_{1, t} + 0.4 x_{2, t-1} + \varepsilon_{2t}.$$

Здесь μ и матричный полином $A(L)$ – те же, что и ранее, а

$$B(L) = B_0 + B_1 L = \begin{pmatrix} 0 & 0.2 \\ 0.2 & 0 \end{pmatrix} + \begin{pmatrix} 0.1 & 0 \\ 0 & 0.4 \end{pmatrix} L = \begin{pmatrix} 0.1L & 0.2 \\ 0.2 & 0.4L \end{pmatrix},$$

так что

$$B(1) = B_0 + B_1 = \begin{pmatrix} 0.1 & 0.2 \\ 0.2 & 0.4 \end{pmatrix}.$$

Матрица долгосрочных мультипликаторов равна

$$C(1) = A^{-1}(1)B(1) = \begin{pmatrix} 6 & 4 \\ 4 & 6 \end{pmatrix} \begin{pmatrix} 0.1 & 0.2 \\ 0.2 & 0.4 \end{pmatrix} = \begin{pmatrix} 1.4 & 2.8 \\ 1.6 & 3.2 \end{pmatrix},$$

так что стабильное решение есть

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 5.2 \\ 4.8 \end{pmatrix} + \begin{pmatrix} 1.4 & 2.8 \\ 1.6 & 3.2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix},$$

т.е.

$$y_1 = 5.2 + 1.4x_1 + 2.8x_2,$$

$$y_2 = 4.8 + 1.6x_1 + 3.2x_2.$$

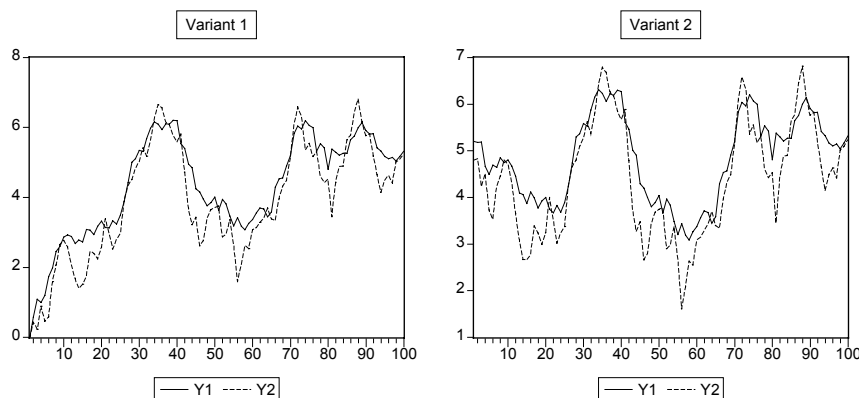
Ниже приведены графики смоделированных реализаций этой открытой системы в случае, когда $x_{1,t}$ и $x_{2,t}$ – независимые друг от друга AR(1) ряды,

$$x_{1,t} = 0.7 x_{1,t-1} + v_{1t}, \quad x_{2,t} = 0.5 x_{2,t-1} + v_{2t}; \quad v_{1t} \text{ и } v_{2t} \sim i.i.d. N(0, 1).$$

В качестве начальных значений при моделировании взяты

- $x_{11} = x_{21} = 0, y_{11} = y_{21} = 0$ (вариант 1)
- $x_{11} = x_{21} = 0, y_{11} = 5.2, y_{21} = 4.8$ (вариант 2).

В результате получаем:



В первом случае, из-за несоответствия начальных значений переменных стабильным соотношениям, системе требуется некоторое время, чтобы выйти на стабильный режим. Во втором случае начальные значения переменных согласованы с долговременными соотношениями между переменными.

Рассмотрим следующую замкнутую VAR(1) для двух переменных:

$$y_{1t} = 0.8 y_{1,t-1} + 0.2 y_{2,t-1} + \varepsilon_{1t},$$

$$y_{2t} = 0.2 y_{1,t-1} + 0.8 y_{2,t-1} + \varepsilon_{2t}.$$

Для этой системы

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} 0.8L & 0.2L \\ 0.2L & 0.8L \end{pmatrix} \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix},$$

так что

$$A(L) = \begin{pmatrix} 1-0.8L & -0.2L \\ -0.2L & 1-0.8L \end{pmatrix}.$$

При этом

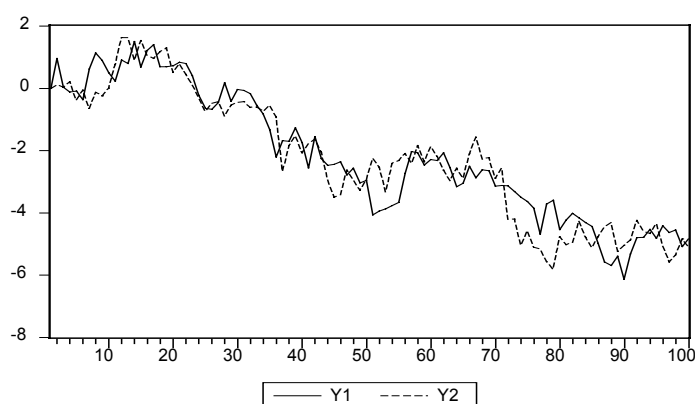
$$A(1) = \begin{pmatrix} 0.2 & 0.2 \\ 0.2 & 0.2 \end{pmatrix},$$

определитель этой матрицы равен нулю, и матрица $A^{-1}(1)$ не определена.

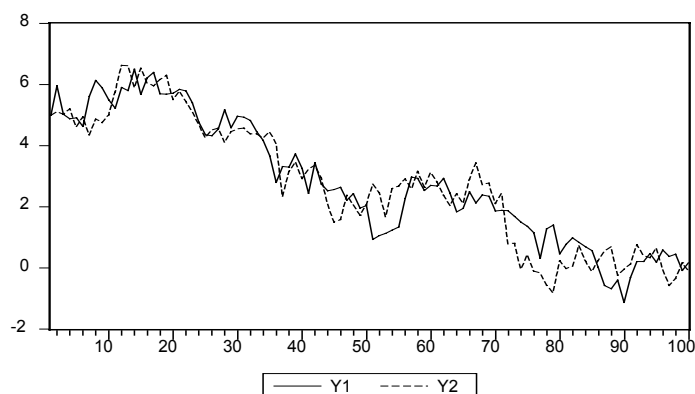
Уравнение $\det A(z) = 0$ имеет здесь вид $(1 - 0.8z)^2 - (0.2z)^2 = 0$, т.е.

$$(1 - z)(1 - 0.6z) = 0.$$

Корни этого уравнения равны $(1/0.6)$ и 1 . Наличие корня, равного 1 , нарушает условие стабильности системы. Как ведут себя в этом случае реализации системы? Ниже приводится соответствующий график.



Здесь стабилизации системы не наблюдается. Можно предположить, что это происходит из-за неудачного выбора начальных значений $y_{11} = y_{21} = 0$. Перемоделируем реализации, полагая начальные значения приблизительно равными наблюдаемому “конечному” уровню: $y_{11} = y_{21} = 5$. Новые реализации



по-прежнему не стабилизируются, и это отражает фундаментальное отличие рассматриваемой нестационарной модели VAR от стабильной.

4.4. Некоторые частные случаи динамических моделей

Чтобы не загромождать изложение, мы ограничимся здесь рассмотрением моделей, входящих в качестве частных случаев в модель ADL(1,1;1)

$$y_t = \mu + a_1 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t.$$

Эти частные случаи, несмотря на свою простоту, дают схематические представления девяяти широко используемых типов моделей.

Различные типы моделей соответствуют различным ограничениям на вектор коэффициентов $\theta = (a_1, \beta_0, \beta_1)$. При наличии двух ограничений мы говорим об *однопараметрической* модели, а при наличии одного ограничения – о *двухпараметрической модели*. Полная модель ADL(1,1;1) является *трехпараметрической*. Ниже мы рассматриваем 9 различных типов моделей.

(1) Статическая регрессия ($a_1 = \beta_1 = 0$): $y_t = \mu + \beta_0 x_t + \varepsilon_t$.

Здесь на значение y_t влияет только значение x_t в тот же момент времени; предшествующие значения y_{t-1} и x_{t-1} не влияют на y_t .

Такая модель обычно не характерна для данных, получаемых последовательно во времени, поскольку в таких ситуациях, как правило, случайные величины ε_t автокоррелированы.

(2) Процесс авторегрессии ($\beta_0 = \beta_1 = 0$): $y_t = \mu + a_1 y_{t-1} + \varepsilon_t$.

Здесь значение y_t зависит только от значения y_{t-1} ; значения переменной x_t в моменты t и $(t-1)$ не влияют на y_t .

Подобные ситуации затрудняют экономический анализ и проведение соответствующей экономической политики из-за того, что в этом случае нет “управляющей” переменной, значения которой можно было бы устанавливать самостоятельно с целью управления значениями переменной y_t .

(3) Модель опережающего показателя ($a_1 = \beta_0 = 0$):

$$y_t = \mu + \beta_1 x_{t-1} + \varepsilon_t.$$

Такие модели могут использоваться для прогнозирования, если изменения показателя y следуют с запаздыванием за изменениями показателя x с достаточной надежностью. Однако, при отсутствии серьезных теоретических оснований, коэффициент β_1 вовсе не обязан быть постоянным. В последнем случае это может приводить к некачественным прогнозам, особенно в периоды структурных изменений, когда хороший прогноз особенно необходим. Кроме того, не видно каких-то особых причин для исключения из правой части запаздывающих значений переменной y .

(4) Модель скорости роста ($a_1 = 1, \beta_1 = -\beta_0$)

$$\Delta y_t = \mu + \beta_0 \Delta x_t + \varepsilon_t,$$

($\Delta = 1 - L$, так что $\Delta y_t = y_t - y_{t-1}$, $\Delta x_t = x_t - x_{t-1}$), соответствует модели статической регрессии, но не для рядов в уровнях, а для рядов в разностях (для *продифференцированных* данных). Однако переход к рядам разностей оправдан только если исходные ряды имеют стохастический тренд и коинтегрированы. Об этом мы будем подробно говорить в последующих главах. А пока укажем только на то, что при неоправданном переходе к рядам разностей теряется информация о характере долговременной экономической связи между рядами в уровнях.

(5) Модель распределенных запаздываний ($a_1 = 0$)

$$y_t = \mu + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t$$

не содержит в правой части запаздываний переменной y . Она страдает теми же недостатками, что и статическая регрессия, но к ним еще может добавиться и проблема мультиколлинеарности переменных x_t и x_{t-1} .

(6) Модель частичной корректировки ($\beta_1 = 0$)

$$y_t = \mu + a_1 y_{t-1} + \beta_0 x_t + \varepsilon_t$$

не содержит в правой части запаздывающих значений переменной x . К такой модели приводят, например, следующие соображения.

Пусть $y_t^* = \alpha + \beta x_t$ – целевой уровень переменной y , а фактически приращение $\Delta y_t = y_t - y_{t-1}$ описывается моделью

$$y_t - y_{t-1} = (1 - \lambda)(y_t^* - y_{t-1}) + \varepsilon_t, \quad 0 \leq \lambda \leq 1,$$

т.е.

$$y_t = (1 - \lambda)y_t^* + \lambda y_{t-1} + \varepsilon_t,$$

так что, с точностью до случайной ошибки ε_t , текущее значение y_t равно взвешенному среднему целевого y_t^* и предыдущего значения переменной y . (Например, y_t – уровень запасов, x_t – уровень продаж.) Тогда

$$y_t = y_{t-1} + (1 - \lambda)(\alpha + \beta x_t - y_{t-1}) + \varepsilon_t = (1 - \lambda)\alpha + \lambda y_{t-1} + (1 - \lambda)\beta x_t + \varepsilon_t,$$

или

$$y_t = \mu + a_1 y_{t-1} + \beta_0 x_t + \varepsilon_t,$$

где

$$\mu = (1 - \lambda)\alpha, \quad a_1 = \lambda, \quad \beta_0 = (1 - \lambda)\beta.$$

Во многих случаях вывод подобных уравнений приводит к автокоррелированным ошибкам, а игнорирование x_{t-1} часто порождает оценку коэффициента a_1 , существенно отличающуюся от оценки a_1 в полной модели.

(7) Фальстарт, или приведенная форма ($\beta_0 = 0$):

$$y_t = \mu + a_1 y_{t-1} + \beta_1 x_{t-1} + \varepsilon_t.$$

К такой модели можно придти, например, если $x_t = \lambda x_{t-1} + u_t$. Тогда подстановка выражения для x_t в полное уравнение ADL(1,1;1) дает:

$$\begin{aligned} y_t &= \mu + a_1 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t \\ &= \mu + a_1 y_{t-1} + (\beta_0 \lambda + \beta_1) x_{t-1} + (\varepsilon_t + \beta_0 u_t), \end{aligned}$$

или

$$y_t = \mu + a_1 y_{t-1} + \beta_1^* x_{t-1} + \varepsilon_t^*.$$

По одному последнему уравнению (**приведенная форма** исходного уравнения) невозможно восстановить значения β_0 и β_1 , не зная значения λ . Т.е. мы можем оценить коэффициенты приведенной формы, но не коэффициенты **структурной формы** (исходного представления ADL(1,1;1)).

(8) Авторегрессионные ошибки ($\beta_1 = -a_1 \beta_0$):

$$y_t = \mu + a_1 y_{t-1} + \beta_0 x_t - a_1 \beta_0 x_{t-1} + \varepsilon_t.$$

Запишем это уравнение в виде

$$y_t - a_1 y_{t-1} = (1 - a_1)\alpha + \beta_0(x_t - a_1 x_{t-1}) + \varepsilon_t.$$

В последнем уравнении легко узнается известное **преобразование Кохрейна – Оркатта**, используемое для преодоления проблемы автокоррелированности ошибок в модели парной регрессии

$$y_t = \alpha + \beta_0 x_t + u_t, \quad u_t = a_1 u_{t-1} + \varepsilon_t, \quad |a_1| < 1.$$

(9) Модель коррекции ошибок ($|a_1| < 1$, $\beta_0 + \beta_1 = b(1 - a_1)$, $b \neq 0$):

$$\Delta y_t = \mu + \beta_0 \Delta x_t - (1 - a_1)(y_{t-1} - b x_{t-1}) + \varepsilon_t,$$

или

$$\Delta y_t = \beta_0 \Delta x_t - (1 - a_1)(y_{t-1} - a - b x_{t-1}) + \varepsilon_t,$$

где $a = \mu / (1 - a_1)$, $b = (\beta_0 + \beta_1) / (1 - a_1)$.

Модели такого вида будут очень часто встречаться у нас при рассмотрении связей между нестационарными временными рядами. В этих случаях такая модель описывает механизм поддержания *долговременной связи*

$$y = a + b x$$

между переменными y_t и x_t в форме коррекций отклонений

$$y_{t-1} - a - b x_{t-1}$$

от долговременной связи в предыдущий момент времени.

Замечание

Исходную (полную) модель ADL(1,1;1)

$$y_t = \mu + a_1 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t$$

всегда можно преобразовать к виду

$$y_t - y_{t-1} = \mu - (1 - a_1) y_{t-1} + \beta_0 (x_t - x_{t-1}) + (\beta_0 + \beta_1) x_{t-1} + \varepsilon_t.$$

Если выполнено условие $|a_1| < 1$ (условие стабильности модели), то

$$\Delta y_t = \mu + \beta_0 \Delta x_t - (1 - a_1)(y_{t-1} - ((\beta_0 + \beta_1)/(1 - a_1)) x_{t-1}) + \varepsilon_t,$$

и при $\beta_0 + \beta_1 \neq 0$ мы получаем модель коррекции ошибок.

Таким образом, модели с $|a_1| < 1$ и $\beta_0 + \beta_1 \neq 0$ могут быть представлены в равносильной форме в виде модели коррекции ошибок.

Обратим теперь внимание на следующее. На практике мы имеем дело только со статистическими данными и не можем знать точно, какая именно модель лежала в основе *процесса порождения данных (data generating process – DGP)*. Мы можем только, привлекая какие-то теоретические положения или результаты ранее проведенных исследований с другими множествами данных, выбрать некоторую *статистическую модель (statistical model – SM)*, которую, по нашему мнению, можно использовать для описания процесса порождения данных. Выбрав такую модель, мы производим ее оценивание и затем можем по оцененной модели проверять различные гипотезы о ее коэффициентах, строить доверительные интервалы для коэффициентов и прогнозировать значения объясняемых переменных для нового набора объясняющих переменных. Между тем, здесь решающее значение имеет соотношение между истинным процессом порождения данных и выбранной статистической моделью.

Если статистическая модель SM оказывается более полной по сравнению с DGP, то тогда оценивание SM приводит к менее эффективным оценкам. С другой стороны, если процесс порождения данных оказывается полнее, чем выбранная SM, то это приводит к более неприятным последствиям – смещению оценок. Вследствие этого, обычно рекомендуется следовать принципу “от общего к частному”, т.е. первоначально выбирать в качестве статистической модели достаточно полную модель, а затем, производя последовательное тестирование статистической модели, редуцировать исходную статистическую модель к более экономной форме.

Пример

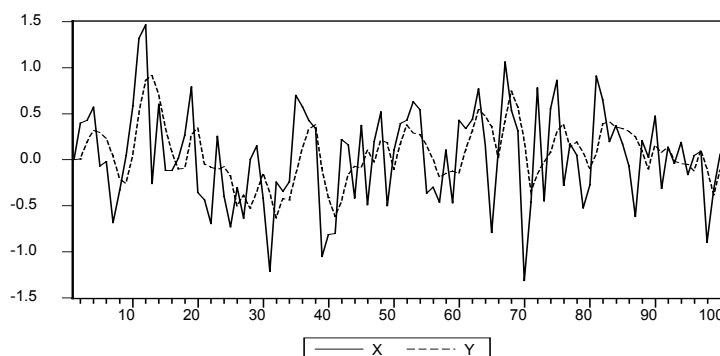
Статистические данные ($n = 100$) порождены стабильной моделью ADL(1,1,1)

$$y_t = 0.5 y_{t-1} + 0.2 x_t + 0.3 x_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim i.i.d. N(0, 0.1^2),$$

$$x_t = 0.5 x_{t-1} + v_t, \quad v_t \sim i.i.d. N(0, 0.5^2),$$

причем ряды ε_t и v_t порождаются независимо друг от друга.

Смоделированные реализации имеют вид



Оценивание по этим реализациям полной модели $ADL(1,1;1)$ в качестве статистической модели дает следующие результаты:

Dependent Variable: Y

Sample(adjusted): 2 100

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.014122	0.009556	1.477773	0.1428
Y(-1)	0.555208	0.034143	16.26107	0.0000
X	0.188567	0.018421	10.23666	0.0000
X(-1)	0.258377	0.020673	12.49808	0.0000
R-squared	0.913395	Mean dependent var	0.062869	
Adjusted R-squared	0.910660	S.D. dependent var	0.310554	
S.E. of regression	0.092824	Akaike info criterion	-1.876660	
Sum squared resid	0.818547	Schwarz criterion	-1.771806	

Исключая из правой части статистической модели константу, получаем:

Dependent Variable: Y

Sample(adjusted): 2 100

Variable	Coefficient	Std. Error	t-Statistic	Prob.
Y(-1)	0.565569	0.033621	16.82186	0.0000
X	0.190325	0.018495	10.29043	0.0000
X(-1)	0.256578	0.020764	12.35668	0.0000
R-squared	0.911404	Mean dependent var	0.062869	
Adjusted R-squared	0.909558	S.D. dependent var	0.310554	
S.E. of regression	0.093394	Akaike info criterion	-1.874134	
Sum squared resid	0.837363	Schwarz criterion	-1.795494	
Log likelihood	95.76965	Durbin-Watson stat	2.218619	

Редуцированная модель признается лучшей по критерию Шварца. Проверка ее на адекватность дает следующие результаты.

- Коррелограмма ряда остатков соответствует процессу белого шума.
- Критерий Бройша – Годфри указывает на отсутствие автокоррелированности у ряда ε_t (P-значение = 0.375 при AR(1) альтернативе и 0.165 при AR(2) альтернативе).
- Критерий Jarque – Bera не обнаруживает значимых отклонений от нормальности (P-значение = 0.689).
- Критерий Уайта не обнаруживает гетероскедастичности (P-значение = 0.285).

Иными словами, применение критериев адекватности к оцененной модели дает удовлетворительные результаты.

Посмотрим теперь, что дает оценивание по тем же данным выбираемых в качестве SM перечисленных ранее 8 редуцированных моделей.

SM₁ (статическая регрессия): $y_t = \mu + \beta_0 x_t + \varepsilon_t$.

Оцененная модель:

Dependent Variable: Y

Sample: 1 100

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X	0.271208	0.053356	5.082965	0.0000
R-squared	0.174472	Mean dependent var	0.062241	
Adjusted R-squared	0.174472	S.D. dependent var	0.309046	
S.E. of regression	0.280794	Akaike info criterion	0.307561	
Sum squared resid	7.805700	Schwarz criterion	0.333613	
Log likelihood	-14.37805	Durbin-Watson stat	0.839862	

В правой части этой статистической модели нет запаздывающих значений объясняемой переменной. Поэтому здесь можно ориентироваться на значения статистики Дарбина – Уотсона. Низкое значение этой статистики указывает на автокоррелированность ряда ε_t , т.е. на неправильную спецификацию выбранной статистической модели.

SM₂ Процесс авторегрессии: $y_t = \mu + a_1 y_{t-1} + \varepsilon_t$.

Оцененная модель:

Dependent Variable: Y

Sample(adjusted): 2 100

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.013941	0.020679	0.674149	0.5018
Y(-1)	0.764874	0.065621	11.65594	0.0000

Поскольку в этой статистической модели правая часть содержит запаздывающее значение объясняемой переменной, ориентироваться на статистику Дарбина – Уотсона не следует. Проверку на отсутствие автокоррелированности для ряда ε_t выполняем, используя критерий Бройша – Годфри. При AR(1) альтернативе Р-значение этого критерия равно 0.00003, так что гипотеза некоррелированности случайных величин ε_t отвергается. Следовательно, выбранная статистическая модель специфицирована неправильно.

SM₃ Модель опережающего показателя: $y_t = \mu + \beta_1 x_{t-1} + \varepsilon_t$.

Оцененная модель:

Dependent Variable: Y

Sample(adjusted): 2 100

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.049508	0.019722	2.510238	0.0137
X(-1)	0.455497	0.037291	12.21457	0.0000

При AR(1) альтернативе Р-значение критерия Бройша – Годфри равно 0.0002, гипотеза некоррелированности случайных величин ε_t отвергается. Выбранная статистическая модель специфицирована неправильно.

SM₄ Модель скорости роста: $\Delta y_t = \mu + \beta_0 \Delta x_t + \varepsilon_t$

Оцененная модель

Dependent Variable: D(Y)

Sample(adjusted): 2 100

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.001126	0.021384	-0.052674	0.9581
D(X)	0.040538	0.033362	1.215078	0.2273
Log likelihood	13.74152	F-statistic		1.476415
Durbin-Watson stat	1.574116	Prob(F-statistic)		0.227286

При AR(1) альтернативе Р-значение критерия Бройша – Годфри равно 0.029, гипотеза некоррелированности случайных величин ε_t отвергается. Выбранная статистическая модель специфицирована неправильно.

SM₅ Модель распределенных запаздываний:

$$y_t = \mu + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t$$

Оцененная модель

Dependent Variable: Y

Sample(adjusted): 2 100

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.046032	0.018096	2.543741	0.0126
X	0.156214	0.035435	4.408526	0.0000
X(-1)	0.414363	0.035435	11.69370	0.0000

При AR(1) альтернативе Р-значение критерия Бройша – Годфри равно 0.00000, гипотеза некоррелированности случайных величин ε_t отвергается. Выбранная статистическая модель специфицирована неправильно.

SM₆ Модель частичной корректировки:

$$y_t = \mu + a_1 y_{t-1} + \beta_0 x_t + \varepsilon_t$$

Оцененная модель

Dependent Variable: Y

Sample(adjusted): 2 100

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.007088	0.015431	0.459354	0.6470
Y(-1)	0.753212	0.048925	15.39514	0.0000
X	0.253493	0.028588	8.867013	0.0000

При AR(1) альтернативе Р-значение критерия Бройша – Годфри равно 0.012, гипотеза некоррелированности случайных величин ε_t отвергается. Выбранная статистическая модель специфицирована неправильно.

SM₇ Приведенная форма: $y_t = \mu + a_1 y_{t-1} + \beta_1 x_{t-1} + \varepsilon_t$.

Оцененная модель

Dependent Variable: Y

Sample(adjusted): 2 100

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.020438	0.013757	1.485648	0.1406
Y(-1)	0.517457	0.048968	10.56734	0.0000
X(-1)	0.318058	0.028613	11.11579	0.0000

S.E. of regression 0.133909 Akaike info criterion -1.153476

Sum squared resid 1.721440 Schwarz criterion -1.074836

Здесь Р-значение критерия Бройша – Годфри при AR(1) альтернативе равно 0.499, а при AR(2) альтернативе равно 0.538. Гипотеза некоррелированности случайных величин ε_t не отвергается, и можно перейти к проверке адекватности другими критериями. Критерий Jarque – Bera не обнаруживает значимых отклонений от нормальности (Р-значение = 0.937). Критерий Уайта не обнаруживает гетероскедастичности (Р-значение = 0.348).

Иными словами, применение критериев адекватности к оцененной модели дает удовлетворительные результаты. Поэтому возможно осуществить редукцию модели, основываясь на статистической незначимости константы в правой части уравнения. Исключение константы из правой части дает:

Dependent Variable: Y

Variable	Coefficient	Std. Error	t-Statistic	Prob.
Y(-1)	0.532005	0.048276	11.02001	0.0000
X(-1)	0.316252	0.028765	10.99445	0.0000

S.E. of regression 0.134740 Akaike info criterion -1.150947

Sum squared resid 1.761018 Schwarz criterion -1.098520

Модель без константы предпочтительнее по критерию Шварца.

С точки зрения анализа остатков, последняя модель вполне может быть использована для описания процесса порождения данных. Однако если мы сравним результаты ее оценивания с полученными ранее результатами оценивания модели $y_t = \alpha_1 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t$, то обнаружим, что в модели с включением x_t в правую часть значения критериев Акаике (- 1.874) и Шварца (- 1.795) гораздо предпочтительнее.

SM₈ Авторегрессионные ошибки:

$$y_t = \mu + \alpha_1 y_{t-1} + \beta_0 x_t - \alpha_1 \beta_0 x_{t-1} + \varepsilon_t.$$

Использование нелинейного (итерационного) метода наименьших квадратов для оценивания параметров этого уравнения дает следующие результаты

Dependent Variable: Y

Sample(adjusted): 2 100

Convergence achieved after 19 iterations

Y=C(1)+C(2)*Y(-1)+C(3)*X-(C(2)*C(3))*X(-1)

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	0.014489	0.020617	0.702743	0.4839
C(2)	0.749747	0.070184	10.68267	0.0000
C(3)	0.052577	0.036535	1.439066	0.1534

R-squared 0.592630 Mean dependent var 0.062869

Adjusted R-squared 0.584144 S.D. dependent var 0.310554

S.E. of regression	0.200267	Akaike info criterion	-0.348499
Sum squared resid	3.850250	Schwarz criterion	-0.269859
Log likelihood	20.25069	Durbin-Watson stat	1.447077

При AR(1) альтернативе Р-значение критерия Бройша – Годфри равно 0.0002, гипотеза некоррелированности случайных величин ε_t отвергается. Выбранная статистическая модель специфицирована неправильно.

Рассмотрим также оценивание SM в форме модели коррекции ошибок (хотя эта модель и не является редуцированной).

SM₉ Модель коррекции ошибок:

$$\Delta y_t = \mu + \beta_0 \Delta x_t - (1 - a_1)(y_{t-1} - b x_{t-1}) + \varepsilon_t.$$

Оцененная модель (нелинейный метод наименьших квадратов):

Dependent Variable: D(Y)

Sample(adjusted): 2 100

Convergence achieved after 4 iterations

$$D(Y) = C(1) + C(2)*D(X) + (C(3)-1)*(Y(-1)-C(4)*X(-1))$$

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	0.014122	0.009556	1.477773	0.1428
C(2)	0.188567	0.018421	10.23666	0.0000
C(3)	0.555208	0.034143	16.26107	0.0000
C(4)	1.004839	0.078119	12.86299	0.0000
R-squared	0.816395	Mean dependent var	-0.001100	
Adjusted R-squared	0.810597	S.D. dependent var	0.213288	
S.E. of regression	0.092824	Akaike info criterion	-1.876660	
Sum squared resid	0.818547	Schwarz criterion	-1.771806	
Log likelihood	96.89465	Durbin-Watson stat	2.248395	

Р-значение критерия Бройша – Годфри при AR(1) альтернативе равно 0.130, а при AR(2) альтернативе равно 0.318; гипотеза некоррелированности случайных величин ε_t не отвергается. Критерий Jarque – Bera не обнаруживает значимых отклонений от нормальности (Р-значение = 0.711). Критерий Уайта не обнаруживает гетероскедастичности (Р-значение = 0.380).

Иными словами, применение критериев адекватности к оцененной модели дает удовлетворительные результаты. Опираясь на них, редуцируем модель, исключая из правой части константу; при этом получаем:

Dependent Variable: D(Y)

Convergence achieved after 3 iterations

$$D(Y) = C(2)*D(X) + (C(3)-1)*(Y(-1)-C(4)*X(-1))$$

	Coefficient	Std. Error	t-Statistic	Prob.
C(2)	0.190325	0.018495	10.29043	0.0000
C(3)	0.565569	0.033621	16.82186	0.0000
C(4)	1.028710	0.080225	12.82279	0.0000
R-squared	0.812174	Mean dependent var	-0.001100	
Adjusted R-squared	0.808261	S.D. dependent var	0.213288	
S.E. of regression	0.093394	Akaike info criterion	-1.874134	
Sum squared resid	0.837363	Schwarz criterion	-1.795494	

Log likelihood	95.76965	Durbin-Watson stat	2.218619
----------------	----------	--------------------	----------

т.е.

$$\Delta y_t = 0.190 \Delta x_t - 0.434(y_{t-1} - 1.029 x_{t-1}) + e_t.$$

Модель без константы предпочтительнее по критерию Шварца.

Уединяя y_t в левой части уравнения, получаем:

$$y_t = 0.566 y_{t-1} + 0.190 x_t + 0.253 x_{t-1} + e_t.$$

Сравним это уравнение с реально использованным для моделирования

$$\text{DGP: } y_t = 0.5 y_{t-1} + 0.2 x_t + 0.3 x_{t-1} + \varepsilon_t$$

и с результатом оценивания соответствующей ему статистической модели:

$$y_t = 0.565 y_{t-1} + 0.190 x_t + 0.257 x_{t-1} + \varepsilon_t.$$

Найдем долговременное соотношение между переменными y_t и x_t , соответствующее теоретическому процессу порождения данных:

$$y_t = 0.5 y_t + 0.2 x_t + 0.3 x_t \rightarrow y = x.$$

В то же время, долговременное соотношение, получаемое по оцененной SM, соответствующей этому DGP:

$$y_t = 0.565 y_t + 0.190 x_t + 0.257 x_t \rightarrow y = 1.002 x.$$

Далее, долговременное соотношение, получаемое по оцененной SM₉ (в варианте без константы в правой части):

$$y_t = 0.566 y_t + 0.190 x_t + 0.253 x_t \rightarrow y = 1.021 x.$$

Наконец, если взять результаты оценивания модели SM₇ (приведенная форма) без включения константы, то для этого случая получаем:

$$y_t = 0.532 y_t + 0.316 x_t \rightarrow y = 0.675 x.$$

Эти результаты указывают на возможность серьезных последствий, проистекающих из неправильной спецификации SM, когда эта спецификация оказывается уже спецификации DGP. Заметим, что в рамках такой SM отнюдь не всегда удастся обнаружить статистическими методами “узость” выбранной спецификации. Мы смогли это сделать в рамках оцененных статистических моделей SM₁ – SM₆ и SM₈, но не в модели SM₇.

Рассмотрим теперь обратную ситуацию, когда, напротив, выбранная для оценивания статистическая модель SM оказывается полнее (“шире”) модели DGP, так что модель, соответствующая DGP, является частным случаем статистической модели, выбранной для оценивания.

В качестве DGP будем последовательно брать модели (1) – (8), а в качестве SM – полную модель ADL(1,1;1) без ограничений на коэффициенты:

$$y_t = \mu + a_1 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t.$$

Значения коэффициентов при переменных в моделях (1) – (8) будем брать такими же, как и в исходной модели ADL(1,1;1)

$$y_t = 0.5 y_{t-1} + 0.2 x_t + 0.3 x_{t-1} + \varepsilon_t.$$

При моделировании DGP во всех случаях берется $\varepsilon_t \sim i.i.d. N(0, 0.1^2)$.

DGP₁ : Статическая регрессия

$$y_t = 0.2 x_t + \varepsilon_t.$$

Оцененная статистическая модель

Dependent Variable: Y

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.004647	0.010300	-0.451175	0.6529
Y(-1)	0.102848	0.101833	1.009966	0.3151
X	0.186813	0.020222	9.238033	0.0000
X(-1)	0.000201	0.028272	0.007101	0.9943
R-squared	0.507795	Mean dependent var	0.001230	
Adjusted R-squared	0.492252	S.D. dependent var	0.143412	
S.E. of regression	0.102190	Akaike info criterion	-1.684398	
Sum squared resid	0.992068	Schwarz criterion	-1.579545	

Р-значение критерия Бройша – Годфри при AR(1) альтернативе равно 0.760, а при AR(2) альтернативе равно 0.951, гипотеза некоррелированности случайных величин ε_t не отвергается. Критерий Jarque – Bera не обнаруживает значимых отклонений от нормальности (Р-значение = 0.733). Критерий Уайта не обнаруживает гетероскедастичности (Р-значение = 0.770).

Иными словами, применение критериев адекватности к оцененной модели дает удовлетворительные результаты. Опираясь на них, можно перейти к проверке гипотез о значениях коэффициентов. При проверке гипотезы о занулении константы и коэффициентов при y_{t-1} и x_{t-1} получаем значение обычной F -статистики, равное $F = 0.738$, и $qF = 2.214$. Исходя из F -распределения для статистики F , получаем Р-значение 0.532. Использование асимптотического распределения χ^2 (3) для qF приводит к Р-значению 0.529. При обоих вариантах гипотеза о занулении трех указанных коэффициентов не отвергается. Тем самым, можно перейти к оцениванию редуцированной модели $y_t = \mu + \beta_0 x_t + \varepsilon_t$, и это дает:

Dependent Variable: Y

Sample: 1 100

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X	0.190067	0.019291	9.852604	0.0000
R-squared	0.494995	Mean dependent var	0.001935	
Adjusted R-squared	0.494995	S.D. dependent var	0.142860	
S.E. of regression	0.101522	Akaike info criterion	-1.727139	
Sum squared resid	1.020359	Schwarz criterion	-1.701087	

Редуцированная модель лучше полной и по критерию Акаике и по критерию Шварца. Остатки от оцененной редуцированной модели проходят тесты на нормальность, отсутствие автокоррелированности и гетероскедастичности.

DGP₂ : Процесс авторегрессии

$$y_t = 0.5y_{t-1} + \varepsilon_t.$$

Оцененная статистическая модель

Dependent Variable: Y

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.004519	0.010315	-0.438075	0.6623
Y(-1)	0.576756	0.084134	6.855173	0.0000
X	-0.013220	0.020253	-0.652774	0.5155
X(-1)	0.021476	0.020228	1.061719	0.2911
R-squared	0.338422	Mean dependent var	-0.007891	

Adjusted R-squared	0.317530	S.D. dependent var	0.123820
S.E. of regression	0.102290	Akaike info criterion	-1.682441
Sum squared resid	0.994011	Schwarz criterion	-1.577588

Р-значение критерия Бройша – Годфри при AR(1) альтернативе равно 0.600, а при AR(2) альтернативе равно 0.773, гипотеза некоррелированности случайных величин ε_t не отвергается. Критерий Jarque – Bera не обнаруживает значимых отклонений от нормальности (Р-значение = 0.654). Критерий Уайта не обнаруживает гетероскедастичности (Р-значение = 0.956).

При проверке гипотезы о занулении константы и коэффициентов при x_t и x_{t-1} получаем значение обычной F -статистики, равное $F = 0.641$, и $qF = 1.283$. Исходя из F -распределения для статистики F , получаем Р-значение 0.529. Использование асимптотического распределения $\chi^2(3)$ для qF приводит к Р-значению 0.527. При обоих вариантах гипотеза о занулении трех указанных коэффициентов не отвергается. Тем самым, можно перейти к оцениванию редуцированной модели $y_t = a_1 y_{t-1} + \varepsilon_t$, и это дает:

Dependent Variable: Y

Variable	Coefficient	Std. Error	t-Statistic	Prob.
Y(-1)	0.575922	0.082705	6.963585	0.0000
R-squared	0.328274	Mean dependent var	-0.007891	
Adjusted R-squared	0.328274	S.D. dependent var	0.123820	
S.E. of regression	0.101482	Akaike info criterion	-1.727825	
Sum squared resid	1.009258	Schwarz criterion	-1.701612	

Редуцированная модель предпочтительнее и по критерию Акаике и по критерию Шварца. Анализ остатков не выявляет значимых отклонений от сделанных предположений в отношении ряда ε_t .

DGP₃ : Модель опережающего показателя

$$y_t = 0.3 x_{t-1} + \varepsilon_t.$$

Оцененная модель

Dependent Variable: Y

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.005202	0.010305	-0.504767	0.6149
Y(-1)	0.052373	0.054196	0.966363	0.3363
X	-0.012475	0.020310	-0.614213	0.5405
X(-1)	0.315962	0.020645	15.30433	0.0000
R-squared	0.736662	Mean dependent var	0.004035	
Adjusted R-squared	0.728346	S.D. dependent var	0.196154	
S.E. of regression	0.102236	Akaike info criterion	-1.683501	
Sum squared resid	0.992959	Schwarz criterion	-1.578647	

Р-значение критерия Бройша – Годфри при AR(1) альтернативе равно 0.614, а при AR(2) альтернативе равно 0.868, гипотеза некоррелированности случайных величин ε_t не отвергается. Критерий Jarque – Bera не обнаруживает значимых отклонений от нормальности (Р-значение = 0.740). Критерий Уайта не обнаруживает гетероскедастичности (Р-значение = 0.804).

При проверке гипотезы о занулении константы и коэффициентов при x_t и y_{t-1} получаем значение обычной F -статистики, равное $F = 0.577$, и $qF = 1.730$. Исходя из F -распределения для статистики F , получаем Р-значение 0.632. Использование

асимптотического распределения $\chi^2(3)$ для qF приводит к Р-значению 0.630. При обоих вариантах гипотеза о занулении трех указанных коэффициентов не отвергается. Тем самым, можно перейти к оцениванию редуцированной модели $y_t = \beta_1 x_{t-1} + \varepsilon_t$, и это дает:

Dependent Variable: Y

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X(-1)	0.315777	0.019302	16.35987	0.0000
R-squared	0.731866	Mean dependent var		0.004035
Adjusted R-squared	0.731866	S.D. dependent var		0.196154
S.E. of regression	0.101572	Akaike info criterion		-1.726058
Sum squared resid	1.011044	Schwarz criterion		-1.699844

Редуцированная модель предпочтительнее и по критерию Акаике и по критерию Шварца. Анализ остатков не выявляет значимых отклонений от сделанных предположений в отношении ряда ε_t .

DGP₄: Модель скорости роста

$$\Delta y_t = \mu + \beta_0 \Delta x_t + \varepsilon_t.$$

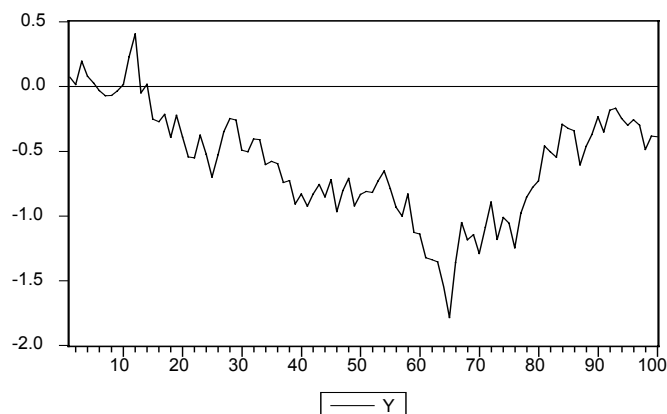
Оцененная модель

Dependent Variable: Y

Included observations: 99 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.029026	0.017741	-1.636049	0.1051
Y(-1)	0.959750	0.024228	39.61303	0.0000
X	0.184064	0.019993	9.206520	0.0000
X(-1)	-0.173461	0.020340	-8.528162	0.0000
R-squared	0.944972	Mean dependent var		-0.599911
Adjusted R-squared	0.943234	S.D. dependent var		0.425076
S.E. of regression	0.101277	Akaike info criterion		-1.702356
Sum squared resid	0.974412	Schwarz criterion		-1.597502
Log likelihood	88.26661	F-statistic		543.7988
Durbin-Watson stat	1.745026	Prob(F-statistic)		0.000000

Отметим близкое к единице оцененное значение коэффициента при y_{t-1} , что может говорить о том, что в DGP “истинное” значение коэффициента равно $a_1 = 1$. Но тогда нарушается условие стабильности системы. И действительно, график ряда y_t , полученного в результате моделирования, имеет вид

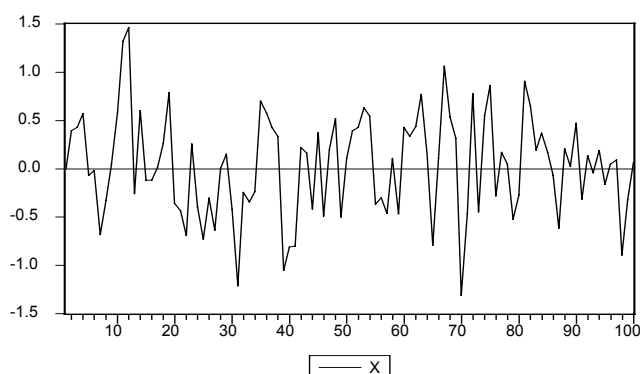


явно указывающий на нестационарность ряда.

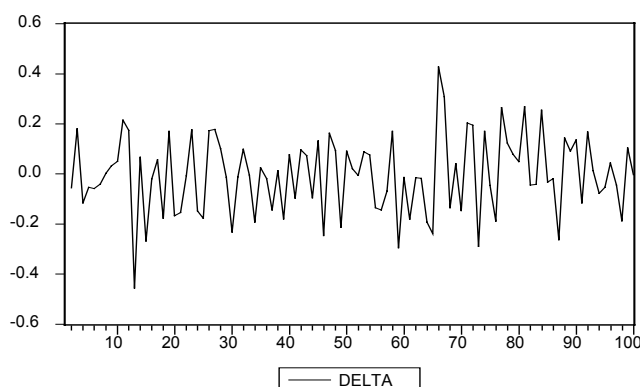
Вопрос о проверке гипотез типа $H_0: a_1 = 1$ требует специального рассмотрения, и мы будем рассматривать его в последующих главах. Сейчас же, исходя из наблюдаемого поведения ряда y_t и близости оцененного значения коэффициента к 1, займемся оценением модели

$$\Delta y_t = \mu + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t.$$

Мы можем использовать для нее стандартную (асимптотическую) технику статистических выводов, поскольку реализация ряда x_t имеет вид



указывающий на стационарность этого ряда, и реализация ряда разностей Δy_t имеет вид



говорящий в пользу стационарности ряда Δy_t .

В результате оценивания получаем:

Dependent Variable: D(Y)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.004915	0.010297	-0.477336	0.6342
X	0.185224	0.020163	9.186333	0.0000
X(-1)	-0.179782	0.020163	-8.916410	0.0000

Здесь мы, конечно, обращаем внимание на статистическую незначимость константы, а также на то, что оцененные коэффициенты при переменных x_t и x_{t-1} близки по абсолютной величине и противоположны по знаку. В связи с этим, мы в рамках статистической модели

$$\Delta y_t = \mu + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t$$

проверяем гипотезу $H_0: \mu = 0, \beta_0 = -\beta_1$. Используя F -распределение для F -статистики и распределение хи-квадрат $\chi^2(2)$ для статистики $qF = 2F$, получаем в обоих случаях P -значение 0.876. Гипотеза H_0 не отвергается, и можно перейти к

оцениванию модели с такими ограничениями, т.е. модели $\Delta y_t = \beta_0 \Delta x_t + \varepsilon_t$. Оцененная модель с ограничениями:

Dependent Variable: D(Y)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(X)	0.182495	0.015882	11.49045	0.0000

P -значение критерия Бройша – Годфри при AR(1) альтернативе равно 0.328, а при AR(2) альтернативе равно 0.605; гипотеза некоррелированности случайных величин ε_t не отвергается. Критерий Jarque – Bera не обнаруживает значимых отклонений от нормальности (P -значение = 0.673). Критерий Уайта не обнаруживает гетероскедастичности (P -значение = 0.988). Таким образом, и в этом случае мы вышли в результате тестирования на статистическую модель, имеющую ту же спецификацию, что и DGP.

DGP₅ Модель распределенных запаздываний

$$y_t = \mu + 0.2 x_t + 0.3 x_{t-1} + \varepsilon_t.$$

Оцененная модель

Dependent Variable: Y

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.003282	0.010206	-0.321584	0.7485
Y(-1)	0.000523	0.058294	0.008975	0.9929
X	0.181735	0.019977	9.097166	0.0000
X(-1)	0.289502	0.024922	11.61638	0.0000
R-squared	0.804020	Mean dependent var		0.010665
Adjusted R-squared	0.797831	S.D. dependent var		0.225113
S.E. of regression	0.101218	Akaike info criterion		-1.703515
Sum squared resid	0.973283	Schwarz criterion		-1.598661

P -значение критерия Бройша – Годфри при AR(1) альтернативе равно 0.972, а при AR(2) альтернативе равно 0.826; гипотеза некоррелированности случайных величин ε_t не отвергается. Критерий Jarque – Bera не обнаруживает значимых отклонений от нормальности (P -значение = 0.689). Критерий Уайта не обнаруживает гетероскедастичности (P -значение = 0.433).

Проверяем гипотезу $H_0: \mu = 0, a_1 = 0$. Используя F -распределение для F -статистики и распределение хи-квадрат $\chi^2(2)$ для статистики $qF = 2F$, получаем в обоих случаях P -значение 0.950. Гипотеза H_0 не отвергается, и можно перейти к оцениванию модели с $\mu = 0, a_1 = 0$, т.е. модели $y_t = \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t$. Оцененная модель с ограничениями:

Dependent Variable: Y

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X	0.181461	0.019754	9.185972	0.0000
X(-1)	0.289367	0.019756	14.64736	0.0000

DGP₆ Модель частичной корректировки

$$y_t = \mu + 0.5 y_{t-1} + 0.2 x_t + \varepsilon_t.$$

Оцененная модель:

Dependent Variable: Y

Variable	Coefficient	Std. Error	t-Statistic	Prob.
----------	-------------	------------	-------------	-------

C	-0.005099	0.010263	-0.496869	0.6204
Y(-1)	0.592041	0.071296	8.304016	0.0000
X	0.188766	0.020280	9.308077	0.0000
X(-1)	0.000973	0.025019	0.038898	0.9691

P-значение критерия Бройша – Годфри при AR(1) альтернативе равно 0.904, а при AR(2) альтернативе равно 0.723; гипотеза некоррелированности случайных величин ε_t не отвергается. Критерий Jarque – Bera не обнаруживает значимых отклонений от нормальности (P-значение = 0.691). Критерий Уайта не обнаруживает гетероскедастичности (P-значение = 0.533).

Проверяем гипотезу $H_0: \mu = 0, \beta_1 = 0$. Используя F-распределение для F-статистики и распределение хи-квадрат $\chi^2(2)$ для статистики $qF = 2F$, получаем в обоих случаях P-значение 0.884. Гипотеза H_0 не отвергается, и можно перейти к оцениванию модели с $\mu = 0, \beta_1 = 0$, т.е. модели $y_t = \mu + a_1 y_{t-1} + \beta_0 x_t + \varepsilon_t$. Оцененная модель с ограничениями:

Dependent Variable: Y

Variable	Coefficient	Std. Error	t-Statistic	Prob.
Y(-1)	0.592666	0.056702	10.45233	0.0000
X	0.188468	0.019202	9.814814	0.0000
R-squared	0.690477	Mean dependent var	0.004719	
Adjusted R-squared	0.687286	S.D. dependent var	0.180476	
S.E. of regression	0.100924	Akaike info criterion	-1.728911	
Sum squared resid	0.988000	Schwarz criterion	-1.676484	

DGP₇ Приведенная форма

$$y_t = 0.5 y_{t-1} + 0.3 x_{t-1} + \varepsilon_t.$$

Оцененная модель

Dependent Variable: Y

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.005886	0.010271	-0.573100	0.5679
Y(-1)	0.559497	0.042762	13.08387	0.0000
X	-0.010318	0.020320	-0.507807	0.6128
X(-1)	0.316645	0.020229	15.65291	0.0000

P-значение критерия Бройша – Годфри при AR(1) альтернативе равно 0.701, а при AR(2) альтернативе равно 0.827; гипотеза некоррелированности случайных величин ε_t не отвергается. Критерий Jarque – Bera не обнаруживает значимых отклонений от нормальности (P-значение = 0.740). Критерий Уайта не обнаруживает гетероскедастичности (P-значение = 0.586).

Проверяем гипотезу $H_0: \mu = 0, \beta_0 = 0$. Используя F-распределение для F-статистики и распределение хи-квадрат $\chi^2(2)$ для статистики $qF = 2F$, получаем в обоих случаях P-значение 0.734. Гипотеза H_0 не отвергается, и можно перейти к оцениванию модели с $\mu = 0, \beta_0 = 0$, т.е. к оцениванию модели $y_t = a_1 y_{t-1} + \beta_1 x_{t-1} + \varepsilon_t$. Оцененная модель с ограничениями:

Dependent Variable: Y

Variable	Coefficient	Std. Error	t-Statistic	Prob.
Y(-1)	0.561389	0.041849	13.41452	0.0000

X(-1)	0.313207	0.019269	16.25422	0.0000
-------	----------	----------	----------	--------

DGP₈ Авторегрессионные ошибки

$$y_t = 0.5 y_{t-1} + 0.2 x_t - 0.1 x_{t-1} + \varepsilon_t.$$

Оцененная модель:

Dependent Variable: Y

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.004519	0.010315	-0.438075	0.6623
Y(-1)	0.576756	0.084134	6.855173	0.0000
X	0.186780	0.020253	9.222532	0.0000
X(-1)	-0.093875	0.025414	-3.693770	0.0004

P-значение критерия Бройша – Годфри при AR(1) альтернативе равно 0.600, а при AR(2) альтернативе равно 0.773; гипотеза некоррелированности случайных величин ε_t не отвергается. Критерий Jarque – Bera не обнаруживает значимых отклонений от нормальности (P-значение = 0.654). Критерий Уайта не обнаруживает гетероскедастичности (P-значение = 0.682).

Исключим из статистической модели статистически незначимую константу:

Dependent Variable: Y

Variable	Coefficient t	Std. Error	t-Statistic	Prob.
Y(-1)	0.578309	0.083705	6.908866	0.0000
X	0.186426	0.020151	9.251413	0.0000
X(-1)	-0.094531	0.025263	-3.741827	0.0003

Заметим, что произведение оцененных коэффициентов при y_{t-1} и x_t равно 0.108, т.е. близко по абсолютной величине и противоположно по знаку коэффициенту при x_{t-1} . В связи с этим наблюдением, проверим гипотезу $H_0: \beta_1 = -a_1\beta_0$. Здесь мы имеем дело с нелинейной гипотезой, и результаты проверки могут зависеть от формы записи этого ограничения на коэффициенты. Поэтому мы проверяем указанную гипотезу в трех формах:

$$H_0: \beta_1 = -a_1\beta_0; H_0: \beta_0 = -(\beta_1/a_1); H_0: a_1 = -(\beta_1/\beta_0).$$

Соответствующие этим формам P-значения $\chi^2(1)$ -критериев равны 0.515, 0.514 и 0.506, так что выводы в отношении гипотезы H_0 согласуются: эта гипотеза не отвергается. Последнее означает, что можно перейти к оцениванию модели с ограничением на коэффициенты в виде $\beta_1 = -a_1\beta_0$, т.е. к модели $y_t = a_1 y_{t-1} + \beta_0 x_t - a_1\beta_0 x_{t-1} + \varepsilon_t$. В итоге получаем оцененную модель

Dependent Variable: Y

Convergence achieved after 3 iterations

$$Y = C(1)*Y(-1) + C(2)*X - (C(1)*C(2))*X(-1)$$

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	0.575812	0.083369	6.906747	0.0000
C(2)	0.182370	0.019110	9.543254	0.0000

которую можно записать в виде

$$y_t = 0.576 y_{t-1} + 0.182 x_t - 0.105 x_{t-1} + \varepsilon_t.$$

Таким образом, во всех рассмотренных случаях, когда DGP являлся частным случаем выбранной для оценивания статистической модели, последовательное применение метода редукции модели от общего к частному (с предварительной проверкой SM на адекватность) выводило нас на редуцированные модели, спецификация которых соответствовала спецификации DGP. В то же время, как мы видели перед этим, при движении “от частного к общему” возможны ситуации, когда остановка происходит на модели, существенно отличающейся от DGP, хотя и проходящей стандартные тесты на адекватность имеющимся статистическим данным.

Это еще раз подчеркивает предпочтительность использования при подборе моделей по статистическим данным метода “от общего к частному”, т.е. первоначальному выбору достаточно общей модели, проверки ее на адекватность имеющимся статистическим данным, и, в случае признания выбранной модели адекватной данным, последующей редукции этой модели с использованием стандартных критериев спецификации.

В связи с последними замечаниями, еще раз обратимся к модели линейной регрессии с автокоррелированными ошибками, образующими стационарный процесс авторегрессии первого порядка. В учебной литературе по эконометрике довольно часто делается упор на эту модель как средство преодоления проблемы автокоррелированности ошибок в рамках известных процедур Кохрейна – Оркатта или Прайса – Уинстена. Однако, как ясно из предыдущего изложения, такая модель (в нашей нумерации – модель 8) является всего лишь весьма частным случаем общей динамической модели $ADL(1,1;1)$. В рамках этой общей модели

$$y_t = \mu + a_1 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t$$

модель, о которой идет речь, выделяется выполнением соотношения

$$\beta_1 = -a_1 \beta_0.$$

В то же время, при $\beta_0 \neq 0$ общую модель $ADL(1,1;1)$ можно представить в виде

$$(1 - a_1 L)y_t = \beta_0 \left(1 + \frac{\beta_1}{\beta_0} L \right) x_t + \varepsilon_t,$$

или

$$a(L) y_t = b(L) x_t + \varepsilon_t,$$

где

$$a(L) = 1 - a_1 L, \quad b(L) = \beta_0 \left(1 + \frac{\beta_1}{\beta_0} L \right).$$

(Для простоты мы полагаем $\mu = 0$.)

Если $a_1 = -\frac{\beta_1}{\beta_0}$, то модель принимает вид

$$(1 - a_1 L)y_t = \beta_0 (1 - a_1 L) x_t + \varepsilon_t,$$

так что многочлены $a(L)$ и $b(L)$ имеют **общий множитель** $(1 - a_1 L)$. Разделив обе части уравнения на этот общий множитель, получаем

$$y_t = \beta_0 x_t + u_t,$$

где

$$u_t = \frac{\varepsilon_t}{1 - a_1 L},$$

так что $(1 - a_1 L) u_t = \varepsilon_t$ и $u_t = a_1 u_{t-1} + \varepsilon_t$.

В связи с наличием общего множителя, модель авторегрессионных ошибок относят к классу моделей, носящему название COMFAC (common factors). Рассматриваемая модель обязана принадлежностью к этому классу именно наличию ограничения $\beta_1 = -a_1\beta_0$. Класс COMFAC является весьма частным случаем моделей с авторегрессионно распределенными запаздываниями. Поэтому применение обычной процедуры проверки автокоррелированности ошибок в модели регрессии $y_t = \beta_0 x_t + u_t$ и коррекции обнаруженной автокоррелированности посредством авторегрессионного преобразования переменных, вообще говоря, некорректно. Правильный порядок действий должен состоять

- В установлении пригодности модели $a(L) y_t = b(L) x_t + \varepsilon_t$ с помощью различных критериев адекватности; гипотеза о том, что ряд ε_t является гауссовским белым шумом, не должна отвергаться – в противном случае следует говорить о непригодности уже этой общей модели.
- В проверке гипотезы о том, что многочлены $a(L)$ и $b(L)$ имеют общие корни.
- Наконец, в случае подтверждения обеих гипотез следует проверить гипотезу $H_0: a_1 = 0$ (она соответствует модели статической регрессии). Заметим, что отвержение этой гипотезы непосредственно в модели с автокоррелированными ошибками вовсе не доказывает наличия указанных общих множителей.

Однако здесь имеются некоторые сложности.

На первом шаге гипотеза H_0 : “ ε_t – белый шум” проверяется, в частности, против альтернативы $H_A: \varepsilon_t \sim \text{AR}(k)$ с $k \leq p$, т.е.

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \dots + \rho_p \varepsilon_{t-p} + v_t,$$

где $v_t \sim i.i.d.$ и хотя бы одно $\rho_j \neq 0$. Модель, соответствующая альтернативе, имеет вид

$$y_t = a_1 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \rho_1 \varepsilon_{t-1} + \dots + \rho_p \varepsilon_{t-p} + v_t,$$

и, фактически, речь идет о проверке гипотезы $H_0: \rho_1^2 = \dots = \rho_p^2 = 0$ против $H_A: \rho_1^2 + \dots + \rho_p^2 \neq 0$. Такую проверку можно осуществить, используя стандартный LM критерий Бройша – Годфри. В то же время не рекомендуется использовать для этой цели критерии, основанные на статистиках Бокса – Пирса и Лjung – Бокса из разд. 3.1 и предназначенные для анализа “сырых” рядов. (См., например, статью [Kwan, Sim (1996)].)

Проблемы возникают и с применением стандартного критерия Вальда для проверки гипотезы $H_1: \beta_1 = -a_1\beta_0$ против альтернативы $H_A: \beta_1 \neq -a_1\beta_0$. Дело в том, что эта гипотеза не является линейной, а в таких случаях результаты применения критерия Вальда зависят от того, в какой форме записано ограничение: $\beta_1 = -a_1\beta_0$, $a_1 = -\beta_1/\beta_0$ или $\beta_0 = -\beta_1/a_1$, что может приводить к противоречивым выводам.

Отметим также проблему, связанную с последовательным использованием нескольких критериев проверки гипотез. В рамках рассмотренной процедуры приходится, по крайней мере, сначала проверять гипотезу H_1 о наличии общих множителей, а затем, если она не отвергается, проверять гипотезу $H_2: a_1 = 0$ о некоррелированности ошибок в статической модели регрессии.

Пусть каждая из этих гипотез проверяется на уровне значимости α , скажем, $\alpha = 0.05$. Какова в такой ситуации вероятность ошибочного отвержения модели статической регрессии? Имеем:

$$\begin{aligned} P\{\text{ошибочно отвергается хотя бы одна из гипотез } H_1, H_2\} &\leq \\ &\leq P\{\text{ошибочно отвергается } H_1\} + P\{\text{ошибочно отвергается } H_2\} = \end{aligned}$$

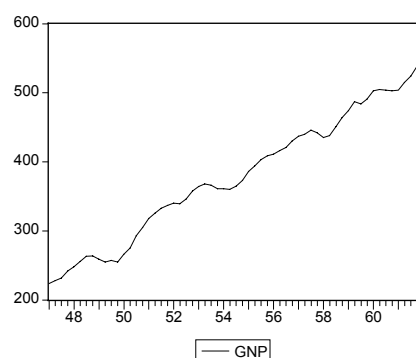
$$= \alpha + \alpha = 2\alpha.$$

Следовательно, если положить $\alpha = 0.025$, то вероятность отвержения модели статической регрессии в рамках двухступенчатой процедуры не будет превышать 0.05. Заметим, что при этом мы еще не принимали в расчет ошибки, связанные с возможностью неправильной диагностики общей модели.

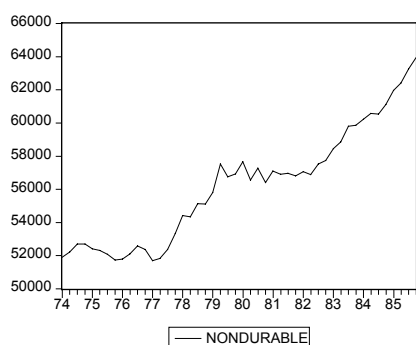
Глава 5. Нестационарные временные ряды

В этой главе нам удобно временно вернуться к прописным и строчным обозначениям для случайных величин и их реализаций, соответственно.

Мы начнем изложение с рассмотрения двух временных рядов. Первый из них представляет статистические данные о величине валового национального продукта (GNP – gross national product) в США за период с первого квартала 1947 г. по четвертый квартал 1961 г. (сезонно скорректированные квартальные данные в пересчете на год – 60 наблюдений, *млрд* долл., в текущих ценах). График этого ряда



имеет выраженный линейный тренд. Второй ряд (NONDURABLE) представляет статистические данные об объеме потребительских расходов на товары кратковременного пользования и услуги в Великобритании за период с первого квартала 1974 г. по четвертый квартал 1985 г. (сезонно скорректированные квартальные данные – 48 наблюдений, *млн* фунтов стерлингов, в текущих ценах):



Этот ряд также обнаруживает линейный тренд.

Хотя выборочные автокорреляционные и частные автокорреляционные функции определялись у нас для стационарных рядов, посмотрим на коррелограммы, построенные по представленным данным. Для ряда GNP коррелограмма имеет вид

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
. *****	. *****	1	0.946	0.946	56.419	0.000
. *****	. .	2	0.893	-0.021	107.52	0.000
. *****	. .	3	0.840	-0.024	153.55	0.000
. *****	. .	4	0.791	0.013	195.14	0.000
. *****	. .	5	0.743	-0.021	232.52	0.000
. *****	. .	6	0.696	-0.022	265.90	0.000

. *****	. . .	7	0.648	-0.030	295.41	0.000
. *****	. . .	8	0.599	-0.044	321.09	0.000
. *****	. . .	9	0.550	-0.033	343.13	0.000
. *****	. . .	10	0.498	-0.052	361.57	0.000
. ****	* . .	11	0.442	-0.073	376.44	0.000
. ****	. . .	12	0.388	-0.034	388.08	0.000
. ****	. . .	13	0.337	-0.002	397.06	0.000
. **	. . .	14	0.291	0.007	403.91	0.000
. **	. . .	15	0.253	0.041	409.21	0.000
. **	. . .	16	0.218	-0.002	413.22	0.000
. *	. . .	17	0.182	-0.034	416.08	0.000
. *	. . .	18	0.143	-0.044	417.90	0.000
. *	. . .	19	0.106	-0.021	418.92	0.000
. *	. . .	20	0.069	-0.039	419.36	0.000
.	21	0.032	-0.028	419.45	0.000
.	22	-0.003	-0.030	419.45	0.000
.	23	-0.037	-0.021	419.59	0.000
*	24	-0.066	-0.005	420.04	0.000

А для ряда NONDURABLE

ACF	PACF	AC	PAC	Q-Stat	Prob
. *****	. *****	1	0.917	0.917	42.921 0.000
. *****	. . .	2	0.843	0.014	79.976 0.000
. *****	. . .	3	0.774	-0.004	111.92 0.000
. *****	. . .	4	0.704	-0.040	138.99 0.000
. *****	. . .	5	0.643	0.013	162.09 0.000
. ****	. . .	6	0.581	-0.041	181.36 0.000
. ****	* . .	7	0.510	-0.090	196.57 0.000
. ***	. . .	8	0.439	-0.050	208.14 0.000
. ***	* . .	9	0.366	-0.066	216.39 0.000
. **	. . .	10	0.300	-0.012	222.07 0.000
. **	. . .	11	0.246	0.026	225.99 0.000
. *	. . .	12	0.196	-0.006	228.56 0.000
. *	. . .	13	0.150	-0.013	230.11 0.000
. *	. . .	14	0.106	-0.025	230.91 0.000
. *	. . .	15	0.072	0.036	231.28 0.000
.	16	0.041	-0.016	231.41 0.000
.	17	0.019	0.028	231.43 0.000
. . .	* . .	18	-0.007	-0.063	231.44 0.000
.	19	-0.032	-0.018	231.52 0.000
* . .	* . .	20	-0.062	-0.073	231.85 0.000

Если отвлечься от видимой нестационарности этих рядов, то поведение выборочных ACF и PACF предполагает идентификацию обоих рядов как рядов типа AR(1). Имея в виду наличие у рядов выраженного линейного тренда, произведем оценивание моделей

$$X_t = \alpha + \beta t + a_1 X_{t-1} + u_t .$$

(Здесь мы используем обозначение u_t , а не ε_t , поскольку ряд u_t на этот раз может и не являться белым шумом.)

Это приводит к следующим результатам.¹

Для ряда GNP:

Dependent Variable: GNP

Method: Least Squares

Sample(adjusted): 1947:2 1961:4

Included observations: 59 after adjusting endpoints

Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	216.0630	11.30237	19.11661	0.0000
T	5.269279	0.281754	18.70170	0.0000
AR(1)	0.846976	0.072723	11.64665	0.0000
Inverted AR Roots	.85			

Остатки обнаруживают явную автокоррелированность : Р-значение критерия Бройша – Годфри при альтернативе AR(1) равно 0.0000. Переоценивание с включением в модель запаздывания на два квартала, дает:

Dependent Variable: GNP

Method: Least Squares

Sample(adjusted): 1947:3 1961:4

Included observations: 58 after adjusting endpoints

Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
	t			
C	217.7399	5.054473	43.07865	0.0000
T	5.221538	0.140436	37.18089	0.0000
AR(1)	1.380274	0.109452	12.61078	0.0000
AR(2)	-0.630066	0.109453	-5.756490	0.0000
Inverted AR Roots	.69 -.39i .69+.39i			

Модули комплексных чисел, обратных авторегрессионным корням, равны 0.7926, что говорит в пользу стационарности детрендрованного ряда

$$X_t^0 = X_t - \mu - \gamma t.$$

К построению модели для ряда GNP можно подойти и иначе. Сначала произвести детрендрование ряда, оценивая модель

$$X_t = \mu + \gamma t + u_t:$$

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	218.4825	2.640153	82.75373	0.0000
T	5.181995	0.075274	68.84144	0.0000
Durbin-Watson stat	0.316211	Prob(F-statistic)		0.000000

¹ Приводимые в таблицах оценки константы (C) и коэффициента при переменной t (T) соответствуют оценкам μ и γ в представлении

$$(X_t - \mu - \gamma t) = a_1(X_{t-1} - \mu - \gamma(t-1)) + a_2(X_{t-2} - \mu - \gamma(t-2)) + u_t$$

($a_2 = 0$ для первой из двух таблиц). Эти оценки получаются применением нелинейного метода наименьших квадратов. При этом обозначение AR(1) указывает на оценку для a_1 , а AR(2) – на оценку для a_2 .

Остатки, полученные при оценивании этой модели, образуют оцененный детрендрованный ряд, коррелограмма которого

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
. *****	. *****	1	0.836	0.836	44.028	0.000
. ****	**** .	2	0.531	-0.554	62.115	0.000
. *.	** .	3	0.183	-0.210	64.294	0.000
* .	. .	4	-0.100	0.044	64.960	0.000
** .	. .	5	-0.272	-0.004	69.949	0.000
*** .	* .	6	-0.339	-0.082	77.846	0.000
*** .	* .	7	-0.350	-0.169	86.446	0.000
*** .	* .	8	-0.332	-0.072	94.332	0.000
** .	. .	9	-0.281	0.058	100.07	0.000
** .	* .	10	-0.234	-0.177	104.16	0.000
** .	*** .	11	-0.234	-0.321	108.32	0.000
** .	. *.	12	-0.226	0.103	112.26	0.000

позволяет идентифицировать этот ряд как AR(2). После этого можно строить AR(2) модель для (оцененного) детрендрованного ряда

$$X_{t_detrended} = X_t - 218.4825 - 5.181995 t :$$

Variable	Coeff.	Std. Error	t-Statistic	Prob.
AR(1)	1.379966	0.107605	12.82435	0.0000
AR(2)	-0.630426	0.107605	-5.858722	0.0000

Объединяя результаты последних двух оцениваний, получаем оцененную модель $X_t - 218.4825 - 5.181995 t =$

$$= 1.379966 (X_{t-1} - 218.4825 - 5.181995(t-1)) - \\ - 0.630426 (X_{t-2} - 218.4825 - 5.181995(t-2)),$$

или

$$X_t = [(1 - 1.379966 + 0.630426) \cdot 218.4825 \\ + 1.379966 \cdot 5.181995 - 0.630426 \cdot 5.181995 \cdot 2] \\ + (1 - 1.379966 + 0.630426) \cdot 5.181995 t \\ + 1.379966 X_{t-1} - 0.630426 X_{t-2} + e_t \\ = 55.338375 + 1.297882 t + 1.379966 X_{t-1} - 0.630426 X_{t-2} + e_t .$$

В то же время, по приведенным результатам оценивания модели

$$X_t = \alpha + \beta t + a_1 X_{t-1} + a_2 X_{t-2} + u_t$$

получаем

$$X_t - 217.7399 - 5.221538 t = \\ = 1.380274 (X_{t-1} - 217.7399 - 5.221538(t-1)) \\ - 0.630066 (X_{t-2} - 217.7399 - 5.221538(t-2)),$$

или

$$X_t = [(1 - 1.380274 + 0.630066) \cdot 217.7399 \\ + 1.380274 \cdot 5.221538 - 0.630066 \cdot 5.221538 \cdot 2] \\ + (1 - 1.380274 + 0.630066) \cdot 5.221538 t \\ + 1.380274 X_{t-1} - 0.630066 X_{t-2} + e_t \\ = 55.017011 + 1.304298 t + 1.380274 X_{t-1} - 0.630066 X_{t-2} + e_t ,$$

так что результаты, полученные при использовании обоих вариантов построения модели, практически совпадают. Поэтому можно использовать смешанный вариант – использовать детрендированный ряд для идентификации порядка модели, а оценивать идентифицированную модель вместе с включенным в нее трендом, в данном случае – оценивать модель $X_t = \alpha + \beta t + a_1 X_{t-1} + a_2 X_{t-2} + u_t$.

Заметим, что диагностика рядов остатков в обеих оцененных моделях говорит в пользу их адекватности.

Перейдем теперь к ряду NONDURABLE. Коррелограмма детрендированного ряда (ряда остатков от оцененной модели регрессии X_t на константу и t)

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
. *****	. *****	1	0.793	0.793	32.083	0.000
. *****	. .	2	0.632	0.011	52.942	0.000
. ***	** . .	3	0.432	-0.195	62.887	0.000
. **	** . .	4	0.219	-0.193	65.515	0.000
. *	. .	5	0.090	0.062	65.965	0.000
* . .	* . .	6	-0.067	-0.152	66.218	0.000
** . .	** . .	7	-0.242	-0.277	69.647	0.000
*** . .	* . .	8	-0.362	-0.084	77.505	0.000
**** . .	** . .	9	-0.510	-0.211	93.500	0.000

обнаруживает резко выделяющийся пик на лаге 1, так что можно попробовать оценить модель $X_t = \alpha + \beta t + a_1 X_{t-1} + u_t$. Это дает следующие результаты:

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	47962.75	2862.678	16.75451	0.0000
T	315.1909	76.44770	4.122961	0.0002
AR(1)	0.884803	0.080824	10.94727	0.0000

Наблюдаемые Р-значения статистик Лjung – Бокса для ряда остатков превышают 0.05 при всех выборах M от 1 до 20. Проверка на отсутствие автокоррелированности по критерию Бройша – Годфри дает Р-значения, большие 0.05, как при AR(1) альтернативе, так и при альтернативах AR(2), AR(3) и т.д. Наконец, при проверке нормальности Р-значение статистики Jarque – Bera равно 0.648, так что по совокупности этих результатов мы могли бы говорить о пригодности оцененной модели

$$X_t - 47962.75 - 315.1909 t = \\ = 0.884803 (X_{t-1} - 47962.75 - 315.1909 (t-1)) + e_t.$$

Обратим, однако, внимание на следующее обстоятельство. Оцененное значение 0.884803 коэффициента при X_{t-1} достаточно близко к единице, и если ориентироваться на вычисленное значение 0.080824 стандартной ошибки для этого коэффициента, то при допущении отклонений от оцененного значения в пределах двух стандартных ошибок, в интервал допустимых значений $0.884803 \pm 2 \cdot 0.080824$ попадают и значения, большие или равные 1. Но последние соответствуют нестационарному процессу авторегрессии.

Таким образом, несмотря на то, что при полученной точечной оценке 0.884803 коэффициента при X_{t-1} построенная модель формально оказывается стационарной относительно детерминированного линейного тренда (т.е. детрендированный процесс следует стационарной AR(1) модели), мы не можем с достаточной степенью

уверенности гарантировать, что истинная модель порождения наблюдений также стационарна относительно линейного тренда.

Между тем вопрос о стационарности или нестационарности модели, порождающей наблюдаемый ряд, привлекает к себе постоянное внимание уже в течение нескольких десятков лет. Особенно это внимание усилилось после серии работ 80-х годов 20 века, в которых было введено понятие коинтеграции. С помощью этого понятия была обоснована методика построения “моделей коррекции ошибок”, в рамках которых удастся моделировать наличие долговременных связей между переменными вместе с указанием краткосрочной динамики, обеспечивающей поддержание этих долговременных связей.

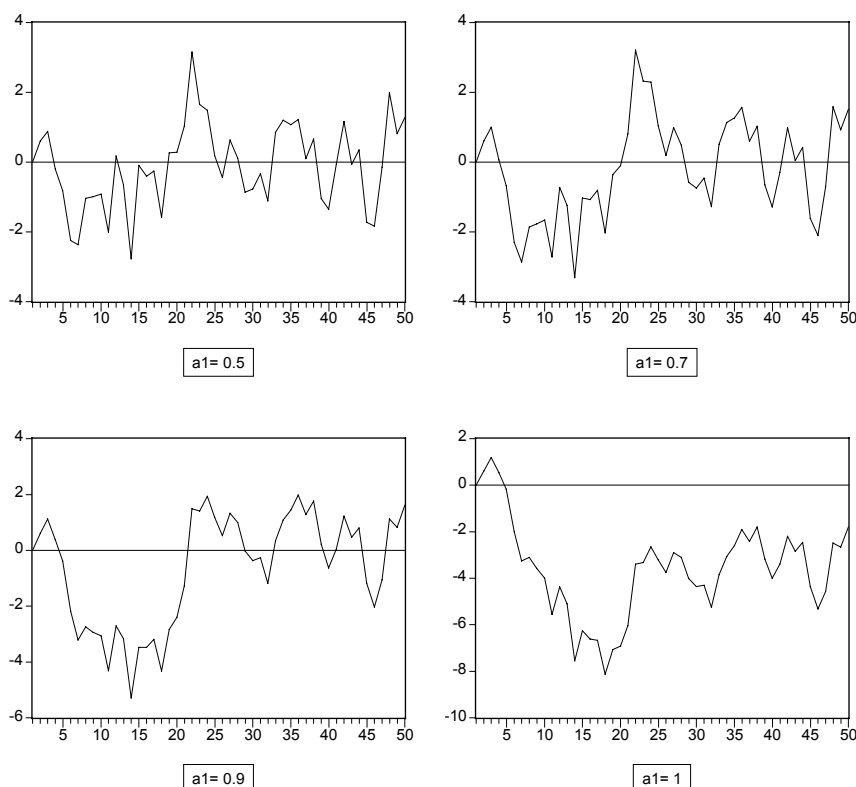
В последующем изложении мы рассмотрим вопросы, связанные с методами различения стационарных (стационарных относительно детерминированного тренда) и нестационарных рядов в рамках ARMA моделей, а также вопросы построения моделей связи между временными рядами.

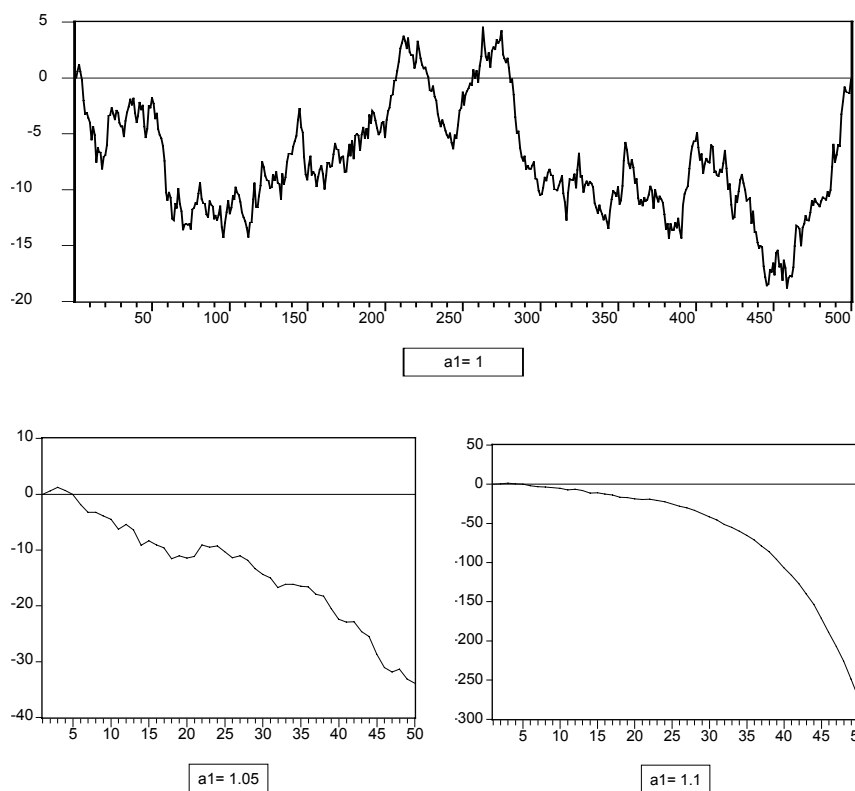
5.1. Нестационарные ARMA модели

Начнем рассмотрение с наиболее простой модели – процесса AR(1)

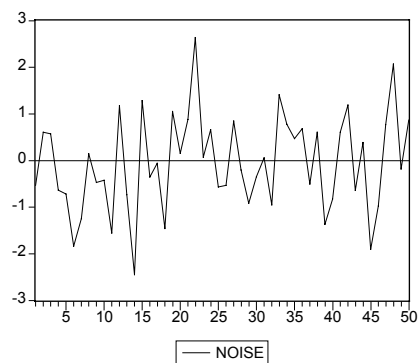
$$X_t = a_1 X_{t-1} + \varepsilon_t.$$

Мы уже знаем, что такой процесс является стационарным при выполнении условия $-1 < a_1 < 1$. А как проявляется нестационарность ряда X_t при нарушении этого условия? Приведем смоделированные реализации такого ряда при $a_1 = 0.5$, $a_1 = 0.7$, $a_1 = 0.9$, $a_1 = 1$, $a_1 = 1.05$, $a_1 = 1.1$.





Во всех случаях в качестве начального значения X_1 взято $X_1 = 0$ и использовалась одна и та же последовательность значений $\varepsilon_1, \dots, \varepsilon_T$, имитирующая гауссовский белый шум с дисперсией, равной единице:



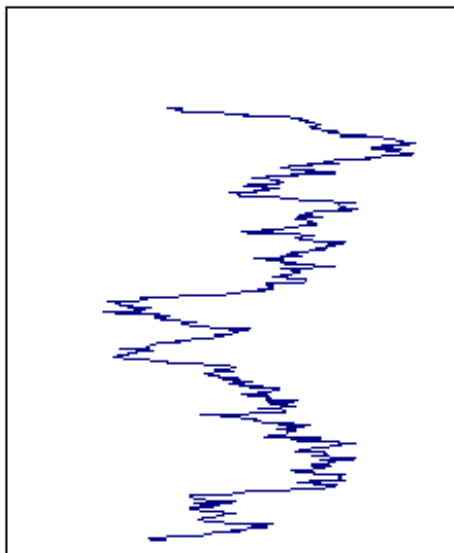
Однако поведение смоделированных рядов оказалось качественно различным. Полезно проследить, как изменяется характер траектории ряда с возрастанием значений коэффициента a_1 от $a_1 = 0.5$ до $a_1 = 1.1$. Заметим при этом, что в порождающих моделях математические ожидания X_t равны нулю.

Модель		Кол-во пересечений нулевого уровня	Среднее значение
Noise шум)	(белый	25	– 0.046
AR(1) 0.5	$a_1 =$	14	– 0.097
AR(1) 0.7	$a_1 =$	8	– 0.191
AR(1) 0.9	$a_1 =$	8	– 0.649
AR(1) 1.0	$a_1 =$	1	– 3.582
AR(1) 1.05	$a_1 =$	1	– 13.511
AR(1) 1.1	$a_1 =$	1	– 59.621

При возрастании значения a_1 от $a_1 = 0$ (белый шум) до $a_1 = 1$ количество пересечений нулевого уровня уменьшается, все более длинными становятся периоды, в течение которых значения ряда находятся по одну сторону от нулевого уровня. Расширенный график ряда при $a_1 = 1$, продленный до 500 наблюдений, иллюстрирует характерное свойство реализаций процесса

$$X_t = X_{t-1} + \varepsilon_t,$$

состоящее в том, что такой процесс, начавшись в момент $t = 1$ со значения $X_t = x_1$ (в данном случае $x_1 = 0$), в дальнейшем очень редко пересекает уровень x_1 (“возвращается к уровню x_1 ”) и, находясь в течение длительного времени по одну сторону от этого уровня (выше или ниже), может удаляться от этого уровня на значительные расстояния.



“Повернутая вертикально”, траектория ряда напоминает траекторию движения сильно нетрезвого человека, пытающегося продвигаться вперед по прямой, но не имеющего возможности успешно выдерживать нужное направление. И это служит некоторым оправданием термина, используемого для AR(1) процесса с $a_1 = 1$:

$X_t = X_{t-1} + \varepsilon_t$ – случайное блуждание (процесс случайного блуждания – *random walk*).

Далее мы рассмотрим этот процесс подробнее, а сейчас обратим внимание на поведение реализаций процесса

$$X_t = a_1 X_{t-1} + \varepsilon_t$$

при $a_1 = 1.05$ и $a_1 = 1.1$. Обе реализации иллюстрируют “взрывной” (“*explosive*”) характер поведения AR(1) процесса при $a_1 > 0$: траектории процесса очень быстро удаляются от начального уровня на все возрастающие расстояния. В связи с этим “взрывные” модели непригодны для описания поведения макроэкономических рядов на сколь-нибудь протяженных интервалах времени.

Пониманию столь различного поведения реализаций AR(1) процесса помогает представление модели в виде

$$X_t - X_{t-1} = a_1 X_{t-1} - X_{t-1} + \varepsilon_t = (a_1 - 1)X_{t-1} + \varepsilon_t,$$

или

$$\Delta X_t = \varphi X_{t-1} + \varepsilon_t,$$

где

$$\Delta X_t = X_t - X_{t-1}, \quad \varphi = a_1 - 1.$$

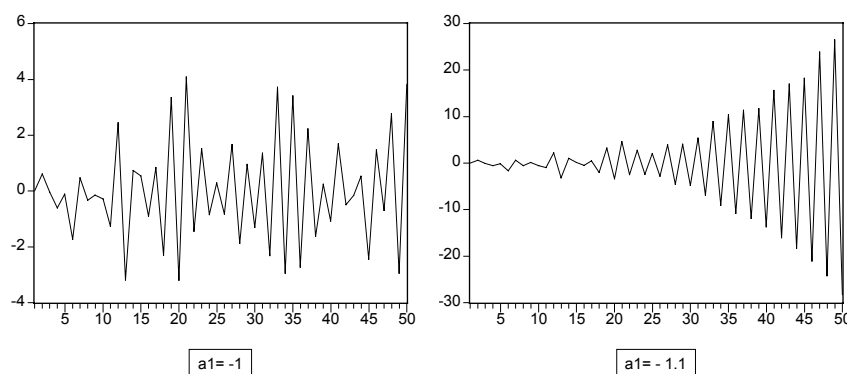
При $a_1 = 1$ имеем $\varphi = a_1 - 1 = 0$, и приращения ΔX_t ряда X_t образуют процесс белого шума, так что условное математическое ожидание ΔX_t при фиксированном (наблюдаемом) значении $X_{t-1} = x_{t-1}$ не зависит от x_{t-1} и равно 0. Соответственно, при фиксированном (наблюдаемом) значении $X_{t-1} = x_{t-1}$, условное математическое ожидание случайной величины $X_t = \Delta X_t + X_{t-1}$ равно x_{t-1} . Если распределение случайной величины ε_t симметрично относительно нуля (а именно таково и гауссовское распределение, которое использовалось нами при моделировании), то

наблюдаемое значение $X_t = x_t$ может с равным успехом оказаться как больше, так и меньше x_{t-1} . Именно это и определяет “блуждающий” характер траектории ряда.

При $a_1 > 1$ имеем $\varphi = a_1 - 1 > 0$, и условное математическое ожидание ΔX_t при фиксированном (наблюдаемом) значении $X_{t-1} = x_{t-1}$, равное $E(\Delta X_t | X_{t-1} = x_{t-1}) = \varphi x_{t-1}$, имеет знак, совпадающий со знаком x_{t-1} . Таким образом, если $x_{t-1} > 0$, то ожидаемое значение следующего наблюдения $X_t = x_t$ больше значения x_{t-1} , а если $x_{t-1} < 0$, то ожидаемое значение следующего наблюдения $X_t = x_t$ меньше значения x_{t-1} . Наличие такого механизма приводит к быстрому и прогрессирующему удалению траектории процесса от начального уровня, что и наблюдалось нами для реализаций AR(1) модели при $a_1 = 1.05$ и $a_1 = 1.1$.

Наконец, при $0 < a_1 < 1$ имеем $\varphi = a_1 - 1 < 0$, и условное математическое ожидание ΔX_t при фиксированном (наблюдаемом) значении $X_{t-1} = x_{t-1}$, равное $E(\Delta X_t | X_{t-1} = x_{t-1}) = \varphi x_{t-1}$, имеет знак, противоположный знаку x_{t-1} . Таким образом, если $x_{t-1} > 0$, то ожидаемое значение следующего наблюдения $X_t = x_t$ меньше значения x_{t-1} , а если $x_{t-1} < 0$, то ожидаемое значение следующего наблюдения $X_t = x_t$ больше значения x_{t-1} . Наличие такого механизма обеспечивает удержание траектории процесса в относительной близости от уровня, равного безусловному математическому ожиданию $E(X_t) = \mu$ ряда (в данном случае, $\mu = 0$), и достаточно частое пересечение траекторией ряда этого уровня.

Мы ограничились здесь рассмотрением ситуаций с $a_1 > 0$, поскольку они наиболее типичны для экономических временных рядов. Для полноты приведем также и смоделированные реализации процесса $X_t = a_1 X_{t-1} + \varepsilon_t$ при $a_1 = -1$ и $a_1 = -1.1$.



Обратимся теперь к процессу случайного блуждания

$$X_t = X_{t-1} + \varepsilon_t, \quad t = 1, \dots, T,$$

со стартовым значением $X_0 = x_0$. Мы можем представить X_t в виде

$$\begin{aligned} X_t &= X_{t-1} + \varepsilon_t = (X_{t-2} + \varepsilon_{t-1}) + \varepsilon_t = X_{t-2} + \varepsilon_{t-1} + \varepsilon_t = (X_{t-3} + \varepsilon_{t-2}) + \varepsilon_{t-1} + \varepsilon_t = \\ &= X_{t-3} + \varepsilon_{t-2} + \varepsilon_{t-1} + \varepsilon_t = \dots = X_0 + (\varepsilon_1 + \dots + \varepsilon_t), \end{aligned}$$

$$X_t = X_0 + \sum_{j=1}^t \varepsilon_j.$$

Отсюда сразу получаем:

$$E(X_t | X_0 = x_0) = x_0 ,$$

$$D(X_t | X_0 = x_0) = D(\varepsilon_1 + \dots + \varepsilon_t) = D(\varepsilon_1) + \dots + D(\varepsilon_t) = tD(\varepsilon_1) = t\sigma_\varepsilon^2 .$$

Далее,

$$\begin{aligned} Cov(X_t, X_{t-1} | X_0 = x_0) &= E[(X_t - x_0)(X_{t-1} - x_0) | X_0 = x_0] = \\ &= E[(\varepsilon_1 + \dots + \varepsilon_t)(\varepsilon_1 + \dots + \varepsilon_{t-1})] = (t-1) \sigma_\varepsilon^2 \end{aligned}$$

не зависит от значения x_0 , так что

$$\begin{aligned} Corr(X_t, X_{t-1}) &= \frac{(t-1)\sigma_\varepsilon^2}{\sqrt{D(X_t)}\sqrt{D(X_{t-1})}} = \frac{(t-1)\sigma_\varepsilon^2}{\sqrt{\sigma_\varepsilon^2 t}\sqrt{\sigma_\varepsilon^2 (t-1)}} = \\ &= \frac{\sqrt{t-1}}{\sqrt{t}} = \sqrt{1 - \frac{1}{t}} . \end{aligned}$$

Отсюда находим:

t	$Corr(X_t, X_{t-1})$
1	0
2	0.707
3	0.806
4	0.866
5	0.894
6	0.913
7	0.925
8	0.935
9	0.943
10	0.949

т.е. соседние значения X_t и X_{t-1} очень сильно коррелированы, притом положительно и тем более сильно, чем больше t . И это приводит к уже наблюдавшемуся нами характеру поведения траекторий случайного блуждания. На первых нескольких шагах траектория как бы “определяется”, где она будет находиться затем в течение довольно длительного периода – выше или ниже начального уровня x_0 . Так что если после нескольких первых шагов траектория случайного блуждания оказалась ниже уровня x_0 (как это было у смоделированной нами реализации), то она может оставаться там в течение весьма продолжительного времени. Если смоделировать очень длинную реализацию случайного блуждания, то она будет состоять из чередующихся длинных участков, на которых функция находится, соответственно, выше или ниже уровня x_0 .

При $X_0 = 0$ получаем

$$X_t = \sum_{j=1}^t \varepsilon_j , \quad t = 1, \dots, T .$$

Рассматривая последний ряд сам по себе (не связывая его со стартовым значением), имеем:

$$E(X_t) = 0, \quad D(X_t) = t\sigma_\varepsilon^2,$$

так что этот ряд – нестационарный.

Этот ряд является моделью *стохастического тренда*, который обнаруживается во многих экономических временных рядах и должен обязательно приниматься во внимание при построении регрессионных моделей связи между двумя или несколькими рядами, имеющими стохастический тренд.

Поясним фундаментальное различие между временными рядами, имеющими только детерминированный тренд, и рядами, которые (возможно, наряду с детерминированным) имеют стохастический тренд.

Для этого рассмотрим следующие две простые модели нестационарных рядов. В первой пусть

$$X_t = \alpha + \beta t + \varepsilon_t, \quad t = 1, \dots, T,$$

т.е. на детерминированный линейный тренд накладываются случайные ошибки в виде белого шума. А вторая пусть представляет *случайное блуждание со сносом*, т.е. процесс

$$X_t = \alpha + X_{t-1} + \varepsilon_t, \quad t = 1, \dots, T, \quad X_0 = x_0,$$

приращения которого имеют ненулевое математическое ожидание

$$E(\Delta X_t) = \alpha \neq 0.$$

Процесс X_t во второй модели можно представить в виде

$$\begin{aligned} X_t &= \alpha + X_{t-1} + \varepsilon_t = \alpha + (\alpha + X_{t-2} + \varepsilon_{t-1}) + \varepsilon_t = 2\alpha + X_{t-2} + \varepsilon_{t-1} + \varepsilon_t = \\ &= 3\alpha + X_{t-3} + \varepsilon_{t-2} + \varepsilon_{t-1} + \varepsilon_t = \dots = x_0 + \alpha t + (\varepsilon_1 + \dots + \varepsilon_t), \end{aligned}$$

$$X_t = x_0 + \alpha t + \sum_{j=1}^t \varepsilon_j,$$

так что ряд X_t имеет и детерминированный и стохастический тренды.

Детрендирование первого ряда приводит к ряду

$$X_t^0 = X_t - (\alpha + \beta t) = \varepsilon_t,$$

который является стационарным. Детрендирование второго ряда приводит к ряду

$$X_t^0 = X_t - (x_0 + \alpha t) = \sum_{j=1}^t \varepsilon_j,$$

который не является стационарным.

Пытаться остационарить ряд можно и другим способом. Именно, можно перейти от ряда уровней X_t к ряду разностей

$$\Delta X_t = X_t - X_{t-1},$$

(такой переход в теории временных рядов называют *дифференцированием*).

При таком переходе получаем для первого ряда

$$\Delta X_t = X_t - X_{t-1} = (\alpha + \beta t + \varepsilon_t) - (\alpha + \beta(t-1) + \varepsilon_{t-1}) = \beta + \varepsilon_t - \varepsilon_{t-1},$$

а для второго ряда

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

Оба продифференцированных ряда ΔX_t стационарны. Первый продифференцированный ряд относится к классу MA(1) и имеет математическое ожидание β . Второй продифференцированный ряд относится к классу MA(0) и имеет математическое ожидание α .

Таким образом, в отличие от детрендинга, операция дифференцирования приводит к стационарному ряду в обоих случаях. Однако в результате дифференцирования первого ряда получается процесс скользящего среднего, который не является обратимым. И это имеет некоторые нежелательные последствия при подборе модели по статистическим данным и использовании подобранной модели для целей прогнозирования будущих значений ряда. (См., например, [Hamilton (1994), главы 4 и 5].) В случае необратимости MA-составляющей продифференцированного ряда становится невозможным использование обычных алгоритмов идентификации модели, оценивания модели и диагностики оцененной модели, рассмотренных ранее в главе 3.

Обобщая рассмотренную ситуацию, рассмотрим ряды

$$X_t = \theta_0 + \theta_1 t + \theta_2 t^2 + \varepsilon_t$$

и

$$Y_t = \alpha + \beta t + \gamma t^2 + Z_t,$$

где Z_t - процесс, определяемый соотношениями

$$Z_t = \varepsilon_t + 2\varepsilon_{t-1} + 3\varepsilon_{t-2} + \dots + t\varepsilon_1, \quad t = 1, \dots, T.$$

Детрендинг первого ряда приводит к стационарному ряду

$$X_t^0 = X_t - (\theta_0 + \theta_1 t + \theta_2 t^2) = \varepsilon_t.$$

Детрендинг второго ряда приводит к ряду

$$Y_t^0 = Y_t - (\alpha + \beta t + \gamma t^2) = Z_t,$$

у которого

$$D(Z_t) = D(\varepsilon_t + 2\varepsilon_{t-1} + 3\varepsilon_{t-2} + \dots + t\varepsilon_1) = \sigma_\varepsilon^2(1 + 2 + \dots + t) = \sigma_\varepsilon^2 t(t+1)/2,$$

так что детрендинговый ряд не является стационарным.

Если вместо вычитания линейного тренда произвести дифференцирование рядов, то для ряда X_t это приводит к ряду

$$\Delta X_t = \theta_1 - \theta_2 + 2\theta_2 t + \varepsilon_t - \varepsilon_{t-1},$$

стационарному относительно линейного тренда: ряд

$$\Delta X_t - (\theta_1 - \theta_2 + 2\theta_2 t) = \varepsilon_t - \varepsilon_{t-1}$$

является стационарным, но необратимым МА(1) рядом. Если же продифференцировать ряд Y_t , то в этом случае продифференцированный ряд

$$\begin{aligned} \Delta Y_t &= (\beta - \gamma + 2\gamma t) + Z_t - Z_{t-1} = \\ &= \beta - \gamma + 2\gamma t + (\varepsilon_t + 2\varepsilon_{t-1} + 3\varepsilon_{t-2} + \dots + t\varepsilon_1) \\ &\quad - (\varepsilon_{t-1} + 2\varepsilon_{t-2} + 3\varepsilon_{t-3} + \dots + (t-1)\varepsilon_1) = \\ &= \beta - \gamma + 2\gamma t + (\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_{t-1} + \varepsilon_t) \end{aligned}$$

уже не является стационарным относительно линейного тренда: ряд

$$\Delta Y_t - (\beta - \gamma + 2\gamma t) = (\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_{t-1} + \varepsilon_t)$$

Нестационарен и представляет собой стохастический тренд.

С другой стороны, осуществляя двукратное дифференцирование ряда Y_t , т.е. переходя к ряду $\Delta^2 Y_t$, где $\Delta^2 = (1 - L)^2 = 1 - 2L + L^2$, получаем стационарный МА(0) ряд

$$\begin{aligned} \Delta^2 Y_t &= 2\gamma + \Delta^2 Z_t = 2\gamma + Z_t - 2Z_{t-1} + Z_{t-2} = \\ &= 2\gamma + (\varepsilon_t + 2\varepsilon_{t-1} + 3\varepsilon_{t-2} + \dots + t\varepsilon_1) \\ &\quad - 2(\varepsilon_{t-1} + 2\varepsilon_{t-2} + 3\varepsilon_{t-3} + \dots + (t-1)\varepsilon_1) + \\ &\quad + (\varepsilon_{t-2} + 2\varepsilon_{t-3} + 3\varepsilon_{t-4} + \dots + (t-2)\varepsilon_1) = 2\gamma + \varepsilon_t. \end{aligned}$$

При двукратном дифференцировании ряда X_t , приходим к ряду

$$\begin{aligned} \Delta^2 X_t &= \Delta(\Delta X_t) = \Delta(X_t - X_{t-1}) = (X_t - X_{t-1}) - (X_{t-1} - X_{t-2}) = \\ &= X_t - 2X_{t-1} + X_{t-2} = 2\theta_2 + \varepsilon_t - 2\varepsilon_{t-1} + \varepsilon_{t-2}, \end{aligned}$$

который представляет собой стационарный процесс скользящего среднего МА(2) с математическим ожиданием $2\theta_2$, не удовлетворяющий условию обратимости. Действительно, уравнение $b(z) = 0$ здесь принимает вид

$$1 - 2z + z^2 = 0$$

и имеет двойной корень $z = 1$.

Таким образом, двукратное дифференцирование “остационаризует” и ряд X_t и ряд Y_t , но в случае ряда X_t результирующий ряд не обладает свойством обратимости.

Обобщение подобных примеров приводит к следующим понятиям.

Временной ряд X_t называется **стационарным относительно детерминированного тренда $f(t)$** , если ряд $X_t - f(t)$ стационарный. Если ряд X_t стационарен относительно некоторого детерминированного тренда, то говорят, что этот ряд принадлежит **классу рядов, стационарных относительно детерминированного тренда**, или что он является **TS рядом (TS – time stationary)**.

В класс TS рядов включаются также стационарные ряды, не имеющие детерминированного тренда.

Временной ряд X_t называется **интегрированным порядка k** , $k = 1, 2, \dots$, если

- ряд X_t не является стационарным или стационарным относительно детерминированного тренда, т.е. не является TS рядом;
- ряд $\Delta^k X_t$, полученный в результате k -кратного дифференцирования ряда X_t , является стационарным рядом;
- ряд $\Delta^{k-1} X_t$, полученный в результате $(k-1)$ -кратного дифференцирования ряда X_t , не является TS рядом.

Если полагать $\Delta^0 X_t = X_t$, то при $k = 1$ третье условие дублирует первое.

Для интегрированного ряда порядка k используют обозначение $\mathbf{I}(k)$. Если ряд X_t является интегрированным порядка k , то мы будем обозначать это для краткости как $X_t \sim \mathbf{I}(k)$. В этой системе обозначений соотношение $X_t \sim \mathbf{I}(0)$ соответствует ряду, который является стационарным и при этом не является результатом дифференцирования TS ряда.

Пусть TS ряд имеет вид $X_t = \alpha + \beta t + Y_t$, где Y_t – стационарный ряд, имеющий нулевое математическое ожидание. Тогда X_t можно представить в виде

$$X_t = \alpha + \beta t + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \quad \psi_0 = 1, \quad \sum_{j=0}^{\infty} \psi_j^2 < \infty,$$

где ε_t – процесс белого шума. (Мы не затрагиваем здесь теоретическую возможность наличия в правой части еще и так называемой линейно детерминированной стохастической компоненты.) Переходя к ряду разностей, получаем:

$$\begin{aligned} \Delta X_t &= \Delta f(t) + \sum_{j=0}^{\infty} \psi_j (\varepsilon_{t-j} - \varepsilon_{t-1-j}) = \\ &= \beta + \sum_{j=0}^{\infty} b_j \varepsilon_{t-j} = \beta + b(L) \varepsilon_t, \end{aligned}$$

где

$$b(L) = \sum_{j=0}^{\infty} b_j L^j, \quad b_0 = 1, \quad b_j = \psi_j - \psi_{j-1}, \quad j = 1, 2, \dots,$$

Отсюда вытекает, что

$$b(1) = \sum_{j=0}^{\infty} b_j = 0,$$

т.е. уравнение $b(z) = 0$ имеет единичный корень.

Таким образом, если для некоторого стационарного ряда Z_t ,

$$Z_t = \mu + b(L) \varepsilon_t, \text{ где } b(L) = \sum_{j=0}^{\infty} b_j L^j, \quad b_0 = 1,$$

выполнено соотношение

$$b(1) = 0,$$

то последнее указывает на “передифференцированность” ряда Z_t : этот TS ряд получен в результате дифференцирования некоторого TS ряда.

Совокупность интегрированных рядов различных порядков $k = 1, 2, \dots$ образует класс **разностно стационарных**, или **DS рядов** (DS – difference stationary). Если некоторый ряд X_t принадлежит этому классу, то мы говорим о нем как о **DS ряде**.

Пусть ряд X_t – интегрированный порядка k . Подвергнем этот ряд k -кратному дифференцированию. Если в результате получается стационарный ряд типа $ARMA(p, q)$, то говорят, что исходный ряд X_t **является рядом типа $ARIMA(p, k, q)$** , или **k раз проинтегрированным $ARMA(p, q)$ рядом ($ARIMA$ – *autoregressive integrated moving average*)**. Если при этом $p = 0$ или $q = 0$, то тогда употребляются и более короткие обозначения:

$$ARIMA(p, k, 0) = ARI(p, k), \quad ARIMA(0, k, q) = IMA(k, q),$$

$$ARIMA(0, k, 0) = ARI(0, k) = IMA(k, 0).$$

Возвращаясь к только что рассмотренным примерам, получаем:

$$X_t = \alpha + \beta t + \varepsilon_t \sim I(0);$$

$$X_t = \alpha + X_{t-1} + \varepsilon_t \sim I(1), \quad X_t \text{ – ряд типа } ARIMA(0, 1, 0);$$

$$X_t = \theta_0 + \theta_1 t + \theta_2 t^2 + \varepsilon_t \sim I(0);$$

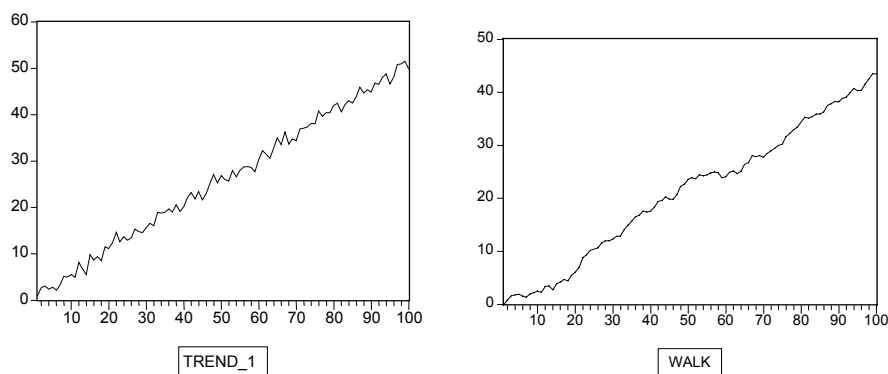
$$X_t = \alpha + \beta t + \gamma t^2 + \varepsilon_t + 2\varepsilon_{t-1} + 3\varepsilon_{t-2} + \dots + t\varepsilon_1 \sim I(2), \quad X_t \text{ – ряд типа } ARIMA(0, 2, 0).$$

Первый и третий из этих рядов являются TS рядами, а второй и четвертый – DS рядами.

Используя ту же самую имитацию белого шума, что и в предыдущих примерах, получаем смоделированные реализации для двух первых процессов

$$TREND_1_t = 1 + 0.5 t + \varepsilon_t, \quad \varepsilon_t \sim N(0, 1),$$

$$WALK_t = 0.5 + WALK_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, 0.5^2), \quad WALK_0 = 0:$$



Оценим для каждой из этих реализаций модель прямолинейной зависимости:

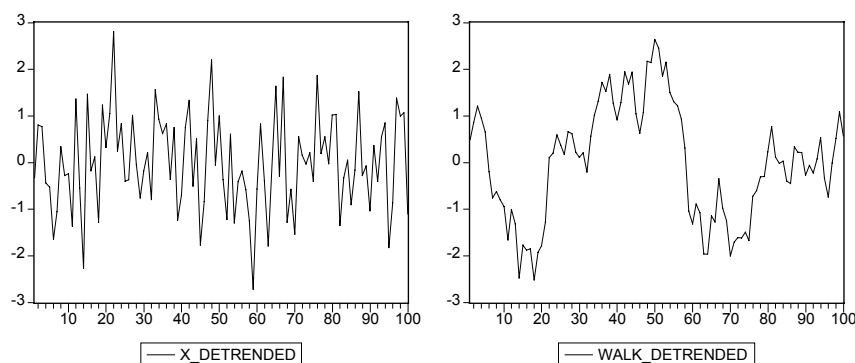
Dependent Variable: TREND_1

Variable	Coef.	Std. Error	t-Statistic	Prob.
C	0.796390	0.208085	3.827226	0.0002
T	0.501522	0.003577	140.1946	0.0000

Dependent Variable: WALK

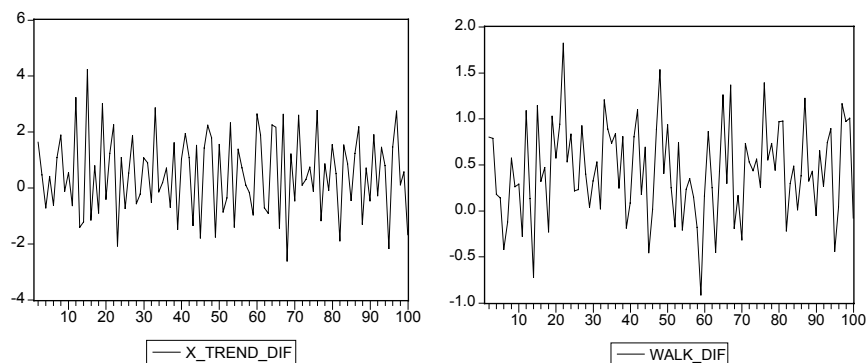
Variable	Coef.	Std. Error	t-Statistic	Prob.
C	-0.930832	0.249804	-3.726248	0.0003
T	0.437818	0.004295	101.9477	0.0000

Полученные при этом детрендированные ряды ведут себя следующим образом:



Поведение первого графика характерно для стационарного ряда, а поведение второго графика – для стохастического тренда. Отметим наличие видимой цикличности с длинным периодом у второго графика. На эту особенность было указано в работах [Chan, Nanyu, Ord (1977)] и [Nelson, Kang (1981)]: в результате очистки ряда от детерминированного тренда могут возникать систематические колебания – длиннопериодные циклы, которых не было у исходного ряда (*ложная периодичность*) и которые могут быть ошибочно истолкованы как проявление некоторого экономического цикла.

В то же время, первые разности реализаций исходных рядов ведут себя следующим образом:



Коррелограммы рядов первых разностей имеют следующий вид.

Для первого ряда (X_TREND_DIF)

ACF	PACF	AC	PAC	Q-Stat	Prob
*** .	*** .	1	-0.449	-0.449	20.527 0.000
. . .	** . .	2	-0.045	-0.308	20.736 0.000
. . .	** . .	3	-0.006	-0.236	20.740 0.000
. . .	** . .	4	-0.052	-0.266	21.021 0.000
. * .	* . .	5	0.078	-0.157	21.676 0.001
* . .	** . .	6	-0.074	-0.221	22.258 0.001
. * .	. * .	7	0.194	0.073	26.358 0.000
** . .	* . .	8	-0.216	-0.120	31.473 0.000
. *	9	0.079	-0.047	32.159 0.000
. . .	* . .	10	-0.056	-0.143	32.513 0.000

Для второго ряда (WALK_DIF)

ACF	PACF	AC	PAC	Q-Stat	Prob
.	1	0.035	0.035	0.1271 0.721
.	2	-0.044	-0.045	0.3271 0.849
.	3	-0.042	-0.039	0.5099 0.917
.	4	-0.025	-0.024	0.5766 0.966
.	5	0.065	0.063	1.0215 0.961
.	6	-0.004	-0.012	1.0231 0.985
. * .	. * .	7	0.101	0.107	2.1405 0.952
* . .	* . .	8	-0.173	-0.181	5.4243 0.711
.	9	-0.041	-0.013	5.6087 0.778
. . .	* . .	10	-0.057	-0.073	5.9707 0.818

Вторая коррелограмма соответствует процессу белого шума. Что касается первой коррелограммы, то наличие единственного значимого пика у автокорреляционной функции говорит в пользу идентификации наблюдаемого ряда

разностей как реализации MA(1) процесса. Подставляя значение $r(1) = -0.449$ вместо $\rho(1)$ в формулу $\rho(1) = b_1/(1 + b_1^2)$, получаем уравнение $-0.449 = b_1/(1 + b_1^2)$. Это уравнение имеет два решения: -1.6036 и -0.6236 . Первое соответствует необратимому, а второе – обратимому MA(1) процессу. Выбирая обратимую версию, получаем в качестве предварительной оценки коэффициента b_1 значение -0.6236 . Уточнение этой оценки в процессе оценивания модели MA(1), приводит к следующему результату.

Dependent Variable: X_TREND_DIF
 Sample(adjusted): 2 100
 Included observations: 99 after adjusting endpoints
 Convergence achieved after 10 iterations
 Backcast: 1

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.501429	0.004483	111.8447	0.0000
MA(1)	-0.977743	0.015377	-63.58459	0.0000

– при использовании backcasting (процедуры обратного прогноза – см. разд. 3.2) – и

Dependent Variable: X_TREND_DIF
 Method: Least Squares
 Sample(adjusted): 2 100
 Included observations: 99 after adjusting endpoints
 Convergence achieved after 18 iterations
 Backcast: OFF

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.518807	0.007616	68.12224	0.0000
MA(1)	-1.062852	0.042816	-24.82378	0.0000

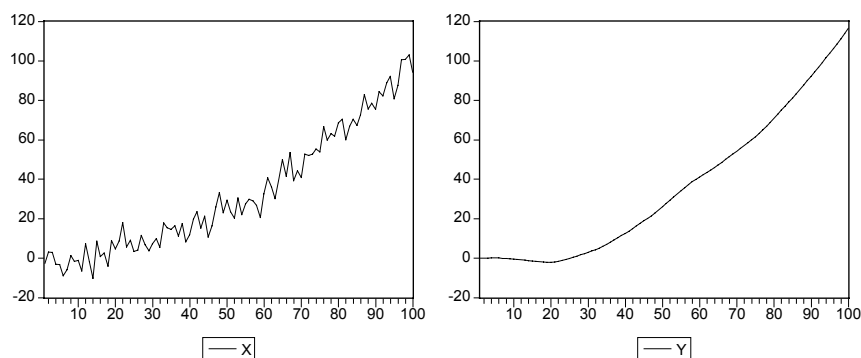
без использования backcasting .

Для обоих вариантов оценивания мы получаем в качестве оценок для коэффициента b_1 значения, очень близкие к 1, что отражает необратимость MA(1) модели, порождающей ряд. Обратим внимание на то, что в этом случае продифференцированный ряд оказывается автокоррелированным несмотря на то, что исходный ряд представляет собой сумму детерминированного линейного тренда и белого шума. Это явление известно как *эффект Служского* ([Slutsky (1937)]).

Смоделируем теперь реализации двух оставшихся рядов:

$$X_t = 0.01 t^2 + \varepsilon_t, \quad \varepsilon_t \sim N(0, 5^2),$$

$$Y_t = 0.04 t^2 + \varepsilon_t + 2\varepsilon_{t-1} + 3\varepsilon_{t-2} + \dots + t \varepsilon_1, \quad \varepsilon_t \sim N(0, 0.1^2):$$



Оценим для каждой из этих реализаций модель квадратичной зависимости:

Dependent Variable: X

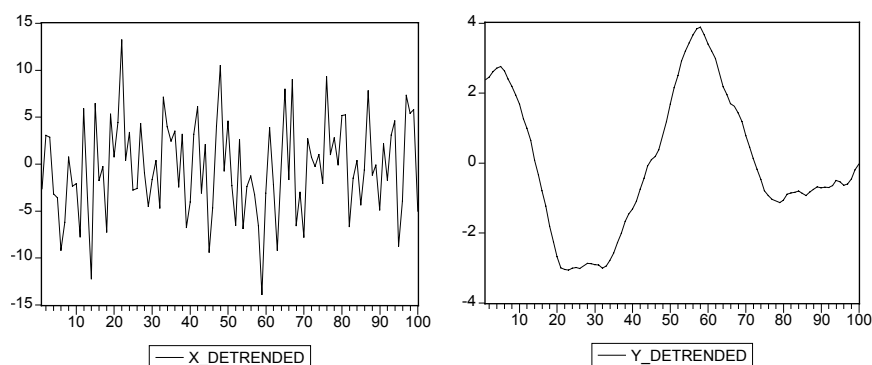
Variable	Coef.	Std. Error	t-Statistic	Prob.
T^2	0.009926	0.000114	86.93333	0.0000

(Константа и линейная составляющая статистически незначимы.)

Dependent Variable: Y

Variable	Coef.	Std. Error	t-Statistic	Prob.
C	-2.273387	0.621988	-3.655036	0.0004
T	-0.119781	0.028427	-4.213709	0.0001
T^2	0.013087	0.000273	47.99344	0.0000

Детрендированные реализации:



Коррелограмма ряда X_DETRENDED

ACF	PACF	AC	PAC	Q-Stat	Prob
.	1	0.030	0.030	0.0936 0.760
.	2	-0.049	-0.050	0.3429 0.842
.	3	-0.042	-0.039	0.5278 0.913
.	4	-0.024	-0.024	0.5867 0.965
. *	. *	5	0.071	0.068	1.1208 0.952
.	6	0.000	-0.008	1.1208 0.981

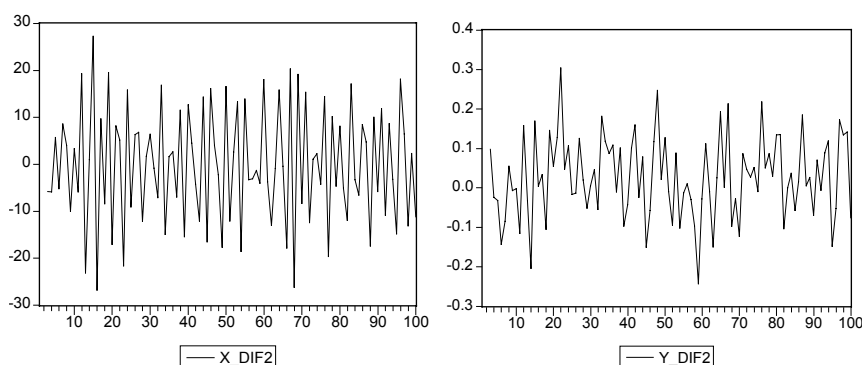
.	*	.	*	7	0.101	0.107	2.2429	0.945
*	.	*	.	8	-0.171	-0.177	5.4698	0.706
.	.	.	.	9	-0.039	-0.011	5.6366	0.776
.	.	*	.	10	-0.050	-0.068	5.9165	0.822

соответствует процессу белого шума, а коррелограмма ряда Y_DETRENDED

ACF	PACF	AC	PAC	Q-Stat	Prob
.	*****	.	*****	1	0.985 0.985 99.989 0.000
.	*****	****	.	2	0.956 -0.507 195.02 0.000
.	*****	**	.	3	0.912 -0.314 282.52 0.000
.	*****	**	.	4	0.857 -0.196 360.52 0.000
.	*****	*	.	5	0.791 -0.136 427.67 0.000
.	*****	*	.	6	0.716 -0.073 483.36 0.000
.	*****	.	.	7	0.635 -0.044 527.61 0.000
.	****	*	.	8	0.549 -0.067 560.97 0.000
.	***	.	.	9	0.458 -0.043 584.49 0.000
.	***	.	.	10	0.365 -0.051 599.58 0.000

характерна для нестационарного AR(2) процесса.

Вторые разности рядов X_t и Y_t ведут себя следующим образом:



Коррелограмма ряда X_DIF2

ACF	PACF		AC	PAC	Q-Stat	Prob
***** . .	***** . .	1	-0.635	-0.635	40.777	0.000
. *	**** . .	2	0.128	-0.463	42.440	0.000
. .	** . .	3	0.038	-0.294	42.586	0.000
* . .	** . .	4	-0.064	-0.273	43.014	0.000
. *	* . .	5	0.081	-0.141	43.710	0.000
* . .	** . .	6	-0.137	-0.292	45.710	0.000
. **	. .	7	0.243	0.037	52.065	0.000
**	8	-0.254	-0.037	59.064	0.000
. *	. .	9	0.160	0.057	61.891	0.000

* .	.	10	-0.099	-0.056	62.976	0.000
-----	---	----	--------	--------	--------	-------

отражает свойство необратимости МА модели. В этом случае значение $r(1) = -0.635$ даже выходит за пределы интервала возможных значений $\rho(1)$ процесса $MA(1)$, т.е. за пределы интервала $-0.5 \leq \rho(1) \leq 0.5$. (Подстановка значения -0.635 вместо $\rho(1)$ в формулу $\rho(1) = b_1/(1 + b_1^2)$ приводит к квадратному уравнению $0.635 b_1^2 + b_1 + 0.635 = 0$, которое не имеет действительных решений.)

Коррелограмма ряда Y_DIF2

ACF	PACF		AC	PAC	Q-Stat	Prob
. .	. .	1	0.031	0.031	0.0942	0.759
. .	. .	2	-0.041	-0.042	0.2636	0.877
. .	. .	3	-0.038	-0.036	0.4127	0.938
. .	. .	4	-0.013	-0.013	0.4313	0.980
. *	. *	5	0.073	0.071	0.9925	0.963
. .	. .	6	-0.006	-0.012	0.9958	0.986
. *	. *	7	0.104	0.111	2.1713	0.950
* .	* .	8	-0.172	-0.178	5.3769	0.717
. .	. .	9	-0.031	-0.006	5.4804	0.791
* .	* .	10	-0.066	-0.084	5.9673	0.818

соответствует процессу белого шума.

Рассмотренные примеры отражают общую ситуацию:

- вычитание детерминированной составляющей TS ряда приводит к стационарному ряду;
- вычитание детерминированной составляющей DS ряда приводит к DS ряду;
- дифференцирование TS ряда приводит к TS ряду; если стохастическая составляющая исходного TS ряда описывается стационарной моделью ARMA, то дифференцирование приводит к TS ряду с необратимой МА составляющей, имеющей единичный корень;
- k - кратное дифференцирование ряда $X_t \sim I(k)$ приводит к стационарному ряду; если стохастическая составляющая исходного $I(k)$ ряда описывается моделью ARIMA, то k - кратное дифференцирование приводит к стационарному ряду с обратимой МА составляющей.

Важным обстоятельством является также то, что в TS-рядах влияние предыдущих шоковых воздействий затухает с течением времени, а в DS-рядах такое затухание отсутствует и каждый отдельный шок влияет с одинаковой силой на все последующие значения ряда. Поясним это на примере простой AR(1) модели $X_t = a X_{t-1} + \varepsilon_t$. Для нее (см. разд. 2.3)

$$X_t = a^t X_0 + a^{t-1} \varepsilon_1 + a^{t-2} \varepsilon_2 + \dots + \varepsilon_t,$$

так что

$$X_{t+h} = a^{t+h} X_0 + a^{t+h-1} \varepsilon_1 + a^{t+h-2} \varepsilon_2 + \dots + a^h \varepsilon_t + \dots + \varepsilon_{h+t}.$$

Отсюда получаем значения импульсных мультипликаторов, показывающих влияние единовременного (импульсного) изменения (“шока”) инновации ε_t на текущее и последующие значения ряда:

$$\partial X_t / \partial \varepsilon_t = 1, \partial X_{t+1} / \partial \varepsilon_t = a, \partial X_{t+2} / \partial \varepsilon_t = a^2, \dots, \partial X_{t+h} / \partial \varepsilon_t = a^h, \dots$$

Таким образом, при $h \rightarrow \infty$

- $\partial X_{t+h} / \partial \varepsilon_t \rightarrow 0$ для $|a| < 1$,
- $\partial X_{t+h} / \partial \varepsilon_t \equiv 1$ при $a = 1$.

Попутно заметим, что если $|a| > 1$ (“взрывной процесс”), то в таком случае

$$\partial X_{t+h} / \partial \varepsilon_t \rightarrow \infty,$$

так что влияние прошлых шоков геометрически возрастает по мере удаления в прошлое. Это обстоятельство служит определенным аргументом против использования взрывных моделей для описания экономических временных рядов.

5.2. Проблема определения принадлежности временного ряда классу TS рядов или классу DS рядов

При построении моделей связей между временными рядами в долгосрочной перспективе необходимо учитывать факт наличия или отсутствия у анализируемых макроэкономических рядов стохастического (недетерминированного) тренда. Иначе говоря, приходится решать вопрос об отнесении каждого из рассматриваемых рядов к классу рядов, стационарных относительно детерминированного тренда (или просто стационарных) – TS (trend stationary) ряды, или к классу рядов, имеющих стохастический тренд (возможно, наряду с детерминированным трендом) и приводящихся к стационарному ряду только путем однократного или k -кратного² дифференцирования ряда – DS (difference stationary) ряды.

Принципиальное различие между этими двумя классами рядов выражается в том, что в случае TS ряда вычитание из ряда соответствующего детерминированного тренда приводит к стационарному ряду, тогда как в случае DS ряда вычитание детерминированной составляющей ряда оставляет ряд нестационарным из-за наличия у него стохастического тренда.

Проблема отнесения макроэкономических рядов динамики, имеющих выраженный тренд, к одному из двух указанных классов активно обсуждалась в последние два десятилетия в мировой эконометрической и экономической литературе. Помимо того, что траектории TS и DS рядов отличаются друг от друга кардинальным образом, определение принадлежности рядов классам TS или DS весьма важно для правильного

² Мы не затрагиваем здесь вопрос о возможной дробной интегрированности рядов.

построения долгосрочных регрессионных моделей, в которых объясняемыми и объясняющими переменными являются макроэкономические временные ряды (модели коинтеграции, модели коррекции ошибок, векторные авторегрессии).

Если выявляется группа макроэкономических рядов, принадлежащих классу DS-рядов, то между этими рядами возможна так называемая **коинтеграционная связь** (см. главу 7), анализ которой позволяет, например,

- проверять гипотезу эффективности финансовых рынков (см., например, [Dwyer, Wallace (1992), Dutt, Ghosh (1999)]),
- проверять выполнение на практике теории паритета покупательной способности ([Ardeni, Lubian (1991), Dutt(1998)]),
- проверять выполнение в долгосрочной перспективе уравнения спроса на деньги ([Johansen, Juselius (1990)], [Hafer, Jansen (1991)], [Funke, Thornton (1999)]).

Более того, при наличии коинтеграционной связи между DS-рядами имеется возможность построения комбинации краткосрочной и долгосрочной динамических регрессионных моделей в форме так называемой **модели коррекции ошибок**, что открывает возможность построения на основании подобранной модели как краткосрочных, так и долгосрочных прогнозов.

Литература по этому вопросу весьма обширна. В качестве обзорных работ можно сослаться на монографии [Maddala, Kim (1998)], [Enders (1995)], [Hamilton (1994)], [Hatanaka (1996)]. Отметим лишь (чрезвычайно малую) часть работ, посвященных построению моделей связи между конкретными макроэкономическими рядами:

- Денежные агрегаты: [Hasan (1998)].
- Инфляция: [Metin (1995)]
- Валовой внутренний продукт: [Christiano, Eichenbaum (1990)]; [Murray, Nelson (2000)].
- Уровень безработицы: [Clark (1989)]; [Woodward, Pillariseti (1999)].
- Обменный курс национальной валюты: [Copeland (1991)], [Kim, Mo (1995)], [Nadal-De Simone, Razzak (1999)].
- Импорт: [Milas (1998)].
- Налоговые ряды: [Molana (1994)] .
- Производство: [Cheung, Chinn (1996)], [den Haan (2000)].
- Биржевые индексы: [Fama, French (1988)].

TS ряды имеют линию тренда в качестве некоторой “центральной линии”, которой следует траектория ряда, находясь то выше, то ниже этой линии, с достаточно частой сменой положений выше-ниже. DS ряды помимо детерминированного тренда (если таковой имеется) имеют еще и стохастический тренд, из-за присутствия которого траектория DS ряда весьма долго пребывает по одну сторону от линии детерминированного тренда (выше или ниже этой линии), удаляясь от нее на

значительные расстояния, так что по-существу в этом случае линия детерминированного тренда перестает играть роль “центральной” линии, вокруг которой колеблется траектория процесса.

В TS-рядах влияние предыдущих шоковых воздействий затухает с течением времени, а в DS-рядах такое затухание отсутствует и каждый отдельный шок влияет с одинаковой силой на все последующие значения ряда. Поэтому наличие стохастического тренда требует проведения определенной экономической политики для возвращения макроэкономической переменной к ее долгосрочной преспективе, тогда как при отсутствии стохастического тренда серьезных усилий для достижения такой цели не требуется – в этом случае макроэкономическая переменная “скользит” вдоль линии тренда как направляющей, пересекая ее достаточно часто и не уклоняясь от этой линии сколь-нибудь далеко.

В течение довольно долгого времени было принято при анализе рядов с выраженным трендом производить оценивание и выделение детерминированного тренда, после чего производить подбор динамической модели (например, ARMA) к ряду, “очищенному от тренда”, т.е. к ряду остатков от соответствующей оцененной регрессионной модели. После введения Боксом и Дженкинсом ([Бокс, Дженкинс (1974)]) в обиход моделей ARIMA стало модным остационирование рядов с выраженным трендом и медленным убыванием (оцененной) автокорреляционной функции путем перехода к рядам первых или вторых разностей. Однако, как показали дальнейшие исследования, произвольный выбор одного из этих двух способов остационирования ряда вовсе не так безобиден, как это казалось поначалу.

В работах [Chan, Nanyu, Ord (1977)], [Nelson, Kang (1981)] было показано, что остационирование DS рядов путем перехода к очищенному ряду (детрендрование) изменяет спектр ряда, приводя к появлению ложной периодичности (ложные длиннопериодные циклы), которая может быть ошибочно истолкована как проявление некоторого экономического цикла. С другой стороны, дифференцирование TS ряда приводит к “передифференцированному ряду”, который хотя и является стационарным, но обладает некоторыми нежелательными свойствами, связанными с необратимостью его МА-составляющей; при этом возникает паразитная автокоррелированность соседних значений продифференцированного ряда (в спектре доминируют короткие циклы). Более того, в случае необратимости МА-составляющей продифференцированного ряда становится невозможным использование обычных алгоритмов оценивания параметров и прогнозирования ряда. (См., например, [Hamilton (1994), главы 4 и 5].)

Итак, построение адекватной модели макроэкономического ряда, которую можно использовать для описания динамики ряда и прогнозирования его будущих значений, и адекватных моделей связей этого ряда с другими макроэкономическими рядами невозможно без выяснения природы этого ряда и природы рядов, с ним связываемых, т.е. без выяснения принадлежности ряда к одному из двух указанных классов (TS или DS). В этом разделе мы займемся проблемой такой классификации.

Как показывает огромное количество работ, подробный обзор которых можно найти, например, в книге [Maddala, Kim (1998)], проблема отнесения ряда к одному из указанных двух классов на основании наблюдения реализации ряда на некотором интервале времени оказалась весьма сложной. Было предложено множество процедур такой классификации, но и по настоящее время предлагаются все новые и новые процедуры, которые либо несколько превосходят старые в статистической эффективности (по крайней мере, теоретически) либо могут составить конкуренцию старым процедурам и служить дополнительным средством подтверждения

классификации, произведенной другими методами. Описание многих таких процедур и ссылки на статьи с подробным описанием и теоретическим обоснованием этих процедур можно найти, например, в упоминавшихся выше книгах [Maddala, Kim (1998)], [Enders (1995)], [Hamilton (1994)], [Hatanaka (1996)].

Здесь мы заметим только, что использование различных процедур может приводить к противоположным выводам о принадлежности наблюдаемого ряда классу TS-рядов или классу DS-рядов. В этом отношении весьма показательным является сопоставление выводов, полученных при анализе 14 макроэкономических рядов США (имеющих протяженность от 62 до 111 лет) в работе [Nelson, Plosser (1982)] и в более поздней работе Перрона ([Perron (1989a)]). Если в первой работе лишь один из 14 рассмотренных рядов был отнесен к классу TS, то во второй работе, напротив, к этому классу было отнесено уже 11 из этих рядов. Правда, подобное кардинальное изменение результатов классификации было связано с расширением понятия TS рядов. В класс TS-рядов стали включать и ряды, стационарные относительно трендов, имеющих “излом” в известный момент времени. Отказ от предположения об известной дате излома тренда, в свою очередь, привел к некоторому изменению классификации, полученной Перроном (см. [Zivot, Andrews (1992)]). Допущение еще более гибких форм функции тренда изменило и последнюю классификацию, см. [Bierens (1997)]. Наконец, работа [Nunes, Newbold, Kuan (1997)] “замкнула круг”: изменение предположения о характере процесса порождения данных по сравнению с работой [Zivot, Andrews (1992)] привело к той же самой классификации 14 рядов, которая была получена в работе [Nelson, Plosser (1982)].

В связи с такими результатами, при анализе конкретных макроэкономических рядов теперь обычно применяют несколько разных статистических процедур, что позволяет несколько укрепить выводы, сделанные в пользу одной из двух (TS или DS) конкурирующих гипотез.

5.3. Различение TS и DS рядов в классе моделей ARMA. Гипотеза единичного корня.

Как уже отмечалось выше, для решения вопроса об отнесении исследуемого ряда X_t к классу TS (стационарных или стационарных относительно тренда) или DS (разностно стационарных) процессов имеется целый ряд различных процедур. Однако все эти процедуры страдают теми или иными недостатками. Процедуры, оформленные в виде формальных статистических критериев, как правило, имеют достаточно низкую мощность, что ведет к весьма частому неопровержению исходной (нулевой гипотезы), когда она в действительности не выполняется. В то же время невыполнение теоретических предпосылок, на которых основывается критерий, при применении его к реальным данным приводит к отличию реально наблюдаемого размера критерия от заявленного уровня значимости. Вследствие последнего обстоятельства теряется контроль над вероятностью ошибки первого рода, и это может приводить к слишком частому отвержению нулевой гипотезы, когда она в действительности верна. В связи с таким положением вещей исследователи обычно используют при анализе рядов на принадлежность их к классу TS или DS не один, а несколько критериев и подкрепляют выводы, полученные с использованием формальных критериев (с установленными уровнями значимости) графическими процедурами. Мы также будем пользоваться в нашем исследовании несколькими процедурами различения TS и DS рядов и в этом разделе кратко опишем эти процедуры. Более подробное их описание можно найти в цитируемой ниже литературе.

В большинстве критериев, предложенных для различения DS и TS гипотез, эта задача решается в классе моделей ARMA (стационарных и нестационарных).

Если ряд X_t имеет тип $ARIMA(p, k, q)$, то в результате его k -кратного дифференцирования мы получаем ряд стационарный ряд $\Delta^k X_t$ типа $ARMA(p, q)$, скажем,

$$a^*(L) \Delta^k X_t = b(L) \varepsilon_t,$$

где $a^*(L)$ и $b(L)$ – полиномы от оператора обратного сдвига L , имеющие степени p и q , соответственно. Заметим, что $\Delta X_t = (1 - L) X_t$, так что

$$\Delta^k X_t = (1 - L)^k X_t,$$

и

$$a^*(L) (1 - L)^k X_t = b(L) \varepsilon_t,$$

или

$$a(L) X_t = b(L) \varepsilon_t,$$

где $a(L) = a^*(L) (1 - L)^k$ – полином степени $(p + k)$. Поскольку ряд $\Delta^k X_t$ стационарный, все p корней полинома $a^*(z)$ находятся за пределами единичного круга, так что полином $a(z)$ имеет p корней за пределами единичного круга и k корней на границе этого круга, точнее, корень $z = 1$ кратности k .

Таким образом, ряд X_t представляется нестационарной моделью $ARMA(p + k, q)$, в которой авторегрессионный полином $a(L)$ имеет ровно k корней, равных 1, а все остальные корни по модулю больше 1. Поэтому проверка нулевой гипотезы H_0 о том,

что некоторый ARMA ряд X_t является DS-рядом (а не стационарным рядом), может быть сведена к проверке гипотезы о том, что авторегрессионный полином $a(L)$ имеет хотя бы один корень, равный 1. Это оправданно, если исходить из предположения, что $a(z)$ не имеет корней внутри единичного круга, т.е. исключить из рассмотрения “взрывные” модели. При этом о гипотезе H_0 кратко говорят как о ***гипотезе единичного корня (UR – unit root hypothesis)***, хотя точнее было бы говорить о гипотезе авторегрессионного единичного корня. В качестве альтернативной тогда выступает ***TS гипотеза*** о том, что рассматриваемый ARMA ряд – стационарный.

Критерии, в которых за исходную (нулевую) гипотезу берется гипотеза TS, служат скорее для подтверждения результатов проверки DS-гипотезы. В этом случае вместо проверки гипотезы единичного корня у полинома $a(z)$ проверяется гипотеза о наличии единичного корня $z = 1$ у уравнения $b^*(z) = 0$, где $b^*(L)$ – полином от оператора обратного сдвига L в представлении в виде процесса скользящего среднего $\Delta X_t = b^*(z)\varepsilon_t$ ряда разностей $\Delta X_t = X_t - X_{t-1}$ исходного процесса X_t .

Глава 6. Процедуры для различения TS и DS рядов

В этой и в последующих главах мы будем обозначать и случайные величины и их реализации строчными буквами.

6.1. Предварительные замечания

Прежде, чем двигаться дальше, следует обратить внимание на одно важное обстоятельство, которое является иногда причиной недоразумений при практическом истолковании полученных результатов.

Рассмотрим TS ряд

$$x_t = \alpha + \beta t + a_1 x_{t-1} + \varepsilon_t, \quad |a_1| < 1.$$

Относительно какого именно линейного тренда этот ряд является стационарным?

Пусть y_t – детрендрованный ряд, так что $y_t = x_t - \gamma - \delta t$ и

$$y_t = a_1 y_{t-1} + \varepsilon_t.$$

Подставляя выражения для y_t и y_{t-1} в последнее соотношение, находим:

$$x_t - \gamma - \delta t = a_1(x_{t-1} - \gamma - \delta(t-1)) + \varepsilon_t,$$

$$x_t = (\gamma - a_1\gamma + a_1\delta) + \delta(1-a_1)t + a_1x_{t-1} + \varepsilon_t,$$

так что $\alpha = \gamma - a_1\gamma + a_1\delta$ и $\beta = \delta(1-a_1)$, откуда получаем

$$\delta = \beta / (1-a_1), \quad \gamma = (\alpha - a_1(\alpha + \beta)) / (1-a_1)^2.$$

Таким образом, ряд x_t является стационарным относительно линейного тренда

$$\frac{\alpha - a_1(\alpha + \beta)}{(1-a_1)^2} + \frac{\beta}{(1-a_1)}t.$$

В частности, при $\beta = 0$ и $\alpha \neq 0$ процесс стационарен и имеет математическое ожидание

$$\mu = \frac{\alpha}{(1-a_1)}.$$

Если $a_1 = 1$, то последнее представление невозможно, и надо исходить непосредственно из определения ряда x_t . В этом случае

$$\begin{aligned} x_t &= \alpha + \beta t + x_{t-1} + \varepsilon_t = \\ &= (\alpha + \beta t + \varepsilon_t) + (\alpha + \beta(t-1) + \varepsilon_{t-1}) + \dots + (\alpha + \beta + \varepsilon_1) + x_0 = \\ &= x_0 + (\alpha + \beta/2)t + (\beta/2)t^2 + (\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_t). \end{aligned}$$

При $\alpha = \beta = 0$ имеем простое случайное блуждание

$$x_t = x_{t-1} + \varepsilon_t, \quad x_t = x_0 + (\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_t).$$

При $\alpha \neq 0, \beta = 0$ имеем

$$x_t = \alpha + x_{t-1} + \varepsilon_t, \quad x_t = x_0 + \alpha t + (\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_t),$$

т.е. случайное блуждание вокруг детерминированного линейного тренда $x_0 + \alpha t$.

Наконец, при $\alpha \neq 0, \beta \neq 0$

$$x_t = \alpha + \beta t + x_{t-1} + \varepsilon_t, \quad x_t = x_0 + (\alpha + \beta/2)t + (\beta/2)t^2 + (\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_t),$$

так что исходный ряд x_t представляет случайное блуждание вокруг детерминированного квадратичного тренда $x_0 + (\alpha + \beta/2)t + (\beta/2)t^2$.

Таким образом, в модели

$$x_t = \alpha + \beta t + a_1 x_{t-1} + \varepsilon_t$$

если $\alpha \neq 0, \beta = 0$, то

- при $|a_1| < 1$ у ряда x_t тренда нет;

- при $a_1 = 1$ ряд x_t имеет стохастический тренд $\sum_{j=1}^t \varepsilon_j$ и линейный тренд $x_0 + \alpha t$;
если $\beta \neq 0$, то
 - при $|a_1| < 1$ ряд x_t имеет линейный тренд $\frac{\alpha - a_1(\alpha + \beta)}{(1 - a_1)^2} + \frac{\beta}{(1 - a_1)} t$;
 - при $a_1 = 1$ ряд x_t имеет стохастический тренд $\sum_{j=1}^t \varepsilon_j$ и квадратичный тренд $x_0 + (\alpha + \beta/2) t + (\beta/2) t^2$.

Ниже приводятся смоделированные реализации, порожденные моделью

$$x_t = \alpha + \beta t + a_1 x_{t-1} + \varepsilon_t$$

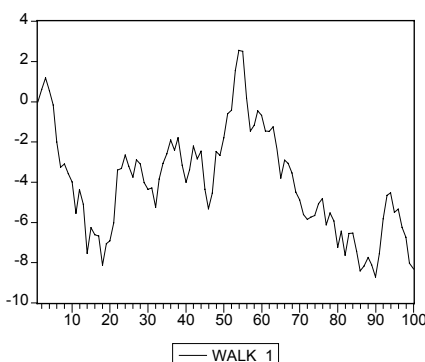
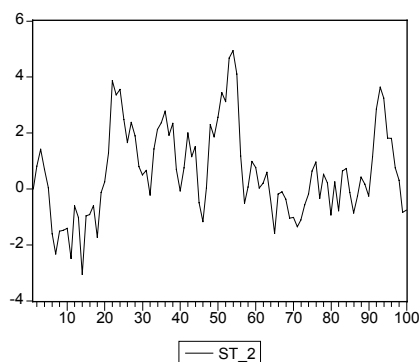
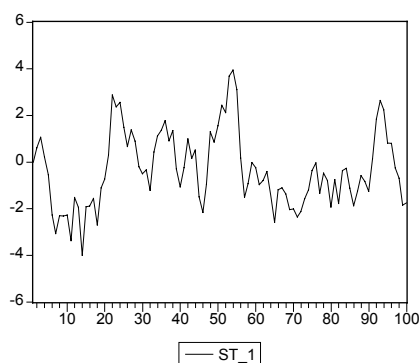
при различных наборах значений параметров α, β, a_1 :

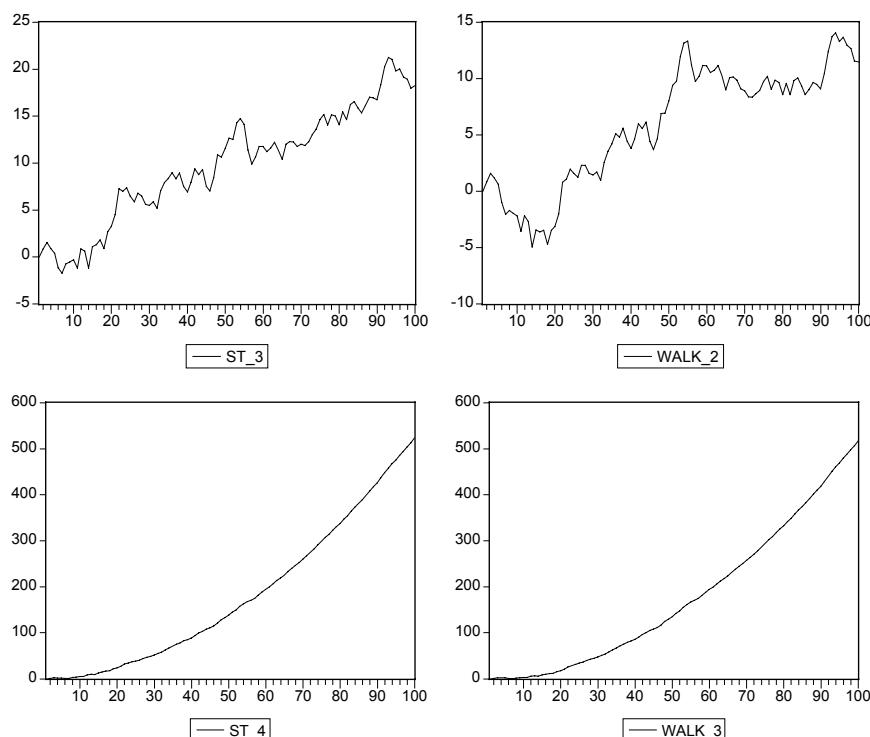
- ST_1 : $a_1 = 0.8, \alpha = 0, \beta = 0$.
- ST_2 : $a_1 = 0.8, \alpha = 0.2, \beta = 0$.
- ST_3 : $a_1 = 0.8, \alpha = 0.16, \beta = 0.04$.
- WALK_1 : $a_1 = 1, \alpha = 0, \beta = 0$.
- WALK_2 : $a_1 = 1, \alpha = 0.2, \beta = 0$.
- WALK_3 : $a_1 = 1, \alpha = 0.2, \beta = 0.1$.

Кроме того, приводится смоделированная реализация, порожденная моделью

$$x_t = \alpha + \beta t + \gamma t^2 + a_1 x_{t-1} + \varepsilon_t$$

- ST_4 : $a_1 = 0.8, \alpha = 0.12, \beta = 0.13, \gamma = 0.01$.





Сравнение первых трех графиков ST_1, ST_2, WALK_1 между собой и сравнение графиков в парах ST_3 – WALK_2, ST_4 – WALK_3 показывает, сколь трудно различить визуально реализации процессов с единичным корнем от реализаций, соответствующих стационарным или стационарным относительно детерминированного тренда процессов.

Перейдем теперь к формальным статистическим критериям наличия (или отсутствия) единичного корня.

6.2. Критерии Дики – Фуллера

Мы уже говорили во Введении о том, что при оценивании по статистическим данным статистической модели в форме процесса авторегрессии

SM: $x_t = a_1 x_{t-1} + \varepsilon_t$, $t = 1, \dots, T$,

в случае, когда истинная модель, порождающая данные (*процесс порождения данных, DGP – data generating process*), имеет вид

DGP: $x_t = x_{t-1} + \varepsilon_t$, где ε_t – гауссовский белый шум,

т.е. $a_1 = 1$, оценка наименьших квадратов \hat{a}_1 коэффициента a_1 не имеет нормального распределения даже асимптотически. В действительности имеет место следующий факт. При $T \rightarrow \infty$ и $a_1 = 1$

$$n(\hat{a}_1 - 1) \rightarrow \frac{(1/2)\{[W(1)]^2 - 1\}}{\int_0^1 [W(r)]^2 dr},$$

где $W(r)$ – *стандартное броуновское движение с непрерывным временем* и сходимость понимается как сходимость распределения случайной величины, стоящей слева, к распределению случайной величины, стоящей справа.

Процесс $W(r)$ является непрерывным аналогом дискретного случайного блуждания $x_t = x_{t-1} + \varepsilon_t$.

Это процесс, для которого

- $W(0) = 0$;
- Приращения $(W(r_2) - W(r_1)), \dots, (W(r_k) - W(r_{k-1}))$ независимы в совокупности, если $0 \leq r_1 < r_2 < \dots < r_k$; $W(s) - W(r) \sim N(0, s - r)$ при $s > r$;
- Реализации $W(r)$ непрерывны с вероятностью 1.

Из определения, в частности, следует, что $W(1) = W(1) - W(0) \sim N(0, 1)$, так что $[W(1)]^2 \sim \chi^2(1)$. Отсюда вытекает, что при больших T

$$P\{\hat{a}_1 < 1\} = P\{\hat{a}_1 - 1 < 0\} \approx P\{[W(1)]^2 - 1 < 0\} = P\{\chi^2(1) < 1\} = 0.68,$$

так что если DGP – простое случайное блуждание (без сноса), DGP: $x_t = x_{t-1} + \varepsilon_t$, то оценивание SM: $x_t = a_1 x_{t-1} + \varepsilon_t$ дает значение $\hat{a}_1 < 1$ примерно в 2/3 случаев.

Критические значения распределения статистики $T(\hat{a}_1 - 1)$ при гипотезе $a_1 = 1$ для конечных T находятся методом статистических испытаний (Монте – Карло); впервые это было сделано Фуллером [Fuller (1976)]. Соответствующие таблицы построены в предположении, что $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$, $\varepsilon_1, \dots, \varepsilon_T$ – независимые случайные величины и $x_0 = 0$. Однако следует заметить, что хотя значение x_0 не влияет на асимптотическое распределение $T(\hat{a}_1 - 1)$, оно влияет на распределение $T(\hat{a}_1 - 1)$ при малых выборках.

Критерий, основанный на статистике $T(\hat{a}_1 - 1)$, отвергает гипотезу $H_0: a_1 = 1$ в пользу альтернативной гипотезы $H_A: a_1 < 1$ на 5% уровне значимости при значениях $T(\hat{a}_1 - 1)$, меньших $T(\hat{a}_1 - 1)_{\text{крит}}$, или при значениях \hat{a}_1 , меньших $\hat{a}_{1 \text{ крит}}$, где

T	$T(\hat{a}_1 - 1)_{\text{крит}}$	$\hat{a}_{1 \text{ крит}}$
25	-7.3	0.708
50	-7.7	0.846
100	-7.9	0.921
250	-8.0	0.968
500	-8.0	0.998
∞	-8.1	

Как видно из этой таблицы, при небольших значениях T гипотеза H_0 отвергается лишь для значений \hat{a}_1 , намного меньших 1. Чувствительность критерия возрастает только при весьма большом количестве наблюдений. Это приводит к тому, что при небольших T отвергнуть гипотезу $H_0: a_1 = 1$ в пользу альтернативной гипотезы $H_A: a_1 < 1$ довольно трудно, даже если \hat{a}_1 существенно меньше 1.

Более привычным являлось бы, конечно, использование для проверки гипотезы $H_0: a_1 = 1$ против $H_A: a_1 < 1$ отношения

$$t = \frac{\hat{a}_1 - 1}{s(\hat{a}_1)} \quad (t\text{-отношение, } t\text{-ratio, } t\text{-статистика}),$$

где $s(\hat{a}_1)$ – оцененная стандартная ошибка оценки \hat{a}_1 . Однако, поскольку при $a_1 = 1$ уже и сама оценка \hat{a}_1 не имеет нормального распределения, то и это отношение не имеет t -распределения Стьюдента. Критические значения этой t -статистики при $T \rightarrow \infty$ и некоторых конечных значениях T также впервые были приведены в работе Фуллера [Fuller (1976)]. Гипотеза $H_0: a_1 = 1$ отвергается в пользу альтернативной гипотезы $H_A: a_1 < 1$ при больших отрицательных значениях указанной статистики. Сравним 5% критические значения, указанные Фуллером, с 5% критическими значениями обычного одностороннего t -критерия, вычисляемыми по распределению Стьюдента $t(T - 1)$ с $(T - 1)$ степенями свободы:

T	Фуллер	Стьюдент
25	-1.95	-1.71
50	-1.95	-1.68

100	– 1.95	– 1.66
250	– 1.95	– 1.65
500	– 1.95	– 1.65
∞	– 1.95	

Таблица иллюстрирует скошенность распределения $(\hat{a}_1 - 1)/s(\hat{a}_1)$ при $a_1 = 1$.

В числе прочего, мы приводили в разд. 6.1 смоделированную реализацию процесса случайного блуждания без сноса $x_t = x_{t-1} + \varepsilon_t$ (WALK_1). Оценим по этой реализации статистическую модель $x_t = a_1 x_{t-1} + \varepsilon_t$, используя первые 50 наблюдений:

Dependent Variable: WALK_1

Sample(adjusted): 2 50

Included observations: 49 after adjusting endpoints

Variable	Coeff.	Std. Error	t-Statistic	Prob.
WALK_1(-1)	0.970831	0.035729	27.17224	0.0000

Значение указанного выше t -отношения равно $(0.970831 - 1)/0.035729 = -0.816$ и превышает критическое значение -1.95 . Поэтому гипотеза о наличии единичного корня не может быть отвергнута на 5% уровне значимости. Это согласуется с полученной оценкой 0.970831 коэффициента a_1 , значительно превышающей критический уровень 0.846 .

С другой стороны, если мы возьмем смоделированную в разд. 6.1 реализацию ST_1 стационарного ряда $x_t = 0.8 x_{t-1} + \varepsilon_t$ (имеющего нулевое математическое ожидание) и оценим по этой реализации статистическую модель $x_t = a_1 x_{t-1} + \varepsilon_t$, используя первые 50 наблюдений, то получим:

Dependent Variable: ST_1

Method: Least Squares

Sample(adjusted): 2 50

Included observations: 49 after adjusting endpoints

Convergence achieved after 2 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.790557	0.090501	8.735349	0.0000

Интересующее нас t -отношение равно $-2.314 < -1.95$, что приводит к отвержению гипотезы единичного корня. Это согласуется с тем, что оцененное значение a_1 здесь равно 0.791 и значимо отличается от 1.

Обратимся теперь к смоделированной реализации ST_2 стационарного AR(1) процесса $x_t = 0.2 + 0.8 x_{t-1} + \varepsilon_t$ (математическое ожидание которого равно 1). Среднеарифметическое значение первых 50 значений ряда равно 0.596 . Поэтому даже если бы мы не знали, как этот ряд моделировался, мы все же могли сказать, что если этот ряд стационарный, то он имеет скорее ненулевое математическое ожидание. Но тогда альтернативой для гипотезы $H_0: x_t = x_{t-1} + \varepsilon_t$ (наличие единичного корня) должна быть гипотеза

$H_A: x_t = \alpha + a_1 x_{t-1} + \varepsilon_t$, $a_1 < 1$, $\alpha \neq 0$.

Поэтому в качестве статистической модели мы берем теперь модель

SM: $x_t = \alpha + a_1 x_{t-1} + \varepsilon_t$.

Предельное распределение статистики $T(\hat{a}_1 - 1)$ не только не является нормальным, но и отличается от распределения этой же статистики при оценивании SM с $\alpha = 0$; при $T = 25$ $P\{\hat{a}_1 < 1\} = 0.95$. В следующей таблице приведены критические значения для этого случая:

T	$T(\hat{a}_1 - 1)_{\text{крит}}$	$\hat{a}_{1 \text{ крит}}$	$t_{\text{крит}}$ (Фуллер)
25	– 12.5	0.500	– 3.00

50	− 13.3	0.734	− 2.92
100	− 13.7	0.863	− 2.89
250	− 14.0	0.944	− 2.88
500	− 14.0	0.972	− 2.87
∞	− 14.1		− 2.86

Анализ в рамках статистической модели SM: $x_t = \alpha + a_1 x_{t-1} + \varepsilon_t$ ряда WALK_1 (по первым 50 наблюдениям) дает значение $t = -2.143 > t_{\text{крит}} = -2.92$,

так что гипотеза о том, что мы имеем дело с реализацией случайного блуждания $x_t = x_{t-1} + \varepsilon_t$, не отвергается.

Анализ в рамках этой же статистической модели ряда ST_2 дает (при $T = 50$) значение $t = -2.245$, так что гипотеза единичного корня не отвергается несмотря на то, что моделировалась реализация стационарного процесса. Последнее связано, конечно, с тем, что оцененное значение $\hat{a}_1 = 0.794$ выше критического уровня 0.734.

Замечание

Если мы проанализируем в рамках все той же SM: $x_t = \alpha + a_1 x_{t-1} + \varepsilon_t$ ряд ST_1 (с $\alpha = 0$), то получим

$$\hat{a}_1 = 0.785 > \hat{a}_{1 \text{ крит}} = 0.734, \quad t = -2.298 > t_{\text{крит}} = -2.92,$$

так что гипотеза единичного корня для ST_1 не отвергается. В то же время, как мы уже видели ранее, если ряд ST_1 анализируется в рамках статистической модели SM: $x_t = a_1 x_{t-1} + \varepsilon_t$, то гипотеза единичного корня отвергается.

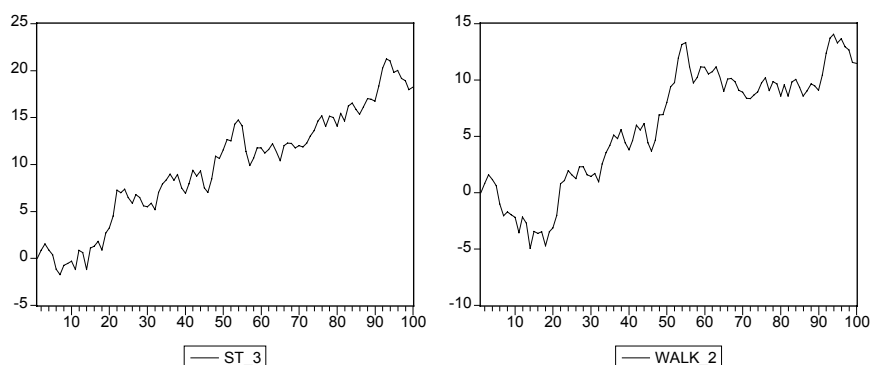
Этот пример иллюстрирует то обстоятельство, что при добавлении в статистическую модель излишних объясняющих переменных (в т.ч. и константы) мощность критерия снижается, и отвергнуть гипотезу единичного корня становится трудно, даже если она не верна. Поэтому важно выбирать статистическую модель “без излишеств”, включая в нее только такие составляющие, которые соответствуют поведению наблюдаемого временного ряда.

Посмотрим теперь на реализацию ST_3 стационарного процесса

$$x_t = 0.16 + 0.04t + 0.8x_{t-1} + \varepsilon_t,$$

стационарного относительно линейного тренда $0.2t$. Эта реализация похожа на реализацию WALK_2 случайного блуждания со сносом 0.2

$$x_t = 0.2 + x_{t-1} + \varepsilon_t:$$



Отсюда возникает проблема различения подобных процессов, и в связи с этим, рассматривается задача проверки гипотезы

$$H_0: x_t = \alpha + x_{t-1} + \varepsilon_t, \quad \alpha \neq 0, \quad (\text{случайное блуждание со сносом})$$

в рамках статистической модели

$$\text{SM: } x_t = \alpha + \beta t + a_1 x_{t-1} + \varepsilon_t.$$

Фуллер затабулировал процентные точки распределений оценки \hat{a}_1 и t -статистики для проверки гипотезы $a_1 = 1$ в такой ситуации.

Если DGP: $x_t = \alpha + x_{t-1} + \varepsilon_t$ с $\alpha \neq 0$, то 5% критические значения статистики $T(\hat{a}_1 - 1)$ и указанной t -статистики таковы:

T	$T(\hat{a}_1 - 1)_{\text{крит}}$	$\hat{a}_1_{\text{крит}}$	$t_{\text{крит}}$ (Фуллер)
25	-17.9	0.284	-3.60
50	-19.8	0.604	-3.50
100	-20.7	0.793	-3.45
250	-21.3	0.914	-3.43
500	-21.5	0.957	-3.42
∞	-21.8		-3.41

Проанализируем в рамках статистической модели SM: $x_t = \alpha + \beta t + a_1 x_{t-1} + \varepsilon_t$ смоделированные реализации (опять берем $T = 50$).

Для WALK_2 имеем $\hat{a}_1 = 0.858 > \hat{a}_{1\text{ крит}} = 0.604$, $t = -2.027 > t_{\text{крит}} = -3.50$,

так что гипотеза единичного корня не отвергается.

Для ST_3 имеем $\hat{a}_1 = 0.733 > \hat{a}_{1\text{ крит}} = 0.604$, $t = -2.687 > t_{\text{крит}} = -3.50$. Здесь значения \hat{a}_1 и t ближе к критическим, чем у WALK_2, но все же еще значительно их превышают, так что гипотеза единичного не отвергается и в этом случае.

Строго говоря, при виде реализаций, подобных ST_3, мы не должны исключать возможность того, что DGP – случайное блуждание без сноса, так что следовало бы знать также и распределения \hat{a}_1 и t в статистической модели

SM: $x_t = \alpha + \beta t + a_1 x_{t-1} + \varepsilon_t$

в предположении, что

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

Исследование этих распределений показало, что они совпадают с распределениями \hat{a}_1 и t , полученными в предположении

DGP: $x_t = \alpha + x_{t-1} + \varepsilon_t$, $\alpha \neq 0$,

так что при использовании статистической модели

SM: $x_t = \alpha + \beta t + a_1 x_{t-1} + \varepsilon_t$

одни и те же таблицы распределений \hat{a}_1 и t годятся и при DGP: $x_t = x_{t-1} + \varepsilon_t$ и при

DGP: $x_t = \alpha + x_{t-1} + \varepsilon_t$, $\alpha \neq 0$.

Проанализируем в рамках статистической модели SM: $x_t = \alpha + \beta t + a_1 x_{t-1} + \varepsilon_t$ реализацию DGP: $x_t = x_{t-1} + \varepsilon_t$, представленную рядом WALK_1. Оценивая эту статистическую модель, получаем (как и при DGP: $x_t = 0.2 + x_{t-1} + \varepsilon_t$): $\hat{a}_1 = 0.858 > \hat{a}_{1\text{ крит}} = 0.604$, $t = -2.027 > t_{\text{крит}} = -3.50$, так что гипотеза единичного корня не отвергается.

Отметим, наконец, замечательный результат Маккиннона [MacKinnon (1991)], который указал простую приближенную формулу для вычисления критических значений t -статистик в критериях Фуллера. Именно, он показал, что если $t_{\text{крит}}(p, T)$ – критическое значение t -статистики по Фуллеру, соответствующее уровню значимости p и количеству наблюдений T , то

$$t_{\text{крит}}(p, T) \approx \beta_\infty + \beta_1 T^{-1} + \beta_2 T^{-2},$$

где β_∞ , β_1 , β_2 – некоторые коэффициенты, зависящие от p и от того, какое из трех распределений Фуллера рассматривается. Маккиннон приводит таблицу этих коэффициентов для $p = 0.01, 0.05, 0.10$.

Пример

В главе 5 мы анализировали статистические данные об объеме потребительских расходов на товары кратковременного пользования и услуги в Великобритании за период с первого квартала 1974 г. по четвертый квартал 1985 г. Этот ряд был идентифицирован как процесс авторегрессии первого порядка относительно линейного тренда:

$$X_t - 47962.75 - 315.1909 t = 0.884803 (X_{t-1} - 47962.75 - 315.1909 (t-1)) + \varepsilon_t ,$$

или

$$X_t = 5804.037 + 36.30898 t + 0.884803 X_{t-1} + \varepsilon_t .$$

Там же мы отметили, что несмотря на то, что при полученной точечной оценке 0.884803 коэффициента при X_{t-1} построенная модель формально оказывается стационарной относительно детерминированного линейного тренда (т.е. детрендрованный процесс следует стационарной AR(1) модели), мы не можем с достаточной степенью уверенности гарантировать, что истинная модель порождения наблюдений также стационарна относительно линейного тренда. Рассмотрим эту проблему с точки зрения критериев единичного корня.

Поскольку исследуемый ряд обладает выраженным линейным трендом, будем действовать в рамках статистической модели

$$SM: x_t = \alpha + \beta t + a_1 x_{t-1} + \varepsilon_t$$

Проверим гипотезу $H_0: x_t = \alpha + x_{t-1} + \varepsilon_t$, пользуясь статистическим пакетом EVIEWS, в котором используется приведенная выше формула Маккиннона:

Test Statistic	-1.425277	1% Critical Value*	-4.1630
		5% Critical Value	-3.5066
		10% Critical Value	-3.1828

*MacKinnon critical values for rejection of hypothesis of a unit root.

В соответствии с этими результатами, гипотеза H_0 не отвергается. Но если считать, что она выполнена, то тогда в конечном счете следует оценивать не модель $x_t = \alpha + \beta t + a_1 x_{t-1} + \varepsilon_t$, а модель $x_t = \alpha + x_{t-1} + \varepsilon_t$. Оценивание последней в форме $\Delta x_t = \alpha + \varepsilon_t$ дает следующий результат:

Variable	Coefficient	Std. Error	t-Statistic	Prob.
α	236.8958	80.80998	2.931517	0.0052
Durbin-Watson stat 2.135959				

Это соответствует модели случайного блуждания со сносом

$$x_t = 236.8958 + x_{t-1} + \varepsilon_t .$$

Все рассмотренные варианты проверки гипотезы о наличии единичного корня в рамках статистических моделей

$$SM: x_t = a_1 x_{t-1} + \varepsilon_t$$

$$SM: x_t = \alpha + a_1 x_{t-1} + \varepsilon_t$$

$$SM: x_t = \alpha + \beta t + a_1 x_{t-1} + \varepsilon_t$$

основывались на распределениях оценки наименьших квадратов коэффициента a_1 и t -статистики для проверки гипотезы $a_1 = 1$ при соответствующих предположениях о процессе порождения данных:

$$DGP: x_t = x_{t-1} + \varepsilon_t \text{ (случайное блуждание),}$$

DGP: $x_t = \alpha + x_{t-1} + \varepsilon_t$ (случайное блуждание со сносом).

С другой стороны, например, если SM: $x_t = \alpha + a_1 x_{t-1} + \varepsilon_t$, то гипотеза $x_t = x_{t-1} + \varepsilon_t$ равносильна гипотезе

$H_0: \alpha = 0, a_1 = 1$.

Если бы мы находились в рамках классической модели линейной регрессии, то проверяли подобную гипотезу с использованием F -статистики

$$\Phi_1 = \frac{(RSS_0 - RSS)/2}{RSS/((T-1)-2)},$$

которая в классическом варианте имеет при гипотезе H_0 F -распределение Фишера $F(2, T-3)$ с двумя и $T-3$ степенями свободы. Поскольку, однако, мы имеем дело при гипотезе H_0 с нестационарным процессом, то, вообще говоря, не следует ожидать, что распределение статистики Φ_1 при гипотезе H_0 будет иметь (хотя бы асимптотически) распределение $F(2, T-3)$. Этот вопрос был исследован Дики и Фуллером [Dickey, Fuller (1981)]. Ими были построены таблицы распределения статистики Φ_1 при гипотезе $H_0: \alpha = 0, a_1 = 1$. Ниже приведены 5% критические значения статистики Φ_1 , рассчитанные Дики и Фуллером, а также (для сравнения) 5% критические значения $F_{\text{крит}}$, рассчитанные по распределению $F(2, n-3)$ (см., также [Hamilton (1994), таблица B.7 Case 2] и [Enders (1995), таблица C]):

T	Φ_1 крит	$F_{\text{крит}}$
25	5.18	3.44
50	4.86	3.20
100	4.71	3.10
250	4.63	3.00
500	4.61	3.00
∞	4.59	3.00

Пример

Опять возьмем для примера ряды WALK_1, ST_2, ST_1 (по 50 наблюдений).

Оценивая SM: $x_t = \alpha + a_1 x_{t-1} + \varepsilon_t$ для ряда WALK_1, получаем $\hat{\alpha} = -0.579$, $\hat{a}_1 = 0.850$, $RSS = 48.0335$. В модели с ограничениями $\alpha = 0, a_1 = 1$ имеем $\hat{x}_t = x_{t-1}$, так что

$$RSS_0 = \sum_{t=2}^T (x_t - x_{t-1})^2 = 52.7939,$$

$$\Phi_1 = \frac{(52.7939 - 48.0335)/2}{48.0335/(50-1-2)} = 2.329 < 4.86 \rightarrow$$

гипотеза H_0 не отвергается.

Для ST_2: $\hat{\alpha} = 0.181$, $\hat{a}_1 = 0.777$, $RSS = 52.6618$. В модели с ограничениями $\alpha = 0, a_1 = 1$ имеем $RSS_0 = 59.0547$,

$$\Phi_1 = \frac{(59.0547 - 52.6618)/2}{52.6618/(50-1-2)} = 2.853 < 4.86 \rightarrow$$

гипотеза H_0 не отвергается.

Для ST_1: $\hat{\alpha} = -0.042$, $\hat{a}_1 = 0.785$, $RSS = 52.7007$. В модели с ограничениями $\alpha = 0, a_1 = 1$ имеем $RSS_0 = 58.0671$,

$$\Phi_1 = 2.662 < 4.86 \rightarrow \text{гипотеза } H_0 \text{ не отвергается.}$$

Таким образом, для всех трех рядов статистические выводы, сделанные на основании t -статистик для коэффициента a_1 , совпали со статистическими выводами, сделанными на основании F -статистики.

В рамках статистической модели

SM: $x_t = \alpha + \beta t + a_1 x_{t-1} + \varepsilon_t$

гипотеза DGP: $x_t = x_{t-1} + \varepsilon_t$ соответствует гипотезе

$H_0: \alpha = \beta = 0, a_1 = 1$,

а гипотеза DGP: $x_t = \alpha + x_{t-1} + \varepsilon_t$ – гипотезе

$H_0: \beta = 0, a_1 = 1$.

F -статистика для первого случая имеет обозначение Φ_2 ; 5% критические значения равны (см. также [Enders (1995), таблица C])

T	$\Phi_{1 \text{ крит}}$
25	5.68
50	5.13
100	4.88
250	4.75
500	4.71
∞	4.68

F -статистика для второго случая имеет обозначение Φ_3 ; 5% критические значения равны (см. также [Hamilton (1994), таблица B.7 Case 4] и [Enders (1995), таблица C]):

T	$\Phi_{1 \text{ крит}}$
25	7.24
50	6.73
100	6.49
250	6.34
500	6.30
∞	6.25

Пример

Рассмотрим ряды WALK_1, WALK_2, ST_3. Оценим статистическую модель

SM: $x_t = \alpha + \beta t + a_1 x_{t-1} + \varepsilon_t$

для каждого из этих рядов и проверим для них

- Гипотезу $H_0: \alpha = \beta = 0, a_1 = 1$, опираясь на статистику Φ_2 ;
- Гипотезу $H_0: \beta = 0, a_1 = 1$, опираясь на статистику Φ_3 .

$H_0: \alpha = \beta = 0, a_1 = 1$

WALK_1: $\hat{\alpha} = -0.854$, $\hat{\beta} = 0.012$, $\hat{a}_1 = 0.858$, $RSS = 46.7158$. В модели с ограничениями $RSS_0 = 52.7939$,

$\Phi_2 = 1.995 < 5.13 \rightarrow$ гипотеза H_0 не отвергается.

WALK_2: $\hat{\alpha} = -0.711$, $\hat{\beta} = 0.040$, $\hat{a}_1 = 0.858$, $RSS = 46.7158$. В модели с ограничениями $RSS_0 = 52.7939$,

Φ_2 опять равно $1.995 < 5.13 \rightarrow$ гипотеза H_0 не отвергается.

ST_3: $\hat{\alpha} = -0.345$, $\hat{\beta} = 0.070$, $\hat{a}_1 = 0.733$, $RSS = 50.3928$, $\Phi_2 = 1.207 < 5.13 \rightarrow$ гипотеза H_0 не отвергается.

$H_0: \beta = 0, a_1 = 1$

WALK_1: $RSS = 46.7158$. В модели с ограничениями $RSS_0 = 52.7282$,

$\Phi_3 = 1.973 < 5.13 \rightarrow$ гипотеза H_0 не отвергается.

WALK_2: Φ_3 опять равно 1.973 \rightarrow гипотеза H_0 не отвергается.

ST_3: $RSS = 50.3928$, $\Phi_2 = 0.711 \rightarrow$ гипотеза H_0 не отвергается.

Рассмотренные примеры указывают на то, что и с помощью формальных статистических критериев бывает практически невозможно отличить реализации

процессов с единичным корнем и без наличия такового при небольшом количестве наблюдений. Это связано с весьма низкой мощностью соответствующих критериев при умеренном количестве наблюдений и “близких” альтернативах. Скажем, достаточная мощность критерия Φ_1 достигается только при $a_1 < 0.8$, а также если α близко к 1 или $\alpha > 1$. Мощности критериев Φ_2 и Φ_3 еще ниже. Те же замечания относятся и к критериям, основанным на “ t -статистиках” и на статистике $T(\hat{a}_1 - 1)$.

Все это приводит к “презумпции наличия единичного корня” в случае, когда в качестве нулевой гипотезы H_0 берется именно гипотеза единичного корня.

В связи с этим, рядом авторов была рассмотрена задача проверки нулевой гипотезы стационарности (стационарности относительно детерминированного тренда) против альтернативной гипотезы единичного корня. В дальнейшем мы коснемся этого вопроса несколько подробнее, а сейчас отметим только, что при таком подходе наблюдается похожая картина. Критерии стационарности имеют низкую мощность, и вследствие этого возникает уже “презумпция отсутствия единичного корня”. Поэтому мы отложим пока знакомство с такими критериями и вернемся опять к рассмотрению ситуации, когда основной (нулевой) является гипотеза наличия единичного корня.

Полученные выше результаты проверки гипотезы единичного корня для смоделированных реализаций стационарных процессов представляются крайне пессимистическими: при использовании первых 50 наблюдений эта гипотеза не отвергается

- Для ST_1 и ST_2 в паре
DGP: $x_t = x_{t-1} + \varepsilon_t$, **SM:** $x_t = \alpha + a_1 x_{t-1} + \varepsilon_t$;
- Для ST_3 в паре
DGP: $x_t = \alpha + x_{t-1} + \varepsilon_t$ (или **DGP:** $x_t = x_{t-1} + \varepsilon_t$),
SM: $x_t = \alpha + \beta t + a_1 x_{t-1} + \varepsilon_t$.

Проследим, что дает проверка гипотезы единичного корня в этих же связках, но при использовании большего количества наблюдений; для этого возьмем теперь $T = 100$. Сравним полученные результаты (в последней строке таблицы использованы 10% критические значения):

	$n = 50$	$n = 100$
ST_1	$t = -2.298 > t_{\text{крит}} = -2.92$, гипотеза единичного корня не отвергается	$t = -3.238 < t_{\text{крит}} = -2.89$, гипотеза единичного корня отвергается
ST_2	$t = -2.245 > t_{\text{крит}} = -2.92$, гипотеза единичного корня не отвергается	$t = -3.217 < t_{\text{крит}} = -2.89$, гипотеза единичного корня отвергается
ST_3	$t = -2.687 > t_{\text{крит}} = -3.18$. гипотеза единичного корня не отвергается на 10% уровне	$t = -3.207 < t_{\text{крит}} = -3.15$, гипотеза единичного корня отвергается на 10% уровне

Последняя серия результатов показывает, что при увеличении количества наблюдений мощность критериев Дики – Фуллера возрастает.

6.3. Расширенные критерии Дики - Фуллера

Обратимся опять к статистическим данным о величине валового национального продукта (GNP) в США за период с первого квартала 1947 г. по четвертый квартал 1961 г. В главе 5 мы идентифицировали этот ряд как процесс авторегрессии второго порядка

$$X_t - 217.740 - 5.222 t = 1.380 (X_{t-1} - 217.740 - 5.222(t-1)) -$$

$$- 0.630 (X_{t-2} - 217.740 - 5.222(t-2)) + \varepsilon_t,$$

или

$$X_t = 55.017 + 1.304 t + 1.380 X_{t-1} - 0.630 X_{t-2} + \varepsilon_t.$$

Как проверить гипотезу о наличии единичного корня в модели авторегрессии, порождающей этот ряд? Ведь в рассмотренных выше критериях Дики – Фуллера проверка такой гипотезы велась в рамках моделей авторегрессии первого порядка. Выход из этого положения оказался достаточно простым.

Рассмотрим статистическую модель

$$SM: x_t = \alpha + \beta t + a_1 x_{t-1} + a_2 x_{t-2} + \dots + a_p x_{t-p} + \varepsilon_t.$$

Путем чисто алгебраических преобразований ее можно преобразовать к виду

$$(\#) \quad x_t = \alpha + \beta t + \rho x_{t-1} + (\theta_1 \Delta x_{t-1} + \dots + \theta_{p-1} \Delta x_{t-p+1}) + \varepsilon_t,$$

где

$$\rho = a_1 + a_2 + \dots + a_p, \quad \theta_j = -(a_{j+1} + \dots + a_p).$$

(В примере с GNP такое преобразование дает

$$x_t = 55.017 + 1.304 t + 1.380 X_{t-1} - 0.630 X_{t-1} + 0.630 X_{t-1} - 0.630 X_{t-2} + \varepsilon_t =$$

$$= 55.017 + 1.304 t + 0.750 X_{t-1} + 0.630 \Delta X_{t-1} + \varepsilon_t.)$$

Если исходить из того, что уравнение $a(z) = 0$ может иметь только один корень $z = 1$, а остальные $p - 1$ корней лежат за пределами единичного круга, то тогда наличие единичного корня равносильно тому, что $a_1 + a_2 + \dots + a_p = 1$, т.е. $\rho = 1$. (См., например, [Hamilton (1994)].) Таким образом, гипотеза существования единичного корня у процесса $AR(p)$ сводится в этом случае к гипотезе $H_0: \rho = 1$ в преобразованном соотношении $(\#)$. Для проверки этой гипотезы можно пользоваться теми же таблицами Фуллера, только на этот раз используются значения статистики $T(\hat{\rho} - 1)$ и t -отношения для проверки гипотезы $\rho = 1$, полученные при оценивании **расширенной (augmented)** статистической модели $(\#)$ (с β и α , равными или не равными нулю). Соответствующие “ t -статистики” обозначают обычно **ADF (augmented Dickey – Fuller)**, в отличие от статистики **DF**, получаемой для модели $AR(1)$.

Для того, чтобы не вычислять самим каждый раз значение t -статистики для гипотезы $\rho = 1$, можно преобразовать $(\#)$ к виду

$$(\#\#) \quad \Delta x_t = \alpha + \beta t + \varphi x_{t-1} + (\theta_1 \Delta x_{t-1} + \dots + \theta_{p-1} \Delta x_{t-p+1}) + \varepsilon_t,$$

где $\varphi = \rho - 1$, так что гипотеза $H_0: \rho = 1$ в $(\#)$ равносильна гипотезе $H_0: \varphi = 0$ в $(\#\#)$.

В качестве альтернативной к $H_0: \rho = 1$ в $(\#)$ выступает гипотеза $H_A: \rho < 1$. При переходе от $(\#)$ к $(\#\#)$ она преобразуется в гипотезу $H_A: \varphi < 0$.

При этом, значение “ t -статистики” для проверки гипотезы $H_0: \rho = 1$ в $(\#)$ численно равно значению “ t -статистики” для проверки гипотезы $H_0: \varphi = 0$ в $(\#\#)$.

Пример

Для ряда GNP оценивание модели

$$\Delta x_t = \alpha + \beta t + \varphi x_{t-1} + \theta_1 \Delta x_{t-1} + \varepsilon_t$$

(обычным методом наименьших квадратов) приводит к следующему результату:

ADF Test Statistic	-4.117782	1% Critical Value*	-4.1219
		5% Critical Value	-3.4875
		10% Critical Value	-3.1718

*MacKinnon critical values for rejection of hypothesis of a unit root.

Augmented Dickey-Fuller Test Equation

Dependent Variable: D(X)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X(-1)	-0.249792	0.060662	-4.117782	0.0001
D(X(-1))	0.630066	0.109453	5.756490	0.0000
C	56.32136	13.18303	4.272264	0.0001
@TREND(1947:1)	1.304300	0.315357	4.135949	0.0001

Гипотеза единичного корня отвергается: значение t -статистики для проверки гипотезы $H_0: \phi = 0$ оказывается ниже 5% критического значения, вычисленного по формуле Маккиннона, и близко к 1% критическому значению.

В связи с последним примером, следует особо отметить, что использование расширенной модели предполагает, что количество запаздывающих разностей, включенных в правую часть, исчерпывает временную зависимость, так что ε_t — независимые случайные величины. В то же время, не следует включать в правую часть излишних запаздывающих разностей, т.к. это снижает мощность критериев как по причине оценивания дополнительных параметров, так и по причине уменьшения используемого количества наблюдений.

Для определения надлежащей глубины запаздываний следует начинать с относительно большого порядка $p = p^*$, а затем опираться на то обстоятельство, что хотя при наличии единичного корня распределения оценки $\hat{\phi}$ и t -статистики для проверки гипотезы $\phi = 0$ нестандартны, распределения оценок коэффициентов $\theta_1, \dots, \theta_{p-1}$ все же являются асимптотически нормальными. Поэтому можно сначала проверить гипотезу о том, что $\theta_{p^*-1} = 0$, используя обычную t -статистику и критические точки соответствующего t -распределения Стьюдента. Если эта гипотеза не отклоняется, то далее проверяем гипотезу $\theta_{p^*-1} = \theta_{p^*-2} = 0$, используя F -критерий и процентные точки F -распределения Фишера, и.т.д. После этого производится обычная диагностика адекватности подобранной модели.

Пример

Продолжим предыдущий пример. Если взять первоначально $p^* = 5$, то получаем:

ADF Test Statistic	-2.873575	1% Critical Value*	-4.1314
		5% Critical Value	-3.4919
		10% Critical Value	-3.1744

*MacKinnon critical values for rejection of hypothesis of a unit root.

Augmented Dickey-Fuller Test Equation

Dependent Variable: D(X)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X(-1)	-0.266169	0.092626	-2.873575	0.0060
D(X(-1))	0.546230	0.133521	4.090958	0.0002
D(X(-2))	0.183918	0.149711	1.228486	0.2253
D(X(-3))	-0.020254	0.152201	-0.133077	0.8947

D(X(-4))	-0.058683	0.148061	-0.396345	0.6936
C	59.45556	19.32396	3.076779	0.0035
@TREND(1947:1)	1.397409	0.482120	2.898469	0.0056

Поскольку здесь $t = -2.873575 > -3.1744$, то гипотеза единичного корня не отвергается даже при выборе 10% уровня значимости. В то же время, статистически незначимыми оказываются коэффициенты при трех последних запаздывающих разностях. P -значение F -статистики критерия для гипотезы о занулении этих трех коэффициентов равно 0.44. Поэтому можно обойтись без трех последних запаздывающих разностей, а такую модель мы только что оценивали, и в ней гипотеза единичного корня была отвергнута.

Критерии Дики - Фуллера фактически предполагают, что наблюдаемый ряд описывается моделью авторегрессии конечного порядка (возможно, с поправкой на детерминированный тренд). Как поступать в случае, когда ряд x_t имеет тип $ARMA(p, q)$ с $q > 0$?

Пусть $x_t \sim ARMA(p, q)$, так что $a(L)x_t = b(L)\varepsilon_t$,

где $a(L), b(L)$ – полиномы порядков p и q , и пусть оператор $b(L)$ обратим, так что процесс можно представить в виде процесса авторегрессии бесконечного порядка $c(L)x_t = \varepsilon_t$,

где

$$c(L)x_t = a(L)/b(L) = 1 + c_1L + c_2L^2 + \dots$$

В этом случае представление (# #) с конечным числом запаздываний в правой части заменяется бесконечным представлением

$$(\# \# \#) \quad \Delta x_t = \alpha + \beta t + \varphi x_{t-1} + (\theta_1 \Delta x_{t-1} + \theta_2 \Delta x_{t-2} + \dots) + \varepsilon_t$$

Однако все коэффициенты последнего невозможно оценить по конечному количеству наблюдений. Как выйти из этого положения?

В работе [Said, Dickey (1984)] было показано, что процесс $ARIMA(p, 1, q)$ с неизвестными p и q можно достаточно хорошо аппроксимировать некоторым процессом $ARI(p^*, 1)$ с $p^* < \sqrt[3]{T}$. Это дает возможность ограничиться в правой части (# # #) конечным числом запаздывающих разностей.

6.4. Краткий обзор критериев Дики – Фуллера

Под критерием Дики – Фуллера в действительности понимается группа критериев, объединенных одной идеей, предложенных и изученных в работах [Dickey (1976)], [Fuller (1976)], [Dickey, Fuller (1979)], [Dickey, Fuller (1981)]. В критериях Дики – Фуллера проверяемой (нулевой) является гипотеза о том, что исследуемый ряд x_t принадлежит классу DS (DS-гипотеза); альтернативная гипотеза – исследуемый ряд принадлежит классу TS (TS-гипотеза). Критерий Дики – Фуллера фактически предполагает, что наблюдаемый ряд описывается моделью авторегрессии первого порядка (возможно, с поправкой на линейный тренд). Критические значения зависят от того, какая статистическая модель оценивается и какая вероятностная модель в действительности порождает наблюдаемые значения. При этом рассматриваются следующие три пары моделей (SM – статистическая модель, statistical model; DGP – модель порождения данных, data generating process).

1) Если ряд x_t имеет детерминированный линейный тренд (наряду с которым может иметь место и стохастический тренд), то в такой ситуации берется пара

$$SM: \Delta x_t = \varphi x_{t-1} + \alpha + \beta t + \varepsilon_t, \quad t = 2, \dots, T,$$

DGP: $\Delta x_t = \alpha + \varepsilon_t, \quad t = 2, \dots, T, \quad \alpha \neq 0.$

В обоих случаях ε_t – независимые случайные величины, имеющие одинаковое нормальное распределение с нулевым математическим ожиданием.

Методом наименьших квадратов оцениваются параметры данной SM и вычисляется значение обычной t -статистики t_φ для проверки гипотезы $H_0 : \varphi = 0$. Полученное значение сравнивается с критическим уровнем t_{crit} , рассчитанным в предположении, что наблюдаемый ряд в действительности порождается данной моделью DGP (случайное блуждание со сносом). DS-гипотеза отвергается, если $t_\varphi < t_{crit}$. Критические уровни, соответствующие выбранным уровням значимости, можно взять из таблиц, приведенных в книгах [Fuller (1976)], [Fuller (1996)], если ряд наблюдается на интервалах длины $T = 25, 50, 100, 250, 500$. Если количество наблюдений T другое, то тогда можно вычислить приближенные критические значения, используя формулы, приведенные в работе [MacKinnon (1991)].

2) Если ряд x_t не имеет детерминированного тренда (но может иметь стохастический тренд) и имеет ненулевое математическое ожидание, то берется пара

SM: $\Delta x_t = \varphi x_{t-1} + \alpha + \varepsilon_t, \quad t = 2, \dots, T,$

DGP: $\Delta x_t = \varepsilon_t, \quad t = 2, \dots, T.$

Методом наименьших квадратов оцениваются параметры данной SM и вычисляется значение t -статистики t_φ для проверки гипотезы $H_0 : \varphi = 0$. Полученное значение сравнивается с критическим уровнем t_{crit} , рассчитанным в предположении, что наблюдаемый ряд в действительности порождается данной моделью DGP (случайное блуждание без сноса). DS-гипотеза отвергается, если $t_\varphi < t_{crit}$. Критические уровни, соответствующие выбранным уровням значимости, можно взять из таблиц, приведенных в книгах [Fuller (1976)], [Fuller (1996)], если ряд наблюдается на интервалах длины $T = 25, 50, 100, 250, 500$. Если количество наблюдений T другое, то тогда можно вычислить приближенные критические значения, используя формулы, приведенные в работе [MacKinnon (1991)].

3) Наконец, если ряд x_t не имеет детерминированного тренда (но может иметь стохастический тренд) и имеет нулевое математическое ожидание, то берется пара

SM: $\Delta x_t = \varphi x_{t-1} + \varepsilon_t, \quad t = 2, \dots, T,$

DGP: $\Delta x_t = \varepsilon_t, \quad t = 2, \dots, T.$

Методом наименьших квадратов оцениваются параметры данной SM и вычисляется значение t -статистики t_φ для проверки гипотезы $H_0 : \varphi = 0$. Полученное значение сравнивается с критическим уровнем t_{crit} , рассчитанным в предположении, что наблюдаемый ряд в действительности порождается данной моделью DGP (случайное блуждание без сноса). DS-гипотеза отвергается, если $t_\varphi < t_{crit}$. Критические уровни, соответствующие выбранным уровням значимости, можно взять из таблиц, приведенных в книгах [Fuller (1976)], [Fuller (1996)], если ряд наблюдается на интервалах длины $T = 25, 50, 100, 250, 500$. Если количество наблюдений T другое, то тогда можно вычислить приближенные критические значения, используя формулы, приведенные в работе [MacKinnon (1991)].

Неправильный выбор оцениваемой статистической модели может существенно отразиться на мощности критерия Дики – Фуллера. Например, если наблюдаемый ряд порождается моделью случайного блуждания со сносом, а статистические выводы производятся на основании результатов оценивания статистической модели без включения в ее правую часть трендовой составляющей, то тогда мощность критерия, основанная на статистике t_φ , стремится к нулю с возрастанием количества наблюдений

(см. [Perron (1988)]). С другой стороны, оцениваемая статистическая модель не должна быть и избыточной, поскольку это также ведет к уменьшению мощности критерия.

Формализованная процедура использования критериев Дики – Фуллера с последовательной проверкой возможности редукции статистической модели приведена в работе [Dolado, Jenkinson, Sosvilla – Rivero (1990)]; см. также [Enders (1995)]. Эта процедура будет рассмотрена ниже в разделе 6.6.

Если наблюдаемый ряд описывается моделью авторегрессии $a(L) x_t = \varepsilon_t$ более высокого (но конечного) порядка p , уравнение $a(z) = 0$ имеет не более одного единичного корня и не имеет корней внутри единичного круга, то тогда можно воспользоваться расширенным (augmented) критерием Дики – Фуллера. В каждой из трех рассмотренных выше ситуаций достаточно дополнить правые части оцениваемых статистических моделей запаздывающими разностями $\Delta x_{t-j}, j = 1, \dots, p-1$, так что оцениваются расширенные статистические модели

$$(1) \quad \text{SM: } \Delta x_t = \alpha + \beta t + \varphi x_{t-1} + \sum_{j=1}^{p-1} \theta_j \Delta x_{t-j} + \varepsilon_t, \quad t = p+1, \dots, T,$$

$$(2) \quad \text{SM: } \Delta x_t = \alpha + \varphi x_{t-1} + \sum_{j=1}^{p-1} \theta_j \Delta x_{t-j} + \varepsilon_t, \quad t = p+1, \dots, T,$$

$$(3) \quad \text{SM: } \Delta x_t = \varphi x_{t-1} + \sum_{j=1}^{p-1} \theta_j \Delta x_{t-j} + \varepsilon_t, \quad t = p+1, \dots, T.$$

Полученные при оценивании расширенных статистических моделей значения t -статистик t_φ для проверки гипотезы $H_0: \varphi = 0$ сравниваются с теми же критическими значениями t_{crit} , что и для нерасширенных моделей. DS-гипотеза отвергается, если $t_\varphi < t_{crit}$.

Расширенный критерий Дики – Фуллера может применяться и тогда, когда ряд x_t описывается смешанной моделью авторегрессии – скользящего среднего. Как было указано в работе [Said, Dickey (1984)], если ряд наблюдений x_1, \dots, x_T порождается моделью $\text{ARIMA}(p, 1, q)$ с $q > 0$, то его можно аппроксимировать моделью $\text{ARI}(p^*, 1) = \text{ARIMA}(p^*, 1, 0)$ с $p^* < \sqrt[3]{T}$ и применять процедуру Дики – Фуллера к этой модели.

Однако даже если ряд наблюдений x_1, \dots, x_T действительно порождается моделью авторегрессии $\text{AR}(p)$ конечного порядка p , то значение p обычно не известно, и его приходится оценивать на основании имеющихся наблюдений, а такое предварительное оценивание влияет на характеристики критерия. Поэтому при анализе данных приходится сначала выбирать значение $p = p_{\max}$ достаточно большим, так чтобы оно было не меньше истинного порядка p_0 авторегрессионной модели, описывающей ряд, или порядка p^* аппроксимирующей авторегрессионной модели, а затем пытаться понизить используемое значение p , апеллируя к наблюдениям.

Такое понижение может осуществляться, например, путем последовательной редукции расширенной модели за счет исключения из нее незначимых (на 10% уровне) запаздывающих разностей (GS-стратегия перехода от общего к частному) или путем сравнения (оцененных) полной и редуцированных моделей с различными $p \leq p_{\max}$ по информационному критерию Шварца (SIC). В работах [Hall (1994)] и [Ng, Perron (1995)] показано, что если $p_{\max} \geq p_0$, то тогда в пределе (при $T \rightarrow \infty$) SIC выбирает правильный порядок модели, а стратегия GS выбирает модель с $p \geq p_0$; при этом факт определения порядка модели на основании имеющихся данных не влияет на асимптотическое распределение статистики Дики – Фуллера. Таблицы критических

значений для конечных значений T , учитывающие порядок модели, приведены в работе [Cheung, Lay (1995)].

При практической реализации указанных двух подходов (GS и SIC), когда мы имеем лишь ограниченное количество наблюдений, эти две процедуры могут приводить к совершенно различным выводам относительно необходимого количества запаздываний в правой части статистической модели, оцениваемой в рамках расширенного критерия Дики – Фуллера. Так, при анализе динамики валового внутреннего продукта (GDP) США по годовым данным на периоде с 1870 по 1994 гг. (см. [Murray, Nelson (2000)]), выбрав $p_{\max} = 8$, авторы получили при использовании GS-стратегии значение $p = 6$, тогда как по SIC было выбрано значение $p = 1$. В подобных конфликтных ситуациях можно для контроля ориентироваться также на достижение некоррелированности по LM-критерию остатков от оцененной модели (см. [Holden, Perman (1994)]). Заметим, однако, что в недавней статье [Taylor (2000)] автор приходит к выводам, отличающимся от выводов Ng и Perron: при конечных выборках расширенные критерии Дики – Фуллера очень чувствительны и к форме детерминистских переменных и к принятой структуре запаздываний. Это, в свою очередь, ведет к отклонениям от номинальных уровней значимости критериев Дики – Фуллера.

6.5. Некоторые другие сочетания DGP и SM

Рассмотрим теперь следующий естественный вопрос: что будет, если мы оцениваем

SM: $x_t = \alpha + a_1 x_{t-1} + \varepsilon_t$,

а процесс порождения данных

DGP: $x_t = \alpha + x_{t-1} + \varepsilon_t$, $\alpha \neq 0$ (случайное блуждание со сносом).

В этом случае при больших t возникающий в DGP детерминированный тренд “забивает” стохастическую составляющую $\sum_{i=1}^t \varepsilon_i$, и поведение переменной x_{t-1} в SM

похоже “в целом” на поведение детерминированной переменной $\alpha(t-1)$. Как результат – распределения оценок для α и a_1 оказываются асимптотически нормальными; но тогда, в принципе, можно было бы в качестве приближения использовать стандартную технику статистического анализа, т.е. использовать критические значения t -отношения, взятые из таблиц распределения Стьюдента. Однако если $\alpha \neq 0$ близко к нулю, то при конечных T распределение t -отношения ближе к распределению, указанному Фуллером для случая $\alpha = 0$, чем к нормальному распределению.

Пример

Для смоделированной реализации WALK_2 случайного блуждания со сносом 0.2 оценивание статистической модели SM: $x_t = \alpha + a_1 x_{t-1} + \varepsilon_t$ дает

Dependent Variable: WALK_2

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.173019	0.160495	1.078031	0.2865
WALK_2(-1)	0.991851	0.045395	21.84958	0.0000

так что интересующая нас t -статистика принимает значение $t = (0.991851 - 1)/0.045395 = -0.180$. 5% критическое значение t -распределения Стьюдента с $(n-p) = (49-2) = 47$ степенями свободы равно -1.68 , тогда как 5% критическое значение по Фуллеру, предназначенное для случая DGP: $x_t = x_{t-1} + \varepsilon_t$ ($\alpha = 0$), равно -2.92 , что приводит к более редкому отвержению гипотезы единичного корня. Впрочем, гипотеза о наличии

единичного корня здесь не отвергается при использовании любого из двух критических значений – 1.68 и – 2.92.

Оценивание смоделированной реализации ST_2 процесса $x_t = 0.2 + 0.8 x_{t-1} + \varepsilon_t$ (стационарный процесс с математическим ожиданием, равным 1) при $T = 50$ дает

Dependent Variable: ST_2

Variable Coefficient Std. Error t-Statistic Prob.

C 0.166899 0.159693 1.045128 0.3013

ST_2(-1) 0.793680 0.091904 8.635959 0.0000

так что интересующая нас t -статистика принимает значение $t = (0.793680 - 1)/0.091904 = -2.245$. Использование критического значения – 1.68 приводит к отвержению гипотезы единичного корня, тогда как использование критического значения – 2.92 не дает возможности отвергнуть эту гипотезу.

Отметим еще одно важное обстоятельство.

Опять рассмотрим смоделированную реализацию ST_1 ряда $x_t = 0.8 x_{t-1} + \varepsilon_t$. Если мы будем проверять для ряда x_t гипотезу единичного корня, то, как теперь ясно, можем исходить либо из SM: $x_t = a_1 x_{t-1} + \varepsilon_t$ (подозревая, что DGP – простое случайное блуждание) либо из SM: $x_t = a + a_1 x_{t-1} + \varepsilon_t$ (подозревая, что DGP – случайное блуждание со сносом).

В первом случае 5% критическое значение t -статистики (при $T = 50$) находим по таблицам Фуллера: оно равно – 1.95. Во втором случае критические значения определяются либо в соответствии с нормальной теорией, так что $t_{\text{крит}} = -1.68$ (в предположении, что снос – ненулевой), либо по Фуллеру – тогда $t_{\text{крит}} = -2.92$.

Оценивание SM в первом случае дает $t = -2.314$. Гипотеза единичного корня отвергается.

Оценивание SM во втором случае дает $t = -2.298$. Если исходить из предположения, что в DGP снос ненулевой, то $t < t_{\text{крит}} = -1.68$, и гипотеза единичного корня отвергается. Если же исходить из предположения, что в DGP сноса нет, то $t > t_{\text{крит}} = -2.92$, и гипотеза единичного корня не отвергается.

Такой неожиданный результат объясняется тем, что пополнение статистической модели (SM) дополнительными регрессорами требует их оценивания, что снижает, в конечном счете, мощность критерия. Поэтому желательно при проверке гипотезы единичного корня оценивать SM, выбираемую “без запаса”. Однако при отсутствии информации о том, равен нулю снос в DGP или нет, при отклонении гипотезы единичного корня следует опираться на консервативное значение, даваемое таблицами Фуллера. Иначе мы можем ошибочно отвергать эту гипотезу более, чем в 5% случаев, если в действительности снос в DGP отсутствует.

6.6. Ряды с квадратичным трендом.

Надо рассмотреть еще и случай, когда по поведению траектории ряда можно подозревать наличие у него детерминированного квадратичного тренда.

Здесь наличие единственного¹ единичного корня может осуществляться уже в форме трех различных DGP:

(а) $x_t = x_{t-1} + \varepsilon_t$,

¹ Квадратичный тренд может возникать и в моделях с *двумя* единичными корнями. Эта ситуация рассматривается далее в разд. 6.8.4.

$$(б) \ x_t = \alpha + x_{t-1} + \varepsilon_t, \ \alpha \neq 0,$$

$$(в) \ x_t = \alpha + \beta t + x_{t-1} + \varepsilon_t, \ \beta \neq 0.$$

Последний случай гарантирует наличие квадратичного тренда; в двух других случаях возможна имитация такого тренда на не очень продолжительном периоде наблюдений.

Если строить проверку гипотезы единичного корня в рамках статистической модели

$$SM: \ x_t = \alpha + \beta t + a_1 x_{t-1} + \varepsilon_t, \ \beta \neq 0,$$

то распределение t -статистики при гипотезе $H_0: a_1 = 1$ будет различным, в зависимости от того, каким в действительности является DGP.

Как уже отмечалось в разд. 6.2, распределение этой t -статистики одно и то же для случаев (а) и (б), т.е. не зависит от того, $\alpha = 0$ или $\alpha \neq 0$.

Если DGP имеет форму (в) с $\beta \neq 0$, то указанная t -статистика имеет распределение, близкое к t -распределению (точнее, асимптотически нормальное $N(0, 1)$). Для конечных T можно обратиться к таблицам [Kwiatkowski, Schmidt (1990)].

Пример

Рассмотрим смоделированную реализацию WALK_3 случайного блуждания вокруг квадратичного тренда:

$$DGP: \ x_t = 0.2 + 0.1 t + x_{t-1} + \varepsilon_t.$$

В качестве статистической модели берем $SM: \ x_t = \alpha + \beta t + a_1 x_{t-1} + \varepsilon_t$; ее оценивание дает значение $\hat{a} = 0.989$ и $t = -0.775$. Использование t -распределения Стьюдента приводит к $t_{\text{крит}} = -1.68$, так что гипотеза единичного корня не отвергается. Использование критического значения Фуллера, соответствующее DGP с $\beta = 0$, дает $t_{\text{крит}} = -3.50$, так что гипотеза единичного корня не отвергается тем более.

При анализе смоделированной траектории ST_4 процесса

$$DGP: \ x_t = 0.12 + 0.13 t + 0.01 t^2 + 0.8 x_{t-1} + \varepsilon_t,$$

стационарного относительно того же самого квадратичного тренда, получаем $\hat{a} = 0.990$ и $t = -0.577$. Гипотеза единичного корня не отвергается как при использовании распределения Стьюдента, так и при использовании таблиц Фуллера.

6.7. Многовариантная процедура проверки гипотезы единичного корня

Долодо и др. ([Dolado, Jenkinson, Sosvilla-Rivero (1990)]) предложили многовариантную процедуру проверки гипотезы единичного корня с использованием критерия Дики – Фуллера, последовательно перебирающую различные комбинации оцениваемой статистической модели (SM) и процесса порождения данных (DGP). Ниже мы объясняем суть этой процедуры, считая для простоты, что рассматриваемый ряд порождается моделью AR(1), быть может, с поправкой на линейный тренд.

На шаге 1 процедуры Долодо оценивается статистическая модель, допускающая наличие тренда, содержащая в правой части уравнения константу и трендовую составляющую:

$$SM: \ \Delta x_t = \alpha + \beta t + \varphi x_{t-1} + \varepsilon_t, \quad t = 2, \dots, T,$$

и при использовании таблицы критических значений предполагается, что данные порождаются моделью

$$DGP: \ \Delta x_t = \alpha + \varepsilon_t, \quad t = 2, \dots, T.$$

Это естественная пара: реализация с видимым трендом (сносом). Критерий принадлежности ряда классу DS формулируется как критерий единичного корня (UR – Unit Root) в авторегрессионном представлении ряда. Проверяемой в рамках данной

статистической модели является гипотеза $H_0 : \varphi = 0$; альтернативная гипотеза $H_A : \varphi < 0$. Получаемое в результате оценивания такой расширенной модели значение t -статистики критерия Дики - Фуллера сравнивается с критическим значением, соответствующим предположению, что данные порождаются моделью случайного блуждания со сносом. Это критическое значение не зависит от того, $\alpha = 0$ или $\alpha \neq 0$.

Если гипотеза $H_0 : \varphi = 0$ отвергается этим критерием, то гипотеза о наличии единичного корня тем самым отвергается окончательно. Дело в том, что если $H_0 : \varphi = 0$ отвергнута при

DGP: $\Delta x_t = \alpha + \varepsilon_t$ (с $\alpha = 0$ или $\alpha \neq 0$),

то она тем более будет отвергнута при

DGP: $\Delta x_t = \alpha + \beta t + \varepsilon_t$, $\beta \neq 0$,

т.к. в последнем случае значение $t_{\text{крит}}$ выше (используется нормальное приближение).

Шаг 2.

Если на шаге 1 гипотеза $H_0 : \varphi = 0$ не была отвергнута, то возможны две причины:

- действительно, $\varphi = 0$;
- $\varphi \neq 0$, но гипотеза $H_0 : \varphi = 0$ не была отвергнута из-за того, что исходили из DGP с $\beta = 0$, тогда как в действительности имел место DGP: $\Delta x_t = \alpha + \beta t + \varepsilon_t$, $\beta \neq 0$.

В связи с последней возможностью, на шаге 2 производится проверка гипотезы

$H_0: \beta = 0$

в рамках

SM: $\Delta x_t = \alpha + \beta t + \varphi x_{t-1} + \varepsilon_t$,

но с

DGP: $\Delta x_t = \alpha + \beta t + \varepsilon_t$, $\beta \neq 0$.

Критические значения соответствующей t -статистики ($\tau_{\beta\tau}$ – в обозначениях Дики – Фуллера) указаны в статье [Dickey, Fuller (1981)]. В следующей таблице приведены 5% критические значения для $|t|$ в случае двухстороннего критерия и для t в случае одностороннего критерия.

n	Двухсторонний критерий	Односторонний критерий (против $\beta > 0$)
25	3.25	2.85
50	3.18	2.81
100	3.14	2.79
250	3.12	2.79

0		
5 0 0	3.11	2.78
∞	3.11	2.78

Если гипотеза $H_0: \beta = 0$ здесь не отвергнута, то это означает для нас, что на первом шаге гипотеза $\varphi = 0$ не была отвергнута не из-за использования критических значений, соответствующих DGP с $\beta = 0$.

Если же гипотеза $H_0: \beta = 0$ оказалась отвергнутой, то следует повторить проверку гипотезы $\varphi = 0$ в рамках

$$SM: \Delta x_t = \alpha + \beta t + \varphi x_{t-1} + \varepsilon_t,$$

но уже опираясь на

$$DGP: \Delta x_t = \alpha + \beta t + \varepsilon_t, \beta \neq 0.$$

Соответствующая t -статистика имеет (при $\beta \neq 0$) асимптотически нормальное $N(0, 1)$ распределение. Для конечных T можно обратиться к таблицам [Kwiatkowski, Schmidt (1990)]. И теперь уже, если гипотеза $\varphi = 0$ будет отвергнута, то отвергнута окончательно. Если же она не отвергнута, то принимаем модель

$$\Delta x_t = \alpha + \beta t + \varepsilon_t, \beta \neq 0.$$

Шаг 3

Мы попадаем на шаг 3, не отвергнув гипотезу единичного корня в рамках статистической модели

$$SM: \Delta x_t = \alpha + \beta t + \varphi x_{t-1} + \varepsilon_t.$$

Возможно, что это связано с пониженной мощностью критериев из-за включения в модель лишней объясняющей переменной t .

В связи с этим, на шаге 3 мы переходим к модели

$$SM: \Delta x_t = \alpha + \varphi x_{t-1} + \varepsilon_t$$

без трендовой составляющей и проверяем гипотезу $\varphi = 0$ (против $\varphi < 0$) в рамках этой SM. Критические значения соответствующей t -статистики берем опять у Фуллера (ситуация 2). Они получены в предположении

$$DGP: \Delta x_t = \varepsilon_t.$$

И опять, если гипотеза $H_0: \varphi = 0$ отвергается, то отвергается окончательно (по тем же причинам, что и на шаге 1).

Шаг 4

Если на шаге 3 гипотеза $\varphi = 0$ не отвергается, то выясняется причастность к этому включения в SM сноса α . С этой целью производится проверка гипотезы $\alpha = 0$ в рамках статистической модели

$$SM: \Delta x_t = \alpha + \varphi x_{t-1} + \varepsilon_t ,$$

но с

$$DGP: \Delta x_t = \varepsilon_t .$$

Критические значения соответствующей t -статистики ($\tau_{\alpha\mu}$ – в обозначениях Дики – Фуллера) указаны в статье [Dickey, Fuller (1981)]. В следующей таблице приведены 5% критические значения для $|t|$ в случае двухстороннего критерия и для t в случае одностороннего критерия.

n	Двухсторонний критерий	Односторонний критерий (против $\alpha > 0$)
25	2.97	2.61
50	2.89	2.56
100	2.86	2.54
250	2.84	2.53
500	2.83	2.52
∞	2.83	2.52

Если при этом гипотеза $\alpha = 0$ не отвергается, то мы не считаем тогда, что неотвержение $\varphi = 0$ на предыдущем этапе было связано с опорой на DGP с $\alpha = 0$.

Если же гипотеза $\alpha = 0$ оказалась отвергнутой, то производится повторная проверка гипотезы $H_0: \varphi = 0$ в рамках

$$SM: \Delta x_t = \alpha + \varphi x_{t-1} + \varepsilon_t ,$$

но с опорой на

$$DGP: \Delta x_t = \alpha + \varepsilon_t \text{ с } \alpha \neq 0 .$$

В этом случае t -статистика для $\alpha = 0$ опять асимптотически нормальна, и опираясь на ее значение, мы либо отвергаем гипотезу $H_0: \varphi = 0$ окончательно либо принимаем модель

$\Delta x_t = \alpha + \varepsilon_t$ с $\alpha \neq 0$. Следует только помнить о том, что при конечных T при значениях α , близких к нулю, распределение этой статистики ближе к

распределению, указанному Фуллером для случая $\alpha = 0$, чем к нормальному распределению.

Шаг 5

Наконец, если и на шаге 4 гипотеза $H_0: \varphi = 0$ не была отвергнута, остается последняя возможность сделать это в рамках статистической модели

$$SM: \Delta x_t = \varphi x_{t-1} + \varepsilon_t.$$

Критические значения t -статистики для $H_0: \varphi = 0$ находятся по таблицам Фуллера (случай 1). И теперь уже, каждое из двух возможных решений – окончательное :

$H_0: \varphi = 0$ отвергается \rightarrow Единичного корня нет;

$H_0: \varphi = 0$ не отвергается $\rightarrow \Delta x_t = \varepsilon_t$. (Точнее, $\Delta x_t = \theta_1 \Delta x_{t-1} + \dots + \theta_{p-1} \Delta x_{t-p+1} + \varepsilon_t$.)

Замечание

Построенный алгоритм отнюдь не лишен недостатков. Помимо того, что здесь не контролируется уровень значимости критерия проверки гипотезы единичного корня, возникают сложности и с интерпретацией результатов, что будет видно из последующих примеров.

Процедуру Доладо и др. можно представить схематически в виде дерева решений, приведенного ниже.

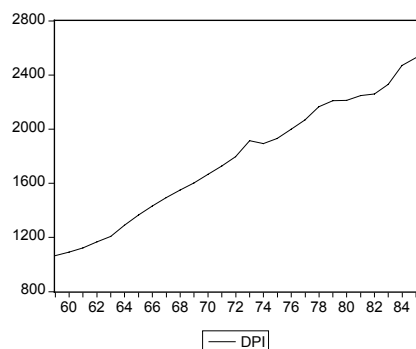
В представленной схеме модели перенумерованы следующим образом.

- Модель 1: $\Delta x_t = \alpha + \beta t + \varphi x_{t-1} \left(+ \sum_{j=1}^{p-1} \theta_j \Delta x_{t-j} \right) + \varepsilon_t$
- Модель 2: $\Delta x_t = \alpha + \varphi x_{t-1} \left(+ \sum_{j=1}^{p-1} \theta_j \Delta x_{t-j} \right) + \varepsilon_t$
- Модель 3: $\Delta x_t = \alpha \left(+ \sum_{j=1}^{p-1} \theta_j \Delta x_{t-j} \right) + \varepsilon_t$

Приведем теперь пример использования процедуры Доладо и др.

Пример

Обратимся к данным о совокупном годовом располагаемом доходе в США за период с 1959 по 1985 годы (в млрд долларов, в ценах 1982 г.). График этого ряда имеет вид



Принимая во внимание все изложенное ранее, здесь очевидна необходимость различения модели случайного блуждания со сносом и процесса, стационарного относительно линейного тренда. Последуем процедуре Доладо и др.

Шаг 1

Оцениваем статистическую модель

$$SM: \Delta x_t = \alpha + \beta t + \phi x_{t-1} + \theta_1 \Delta x_{t-1} + \dots + \theta_{p-1} \Delta x_{t-p+1} + \varepsilon_t.$$

Сравнение по критерию Шварца указывает в пользу исключения из правой части запаздывающих разностей, так что останавливаемся на модели

$$SM: \Delta x_t = \alpha + \beta t + \phi x_{t-1} + \varepsilon_t.$$

Оцененная модель ($T = 26$):

$$\Delta x_t = 461.338 + 25.857 t - 0.448 x_{t-1} + e_t;$$

t -статистика для проверки гипотезы $H_0: \phi = 0$ равна $t_\phi = -2.640$. Критическое (5%) значение выбирается при

$$DGP: \Delta x_t = \alpha + \varepsilon_t, \alpha \neq 0,$$

и равно $t_{\text{крит}} = -3.59$ (для $T = 26$). Наблюдаемое значение t_ϕ больше $t_{\text{крит}}$ → Гипотеза единичного корня не отвергается.

Шаг 2

Статистическая модель та же, но в качестве DGP рассматривается

$$DGP: \Delta x_t = \alpha + \beta t + \varepsilon_t.$$

Проверяемая гипотеза

$$H_0: \beta = 0.$$

Статистика критерия $\tau_{\beta t} = 2.680$. Критическое (5%) значение одностороннего критерия ($H_A: \beta > 0$) равно 2.85 → гипотеза $H_0: \beta = 0$ не отвергается.

Шаг 3

Оцениваем модель

$$SM: \Delta x_t = \alpha + \phi x_{t-1} + \varepsilon_t;$$

в качестве DGP рассматривается

$$DGP: \Delta x_t = \varepsilon_t.$$

Проверяется гипотеза $H_0: \varphi = 0$. Оцененная модель: $\Delta x_t = 47.069 + 0.00522 x_{t-1} + e_t$; t -статистика $t_\varphi = 0.335$. Критическое (5%) значение $t_{\text{крит}} = -2.98 \rightarrow$ гипотеза единичного корня не отвергается.

Шаг 4

SM: $\Delta x_t = \alpha + \varphi x_{t-1} + \varepsilon_t$, DGP: $\Delta x_t = \varepsilon_t$, $H_0: \alpha = 0$. Оцененная модель: $\Delta x_t = 47.069 + 0.00522 x_{t-1} + e_t$; t -статистика $t_\alpha = 1.682 < t_{\text{крит}} = 2.61 \rightarrow$ Гипотеза $H_0: \alpha = 0$ не отвергается.

Шаг 5

SM: $\Delta x_t = \varphi x_{t-1} + \varepsilon_t$, DGP: $\Delta x_t = \varepsilon_t$, $H_0: \varphi = 0$. Оцененная модель: $\Delta x_t = 0.03070 x_{t-1} + e_t$; t -статистика $t_\varphi = 7.987 > t_{\text{крит}} = -1.95 \rightarrow$ Гипотеза $H_0: \varphi = 0$ не отвергается. \rightarrow Окончательная модель:

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

Интерпретация

Если мы соглашаемся с несомненной тенденцией возрастания совокупного располагаемого дохода с течением времени (по крайней мере, в США), то принятую нами на последнем шаге модель вряд ли можно считать удовлетворительной: случайное блуждание без сноса должно со временем обнаружить убывание значений ряда.

Возможная причина этого – неотклонение гипотезы $H_0: \beta = 0$ на шаге 2. (Заметим, что там разница между наблюдаемым и критическим значениями t -статистики была довольно небольшой: $\tau_{\beta t} = 2.680$, $t_{\text{крит}} = 2.85$.) Если возвратиться к шагу 2 и изменить решение в пользу отклонения гипотезы $H_0: \beta = 0$, так что тогда $\beta \neq 0$, то гипотеза $H_0: \varphi = 0$ проверяется в рамках пары

$$\text{SM: } \Delta x_t = \alpha + \beta t + \varphi x_{t-1} + \varepsilon_t, \text{ DGP: } \Delta x_t = \alpha + \beta t + \varepsilon_t, \beta \neq 0.$$

В таком случае статистика t_φ имеет асимптотически нормальное $N(0, 1)$ распределение, 5% критическое значение одностороннего критерия приблизительно равно $t_{\text{крит}} = -1.645$. У нас же наблюдаемое значение $t_\varphi = -2.640$ (см. шаг 1), так что гипотеза единичного корня отвергается, и мы имеем тогда дело с процессом, стационарным относительно линейного тренда:

$$\Delta x_t = 461.338 + 25.857 t - 0.448 x_{t-1} + \varepsilon_t.$$

6.8. Обзор некоторых других процедур

6.8.1. Критерий Филлипса – Перрона

Этот критерий, предложенный в работе [Phillips, Perron (1988)], сводит проверку гипотезы о принадлежности ряда x_t классу DS к проверке гипотезы $H_0: \varphi = 0$ в рамках статистической модели

$$\text{SM: } \Delta x_t = \alpha + \beta t + \varphi x_{t-1} + u_t, \quad t = 2, \dots, T,$$

где, как и в критерии Дики – Фуллера, параметры α и β могут быть взяты равными нулю. Однако, в отличие от критерия Дики – Фуллера, случайные составляющие u_t с нулевыми математическими ожиданиями могут быть автокоррелированными (с достаточно быстрым убыванием автокорреляционной функции), иметь различные

дисперсии (гетероскедастичность) и не обязательно нормальные распределения (но такие, что $E|u_t|^\delta \leq C < \infty$ для некоторого $\delta > 2$). Тем самым, в отличие от критерия Дики – Фуллера, к рассмотрению допускается более широкий класс временных рядов.

Критерий Филлипса – Перрона основывается на t -статистике для проверки гипотезы $H_0: \varphi = 0$ в рамках указанной статистической модели, но использует вариант этой статистики Z_t , скорректированный на возможную автокоррелированность и гетероскедастичность ряда u_t . При вычислении статистики Z_t приходится оценивать так называемую **“долговременную” (“long-run”) дисперсию** ряда u_t , которая определяется как

$$\lambda^2 = \lim_{T \rightarrow \infty} T^{-1} E(u_1 + \dots + u_T)^2.$$

Если u_t^* – остатки от оцененной (методом наименьших квадратов) статистической модели $\Delta x_t = \varphi x_{t-1} + \alpha + \beta t + u_t$, $t = 2, \dots, T$, то в качестве оценки $(\lambda^2)^*$ для λ^2 можно взять оценку [Newey, West (1987)]

$$(\lambda^2)^* = \gamma_0^* + 2 \sum_{j=1}^l \left[1 - \frac{j}{l+1} \right] \gamma_j^*,$$

где

$$\gamma_j^* = T^{-1} \sum_{t=j+1}^l u_t^* u_{t-j}^*$$

j -я выборочная автоковариация ряда u_t . Если и l и T стремятся к бесконечности, но так, что $(l/T^{1/4}) \rightarrow 0$, то тогда $(\lambda^2)^*$ – состоятельная оценка для λ^2 (см. [Phillips (1987)]) и асимптотические распределения статистики Z_t совпадают с соответствующими асимптотическими распределениями статистики t_φ в критерии Дики – Фуллера. Поскольку реально мы имеем лишь конечное количество наблюдений, встает вопрос о выборе количества используемых лагов l в оценке Newey – West (параметр l называют **“шириной окна”** – window size). Этот вопрос достаточно важен, т.к. недостаточная ширина окна ведет к отклонениям от номинального размера критерия (уровня значимости). В то же время, увеличение ширины окна для избежания отклонений от номинального размера критерия ведет к падению мощности критерия. Таким образом, выбор какой-то конкретной ширины окна является компромиссом между двумя этими противоположными тенденциями.

Целый ряд исследований в этом направлении (сюда относятся, например, работы [Phillips, Perron (1988)], [Schwert (1989)]) не привел к какому-либо простому правилу выбора значения l .

Часто при выборе этого параметра пользуются рекомендациями [Schwert (1989)], полагая $l = [K \times (T/100)^{1/4}]$, где $[a]$ – целая часть числа a , а значение K полагается равным 4 для квартальных и равным 12 для месячных данных. Другое правило выбора значения l реализованное, в частности, в пакете EVIEWS, состоит в выборе значения $l = [4 \times (T/100)^{2/9}]$ ([Newey, West (1994)]). Некоторые авторы рекомендуют не опираться только лишь на длину ряда, а учитывать при выборе l количество значимых автокорреляций ряда.

Критические значения для статистики Z_t берутся из тех же таблиц [Fuller (1976)] или вычисляются по формулам [MacKinnon(1991)].

Заметим также, что если ряд x_t представляется моделью IMA(1, q), то тогда это значение q и следует использовать в качестве параметра l в оценке Newey-West. Если

при этом $q = 1$, так что $\Delta x_t = \varepsilon_t + b_1 \varepsilon_{t-1}$, то при $b_1 > 0$ критерий Филлипса-Перрона имеет более высокую мощность, чем критерий Дики – Фуллера, при одновременном уменьшении вероятности ошибки первого рода. В то же время, при $b_1 < 0$ высокая мощность критерия Филлипса-Перрона достигается за счет значительного возрастания ошибки первого рода, так что этот критерий не рекомендуется применять при $b_1 < 0$ (он будет слишком часто ошибочно отвергать гипотезу о принадлежности ряда классу DS).

Пример

В рассмотренном ранее примере с GNP оценивание модели

$$\Delta x_t = \alpha + \beta t + \varphi x_{t-1} + \theta_1 \Delta x_{t-1} + \varepsilon_t$$

привело к следующему результату (см. разд. 6.3):

ADF Test Statistic	-4.117782	1% Critical Value*	-4.1219
		5% Critical Value	-3.4875
		10% Critical Value	-3.1718

Гипотеза единичного корня отвергается: значение t -статистики для проверки гипотезы $H_0: \varphi = 0$ оказывается ниже 5% критического значения, вычисленного по формуле Маккиннона, и близко к 1% критическому значению.

В то же время, если взять первоначально AR модель с $p_{\max} = 5$, то получаем:

ADF Test Statistic	-2.873575	1% Critical Value*	-4.1314
		5% Critical Value	-3.4919
		10% Critical Value	-3.1744

Augmented Dickey-Fuller Test Equation

Dependent Variable: D(X)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X(-1)	-0.266169	0.092626	-2.873575	0.0060
D(X(-1))	0.546230	0.133521	4.090958	0.0002
D(X(-2))	0.183918	0.149711	1.228486	0.2253
D(X(-3))	-0.020254	0.152201	-0.133077	0.8947
D(X(-4))	-0.058683	0.148061	-0.396345	0.6936
C	59.45556	19.32396	3.076779	0.0035
@TREND(1947:1)	1.397409	0.482120	2.898469	0.0056

Поскольку здесь $t = -2.873575 > -3.1744$, то гипотеза единичного корня не отвергается даже при выборе 10% уровне значимости. В то же время, статистически незначимыми оказываются коэффициенты при трех последних запаздывающих разностях. P -значение F -статистики критерия для гипотезы о занулении этих трех коэффициентов равно 0.44. Поэтому можно обойтись без трех последних запаздывающих разностей, а такую модель мы только что рассматривали, и в ней гипотеза единичного корня была отвергнута.

Посмотрим, что дает здесь применение критерия Филлипса – Перрона. Использование рекомендации [Newey, West (1994)] по выбору ширины окна дает значение $l = [4 \times (T/100)^{2/9}] = 3$; в результате получаем

PP Test Statistic	-2.871178	1% Critical Value*	-4.1190
		5% Critical Value	-3.4862
		10% Critical Value	-3.1711

Lag truncation for Bartlett kernel: 3 (Newey-West suggests: 3)

Residual variance with no correction	29.28903
Residual variance with correction	54.87482

Phillips-Perron Test Equation

Dependent Variable: D(GNP)

Method: Least Squares

Sample(adjusted): 1947:2 1961:4

Included observations: 59 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
GNP(-1)	-0.153024	0.072723	-2.104212	0.0399
C	38.33211	15.97824	2.399020	0.0198
@TREND(1947:1)	0.806326	0.378145	2.132322	0.0374

Статистические выводы, полученные при применении критерия Перрона с шириной окна, выбранной в соответствии с рекомендациями [Newey, West (1994)], противоположны выводам, полученным при применении расширенного критерия Дики – Фуллера с включением в правую часть одной запаздывающей разности.

Пример

Сравним результаты применения критериев Дики – Фуллера и Филлипса – Перрона на реализациях ST_1, ST_2, ST_3: для ST_1 и ST_2 – в паре

DGP: $x_t = x_{t-1} + \varepsilon_t$, **SM:** $x_t = \alpha + a_1 x_{t-1} + \varepsilon_t$;

для ST_3 – в паре

DGP: $x_t = \alpha + x_{t-1} + \varepsilon_t$ (или **DGP:** $x_t = x_{t-1} + \varepsilon_t$),

SM: $x_t = \alpha + \beta t + a_1 x_{t-1} + \varepsilon_t$.

Для статистик этих критериев используем обозначения DF и PP(*l*), соответственно, где *l* – ширина окна, используемая при построении статистики Филлипса – Перрона и выбираемая в соответствии с рекомендациями [Newey, West (1994)].

	<i>n</i> = 50	<i>n</i> = 100
ST_1	DF = -2.298 > $t_{\text{крит}10\%} = -2.60$, PP(3) = -2.394 > $t_{\text{крит}10\%} = -2.60$, гипотеза единичного корня не отвергается на 10% уровне	DF = -3.238 < $t_{\text{крит}5\%} = -2.89$, PP(4) = -3.399 < $t_{\text{крит}5\%} = -2.89$, гипотеза единичного корня отвергается на 5% уровне
ST_2	DF = -2.387 > $t_{\text{крит}10\%} = -2.60$, PP(3) = -2.322 > $t_{\text{крит}10\%} = -2.60$, гипотеза единичного корня не отвергается на 10% уровне	DF = -3.217 < $t_{\text{крит}5\%} = -2.89$, PP(4) = -3.364 < $t_{\text{крит}5\%} = -2.89$, гипотеза единичного корня отвергается на 5% уровне
ST_3	DF = -2.687 > $t_{\text{крит}10\%} = -3.18$, PP(3) = -2.755 > $t_{\text{крит}10\%} = -3.18$, гипотеза единичного корня не отвергается на 10% уровне	DF = -3.207 < $t_{\text{крит}10\%} = -3.15$, PP(4) = -3.368 < $t_{\text{крит}10\%} = -3.15$, гипотеза единичного корня отвергается на 10% уровне

Статистические выводы, полученные с применением статистик DF и PP, здесь совпадают и указывают на возрастание мощности критериев при увеличении количества наблюдений.

6.8.2. Критерий Лейбурна

В работе [Leypourne (1995)] предлагается вычислять значения статистики критерия Дики – Фуллера DF для исходного ряда x_t и для ряда, получаемого из исходного обращением времени, и затем взять максимум DF_{\max} из двух полученных значений. Лейбурн изучил асимптотическое распределение статистики DF_{\max} и построил таблицы критических значений при $T = 25, 50, 100, 200, 400$ для моделей с (линейным) трендом и без тренда. Таблицы получены моделированием в предположении независимости и одинаковой распределенности ошибок (инноваций). Однако автор утверждает, что ими можно пользоваться и в рамках расширенного варианта критерия Дики – Фуллера. Критерий Лейбурна обладает несколько большей мощностью по сравнению с критерием Дики – Фуллера.

Пример

При анализе стационарного ряда ST_3 по 100 наблюдениям мы получили значение статистики Дики – Фуллера $DF = -3.207$. Для обращенного ряда значение статистики Дики – Фуллера равно -3.352 . Максимум из этих двух значений, равный -3.207 , остается выше 5% критического уровня -3.45 , рассчитываемого по таблицам Фуллера. Однако 5% критический уровень для максимума приблизительно равен (по Лейбурну) -3.15 , и это дает возможность отвергнуть гипотезу единичного корня для ряда ST_3 уже на 5% уровне.

6.8.3. Критерий Шмидта – Филлипса.

В работе [Schmidt, Phillips (1992)] авторы строят критерий для проверки гипотезы DS (в форме гипотезы единичного корня) в рамках модели

$$x_t = \psi + \xi t + w_t,$$

где

$$w_t = \beta w_{t-1} + \varepsilon_t, \quad t = 2, \dots, T.$$

Это удобно тем, что здесь в любом случае ($\beta = 1$ или $\beta \neq 1$) параметр ψ представляет уровень, а параметр ξ представляет тренд. При этом распределения статистик критерия и при нулевой (DS) и при альтернативной (TS) гипотезах не зависят от мешающих параметров ψ , ξ и σ_ε . Асимптотические распределения выводятся при тех же условиях, что и в критерии Филлипса – Перрона, и при ширине окна l порядка $T^{1/2}$. Вместо линейного тренда в модели можно использовать и полиномиальный тренд. Более полное описание этого критерия и таблицу критических значений можно найти в [Maddala, Kim (1998), стр.85]. Здесь мы ограничимся только рассмотрением примера его применения.

Пример

Опять обращаясь к анализу ряда ST_3 по 100 наблюдениям, находим значение статистики критерия Шмидта – Филлипса: -3.12 . В то же время, 5% критическое значение равно -3.06 . Это дает возможность отвергнуть гипотезу единичного корня на 5% уровне.

6.8.4. Критерий DF-GLS

Этот критерий, асимптотически более мощный, чем критерий Дики – Фуллера, был предложен в работе [Elliott, Rothenberg, Stock (1996)]. Критерий DF-GLS проверяет (см. [Maddala, Kim (1998)]) нулевую гипотезу $a_0 = 0$ в модели

$$\Delta y_t^d = a_0 y_{t-1}^d + a_1 \Delta y_{t-1}^d + \dots + a_p \Delta y_{t-p}^d + error,$$

где y_t^d - “локально детрендрованный” ряд (подробности см. в цитированной работе).

Пример

Продолжая предыдущий пример, вычисляем статистику критерия DF-GLS. Ее значение равно – 3.246, что меньше 5% критического уровня – 2.89. Гипотеза единичного корня отвергается на 5% уровне, причем более уверенно, чем в случаях критериев Лейбурна и Шмидта-Филлипса.

6.8.5. Критерий Квятковского – Филлипса – Шмидта – Шина (KPSS)

Этот критерий, предложенный в работе [Kwiatkowski, Phillips, Schmidt, Shin (1992)], в качестве нулевой берет гипотезу TS. Рассмотрение ведется в рамках модели

Ряд = Детерминированный тренд + Стохастический тренд + Стационарная ошибка.

Стохастический тренд представляется случайным блужданием, и нулевая гипотеза предполагает, что дисперсия инноваций, порождающих это случайное блуждание, равна нулю. Альтернативная гипотеза соответствует предположению о том, что эта дисперсия отлична от нуля, так что анализируемый ряд принадлежит классу DS рядов. В такой формулировке предложенный критерий является LM критерием для проверки указанной нулевой гипотезы.

Как и в критерии Филлипса-Перрона, требования на ошибки здесь менее строгие, чем в критерии Дики - Фуллера. Однако при применении данного критерия возникает проблема выбора ширины окна l в оценке Newey-West, поскольку значения статистики критерия довольно чувствительны к значению l . Сами авторы в цитируемой статье рассматривают варианты выбора ширины окна, следующие рекомендациям Шверта (см. [Schwert (1989)]).

Подробное описание критерия KPSS можно найти вместе с таблицей критических значений в [Maddala, Kim (1998), стр.120-122].

Пример

При анализе ряда ST_3 по 100 наблюдениям значение статистики критерия KPSS с $l = 3$ равно 0.157. В рамках этого критерия нулевая гипотеза о том, что мы имеем дело с TS рядом, отвергается, если наблюдаемое значение статистики критерия превышает критический уровень. По таблицам, предусматривающим наличие линейного тренда, находим: 5% критический уровень равен 0.146, так что TS гипотеза отвергается в пользу DS гипотезы. Такой вывод противоречит статистическим выводам, полученным при применении критериев Лейбурна, Шмидта-Филлипса и DF-GLS, и иллюстрирует трудности с различением TS и DS рядов, имеющих похожие реализации.

6.8.6. Процедура Кохрейна (отношение дисперсий)

Эта процедура, предложенная в работе [Cochrane (1998)], основывается на изучении характера поведения отношения дисперсий

$$VR_k = \frac{V_k}{V_1}$$

(VR – variance ratio), где

$$V_k = \frac{1}{k} D(x_t - x_{t-k}) .$$

Если x_t – случайное блуждание, то тогда $VR_k = 1$, а если x_t – процесс, стационарный относительно линейного тренда (или просто стационарный), то тогда $VR_k \rightarrow 0$ при $k \rightarrow \infty$.

При работе с реальными данными дисперсии заменяются их состоятельными оценками и полученное отношение умножается еще на $T / (T - k + 1)$ для достижения несмещенности полученной оценки для VR_k . Затем строится график значений полученных оценок для VR_k при различных $k = 1, \dots, K$ и по поведению этого графика делаются выводы о принадлежности ряда классу TS или DS, имея в виду различия в поведении этого графика для этих двух классов временных рядов.

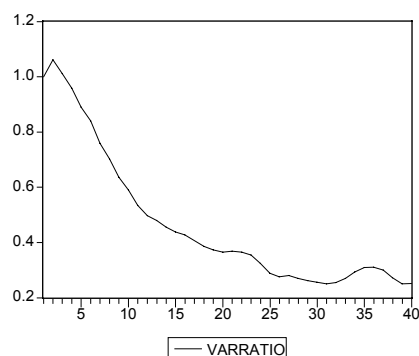
Другой вариант работы с реальными данными состоит в использовании равносильного представления статистики отношения дисперсий VR_k :

$$VR_k = 1 + 2 \sum_{j=1}^k \left(1 - \frac{j}{k+1}\right) r_j ,$$

где r_j – значение на лаге j автокорреляционной функции ряда разностей $\Delta x_t = x_t - x_{t-1}$.

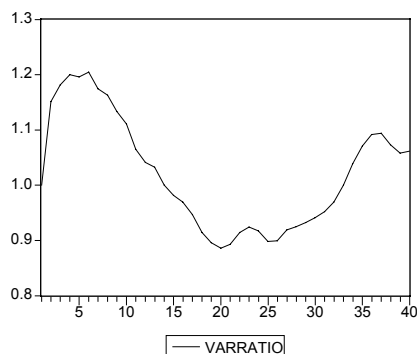
Пример

Обратимся опять к реализации ST_3 ряда, стационарного относительно линейного тренда, по которой оказалось затруднительным вынести определенное решение относительно принадлежности к классу TS или к классу DS модели, порождающей эту реализацию. Привлечем к решению этого вопроса процедуру Кохрейна.



Поведение отношения дисперсий говорит в пользу TS гипотезы.

Для сравнения приведем аналогичный график отношения дисперсий для реализации WALK_2 случайного блуждания со сносом:



Поведение отношения дисперсий указывает на то, что WALK_2 порождается DS моделью.

6.9. Некоторые проблемы, возникающие при различении TS и DS гипотез

6.9.1. Коррекция сезонности

В рассмотренных выше процедурах никак не затрагивался вопрос о коррекции сезонного поведения ряда, не снимаемого ни введением в модель линейного тренда ни путем дифференцирования ряда. Разумеется, данные, поступающие в распоряжение исследователя, уже могли быть подвергнуты сезонной коррекции соответствующими статистическими агентствами. Более того, во многих странах сырые (не скорректированные на сезонность) данные просто недоступны. В то же время, при анализе данных, подвергшихся сезонному сглаживанию с использованием фильтров или с использованием специфических методик правительственных агентств, существенно больше шансов классифицировать исследуемый ряд как DS (см., например, [Ghysels, Perron (1993)]), чем при анализе сырых данных. Поэтому некоторые авторы рекомендуют по возможности вообще избегать использования сезонно-сглаженных данных ([Davidson, MacKinnon (1993)]). Более предпочтительным является использование сырых данных и устранение из них сезонности путем оценивания регрессии сырого ряда на сезонные фиктивные (dummy) переменные D_1, \dots, D_{12} (если данные месячные) или D_1, \dots, D_4 (если данные квартальные). Остатки от оцененной регрессии образуют *очищенный ряд*, к которому можно применять изложенные выше методы. Теоретическое оправдание такого подхода при применении критерия Дики - Фуллера дано в работе [Dickey, Bell, Miller (1986)], где показано, что асимптотическое распределение статистики t_φ не изменяется при исключении из ряда детерминированных сезонных компонент.

6.9.2. Протяженность ряда и мощность критерия

Следует иметь в виду, что мощность критериев единичного корня зависит, в первую очередь, от фактической протяженности ряда во времени, а не от частоты, с которой производятся наблюдения. Соответственно, имея значения ряда за десятилетний период, мы не получаем выигрыша в мощности, анализируя месячные данные, а не квартальные или годовые. Результаты исследований в этом направлении можно найти, например, в статьях [Shiller, Perron (1985)] и [Perron (1989b)].

6.9.3. Проблема согласованности статистических выводов при различении TS и DS гипотез

При решении задачи отнесения рассматриваемого ряда к классу TS или к классу DS рядов двумя статистическими критериями, один из которых берет в качестве нулевой гипотезу TS, а другой – гипотезу DS, возможны следующие ситуации:

	H_0 : TS – не отвергается	H_0 : TS – отвергается
H_0 : DS – не отвергается	Исход 1	Исход 2
H_0 : DS – отвергается	Исход 3	Исход 4

Эти ситуации интерпретируются следующим образом:

- Исход 2 – в пользу DS модели.
- Исход 3 – в пользу TS модели.
- Исход 1 – невозможность принять решение из-за низкой мощности обоих критериев.
- Исход 4 – процесс порождения данных (DGP) не сводится к допускаемым используемыми критериями TS и DS моделям.

6.9.4. Наличие нескольких единичных корней

После появления работ [Fuller (1976)] и [Dickey, Fuller (1981)] было проведено довольно много практических исследований экономических временных рядов с целью решения вопроса о наличии или отсутствии единичных корней в моделях процессов, порождающих эти ряды.

При этом обычно сначала рассматривался сам временной ряд, и проводилась проверка его на нестационарность с использованием критериев Дики – Фуллера .

Если гипотеза единичного корня не отвергалась, то после этого переходили к рассмотрению ряда разностей и проверяли гипотезу единичного корня для этого ряда, применяя к ряду разностей процедуру Дики – Фуллера.

Если при анализе ряда разностей гипотеза единичного корня отвергалась, то принималось решение о том, что исходный ряд – интегрированный порядка 1. В противном случае переходили к рассмотрению ряда вторых разностей и проверяли гипотезу единичного корня для этого ряда. Обычно на этом шаге гипотеза единичного корня отвергалась и исходный ряд определялся как интегрированный порядка 2.

Более поздние исследования показали, что такого рода последовательные процедуры не обладают заявленными уровнями значимости, имея тенденцию к занижению

действительного количества единичных корней. И в таком несоответствии нет ничего удивительного: критерии Дики – Фуллера основаны на предположении, что если единичный корень и имеется, то тогда он единственный. Положение здесь похоже на другие ситуации, когда последовательная проверка гипотез идет не от общего к частному, а от частного к общему.

В связи с этим, для ситуаций, когда предполагаемая модель авторегрессии для анализируемого ряда может иметь порядок p выше первого, $p > 1$, в работе [Dickey, Pantula (1987)] была предложена процедура последовательной проверки гипотез о количестве единичных корней характеристического уравнения, построенная по принципу “от общего к частному”. Сначала проверяется гипотеза о том, что все p корней характеристического многочлена единичные; при ее отвержении проверяется гипотеза о наличии $p - 1$ единичных корней и т.д.

Поясним смысл этой процедуры на примере процесса авторегрессии AR(2)

$$a(L) x_t = \varepsilon_t,$$

т.е.

$$(1 - a_1 L - a_2 L^2) x_t = \varepsilon_t,$$

или

$$(1 - aL)(1 - bL) x_t = \varepsilon_t,$$

где

$$a = 1/z_1, b = 1/z_2, \text{ а } z_1, z_2 - \text{ корни уравнения } a(z) = 0.$$

При этом мы предполагаем, что процесс не носит взрывного характера, так что $|z_1|, |z_2| \geq 1$, а значит, $|a|, |b| \leq 1$.

Раскрывая скобки и перенося все составляющие, кроме x_t , в правую часть уравнения, получаем:

$$x_t = (a + b) x_{t-1} - abx_{t-2} + \varepsilon_t.$$

Вычтем из обеих частей x_{t-1} :

$$\Delta x_t = (a + b - 1) x_{t-1} - abx_{t-2} + \varepsilon_t.$$

Из обеих частей полученного равенства вычтем Δx_{t-1} :

$$\begin{aligned} \Delta^2 x_t &= \Delta x_t - \Delta x_{t-1} = -\Delta x_{t-1} + (a + b - 1) x_{t-1} - abx_{t-2} + \varepsilon_t = \\ &= (a + b - 2) x_{t-1} + (1 - ab) x_{t-2} + \varepsilon_t. \end{aligned}$$

Выделим в правой части первую разность:

$$\begin{aligned} \Delta^2 x_t &= (a + b - 2) x_{t-1} + [-(ab - 1) x_{t-2} + (ab - 1) x_{t-1}] - (ab - 1) x_{t-1} + \varepsilon_t \\ &= (a + b - ab - 1) x_{t-1} + (ab - 1) \Delta x_{t-1} + \varepsilon_t, \end{aligned}$$

так что

$$\Delta^2 x_t = (a - 1)(1 - b) x_{t-1} + (ab - 1) \Delta x_{t-1} + \varepsilon_t.$$

Это базовое соотношение позволяет идентифицировать ситуации, когда имеется 2 единичных корня, когда имеется 1 единичный корень и когда единичных корней нет. Именно,

- Если $a = b = 1$ (два единичных корня), то $\Delta^2 x_t = \varepsilon_t$.
- Если $a = 1, |b| < 1$ (один единичный корень), то $\Delta^2 x_t = (b - 1) \Delta x_{t-1} + \varepsilon_t$, или $\Delta^2 x_t = \varphi \Delta x_{t-1} + \varepsilon_t$ с $\varphi < 0$.
- Если $|a| < 1$ и $|b| < 1$ (нет единичных корней), то $\Delta^2 x_t = \psi x_{t-1} + \varphi \Delta x_{t-1} + \varepsilon_t$ с $\varphi < 0$ и $\psi < 0$.

Соответственно, процедура, предложенная Дики и Пантулой, такова.

Если мы допускаем наличие двух единичных корней, то сначала оцениваем статистическую модель

$$\Delta^2 x_t = \alpha + \varphi \Delta x_{t-1} + u_t$$

и сравниваем значение t -статистики для коэффициента φ с критическим значением соответствующей статистики Дики – Фуллера (случай 1 или 2, в зависимости от того, будем ли мы исходить из $\alpha = 0$ или $\alpha \neq 0$). Здесь u_t – либо просто процесс белого шума либо включает в себя еще и запаздывающие значения второй разности $\Delta^2 x_{t-1}, \dots, \Delta^2 x_{t-p+1}$.

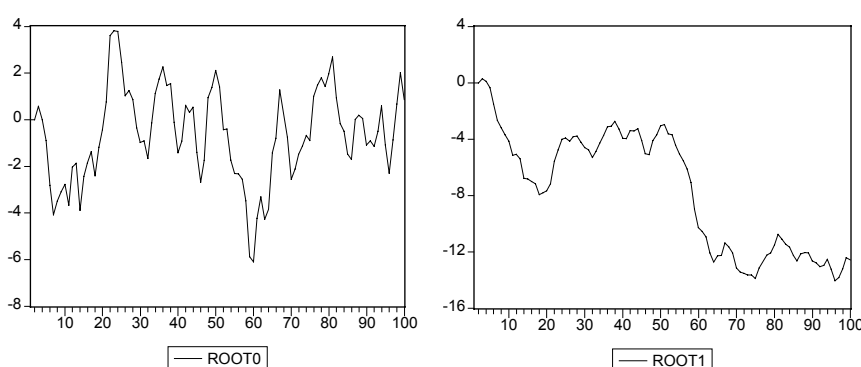
Если гипотеза о наличии двух единичных корней ($\varphi = 0$) отвергается, то тогда следует оценить статистическую модель

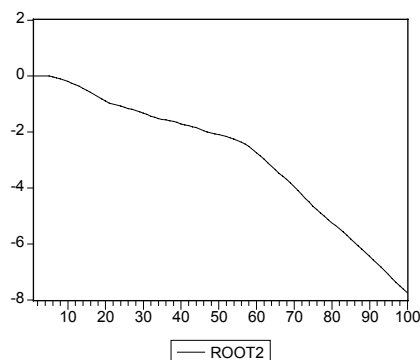
$$\Delta^2 x_t = \psi x_{t-1} + \varphi \Delta x_{t-1} + u_t$$

и проверить гипотезу $\psi = 0$ против альтернативы $\psi < 0$. Отклонение этой гипотезы означает признание того, что у ряда x_t нет единичных корней, а ее неотклонение – что $x_t \sim I(1)$.

Пример

Рассмотрим смоделированные реализации трех моделей AR(2) с различным количеством единичных корней:





Посмотрим, что дает применение процедуры Дики – Пантулы в этой ситуации.

На первом шаге для каждого из рядов оцениваем статистическую модель

$$SM: \Delta^2 x_t = \alpha + \varphi \Delta x_{t-1} + \varepsilon_t$$

и проверяем гипотезу $\varphi = 0$ против альтернативы $\varphi < 0$. (Анализ рядов остатков для обеих оцененных моделей указывает на отсутствие необходимости включения в правую часть статистической модели запаздывающих значений второй разности.)

Для ряда ROOT2 мы используем критические значения Фуллера, соответствующие случаю 2 ($\alpha \neq 0$), ориентируясь на наличие у реализации видимого квадратичного тренда. Для $T = 100$ критическое 5% значение статистики Дики – Фуллера равно -2.89 . Вычисленное значение t -статистики равно -1.64 ; гипотеза о наличии двух единичных корней не отвергается. Для рядов ROOT0 и ROOT1 используем критические значения Фуллера, соответствующие случаю 1 ($\alpha = 0$), принимая во внимание отсутствие у реализаций видимого квадратичного тренда. В этом случае для $T = 100$ критическое 5% значение статистики Дики – Фуллера равно -1.95 . Вычисленные значения t -статистик равны -7.83 для ряда ROOT0 и -5.50 для ряда ROOT1; в обоих случаях гипотеза о наличии двух единичных корней отвергается.

Следующий шаг процедуры выполняется поэтому только для рядов ROOT0 и ROOT1. Для этих рядов мы оцениваем статистическую модель

$$SM: \Delta^2 x_t = \psi x_{t-1} + \varphi \Delta x_{t-1} + \varepsilon_t$$

и проверяем гипотезу $\psi = 0$ против альтернативы $\psi < 0$. Значения соответствующей t -статистики равны -3.89 для ряда ROOT0 и -1.63 для ряда ROOT1, так что гипотеза $\psi = 0$ отвергается для ряда ROOT0 и не отвергается для ряда ROOT1.

Заметим теперь, что в модели DGP для ряда ROOT2 действительно было два единичных корня, в модели DGP для ряда ROOT1 – один единичный корень, а в модели DGP для ряда ROOT0 – ни одного единичного корня:

DGP для ROOT0: $x_t = 1.1 x_{t-1} - 0.3 x_{t-2} + \varepsilon_t$, или

$$(1 - 0.6L)(1 - 0.5L)x_t = \varepsilon_t,$$

DGP для ROOT1: $x_t = 1.5 x_{t-1} - 0.5 x_{t-2} + \varepsilon_t$, или

$$(1 - L)(1 - 0.5L)x_t = \varepsilon_t ;$$

DGP для ROOT2: $x_t = 2 x_{t-1} - x_{t-2} + \varepsilon_t$, или

$$(1 - L)^2 x_t = \varepsilon_t .$$

Более подробно с проблемами, возникающими при проверке гипотез, связанных с наличием нескольких единичных корней, можно ознакомиться, например, в книге [Patterson (2000)].

6.10. Критерий Перрона и его обобщение

6.10.1. Критерий Перрона

Предложенная в работе [Perron (1989a)] процедура проверки нулевой гипотезы о принадлежности ряда классу DS обобщает процедуру Дики – Фуллера на ситуации, когда на периоде наблюдений имеются структурные изменения модели в некоторый момент времени T_B либо в форме сдвига уровня, либо в форме изменения наклона тренда, либо в форме сочетания этих двух изменений. Важность такого обобщения связана с тем обстоятельством, что если DS-критерий не допускает возможности изменения структуры модели, тогда как такое изменение в действительности имеет место, то он имеет очень низкую мощность, т.е. практически всегда не отвергает DS-гипотезу (см., например, [Engle, Granger (1991)]).

Последнее можно лучше всего проиллюстрировать на примере работы Нельсона и Пlossера [Nelson, Plosser (1982)], в которой был проведен статистический анализ 13 основных макроэкономических рядов США по годовым данным за достаточно длинные периоды (от 62 до 111 лет) и квартального ряда GNP, относящегося к периоду после Второй мировой войны (1948 – 1987 г.г.). Все ряды были взяты в логарифмах, за исключением ряда процентных ставок.

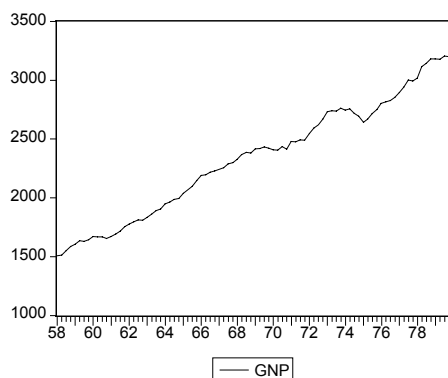
Для этих рядов гипотеза единичного корня проверялась в связке

SM: $\Delta x_t = \alpha + \beta t + \varphi x_{t-1} + u_t$, где u_t – стационарный процесс AR(k),

DGP: $\Delta x_t = \alpha + u_t$ (с $\alpha = 0$ или $\alpha \neq 0$),

и использовались критические значения Фуллера для этой ситуации. При этом Нельсон и Пlossер обнаружили, что для 13 из 14 рядов гипотеза единичного корня не отвергается. Единственным исключением оказался ряд логарифмов уровней занятости.

Полученные Нельсоном и Пlossером результаты сформировали устойчивое мнение о том, что макроэкономические ряды, обнаруживающие тренд, скорее всего могут моделироваться как DS ряды. В то же время, мы уже видели на примерах, что критерии Дики – Фуллера имеют не очень высокую мощность, и последнее может являться причиной неотвержения гипотезы единичного корня для указанных 14 рядов. Вместе с тем, надо учесть и следующее обстоятельство, на которое обратил внимание Перрон в статье [Perron (1989a)]. Рассмотрим, для примера, график ряда GNP для периода с 1 квартала 1958 г. по 4 квартал 1979 г.



В качестве альтернативы процессу случайного блуждания (со сносом или без сноса) критерий Дики – Фуллера предлагает процесс, стационарный относительно линейного тренда. Однако при просмотре приведенного графика возникает впечатление, что линейный тренд ряда имеет различный наклон на подпериодах до 1974 г. и после 1974 г.

Если непосредственно оценивать линейный тренд на всем периоде 1958:1 – 1979:4, то угловой коэффициент тренда оценивается как 19.086. При оценивании на подпериоде 1958:1 – 1973:4 угловой коэффициент тренда оценивается как 19.852. В то же время, при оценивании на подпериоде 1975:1 – 1979:4 угловой коэффициент тренда оценивается как 31.995. Это заставляет усомниться в пригодности выбора в качестве альтернативы случайному блужданию процесса, стационарного относительно именно линейного тренда. Скорее, надо было бы использовать в качестве альтернативы процесса, стационарного относительно ломаной с узлом в районе 1975 г.

Основываясь на подобных наблюдениях и в отношении других макроэкономических рядов, Перрон и предложил в [Peron (1989a)] три модели, допускающие структурные изменения модели ряда.

В критерии Перрона момент изменения структуры предполагается экзогенным, в том смысле, что он выбирается не на основании визуального исследования графика ряда, а связывается с моментом известного масштабного изменения экономической обстановки, существенного отражающегося на поведении рассматриваемого ряда. Трём указанным выше формам изменения структуры модели соответствуют три различных варианта регрессионных моделей, которые строятся путем вбирования в себя моделей, соответствующих нулевой и альтернативной гипотезам.

А. Модель “краха”:

$$x_t = c + \vartheta DMU_t + \beta t + d DTB_t + \alpha x_{t-1} + \varepsilon_t$$

В. Модель “изменения роста”:

$$x_t = c + \vartheta DMU_t + \beta t + \gamma DTS_t + \alpha x_{t-1} + \varepsilon_t$$

С. Модель, допускающая наличие обоих эффектов:

$$x_t = c + \vartheta DMU_t + \beta t + \delta DT_t + d DTB_t + \alpha x_{t-1} + \varepsilon_t$$

Здесь

c – постоянная,

$$DTB_t = \begin{cases} 1 & \text{для } t = T_B + 1 \\ 0 & \text{в противном случае} \end{cases} ;$$

$$DMU_t = \begin{cases} 1 & \text{для } t > T_B \\ 0 & \text{для } t \leq T_B \end{cases} ;$$

$$DTS_t = \begin{cases} t - T_B & \text{для } t > T_B \\ 0 & \text{для } t \leq T_B \end{cases} ;$$

$$DT_t = \begin{cases} t & \text{для } t > T_B \\ 0 & \text{для } t \leq T_B \end{cases} .$$

Нулевые гипотезы единичного корня накладывают следующие ограничения на истинные параметры моделей:

Модель А.

$$\alpha = 1, \beta = \theta = 0, d \neq 0.$$

Модель В.

$$\alpha = 1, \beta = \gamma = 0, \theta \neq 0.$$

Модель С.

$$\alpha = 1, \beta = \delta = 0, d \neq 0, \theta \neq 0.$$

Альтернативные гипотезы накладывают следующие ограничения на истинные параметры моделей

Модель А.

$$\alpha < 1, \beta \neq 0, \theta \neq 0, d = 0$$

Модель В.

$$\alpha < 1, \beta \neq 0, \gamma \neq 0, \theta = 0$$

Модель С.

$$\alpha < 1, \beta \neq 0, \delta \neq 0, d = 0, \theta \neq 0.$$

В такой формулировке нулевая и альтернативная гипотезы являются гнездовыми гипотезами.

Асимптотические критические значения t -статистики критерия Перрона зависят от типа структурных изменений, параметра $\lambda = T/T_B$ и от того, какая из моделей постулируется – **модель с аддитивным выбросом (АО)**, в которой структурное изменение происходит внезапно, или **модель с инновационным выбросом (ИО)**, в которой структурное изменение происходит постепенно. Приведенные в работе [Perron (1989a)] таблицы критических значений соответствуют моделям с инновационным выбросом². Как поступать в случае моделей с аддитивными выбросами, сообщается в работе [Perron, Vogelsang (1993)].

В связи с процедурами, допускающими излом траекторий, надо обратить особое внимание на различие между моделями внезапного и постепенного излома.

В течение нескольких лет в этом вопросе была некоторая путаница, так что даже сам автор первоначальной процедуры, допускающей изломы разных видов ([Perron (1989a)], ошибочно интерпретировал оцененные им модели и критические значения, полученные путем статистического моделирования.

Пусть z_t – стационарный процесс авторегрессии первого порядка с нулевым математическим ожиданием,

$$z_t = a_1 z_{t-1} + \varepsilon_t,$$

и ряд y_t определяется как

$$y_t = f(t) + z_t,$$

где $f(t) = 0$ при $t \leq T_B$ и $f(t) = \mu \neq 0$ при $t > T_B$. Поскольку $E(z_t) = 0$, то $E(y_t) = 0$ при $t \leq T_B$ и $E(y_t) = \mu$ при $t > T_B$. Таким образом, при переходе через дату излома T_B ряд y_t сразу начинает осциллировать вокруг уровня μ (вместо осцилляции вокруг нулевого уровня до этого перехода).

Рассмотрим теперь другую модель, в которой функция скачка “встроена” в уравнение AR(1) для y_t . Именно, пусть

$$y_t = f(t) + a_1 y_{t-1} + \varepsilon_t, |a_1| < 1,$$

где

$$f(t) = 0 \text{ при } t \leq T_B \text{ и } f(t) = \mu(1 - a_1) \text{ при } t > T_B, \mu \neq 0.$$

² В самой этой работе ошибочно полагалось, что приведенные в ней критические значения соответствуют моделям с аддитивным выбросом.

До момента T_B ряд y_t осциллирует вокруг нулевого уровня. Как будут вести себя траектории такого ряда y_t после перехода через дату излома T_B ? Для выяснения этого удобно записать:

$$y_t = a_1 y_{t-1} + (f(t) + \varepsilon_t) = a_1 y_{t-1} + v_t.$$

Тогда для $t = T_B + h$ имеем

$$\begin{aligned} y_{TB+h} &= y_0^{TB+h} + \sum_{k=0}^{TB+h-1} a_1^k v_{t-k} = \\ &= y_0^{TB+h} + \sum_{k=0}^{TB+h-1} a_1^k (f(t-k) + \varepsilon_{t-k}) = \\ &= \left[y_0^{TB+h} + \sum_{k=0}^{TB+h-1} a_1^k \varepsilon_{t-k} \right] + \sum_{k=0}^{h-1} a_1^k \mu(1-a_1). \end{aligned}$$

Первое слагаемое в правой части (выражение в квадратных скобках) соответствует модели с $E(y_t) = 0$. Вторая сумма при $h \rightarrow \infty$ имеет предел

$$\lim_{h \rightarrow \infty} \sum_{k=0}^{h-1} a_1^k \mu(1-a_1) = \mu.$$

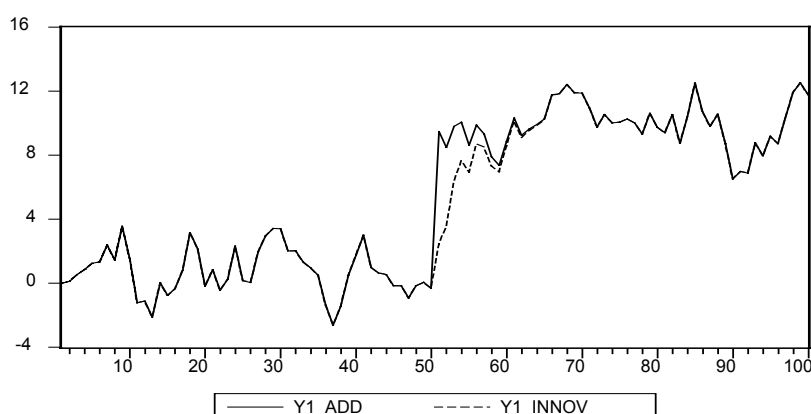
В этой модели после момента $t = T_B$ процесс y_t лишь постепенно выходит на новый уровень μ , вокруг которого начинает происходить осциляция траектории ряда.

Поскольку во второй модели значения $f(t)$ обрабатываются аналогично инновациям ε_t (влияние обоих здесь убывает геометрически), то вторую модель называют моделью **инновационного** выброса. В отличие от нее, первая модель называется моделью **аддитивного** выброса.

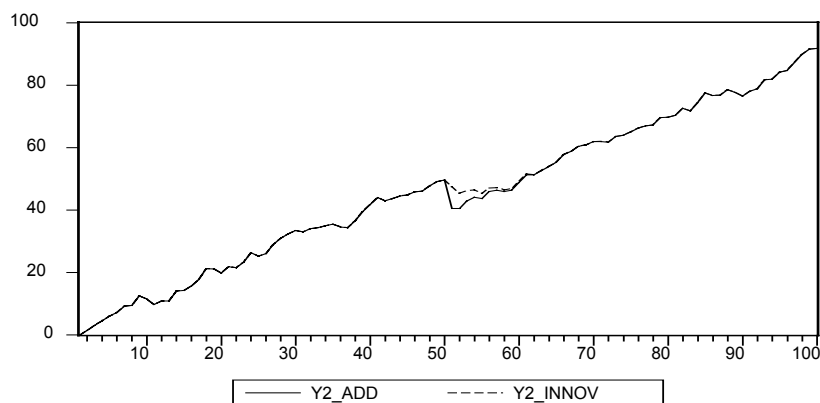
Аналогично, можно рассматривать пары моделей (аддитивная – инновационная), допускающие изменение наклона тренда без изменения уровня ряда или допускающие и изменение наклона тренда и изменение уровня ряда.

Ниже мы приводим графики, иллюстрирующие подобные ситуации.

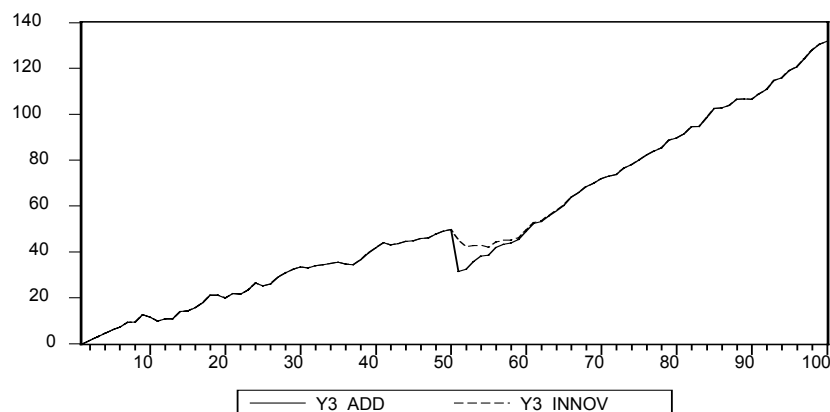
Сдвиг среднего уровня ряда:



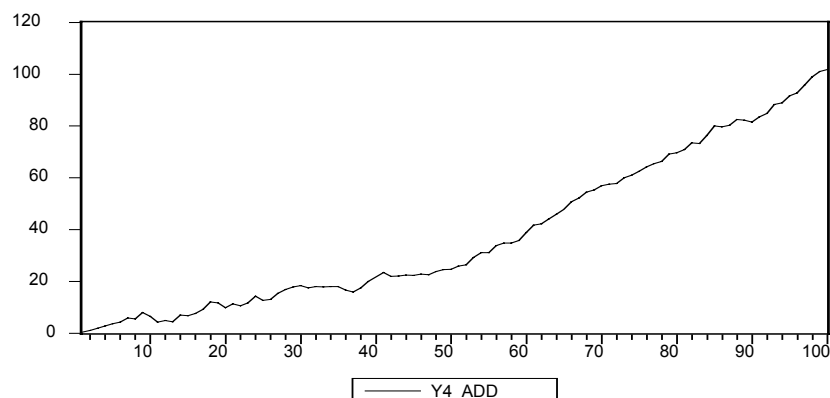
Сдвиг реализации без изменения наклона тренда:



Сдвиг реализации с изменением наклона тренда:



Изменение наклона тренда без сдвига реализации – “сегментированный тренд”:



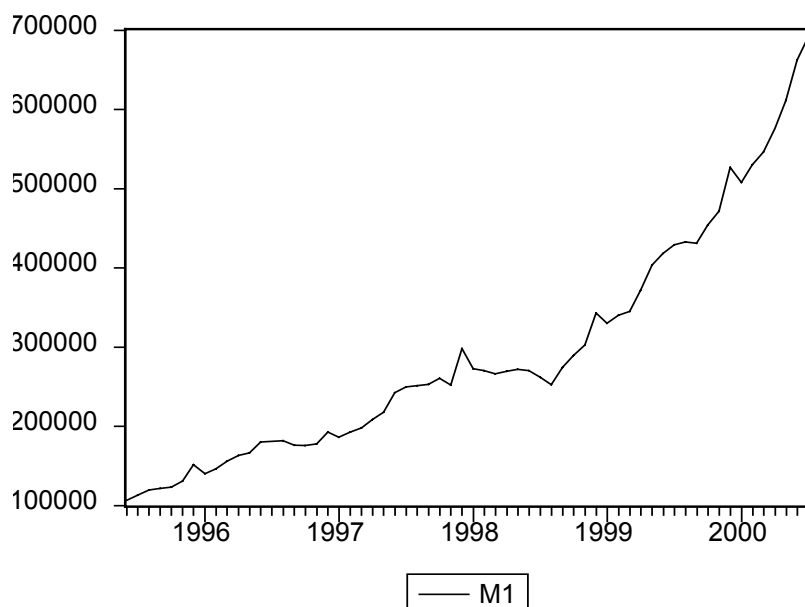
Возвратимся теперь к обсуждению статьи [Perron (1989a)].

Проведя ревизию результатов Нельсона-Плоссера для 14 рядов с допущением структурных изменений модели и экзогенным выбором даты излома, Перрон получил совершенно другие результаты. Теперь уже гипотеза единичного корня была отвергнута для 11 из 14 рядов, т.е. результаты получились практически прямо противоположными результатам Нельсона-Плоссера. Чуть позже мы обсудим это обстоятельство, а сейчас приведем пример применения процедуры Перрона к одному из основных российских макроэкономических рядов.

Пример

В качестве примера использования процедуры Перрона с экзогенной датой излома мы рассмотрим проверку гипотезы о наличии единичного корня в авторегрессионном

представлении модели, порождающей ряд $x_t = M1$, где M1 – денежный агрегат, представляющий все денежные средства в экономике Российской Федерации, которые могут быть использованы как средство платежа. Мы используем месячные данные за период 1995:06 – 2000:07 в номинальных величинах. График ряда $X_t = M1$ имеет следующий вид:



При анализе этого ряда на наличие единичного корня с использованием критериев Дики – Фуллера и Филлипса – Перрона (см. [Эконометрический анализ динамических рядов ... (2001)]) гипотеза единичного корня не была отвергнута, что может быть связано с неудачным выбором альтернативных гипотез. График ряда позволяет предположить, что более подходящей может оказаться модель с изломом тренда в конце 1998 – начале 1999 г., связанным с финансово-экономическим кризисом 1998 года.

Если предполагать, что излом тренда выражается в изменении его наклона после августа 1998 г., то мы можем обратиться к статистической процедуре проверки гипотезы единичного корня, предложенной в упомянутой выше работе Перрона и соответствующей одномоментному (внезапному) изменению наклона тренда (АО модель – модель с аддитивным выбросом).

Согласно этой процедуре, если TB – момент скачка, то сначала следует оценить статистическую модель

$$x_t = \mu + \beta t + \gamma DTS_t + u_t,$$

в которой переменная DTS_t равна $t - TB$ для $t > TB$ и равна 0 для всех других значений t . В результате оценивания этой модели получаем ряд остатков e_t . Затем оценивается модель регрессии e_t на e_{t-1} и запаздывающие разности $\Delta e_{t-1}, \dots, \Delta e_{t-p}$:

$$e_t = \alpha e_{t-1} + \sum_{j=1}^p c_j \Delta e_{t-j} + \varepsilon_t;$$

полученное при этом значение t -статистики для проверки гипотезы $H_0: \alpha = 1$ сравнивается с критическим значением из таблицы, приведенной в статье [Perron, Vogelsang (1993), стр. 249]. В правую часть оцениваемой статистической модели следует включать достаточное количество запаздывающих разностей, чтобы исключить автокоррелированность ошибок в расширенной модели.

В нашем случае $TB = 42$, что соответствует 1998:08. В правую часть уравнения для остатков приходится дополнительно включать 12 запаздывающих разностей, т.к. иначе (при 11 разностях) получаем P -значение критерия Бройша-Годфри (с $AR(1)$ альтернативой), равное 0.0002 и указывающее на автокоррелированность остатков. Для повышения мощности критерия, используя стратегию GS (“от общего к частному”) и критерий Шварца SIC, осуществим редукцию модели, последовательно исключая из нее запаздывающие разности со статистически незначимыми (на 10% уровне значимости) коэффициентами. Результаты такой последовательной редукции сведены в следующую таблицу

Порядок запаздывания исключаемой разности	SIC	P-val LM-автокорр.	P-val White	P-val J-B	t-статистика критерия Петрона
– (полная модель с 12 запаздывающими разностями)	22.236	1 – 0.983 2 – 0.967	0.701	0.281	-1.92
8	22.157				-2.27
11	22.089				-2.60
10	22.018				-2.90
9*	21.986	1 – 0.590 2 – 0.844 3 – 0.954	0.372	0.223	-3.27
4	21.974		0.040		-2.78
5	21.935		0.035		-2.59
3	21.898		0.016		-2.22
1** (выбор по GS)	21.837		0.006	0.518	-2.04
7	21.834		0.002	0.184	-1.37
6	21.793			0.008	-1.31
2 (выбор по SIC)	21.782			0.006	-0.92

В первом столбце таблицы указаны порядки запаздывания разностей, последовательно исключаемых из правой части оцениваемой статистической модели. Запаздывающая разность исключается из уравнения, если коэффициент при этой разности признается статистически незначимым на 10% уровне значимости.

Во втором столбце приведены значения информационного критерия Шварца (SIC), соответствующие редуцированным моделям.

В третьем столбце приведены P -значения (P -values) LM-критерия автокоррелированности ошибок Бройша-Годфри. Цифры, предворяющие эти P -значения, указывают на допустимый (при альтернативе) порядок авторегрессионной модели для ошибок в редуцированном уравнении.

В четвертом столбце приведены P -значения критерия Уайта (White) гетероскедастичности ошибок.

В пятом столбце приведены P -значения критерия Jarque-Bera для проверки нормальности распределения ошибок.

В последнем столбце таблицы приведены значения t -статистики (расширенного) критерия Дики - Фуллера, получаемой при оценивании соответствующей редуцированной (или полной) модели.

При редукции модели методом “от общего к частному” (с 10% уровнем значимости) из расширенной модели с 12 запаздывающими разностями последовательно удаляются разности, запаздывающие на 8, 11, 10, 9 единиц времени (месяцев). Это приводит к модели, содержащей в правой части разности, запаздывающие на 1 – 7 и 12 месяцев; результаты оценивания этой модели приведены в строке таблицы, отмеченной звездочкой. Если продолжать редукцию, отбрасывая запаздывающие разности с коэффициентами, статистически незначимыми на 10% уровне, то остановка происходит на модели, результаты для которой находятся в строке, отмеченной двумя звездочками.

Критерий Шварца выбирает модель, результаты оценивания которой приведены в последнем столбце таблицы.

Поскольку отклонения от нормальности, некоррелированности и гомоскедастичности могут отражаться на критических значениях статистики критерия, то в этом отношении предпочтительнее модель, результаты для которой приведены в строке, помеченной звездочкой.

Асимптотические критические значения статистики критерия Перрона зависят от положения момента излома на интервале наблюдений через параметр $\lambda = TB/T$, где TB – момент, непосредственно после которого происходит излом тренда, а T – количество наблюдений. В нашем случае $\lambda = 42/62 = 0.667$. Соответствующее 5% критическое значение (при сделанном предположении о внезапном изменении наклона тренда) заключено между значениями -3.94 (для $\lambda = 0.6$) и -3.89 (для $\lambda=0.7$). Гипотеза единичного корня не отвергается ни в полной модели и ни в одной из редуцированных моделей.

Отметим также, что момент излома тренда 1998:08 был выбран нами на основании уже имеющейся информации об августовском кризисе 1998 г. и визуального обращения к графику ряда М1. Между тем, выбор даты излома тренда на основании анализа графика ряда влияет на критические значения t -статистики критерия единичного корня.

6.10.2. Обобщенная процедура Перрона

Анализируя результаты Перрона в отношении 14 макроэкономических рядов США, ряд авторов задался вопросом о влиянии метода датировки на критические значения соответствующих статистик. В работе [Zivot, Andrews (1992)] было обращено внимание на то, что при рассмотрении послевоенного GNP в качестве даты структурного сдвига Перрон взял второй квартал 1973 г. (что соответствует мировому топливно-энергетическому кризису). И это можно было бы считать экзогенным выбором, поскольку решение принималось международной организацией (ОПЕК). Однако в послевоенный период имели место и такие крупные события, как снижение налогов (1964 г.), война во Вьетнаме, финансовое разрегулирование в 80-е годы. Тем не менее, Перрон взял за точку сдвига именно 1973 г., обращаясь предварительно к поведению ряда GNP. А если это так, то нарушается условие, согласно которому статистические гипотезы формулируются до любого (даже визуального) анализа данных, на основании которых принимается решение об отклонении или неотклонении нулевой гипотезы. С этой точки зрения, критерий Перрона, предложенный в работе [Peron (1989a)], является условным, при условии, что точка смены режима известна.

Вместо условного критерия Перрона, Zivot и Andrews предложили использовать безусловный критерий (относящийся к инновационным выбросам), в котором датировка точки смены режима производится в “автоматическом режиме”, путем перебора всех возможных вариантов датировки и вычисления для каждого варианта датировки t -статистики t_α для проверки гипотезы $H_0: \alpha = 1$; в качестве оцененной даты берется такая (t_{\min}), для которой значение t_α оказывается минимальным. К чему это приводит?

Возьмем, для примера, ряд, представляющий занятость (1890 – 1970). Этот ряд исследуется в [Zivot, Andrews (1992)] в рамках модели (A) (см. выше Модели A, B, C), но только без включения в правую часть переменной DTB_t . Перрон для всех

рядов, кроме послевоенного GNP, определил в качестве точки смены режима 1929 г. (Великая Депрессия). Для ряда занятости значение t_α для этого года равно $t_\alpha = -4.95$, $TB = 40$, $\lambda = 40/81 = 0.49$. При таком значении λ критическое (5%) значение для t_α приближенно равно -3.76 , так что гипотеза единичного корня отвергается. С другой стороны, выполняя указанный перебор, Zivot и Andrews получили ту же датировку (1929 г.), так что $t_{\min} = -4.95$. Значение t -статистики не изменилось. Однако распределение статистики t_{\min} отличается от распределения статистики t_α для фиксированного года: 5% критическое значение для t_{\min} равно -5.26 . Поскольку $t_{\min} = -4.95 > -5.26$, гипотеза единичного корня ($H_0: \alpha = 1$) теперь не отвергается.

Аналогичный анализ для остальных рядов из работы Нельсона и Пlossера приводит к следующим результатам. Гипотеза единичного корня не отвергается для 11 из 14 рядов. Исключение составляют реальный и номинальный GNP (годовые данные) и промышленное производство (1986 – 1970). И это объясняется консервативностью критических значений при эндогенной датировке (путем перебора): при заданном значении λ последние существенно ниже критических значений, соответствующих экзогенной датировке.

Следует, впрочем, заметить, что при оценивании уравнений для номинального GNP, номинальной заработной платы и биржевого курса обыкновенных акций ряды остатков имели слишком большие значения коэффициента пикообразности – *куртозиса (kurtosis)*: 5.68, 4.658, 4.324, говорящие не в пользу предположения о нормальности инноваций, при котором были получены критические значения статистики t_{\min} . (Куртозис распределения определяется как отношение четвертого центрального момента распределения к квадрату дисперсии. Для нормального распределения значение куртозиса равно 3.³) Перемоделирование критических значений с использованием (вместо нормального) распределения Стьюдента с подходящими числами свободы дало для этих трех рядов следующие 5% критические значения: -5.86 , -5.81 и -5.86 (против -5.38 , -5.33 и -5.63 , соответственно). Значения статистики t_{\min} для этих рядов равны -5.82 , -5.30 и -5.61 , что, в общем, практически не изменяет статистических выводов.

Наконец, если предположить, что распределение инноваций имеет настолько тяжелые хвосты, что $D(\varepsilon_t) = \infty$, то критические значения статистики t_{\min} уменьшаются столь значительно, что отвергнуть гипотезу единичного корня на 5% уровне значимости становится невозможным ни для одного ряда.

Перрон вернулся к проблеме проверки гипотезы единичного корня в работе [Peron (1997)] и, развивая результаты Zivot, Andrews, исследовал зависимость критических значений статистики t_{\min} от выбора количества запаздывающих разностей, включаемых в правые части оцениваемых уравнений. При этом Перрон работал с моделями (A) и (C), содержащими в правых частях (в отличие от Zivot, Andrews) переменную DTB_t .

³ В отечественной литературе в качестве характеристики пикообразности распределения чаще используется *коэффициент эксцесса* $\kappa = (\text{куртозис} - 3)$, равный 0 для нормального распределения. Мы ориентируемся здесь на куртозис из-за того, что в распечатках результатов, получаемых при применении пакета статистического анализа ECONOMETRIC VIEWS, приводятся именно значения (оцененного) куртозиса.

Методика, разработанная в [Perron (1997)], реализована в виде процедуры PERRON97 в пакете статистического анализа RATS.

При этом рассматриваются модели IO1 – с инновационным выбросом с изменением постоянной, IO2 – с инновационным выбросом, изменяющим и постоянную и наклон тренда, АО – с аддитивным выбросом, изменяющим только наклон тренда.

Предусмотрены три метода оптимального выбора даты излома:

UR – по минимуму t -статистики критерия для проверки гипотезы $\alpha = 1$;

STUDABS – по максимуму абсолютной величины t -статистики критерия для проверки гипотезы о равенстве нулю коэффициента при переменной, отвечающей за изменение константы (в модели IO1) или за изменение наклона тренда (в модели IO2);

STUD – по минимуму t -статистики критерия для проверки гипотезы о равенстве нулю коэффициента при переменной, отвечающей за изменение константы (в модели IO1) или за изменение наклона тренда (в модели IO2);

При практической реализации критерия обычно несколько ограничивают интервал возможных дат излома, чтобы исключить слишком ранние или слишком поздние даты излома.

Пример (продолжение примера с рядом M1)

Для учета влияния датировки при проверке гипотезы единичного корня в моделях, допускающих структурное изменение, воспользуемся процедурой PERRON97 из пакета статистического анализа RATS, реализующей методику, приведенную в статье [Perron (1997)]. Имея в виду предыдущие результаты, ограничим максимальное запаздывание разностей, включаемых в правую часть оцениваемых уравнений, тринадцатью.

Сначала рассмотрим модель, допускающую сдвиг траектории и изменение наклона тренда в форме инновационного выброса (IO). Результаты применения процедуры PERRON97 для этой модели таковы:

break date TB = 1999:07; statistic $t(\alpha=1) = -3.34124$			
critical values at	1%	5%	10%
for 70 obs.	-6.32	-5.59	-5.29
number of lag retained : 12			
explained variable : M1			
	coefficient	student	
CONSTANT	124786.79561	3.33345	
DU	-2506239.31872	-3.77751	
D(Tb)	40455.79442	2.72347	
TIME	9769.03708	3.44839	
DT	23866.02686	3.78217	
M1{1}	-0.91050	-1.59235	

Здесь

$DU_t = 1$ для $t > TB$ и $DU_t = 0$ для всех других значений t ;

$D(Tb)_t = 1$ для $t = TB + 1$ и $D(Tb)_t = 0$ для всех других значений t ;

$DT = t$ для $t > TB$ и $DT_t = 0$ для всех других значений t ;

$(M1\{1\})_t = M1_{t-1}$.

(Заметим, что при постулировании инновационного выброса оценивание регрессионной модели при каждой испытываемой дате производится в один этап – в правую часть регрессионной модели в качестве объясняющих включаются сразу все 6

переменных: CONST, DU, D(Tb), TIME, DT и запаздывающая на один шаг переменная $M1\{1\}$.)

Процедура PERRON97 определяет в этом случае дату излома как 1999:07, если выбор даты излома осуществляется по минимуму t -статистики критерия единичного корня $t_{\alpha=1}$, взятому по всем возможным моментам излома. При этом $t_{\alpha=1} = -3.341$, что выше 5% критического уровня -5.59 , и гипотеза единичного корня не отвергается. Наибольшее запаздывание разностей, включаемых в правую часть уравнений, выбирается равным 12 в рамках применения процедуры GS для редукции модели с 10% уровнем значимости.

Если выбор даты излома осуществляется по максимуму абсолютной величины t -статистики для коэффициента d при переменной DT_t , отвечающей за изменение наклона тренда, то выбирается 1998:04. При этом $t_{\alpha=1} = -0.547$, что выше 5% критического значения -5.33 ; гипотеза единичного корня не отвергается. (Наибольшее запаздывание разностей здесь уменьшается до 11). Наконец, если выбор даты излома тренда осуществляется по минимуму t -статистики критерия для проверки гипотезы о равенстве нулю коэффициента при переменной DT , отвечающей за изменение наклона тренда, то выбирается опять 1998:04 с тем же выводом о неотвержении гипотезы единичного корня (UR-гипотезы).

Рассмотрим теперь модель, допускающую только изменение наклона тренда (без сдвига траектории) в форме аддитивного выброса (АО). Результаты применения процедуры PERRON97 для этой модели таковы:

break date TB = 1999:02; statistic t(alpha=1) = -3.59417			
critical values at	1%	5%	10%
for 100 obs.	-5.45	-4.83	-4.48
number of lag retained : 12			
explained variable : M1			
	coefficient	student	
CONSTANT	104939.65455	20.48279	
TIME	4832.56930	26.73200	
DT	14335.07564	21.11189	
M1 {1}	-0.75752	-1.54915	

(Заметим, что при постулировании аддитивного выброса оценивание регрессионной модели при каждой испытываемой дате производится в два этапа. На первом шаге в правую часть регрессионной модели в качестве объясняющих включаются только переменные CONST, TIME, DT; в результате оценивания этой модели получаем ряд остатков e_t . На втором шаге оценивается модель регрессии e_t на e_{t-1} и запаздывающие разности $\Delta e_{t-1}, \dots, \Delta e_{t-p}$).

Датировка момента излома осуществляется по минимуму статистики $t_{\alpha=1}$ для проверки гипотезы о равенстве 1 коэффициента при e_{t-1} в последней модели. При этом дата излома определяется как 1999:02, $t_{\alpha=1} = -3.594$ (используются 12 запаздывающих разностей), 5% критическое значение равно -4.83 , так что UR-гипотеза не отвергается и в этом случае.

Заметим, что распределение ошибок имеет в последней ситуации распределение, отличающееся от нормального: оцененный коэффициент пикообразности распределения – куртозис – превышает на 1.626 значение куртозиса нормального распределения, равного 3. Как следует из работы [Zivot, Andrews (1992)] (мы это уже отмечали ранее), в таких ситуациях критические уровни сдвигаются в сторону больших отрицательных значений, так что если использовать скорректированные на ненормальность критические уровни, то UR-гипотеза не будет отвергнута тем более.

Приведем здесь для полноты итоги анализа ряда M1 на интервале 1995:06 по 2000:07, проведенного в работе [Эконометрический анализ динамических рядов ... (2001)]. Результаты применения различных процедур сведены в одну таблицу.

Используемая процедура (критерий)	Исходная (нулевая) гипотеза	
	DS	TS
Критерий Дики - Фуллера (расширенный)	Не отвергается	
Критерий Филлипса-Перрона	Не отвергается	
Критерий DF-GLS	Не отвергается	
Критерий KPSS		Отвергается
Отношение дисперсий Кохрейна	В пользу DS	
Критерий Перрона (экзогенный выбор даты излома тренда)	Не отвергается	
Обобщенный критерий Перрона (эндогенный выбор даты излома тренда)	Не отвергается	

Статистические выводы, полученные при применении всех перечисленных в таблице процедур, согласуются между собой: нулевая DS-гипотеза не отвергается, тогда как нулевая TS-гипотеза отвергается; поведение отношения дисперсий Кохрейна также говорит в пользу DS-гипотезы.

Глава 7. Регрессионный анализ для нестационарных объясняющих переменных

7.1. Проблема ложной регрессии

Мы начнем обсуждение с проблемы *ложной (фиктивной, паразитной – spurious)* регрессии.

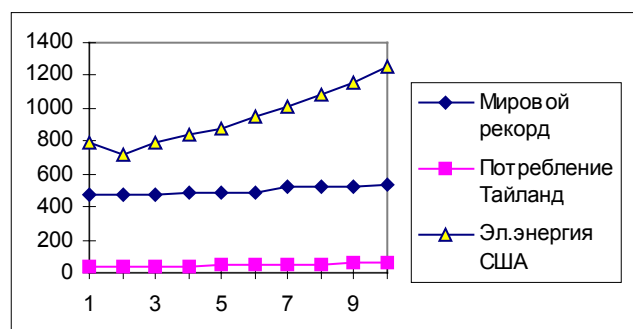
Обратимся в этой связи к следующему примеру, который был рассмотрен ранее в работе [Носко (2000)].

Пример

Рассмотрим динамику изменений в период с 1957 по 1966 годы трех совершенно различных по природе показателей: *E* - суммарного производства электроэнергии в США (в млрд квт-час), *C* - совокупных потребительских расходов в Тайланде (в млрд. бат) и *H* - мирового рекорда на конец года в прыжках в высоту с шестом среди мужчин (в см). Значения этих показателей приведены в следующей таблице:

Год	Потребление Тайланд млрд бат	Эл.энергия США млрд квт-час	Мир.рекорд (прыжки с шестом) см
1957	34.9	716	478
1958	35.9	724	478
1959	37.9	797	478
1960	41.1	844	481
1961	43.5	881	483
1962	46.7	946	493
1963	48.9	1011	520
1964	52.0	1083	528
1965	56.1	1157	528
1966	62.6	1249	534

Динамика изменений показателей показана на графике:



По этим данным мы можем формально, используя метод наименьших квадратов, подобрать модели линейной зависимости каждого из трех показателей от каждого из остальных показателей. Это приводит, например, к моделям

$$E = -2625.5 + 7.131H, \quad R^2 = 0.900;$$

$$C = -129.30 + 0.350H, \quad R^2 = 0.871;$$

$$E = 23.90 + 19.950C, \quad R^2 = 0.993;$$

$$C = -0.860 + 0.0498E, \quad R^2 = 0.993.$$

(Заметим, кстати, что произведение угловых коэффициентов двух последних прямых, соответствующих моделям линейной связи, в которых объясняемая и объясняющая переменная меняются местами, равно $19.950 \cdot 0.0498 = 0.993$ и совпадает со значением коэффициента детерминации R^2 в этих двух подобранных моделях.)

Мы видим, что во всех подобранных моделях значения коэффициента детерминации весьма высоки, и это формально означает, что изменчивость “объясняющих” переменных в этих моделях составляет значительный процент от изменчивости “объясняемой” переменной, стоящей в левой части уравнения. Однако, вряд ли мы всерьез можем полагать, что динамика роста суммарного производства электроэнергии в США действительно объясняется динамикой роста мирового рекорда по прыжкам в высоту с шестом, несмотря на высокое значение 0.9 коэффициента детерминации в первом из четырех уравнений.

В ситуациях, подобных последнему примеру, принято говорить о **фиктивной (ложной, паразитной - spurious)** линейной связи между соответствующими показателями. И такие ситуации часто встречаются при рассмотрении показателей, динамика изменений которых обнаруживает заметный тренд (убывание или возрастание) - именно такой характер имеют исследуемые показатели в последнем примере.

Чтобы понять, почему это происходит, вспомним известное соотношение, связывающее коэффициент детерминации R^2 и квадрат выборочного коэффициента корреляции между переменными y и x :

$$R^2 = r_{yx}^2.$$

Из этого равенства вытекает, что близкие к единице значения коэффициента детерминации соответствуют близким по абсолютной величине к единице значениям коэффициента корреляции между переменными y и x . Но этот коэффициент корреляции равен

$$r_{yx} = \frac{Cov(y, x)}{\sqrt{Var(y)}\sqrt{Var(x)}},$$

где

$$Cov(y, x) = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x}), \quad Var(y) = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2,$$

$$Var(x) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

При фиксированных значениях $Var(x)$ и $Var(y)$, значение r_{yx} будет тем ближе к 1, чем большим будет значение $Cov(y, x) > 0$. Последнее же обеспечивается

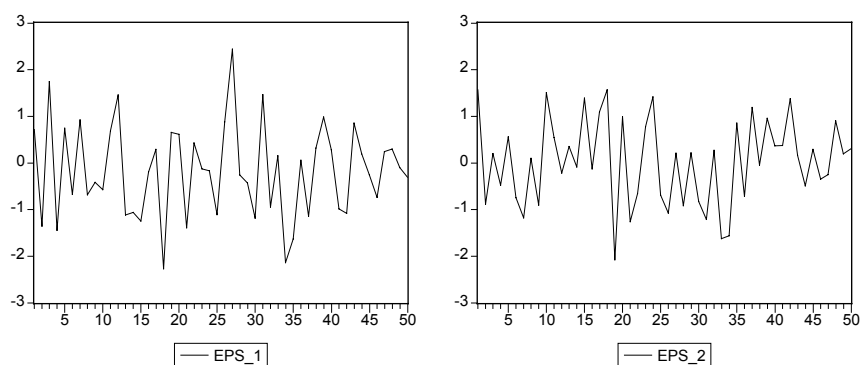
совпадением знаков разностей $y_i - \bar{y}$ и $x_i - \bar{x}$ для максимально возможной доли наблюдений переменных y и x , что как раз и имеет место, когда в процессе наблюдения обе переменные возрастают или обе переменные убывают по величине. (В этом случае превышение одной из переменных своего среднего значения сопровождается, как правило, и превышением второй переменной своего среднего значения. Напротив, если одна из переменных принимает значение, меньшее своего среднего, то и вторая переменная, как правило, принимает значение, меньшее своего среднего.)

Аналогичным образом, значение r_{yx} будет тем ближе к -1 , чем меньшим будет значение $Cov(y, x) < 0$. Последнее же обеспечивается несовпадением знаков разностей $y_i - \bar{y}$ и $x_i - \bar{x}$ для максимально возможной доли наблюдений переменных y и x , что имеет место, когда в процессе наблюдения одна из переменных возрастает, а вторая убывает. (В этом случае, если одна из переменных принимает значение, меньшее своего среднего, то вторая переменная, как правило, принимает значение, большее своего среднего.)

Из сказанного следует, что близость к единице абсолютной величины наблюдаемого значения коэффициента детерминации не обязательно означает наличие причинной связи между двумя рассматриваемыми переменными, а может являться лишь следствием тренда значений переменных.

Пример

Смоделируем реализации двух статистически независимых между собой последовательностей ε_{1t} и ε_{2t} независимых, одинаково распределенных случайных величин, имеющих стандартное нормальное распределение $N(0, 1)$. Смоделированные реализации имеют вид



На их основе построим реализацию линейной модели DGP

$$\text{DGP: } x_t = 1 + 0.2t + \varepsilon_{1t},$$

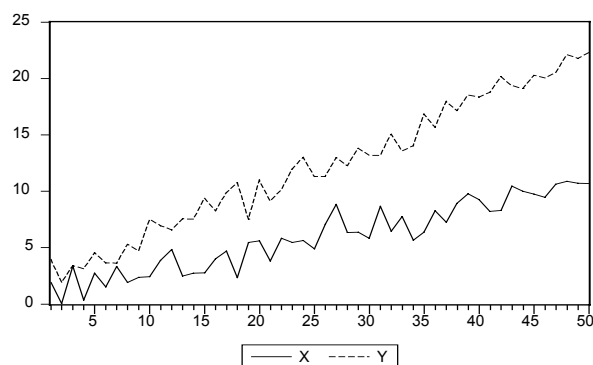
$$y_t = 2 + 0.4t + \varepsilon_{2t},$$

и рассмотрим результаты оценивания статистической модели

$$\text{SM: } y_t = \alpha + \beta x_t + \varepsilon_t$$

по смоделированной реализации.

Графики рядов x_t и y_t имеют в этом случае вид



Оба ряда имеют выраженные линейные тренды.

Оцененная статистическая модель:

Dependent Variable: Y

Sample: 1 50

Included observations: 50

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.553866	0.685771	2.265868	0.0280
X	1.800255	0.102997	17.47878	0.0000
R-squared	0.864218	Mean dependent var	12.22809	
Adjusted R-squared	0.861389	S.D. dependent var	5.925326	
S.E. of regression	2.206028	Akaike info criterion	4.459442	
Sum squared resid	233.5948	Schwarz criterion	4.535923	
Log likelihood	-109.4860	F-statistic	305.5076	
Durbin-Watson stat	2.150060	Prob(F-statistic)	0.000000	

Оцененные коэффициенты статистически значимы, коэффициент детерминации высокий, проверка на адекватность не выявляет нарушений стандартных предположений классической линейной модели регрессии.

Включим в правую часть статистической модели линейный тренд. Оценивание расширенной модели дает следующий результат:

Dependent Variable: Y

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.037450	0.294861	6.909879	0.0000
T	0.412232	0.028055	14.69394	0.0000
X	-0.054186	0.133658	-0.405410	0.6870
R-squared	0.975727	Mean dependent var	12.22809	
Adjusted R-squared	0.974694	S.D. dependent var	5.925326	
S.E. of regression	0.942598	Akaike info criterion	2.777771	
Sum squared resid	41.75908	Schwarz criterion	2.892492	
Log likelihood	-66.44428	F-statistic	944.6386	
Durbin-Watson stat	2.249075	Prob(F-statistic)	0.000000	

Остатки проходят тесты на адекватность, так что можно обратить внимание на протокол оценивания расширенной статистической модели. В соответствии с этим протоколом, коэффициент при переменной x_t статистически незначим, так что эта

переменная по-существу не проявляет себя в качестве переменной, объясняющей изменчивость значений переменной y_t .

Исключение x_t из правой части уравнения приводит к оцененной модели

Dependent Variable: Y

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.990020	0.268291	7.417403	0.0000
T	0.401493	0.009157	43.84727	0.0000
R-squared	0.975642	Mean dependent var	12.22809	
Adjusted R-squared	0.975134	S.D. dependent var	5.925326	
S.E. of regression	0.934357	Akaike info criterion	2.741262	
Sum squared resid	41.90511	Schwarz criterion	2.817743	
Log likelihood	-66.53155	F-statistic	1922.583	
Durbin-Watson stat	2.249658	Prob(F-statistic)	0.000000	

которая предпочтительнее расширенной модели и по критерию Акаике и по критерию Шварца. Более того, по этим критериям последняя модель намного предпочтительнее исходной модели $y_t = \alpha + \beta x_t + \varepsilon_t$, и это связано с тем, что при оценивании исходной SM остаточная сумма квадратов равна 233.59, а при оценивании последней модели остаточная сумма квадратов равна всего лишь 41.91. Это еще более убедительно подтверждает, что изменчивость переменной y_t в действительности не объясняется изменчивостью переменной x_t .

В рассмотренном примере паразитная связь между переменными была обусловлена тем, что в самой модели DGP обе переменные имеют в своем составе детерминированный линейный тренд.

Однако ложная (паразитная) связь между переменными может возникать не только в результате наличия у этих переменных детерминированного тренда. Паразитная связь может возникать и между переменными, имеющими не детерминированный, а стохастический тренд. Приведем соответствующий пример.

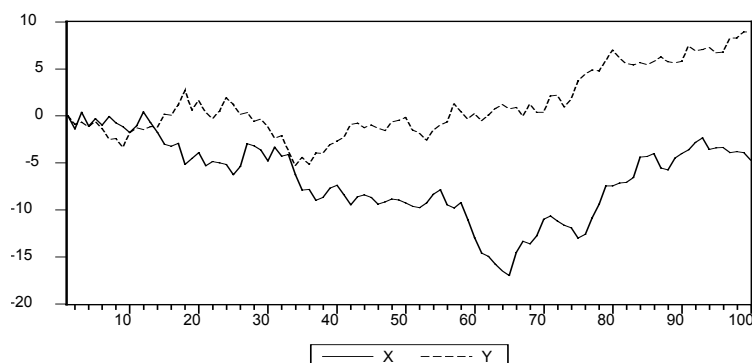
Пример

Возьмем теперь процесс порождения данных в виде

$$\text{DGP: } x_t = x_{t-1} + \varepsilon_{1t},$$

$$y_t = y_{t-1} + \varepsilon_{2t},$$

где ε_{1t} и ε_{2t} — те же, что и в предыдущем примере. Это приводит к смоделированной реализации



Предположим, что нам доступны статистические данные, соответствующие последним 50 наблюдениям (с 51 по 100). Оценивание по этим наблюдениям статистической модели

$$\text{SM: } y_t = \alpha + \beta x_t + \varepsilon_t$$

приводит к следующим результатам:

Dependent Variable: Y

Sample: 51 100

Included observations: 50

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	8.616496	0.748277	11.51512	0.0000
X	0.597513	0.077520	7.707873	0.0000
R-squared	0.553120	Mean dependent var	3.404232	
Adjusted R-squared	0.543810	S.D. dependent var	3.354003	
S.E. of regression	2.265356	Akaike info criterion	4.512519	
Sum squared resid	246.3283	Schwarz criterion	4.589000	
Log likelihood	-110.8130	F-statistic	59.41131	
Durbin-Watson stat	0.213611	Prob(F-statistic)	0.000000	

Несмотря на то, что в DGP ряды y_t и x_t порождаются независимо друг от друга и их модели не содержат детерминированного тренда, мы наблюдаем и здесь довольно высокое значение коэффициента детерминации 0.553. Конечно, это связано с тем, что на рассматриваемом периоде реализации обоих рядов имеют видимый тренд:



Если, однако, обратиться ко всему периоду из 100 наблюдений, то результаты оценивания будут совсем другими:

Dependent Variable: Y

Sample: 1 100

Included observations: 100

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.490206	0.664538	2.242470	0.0272
X	0.055097	0.083978	0.656086	0.5133
R-squared	0.004373	Mean dependent var	1.120548	
Adjusted R-squared	-0.005786	S.D. dependent var	3.513463	
S.E. of regression	3.523613	Akaike info criterion	5.376648	
Sum squared resid	1216.753	Schwarz criterion	5.428752	
Log likelihood	-266.8324	F-statistic	0.430449	
Durbin-Watson stat	0.061638	Prob(F-statistic)	0.513306	

В этом случае значение коэффициента детерминации близко к нулю, а оцененный коэффициент при x_t равен 0.0551 против значения 0.5975, полученного при оценивании по наблюдениям с 51 по 100. Это отражает действительное отсутствие детерминированного тренда в DGP и, в связи с этим, крайнюю нестабильность оценок коэффициента при x_t , полученных на различных интервалах. Последнее сопровождается также крайне низкими значениями статистики Дарбина – Уотсона (0.214 на полном периоде наблюдений и 0.062 на второй половине этого интервала).

Все указанные признаки являются характерными чертами, присущими результатам оценивания линейной модели связи между переменными, имеющими стохастический (но не детерминированный !) тренд и порождаемыми статистически независимыми моделями. Теоретическое исследование подобной ситуации показывает следующее.

Пусть DGP: $x_t = x_{t-1} + \varepsilon_{1t}$, $y_t = y_{t-1} + \varepsilon_{2t}$, где $x_0 = 0$, $y_0 = 0$, а ε_{1t} и ε_{2t} – статистически независимые между собой последовательности одинаково распределенных случайных величин, $\varepsilon_{1t} \sim N(0, \sigma_1^2)$, $\varepsilon_{2t} \sim N(0, \sigma_2^2)$, так что $Cov(x_t, y_t) = 0$. Предположим, что по T наблюдениям (x_t, y_t) , $t = 1, 2, \dots, T$, производится оценивание статистической модели

SM: $y_t = \beta x_t + u_t$, $u_t \sim i.i.d. N(0, \sigma_u^2)$, $Cov(x_t, u_t) = 0$.

Стандартная оценка наименьших квадратов для коэффициента β в этой гипотетической модели имеет вид

$$\hat{\beta}_T = \left(\sum_{t=1}^T y_t x_t \right) / \left(\sum_{t=1}^T x_t^2 \right).$$

При сделанных предположениях относительно DGP, $\hat{\beta}_T$ не сходится по вероятности при $T \rightarrow \infty$ ни к какой константе и имеет предельное распределение, отличное от нормального.

С другой стороны, при выбранной спецификации SM модели, в предположениях этой модели (а не DGP !) имеем:

$$Cov(x_t, y_t) = Cov(x_t, \beta x_t + u_t) = \beta Cov(x_t, x_t) = \beta D(x_t),$$

т.е. оцениваемым параметром является

$$\beta = Cov(x_t, y_t) / D(x_t).$$

Поскольку же в действительности (в DGP) $Cov(x_t, y_t) = 0$, то и это значение $\beta = 0$, так что если бы гипотетическая модель (соответствующая SM) была верна, то тогда мы бы имели

$$\hat{\beta}_T \rightarrow 0 \text{ по вероятности.}$$

Далее, при $T \rightarrow \infty$ значения t -статистики t_β для проверки гипотезы $H_0: \beta = 0$ неограниченно возрастают по абсолютной величине, так что использование таблиц t -распределения будет практически всегда приводить к отклонению этой гипотезы, т.е. к выводу о том, что между переменными x_t и y_t существует линейная регрессионная связь. В действительности, нетривиальное предельное распределение имеет не статистика t_β , а статистика $(1/\sqrt{T}) t_\beta$, причем предельное распределение последней является нестандартным.

Что касается статистики Дарбина – Уотсона (DW), то при $T \rightarrow \infty$

$DW \rightarrow 0$ по вероятности,

и это позволяет распознавать неправильную спецификацию статистической модели в форме паразитной регрессии. Последнее обстоятельство проявляется в поведении остатков от оцененной статистической модели, которое не соответствует поведению стационарного процесса.

Пример

В предыдущем примере мы имели

$$\text{DGP: } x_t = x_{t-1} + \varepsilon_{1t}, \quad y_t = y_{t-1} + \varepsilon_{2t},$$

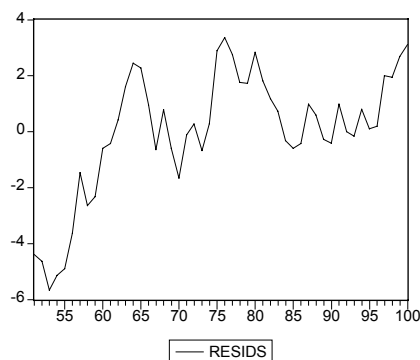
где ε_{1t} и ε_{2t} – последовательности независимых, одинаково распределенных случайных величин, имеющих стандартное нормальное распределение $N(0, 1)$, и оценивали статистическую модель

$$\text{SM: } y_t = \alpha + \beta x_t + \varepsilon_t.$$

При оценивании такой статистической модели по наблюдениям с 51 по 100 мы получили:

$$\hat{\beta}_T = 0.598, \quad t_{\beta} = 7.708, \quad DW = 0.214.$$

При этом график остатков



не похож на график стационарного ряда.

Естественно было бы попытаться использовать для выявления такого “неподобающего” поведения остатков не просто визуальное рассмотрение графика остатков, но и формальные статистические критерии, тем более, что критерии проверки интегрированности временных рядов мы уже обсуждали ранее (критерии Дики – Фуллера, Филлипса – Перрона и др.).

Дело, однако, в том, что теперь мы имеем дело не с “сырым” рядом, а с рядом остатков, которые вычисляются после предварительного оценивания модели (коэффициентов α и β в последнем примере). Это обстоятельство существенно влияет на распределения соответствующих статистик и не дает возможности пользоваться таблицами, которые мы использовали ранее при анализе на интегрированность сырых рядов. С учетом этого обстоятельства, были построены таблицы, позволяющие производить анализ остатков в случае интегрированных объясняемой и объясняющих переменных, о чем мы поговорим в дальнейшем. Сейчас же только отметим, что дает применение соответствующих таблиц к только что рассмотренному примеру.

Пример (продолжение)

При оценивании статистической модели SM: $y_t = \alpha + \beta x_t + \varepsilon_t$ по наблюдениям с 51 по 100 мы получили оцененную модель

$$y_t = 8.616 + 0.598 x_t + e_t.$$

С полученным рядом остатков мы поступим так же, как и в случае применения критерия Дики – Фуллера к сырому ряду, т.е. оценим модель

$$\Delta e_t = \phi e_{t-1} + v_t$$

и вычислим t - статистику t_ϕ для проверки гипотезы $H_0: \phi = 0$, интерпретируя эту гипотезу как гипотезу единичного корня для ряда остатков.

Гипотеза H_0 отвергается в пользу гипотезы $H_A: \phi < 0$ (интерпретируемой как гипотеза стационарности ряда остатков), если $t_\phi < t_{\text{крит}}$. Приближенные значения $t_{\text{крит}}$ можно найти по формуле

$$t_{\text{крит}} \approx \hat{\beta}_\infty + \hat{\beta}_1 T^{-1} + \beta_2 T^{-2},$$

где $\hat{\beta}_\infty, \hat{\beta}_1, \hat{\beta}_2$ зависят от выбранного уровня значимости и указаны в статье [MacKinnon (1991)]. Для 5% уровня значимости

$$t_{\text{крит}} \approx -3.3377 + 5.967 T^{-1} - 8.98 T^{-2},$$

что при $T = 50$ дает $t_{\text{крит}} = -3.46$. Последнее значение существенно меньше 5% критического значения статистики Дики – Фуллера -2.92 , рассчитанного для случая сырого ряда.

В нашем примере оценивание тестового уравнения Дики – Фуллера дает значение $t_\phi = -2.01$. Последнее существенно выше 5% критического уровня, и гипотеза единичного корня не отвергается.

Вообще говоря, вопрос о ложной (паразитной) или неложной (действительной) линейной регрессионной связи между двумя переменными x_t и y_t , представляющими интегрированные ряды первого порядка ($x_t, y_t \sim I(1)$), более точно формулируется следующим образом. **Существует ли вообще такое значение β , при котором ряд $y_t - \beta x_t$ стационарен?**

Если ответ на этот вопрос положительный, то говорят что *ряды x_t и y_t (переменные x_t и y_t) коинтегрированы*. Если же ответ оказывается отрицательным, то *ряды x_t и y_t (переменные x_t и y_t) не являются коинтегрированными*.

В последнем случае непосредственное оценивание модели $y_t = \alpha + \beta x_t + u_t$ бессмысленно, т.к. получаемая оценка $\hat{\beta}_T$, собственно говоря, не является оценкой какого-либо теоретического параметра связи между переменными x_t и y_t . (См., впрочем, замечание в конце разд. 7.3.)

Напротив, если переменные x_t и y_t коинтегрированы, то $\hat{\beta}_T$ является оценкой того единственного значения β , при котором ряд $y_t - \beta x_t$ стационарен

Заметим теперь, что если в

$$\text{DGP: } x_t = x_{t-1} + \varepsilon_{1t},$$

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

допустить коррелированность значений ε_{1t} и ε_{2t} в совпадающие моменты времени, т.е. $\text{Cov}(\varepsilon_{1t}, \varepsilon_{2t}) \neq 0$, то коррелированность ε_{1t} и ε_{2t} вовсе не означает, что ряды x_t и y_t коинтегрированы. Предположим все же, что существует некоторое значение β , при котором $y_t = \beta x_t + u_t$, где u_t – стационарный ряд. Тогда $y_{t-1} - \beta x_{t-1} = u_{t-1}$, так что $\Delta y_t = \beta \Delta x_t + \Delta u_t$, а отсюда $\varepsilon_{2t} = \beta \varepsilon_{1t} + \Delta u_t$. Последнее можно записать в виде $u_t = u_{t-1} + \eta_t$, где

$$\eta_t = \varepsilon_{2t} - \beta \varepsilon_{1t} \sim i.i.d. N(0, \sigma_\eta^2).$$

Но это означает, что u_t – нестационарный процесс.

С другой стороны, если $x_0 = y_0 = 0$, то

$$\text{Cov}(x_t, y_t) = \text{Cov}(\varepsilon_{11} + \dots + \varepsilon_{1t}, \varepsilon_{21} + \dots + \varepsilon_{2t}) = t \text{Cov}(\varepsilon_{11}, \varepsilon_{21}),$$

так что x_t и y_t – коррелированные, но не коинтегрированные случайные блуждания.

Существенно, что распределение статистики Дики – Фуллера в подобной ситуации не зависит от конкретного вида матрицы ковариаций $\Sigma = (\text{Cov}(\varepsilon_{k1}, \varepsilon_{s1}))$, $k, s = 1, 2$.

Тем же свойством обладает и распределение статистики Дарбина – Уотсона, примененной к ряду остатков (**CRDW – cointegrating regression DW**):

$$e_t = y_t - \hat{\alpha}_T - \hat{\beta}_T x_t.$$

При $T = 50$ 5% критическое значение последней статистики равно 0.78. Гипотеза некоинтегрированности рядов отвергается, если наблюдаемое значение этой статистики превышает критическое значение.

В нашем последнем примере значение статистики Дарбина – Уотсона равно 0.214, так что гипотеза некоинтегрированности не отвергается и этим критерием.

Что следует делать в случае обнаружения паразитной связи между интегрированными порядка 1 переменными x_t и y_t ? Имеются три возможных пути обхода возникающих здесь трудностей:

1. Включить в правую часть уравнения запаздывающие значения обеих переменных, точнее, рассмотреть статистическую модель

$$\text{SM: } y_t = \alpha + \beta x_t + \gamma y_{t-1} + \delta x_{t-1} + u_t,$$

где u_t – стационарный ряд и переменная x_t трактуется как экзогенная.

Последнее уравнение можно записать иначе в следующих двух формах:

$$(a) \quad y_t = \alpha + \gamma y_{t-1} + \beta \Delta x_t + (\beta + \delta) x_{t-1} + u_t,$$

$$(б) \quad y_t = \alpha + \gamma y_{t-1} + (\beta + \delta) x_t - \delta \Delta x_t + u_t.$$

В обеих формах слева стоит интегрированная переменная $y_t \sim I(1)$.

В правой части (а) параметр β является коэффициентом при стационарной переменной Δx_t , имеющей нулевое математическое ожидание; $y_{t-1}, x_{t-1} \sim I(1)$, u_t – стационарный ряд. Как было показано в работе [Sims, Stock, Watson (1990)], в такой ситуации оценки наименьших квадратов для всех коэффициентов SM состоятельны, оценка параметра β асимптотически нормальна. Обычная t -статистика для проверки гипотезы $H_0: \beta = 0$ имеет асимптотически нормальное распределение $N(0,1)$, если u_t – белый шум. Аналогично, в правой части (б) параметр $-\delta$ является коэффициентом при стационарной переменной Δx_t , имеющей нулевое математическое ожидание; $y_{t-1}, x_t \sim I(1)$, u_t – стационарный ряд. Поэтому оценка параметра δ в рамках модели SM асимптотически нормальна, и t -статистика для проверки гипотезы $H_0: \delta = 0$ имеет асимптотически нормальное распределение $N(0,1)$, если u_t – белый шум. Оценки для β и δ остаются асимптотически нормальными и если u_t – стационарный ряд, не являющийся белым шумом. Однако при этом асимптотическое распределение $N(0,1)$ имеют

скорректированные варианты t -статистик, в знаменателях которых стандартные оценки дисперсии ряда u_t заменяются состоятельными оценками **долговременной дисперсии** этого ряда, определенной в разд. 6.8.1.

В то же время, статистика $qF = 2F$ для проверки гипотезы $H_0: \beta = \delta = 0$ не имеет асимптотического распределения $\chi^2(2)$, поскольку рассматриваемую SM не удастся линейно репараметризовать таким образом, чтобы в правой части преобразованного уравнения и β и δ одновременно стали коэффициентами при стационарных переменных, имеющих нулевые математические ожидания. (У нас они становятся таковыми при разных репараметризациях.)

2. Перед оцениванием модели связи продифференцировать ряды x_t и y_t ,

т.е. рассмотреть модель в разностях

SM: $\Delta y_t = \alpha + \beta \Delta x_t + u_t$, где u_t – стационарный ряд.

В этой модели оценки наименьших квадратов и для α и для β асимптотически нормальны. Обе t -статистики имеют асимптотически нормальное распределение $N(0,1)$, если u_t – белый шум. Если u_t – стационарный ряд, не являющийся белым шумом, то необходимо произвести коррекцию t -статистик, как и в предыдущем пункте.

3. Использовать для оценивания модель регрессии с автокоррелированными остатками

SM: $y_t = \alpha + \beta x_t + u_t$, $u_t = \rho x_{t-1} + \varepsilon_t$, $u_t \sim i.i.d. N(0, \sigma_\varepsilon^2)$.

При этом предпочтительнее оценивать все три параметра α , β , ρ одновременно, используя представление

$$y_t - \rho y_{t-1} = \alpha(1 - \rho) + \beta(x_t - \rho x_{t-1}) + \varepsilon_t.$$

В случае ложной регрессии $\hat{\rho} \rightarrow 1$ (по вероятности), так что, фактически, при больших T этот метод равносителен предварительному дифференцированию рядов.

Пример

Применим указанные три подхода к анализу реализаций, смоделированных ранее в этом разделе в соответствии с

DGP: $x_t = x_{t-1} + \varepsilon_{1t}$, $y_t = y_{t-1} + \varepsilon_{2t}$,

где ε_{1t} и ε_{2t} – последовательности независимых, одинаково распределенных случайных величин, имеющих стандартное нормальное распределение $N(0,1)$. Для анализа используем последние 50 наблюдений.

(1) Применяя первый подход, получаем оцененную модель

Dependent Variable: Y

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.447271	0.550358	0.812691	0.4206
X	-0.014458	0.123861	-0.116725	0.9076
Y(-1)	0.970105	0.055989	17.32664	0.0000
X(-1)	0.033532	0.120061	0.279290	0.7813

Наблюдаемые Р-значения для коэффициентов при переменных x_t и x_{t-1} указывают на то, что переменная x_t фактически не объясняет изменчивость переменной y_t .

(2) Применяя второй подход, получаем оцененную модель

Dependent Variable: D(Y)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.184523	0.117614	1.568884	0.1232
D(X)	-0.033386	0.116361	-0.286915	0.7754

Оба коэффициента статистически незначимы, и это отражает некоррелированность рядов ε_{1t} и ε_{2t} .

(3) Применяя третий подход, оцениваем модель

$$y_t - \rho y_{t-1} = \alpha(1 - \rho) + \beta(x_t - \rho x_{t-1}) + \varepsilon_t,$$

т.е.

$$y_t = \alpha^* + \rho y_{t-1} + \beta(x_t - \rho x_{t-1}) + \varepsilon_t.$$

При этом получаем:

Dependent Variable: Y

Method: Least Squares

Convergence achieved after 7 iterations

Y=C(1)+C(2)*Y(-1)+C(3)*(X-C(2)*X(-1))

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	0.217398	0.159423	1.363650	0.1792
C(2)	0.988791	0.035801	27.61920	0.0000
C(3)	-0.027306	0.119276	-0.228934	0.8199

R-squared 0.940380 Mean dependent var 3.404232

Как и ожидалось, коэффициент при y_{t-1} оказался очень близким к 1, а два других коэффициента статистически незначимы. Проверка на одновременное зануление этих двух коэффициентов дает P -значение 0.367.

Пример

Изменим процесс порождения данных, оставляя те же формулы для x_t и y_t , т.е.

$$x_t = x_{t-1} + \varepsilon_{1t}, \quad y_t = y_{t-1} + \varepsilon_{2t},$$

но теперь пусть

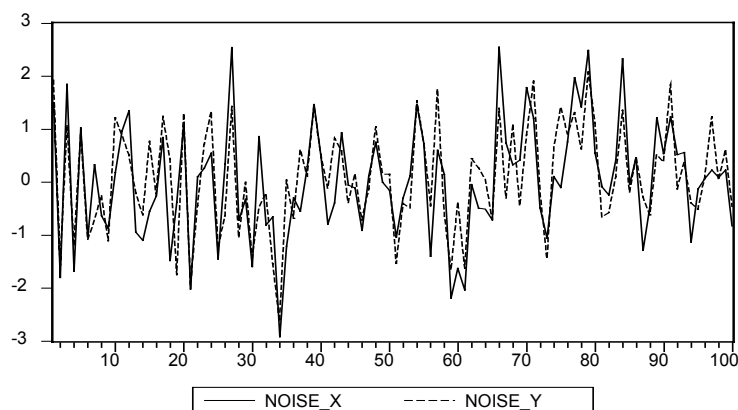
ε_{1t} — последовательность независимых, одинаково распределенных случайных величин, имеющих нормальное распределение $N(0, 1.25)$,

ε_{2t} — последовательность независимых, одинаково распределенных случайных величин, имеющих нормальное распределение $N(0, 1.25)$,

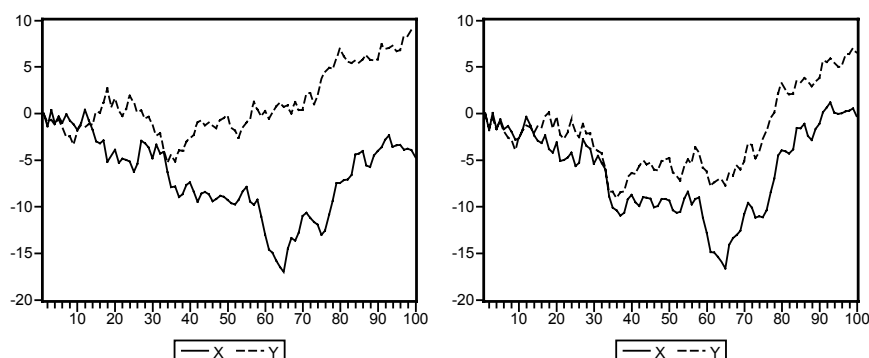
$Cov(\varepsilon_{1t}, \varepsilon_{2s}) = 0$ для $t \neq s$, $Cov(\varepsilon_{1t}, \varepsilon_{2t}) = 1$.

Отсюда, в частности, следует, что $Corr(\varepsilon_{1t}, \varepsilon_{2t}) = 0.8$.

Смоделированные реализации ε_{1t} и ε_{2t} имеют вид



Траектории смоделированной пары рядов ε_{1t} и ε_{2t} ведут себя достаточно согласованным образом; оцененный коэффициент корреляции между этими рядами равен 0.789. Полученные при этом реализации рядов x_t и y_t ведут себя, как показано ниже на правом рисунке. Для сравнения на левом рисунке показано поведение реализаций рядов ε_{1t} и ε_{2t} при их полной статистической независимости.



Для сопоставимости с ранее полученными результатами, опять обратимся ко второй части отрезка наблюдений; здесь оцененный коэффициент корреляции между рядами ε_{1t} и ε_{2t} равен 0.792.

Сначала оценим модель $y_t = \alpha + \beta x_t + u_t$. В результате получаем для ряда остатков значение статистики Дики – Фуллера, равное -2.112 , которое выше 5% критического уровня -3.46 . Соответственно, гипотеза о ложности регрессионной связи не отвергается.

(1) Применяя первый подход, получаем оцененную модель

Dependent Variable: Y

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.548392	0.377080	1.454312	0.1526
X	0.718479	0.079943	8.987361	0.0000
Y(-1)	0.913556	0.067811	13.47210	0.0000
X(-1)	-0.641522	0.088805	-7.223976	0.0000
R-squared	0.987574	Mean dependent var	-0.957402	

По сравнению с ранее рассмотренным случаем, в котором ряды ε_{1t} и ε_{2t} были между собой статистически не связанными, теперь оказываются статистически значимыми и

коэффициенты при переменных x_t и x_{t-1} . Исключая из правой части модели константу, получаем:

Dependent Variable: Y

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X	0.695862	0.079341	8.770553	0.0000
Y(-1)	1.005257	0.025245	39.82053	0.0000
X(-1)	-0.707002	0.077447	-9.128837	0.0000
R-squared	0.987003	Mean dependent var	-0.957402	

т.е. $y_t = 1.005 y_{t-1} + 0.695 x_t - 0.707 x_{t-1} + e_t$. Оценка коэффициента при y_{t-1} близка к 1; оцененные коэффициенты при x_t и x_{t-1} близки по абсолютной величине и противоположны по знаку, что вполне согласуется с реализованной моделью DGP.

(2) Применяя второй подход, получаем оцененную модель

Dependent Variable: D(Y)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.100915	0.083851	1.203508	0.2347
D(X)	0.694000	0.077274	8.981039	0.0000
R-squared	0.626921	Mean dependent var	0.226857	

И здесь, в отличие от ранее использовавшегося DGP, становится значимым коэффициент при переменной Δx_t , что отражает коррелированность случайных величин ε_{1t} и ε_{2t} , т.е. коррелированность Δx_t и Δy_t . Исключая из правой части уравнения статистически незначимую константу, получаем:

Dependent Variable: D(Y)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(X)	0.709553	0.076533	9.271155	0.0000

т.е. $\Delta y_t = 0.710 \Delta x_t + e_t$, или $y_t = y_{t-1} + 0.710 x_t - 0.710 x_{t-1} + e_t$.

(3) Наконец, применяя третий подход, оцениваем модель

$$y_t = \alpha^* + \rho y_{t-1} + \beta (x_t - \rho x_{t-1}) + \varepsilon_t;$$

при этом получаем:

Dependent Variable: Y

Convergence achieved after 8 iterations

Y=C(1)+C(2)*Y(-1)+C(3)*(X-C(2)*X(-1))

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	0.329205	0.250398	1.314726	0.1950
C(2)	0.941984	0.056946	16.54170	0.0000
C(3)	0.723593	0.079341	9.119982	0.0000
R-squared	0.987410	Mean dependent var	-0.957402	

Здесь становится статистически значимым коэффициент β .

Исключение из правой части константы дает:

Dependent Variable: Y

Convergence achieved after 4 iterations

Y=C(2)*Y(-1)+C(3)*(X-C(2)*X(-1))

	Coefficient	Std. Error	t-Statistic	Prob.
C(2)	1.014411	0.020750	48.88608	0.0000
C(3)	0.702102	0.078268	8.970448	0.0000

т.е. $y_t = 1.014 y_{t-1} + 0.702 (x_t - 1.014 x_{t-1}) + e_t$, или
 $y_t = 1.014 y_{t-1} + 0.702 x_t - 0.712 x_{t-1} + e_t$

Отметим близость результатов, полученных тремя методами:

$y_t = 1.005 y_{t-1} + 0.695 x_t - 0.707 x_{t-1} + e_t$ (метод 1),

$y_t = y_{t-1} + 0.710 x_t - 0.710 x_{t-1} + e_t$ (метод 2),

$y_t = 1.014 y_{t-1} + 0.702 x_t - 0.712 x_{t-1} + e_t$ (метод 3).

Фактически, во всех трех случаях воспроизводится одна и та же линейная модель связи между рядами разностей:

$$\Delta y_t = 0.7 \Delta x_t + e_t.$$

Эта регрессионная связь между продифференцированными рядами не является ложной (в отличие от регрессионной связи между рядами уровней): статистика Дарбина – Уотсона принимает значение 1.985; Р-значение критерия Jarque – Bera равно 0.344.

Замечание

В связи с результатами, полученными при рассмотрении последних примеров, естественно возникает следующий вопрос, который поднимался в свое время различными исследователями. Не будет ли разумным, имея дело с рядами, траектории которых обнаруживают выраженный тренд, сразу приступить к оцениванию связей между рядами разностей (между продифференцированными рядами) ?

Против некритичного использования такого подхода говорят два обстоятельства:

(а) Если ряды в действительности стационарны относительно детерминированного тренда, то тогда дифференцирование приводит к ***передифференцированным*** рядам, имеющим необратимую МА составляющую.

(б) Если ряды являются интегрированными порядка 1 и при этом коинтегрированы, то при переходе к продифференцированным рядам теряется информация о долговременной связи между уровнями этих рядов.

Дифференцирование рядов оправданно и полезно, если ряды являются интегрированными, но при этом между ними отсутствует коинтеграционная связь.

7.2. Коинтегрированные временные ряды. Модели коррекции ошибок

Пусть $y_t \sim I(1)$, $x_t \sim I(0)$. Строить регрессию y_t на x_t в этом случае бессмысленно, т.к. для любых a и b в такой ситуации

$$y_t - a - b x_t \sim I(1).$$

Пусть, наоборот, $y_t \sim I(0)$, $x_t \sim I(1)$. Для любых a и $b \neq 0$ здесь опять

$$y_t - a - b x_t \sim I(1),$$

и только при $b = 0$ получаем

$$y_t - a - b x_t \sim I(0),$$

так что и в таком сочетании строить регрессию одного ряда на другой не имеет смысла.

Пусть теперь $y_t \sim I(1)$, $x_t \sim I(1)$ – два интегрированных ряда.

Если для любого b

$$y_t - b x_t \sim I(1),$$

то регрессия y_t на x_t является фиктивной, и мы уже выяснили, как следует действовать в такой ситуации.

Обратимся теперь к случаю, когда при некотором $b \neq 0$

$$y_t - b x_t \sim I(0) \text{ – стационарный ряд.}$$

Если это так, то ряды y_t и x_t называют **коинтегрированными** рядами, а вектор $(1, -b)^T$ – **коинтегрирующим вектором**.

Вообще, ряды $y_t \sim I(1)$, $x_t \sim I(1)$ называют **коинтегрированными (в узком смысле – детерминистская коинтеграция)**, если существует ненулевой (коинтегрирующий) вектор $\beta = (\beta_1, \beta_2)^T \neq 0$, для которого

$$\beta_1 x_t + \beta_2 y_t \sim I(0) \text{ – } \underline{\text{стационарный}} \text{ ряд.}$$

Заметим, что если вектор $\beta = (\beta_1, \beta_2)^T$ является коинтегрирующим вектором для рядов x_t и y_t , то тогда коинтегрирующим для этих рядов будет и любой вектор вида $c\beta = (c\beta_1, c\beta_2)^T$, где $c \neq 0$ – постоянная величина. Чтобы выделить какой-то определенный вектор, приходится вводить условие нормировки, например, рассматривать только векторы вида $(1, -b)^T$ (или только векторы $(-a, 1)^T$).

Поскольку мы предполагаем сейчас, что $x_t, y_t \sim I(1)$, то ряды разностей $\Delta x_t, \Delta y_t$ стационарны. Будем предполагать в дополнение, что стационарен векторный ряд $(\Delta x_t, \Delta y_t)^T$, так что для него существует разложение Вольда в виде скользящего среднего

$$(\Delta x_t, \Delta y_t)^T = \mu + B(L) \varepsilon_t,$$

где

$$\mu = (\mu_1, \mu_2)^T, \mu_1 = E(\Delta x_t), \mu_2 = E(\Delta y_t);$$

$$\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})^T \text{ – векторный белый шум,}$$

т.е.

$\varepsilon_1, \varepsilon_2, \dots$ – последовательность не коррелированных между собой, одинаково распределенных случайных векторов, для которых

$$E(\varepsilon_t) = (0, 0)^T, D(\varepsilon_{1t}) = \sigma_1^2, D(\varepsilon_{2t}) = \sigma_2^2, Cov(\varepsilon_{1t}, \varepsilon_{2t}) = \sigma_{12} \text{ – постоянные величины;}$$

$$B(L) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \sum_{k=1}^{\infty} \begin{pmatrix} b_{11}^{(k)} & b_{12}^{(k)} \\ b_{21}^{(k)} & b_{22}^{(k)} \end{pmatrix} L^k.$$

Знаменитый результат Гренджера ([Granger (1983)], см. также [Engle, Granger (1987)]) состоит в том, что в случае коинтегрированности $I(1)$ рядов x_t и y_t (в узком смысле)

(I) В разложении Вольда $(\Delta x_t, \Delta y_t)^T = \mu + B(L) \varepsilon_t$ матрица $B(1)$ имеет ранг 1.

(II) Система рядов x_t и y_t допускает **векторное ARMA представление**

$$A(L) (x_t, y_t)^T = c + d(L) \varepsilon_t,$$

в котором

$$\varepsilon_t \text{ – тот же векторный белый шум, что и в (I),}$$

$$c = (c_1, c_2)^T, c_1 \text{ и } c_2 \text{ – постоянные,}$$

$$A(L) \text{ – матричный полином от оператора запаздывания,}$$

$$d(L) \text{ – скалярный полином от оператора запаздывания,}$$

причем

$A(0) = I_2$ (единичная матрица размера 2×2),
 $\text{rank } A(1) = 1$ (ранг 2×2 -матрицы $A(1)$ равен 1),
 значение $d(1)$ конечно.

В связи с тем, что в последнем представлении ранг (2×2) -матрицы $A(1)$ меньше двух, об этом представлении часто говорят как о **векторной авторегрессии пониженного ранга (reduced rank VAR)**.

В развернутой форме представление (II) имеет вид

$$\begin{cases} x_t = c_1 + \sum_{j=1}^p (a_{1j}x_{t-j} + b_{1j}y_{t-j}) + \sum_{k=0}^q \theta_k \varepsilon_{1,t-k}, \\ y_t = c_2 + \sum_{j=1}^p (a_{2j}x_{t-j} + b_{2j}y_{t-j}) + \sum_{k=0}^q \theta_k \varepsilon_{2,t-k} \end{cases}$$

При этом верхние пределы p и q у сумм в правых частях могут быть бесконечными.

Если возможно **векторное AR представление**, то в нем

$$d(L) \equiv 1, \quad p < \infty.$$

(III) Система рядов x_t и y_t допускает представление в форме **модели коррекции ошибок (error correction model – ECM)**

$$\Delta x_t = \mu_1 + \alpha_1 z_{t-1} + \sum_{j=1}^{\infty} (\gamma_{1j} \Delta x_{t-j} + \delta_{1j} \Delta y_{t-j}) + \sum_{k=0}^{\infty} \theta_k \varepsilon_{1,t-k},$$

$$\Delta y_t = \mu_2 + \alpha_2 z_{t-1} + \sum_{j=1}^{\infty} (\gamma_{2j} \Delta x_{t-j} + \delta_{2j} \Delta y_{t-j}) + \sum_{k=0}^{\infty} \theta_k \varepsilon_{2,t-k},$$

где

$z_t = y_t - \beta x_t - E(y_t - \beta x_t)$ – стационарный ряд с нулевым математическим ожиданием,

$$z_t \sim I(0),$$

и

$$\alpha_1^2 + \alpha_2^2 > 0.$$

Если в (II) возможно векторное $AR(p)$ представление ($p < \infty$), то тогда ЕСМ принимает вид

$$\Delta x_t = \mu_1 + \alpha_1 z_{t-1} + \sum_{j=1}^{p-1} (\gamma_{1j} \Delta x_{t-j} + \delta_{1j} \Delta y_{t-j}) + \varepsilon_{1,t},$$

$$\Delta y_t = \mu_2 + \alpha_2 z_{t-1} + \sum_{j=1}^{p-1} (\gamma_{2j} \Delta x_{t-j} + \delta_{2j} \Delta y_{t-j}) + \varepsilon_{2,t},$$

Здесь важно отметить следующее:

- Если ряды $x_t, y_t \sim I(1)$ коинтегрированы, то все составляющие в ЕСМ стационарны.
- Если векторный ряд $(x_t, y_t)^T \sim I(1)$ (так что векторный ряд $(\Delta x_t, \Delta y_t)^T$ стационарен) и порождается ЕСМ моделью, то ряды x_t и y_t коинтегрированы. (Действительно, в этом случае все составляющие ЕСМ, отличные от z_{t-1} , стационарны; но тогда стационарна и z_{t-1} .)

- Если ряды $x_t, y_t \sim I(1)$ коинтегрированы, то тогда VAR в разностях не может иметь конечный порядок. (В отличие от случая, когда ряды x_t и y_t не коинтегрированы.)

Абсолютную величину $z_t = y_t - \alpha - \beta x_t$, где $\alpha = E(y_t - \beta x_t)$, можно рассматривать как расстояние, отделяющее систему в момент t от равновесия, задаваемого соотношением $y_t - \alpha - \beta x_t = 0$. Величины и направления изменений x_t и y_t принимают во внимание величину и знак предыдущего отклонения от равновесия z_{t-1} . Ряд z_t , конечно, вовсе не обязательно убывает по абсолютной величине при переходе от одного периода времени к другому, но он является стационарным рядом, и поэтому расположен к движению по направлению к своему среднему.

Замечание 1

Переменная x_t не является причиной по Гренджеру для переменной y_t , если неучет прошлых значений переменной x_t не приводит к ухудшению качества прогноза значения y_t по совокупности прошлых значений этих двух переменных. Переменная y_t не является причиной по Гренджеру для переменной x_t , если неучет прошлых значений переменной y_t не приводит к ухудшению качества прогноза значения x_t по совокупности прошлых значений этих двух переменных. (Качество прогноза измеряется среднеквадратичной ошибкой прогноза.)

Если $x_t, y_t \sim I(1)$ и коинтегрированы, то должна иметь место причинность по Гренджеру, по крайней мере, в одном направлении. Этот факт вытекает из представления такой системы рядов в форме ЕСМ, в которой $\alpha_1^2 + \alpha_2^2 > 0$. Значение x_{t-1} через посредство z_{t-1} помогает в прогнозировании значения y_t (т.е. переменная x_t является причиной по Гренджеру для переменной y_t), если $\alpha_2 \neq 0$. Значение y_{t-1} через посредство z_{t-1} помогает в прогнозировании значения x_t (т.е. переменная y_t является причиной по Гренджеру для переменной x_t), если $\alpha_1 \neq 0$.

Замечание 2

Пусть $x_t, y_t \sim I(1)$ коинтегрированы и $w_t \sim I(0)$. Тогда для любого k коинтегрированы ряды x_t и $\gamma y_{t-k} + w_t$, $\gamma \neq 0$. Формально, если $x_t \sim I(1)$, то коинтегрированы ряды x_t и x_{t-k} . (Действительно, тогда $x_t - x_{t-k} = \Delta x_t + \Delta x_{t-1} + \dots + \Delta x_{t-k} -$ сумма $I(0)$ -переменных, которая также является $I(0)$ -переменной.)

Итак, при коинтегрированности рядов $x_t, y_t \sim I(1)$ мы имеем

- модель долговременной (равновесной) связи $y_t = \alpha + \beta x_t$;
- модель краткосрочной динамики в форме ЕСМ,

и эти модели согласуются друг с другом.

Проблема, однако, состоит в том, что для построения ЕСМ по реальным статистическим данным нам надо знать коинтегрирующий вектор (в данном случае, знать значение β). Хорошо, если этот вектор определяется экономической теорией. К сожалению, чаще его приходится оценивать по имеющимся данным.

Энгл и Гренджер [Engle, Granger (1987)] рассмотрели двухшаговую процедуру, в которой на первом шаге значения α и β оцениваются в рамках модели регрессии y_t на x_t

$$y_t = \alpha + \beta x_t + u_t.$$

Получив методом наименьших квадратов оценки $\hat{\alpha}$ и $\hat{\beta}$ (НК-оценки), мы тем самым находим оцененные значения отклонений от положения равновесия

$$\hat{z}_t = y_t - \hat{\alpha} - \hat{\beta} x_t$$

– это просто остатки от оцененной регрессии.

После этого, на втором шаге, методом наименьших квадратов отдельно (не как система!) оцениваются уравнения

$$\Delta x_t = \mu_1 + \alpha_1 \hat{z}_{t-1} + \sum_{j=1}^{p-1} (\gamma_{1j} \Delta x_{t-j} + \delta_{1j} \Delta y_{t-j}) + v_t,$$

$$\Delta y_t = \mu_2 + \alpha_2 \hat{z}_{t-1} + \sum_{j=1}^{p-1} (\gamma_{2j} \Delta x_{t-j} + \delta_{2j} \Delta y_{t-j}) + w_t,$$

(т.е. предполагается модель VAR(p) для x_t, y_t).

Определяющим в этой процедуре является то обстоятельство, что получаемая на первом шаге оценка $\hat{\beta}$ быстрее обычного приближается (по вероятности) к истинному значению β – второй компоненте коинтегрирующего вектора $(1, \beta)^T$. ($\hat{\beta}$ является **суперсостоятельной оценкой** для β .) Это, в конечном счете, приводит к тому, что оценки в отдельном уравнении ЕСМ, использующие оцененные значения z_{t-1} , имеют то же самое асимптотическое распределение, что и оценка максимального правдоподобия, использующая истинные значения z_{t-1} . (Обычно это асимптотически нормальное распределение.) При этом НК-оценки стандартных ошибок всех коэффициентов являются состоятельными оценками истинных стандартных ошибок.

Заметим, что последние результаты справедливы несмотря на то, что ряд оцененных значений \hat{z}_t формально не является стационарным, поскольку $\hat{\beta} \neq \beta$.

Отметим также, что если мы хотим использовать другую нормировку коинтегрирующего вектора в виде $(\beta, 1)^T$, то нам придется оценивать регрессию x_t на константу и y_t , и это приведет к вектору, не пропорциональному вектору, оцененному в первом случае.

Замечание

Тот факт, что $\hat{\beta}$ быстрее обычного сходится (по вероятности) к β , вовсе не означает, что мы можем пользоваться на первом шаге процедуры Энгла – Гренджера обычными регрессионными критериями. Дело в том, что получаемые на первом шаге оценки и статистики, вообще говоря, имеют нестандартные асимптотические распределения.

Однако первый шаг является в данном контексте вспомогательным, и на этом шаге нет необходимости обращать внимание на сообщаемые в протоколах соответствующих пакетов программ значения статистик.

Напротив, на втором шаге мы можем использовать обычные статистические процедуры (разумеется, если количество наблюдений не мало и если коинтеграция имеется).

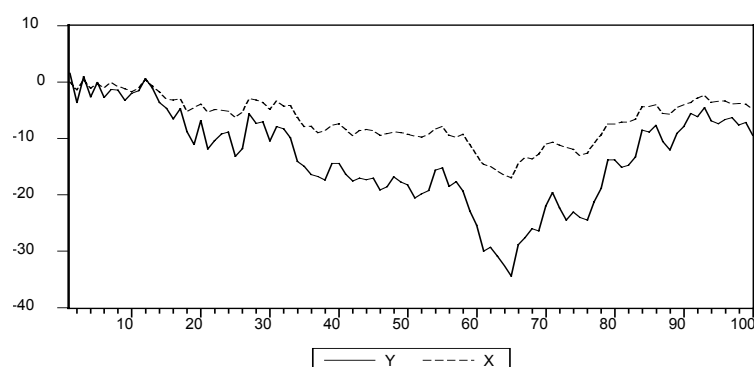
Пример

Рассмотрим реализацию процесса порождения данных

$$\text{DGP: } x_t = x_{t-1} + \varepsilon_t, \quad y_t = 2x_t + v_t,$$

где $x_1 = 0$, а ε_t и v_t – порождаемые независимо друг от друга последовательности независимых, одинаково распределенных случайных величин, имеющих стандартное

нормальное распределение $N(0, 1)$. Графики полученных реализаций рядов x_t и y_t имеют следующий вид



Пара (x_t, y_t) образует векторный процесс авторегрессии

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

$$y_t = 2x_{t-1} + \eta_t,$$

где $\eta_t = v_t + 2\varepsilon_t \sim i.i.d. N(0, 5)$.

В форме ЕСМ пара уравнений принимает вид

$$\Delta x_t = \varepsilon_t,$$

$$\Delta y_t = -(y_{t-1} - 2x_{t-1}) + \eta_t = -z_t + \eta_t,$$

где $z_t = y_t - 2x_t$,

или

$$\Delta x_t = \alpha_1 z_{t-1} + \varepsilon_t,$$

$$\Delta y_t = \alpha_2 z_{t-1} + \eta_t,$$

где $\alpha_1 = 0$, $\alpha_2 = -1$, так что $\alpha_1^2 + \alpha_2^2 > 0$.

На практике, приступая к анализу статистических данных, исследователь не знает точно, какой порядок имеет VAR в DGP. Имея это в виду, выберем для оценивания в качестве статистической модели ЕСМ в виде

$$\Delta x_t = \alpha_1 z_{t-1} + \gamma_{11} \Delta x_{t-1} + \delta_{11} \Delta y_{t-1} + v_t,$$

$$\Delta y_t = \alpha_2 z_{t-1} + \gamma_{21} \Delta x_{t-1} + \delta_{21} \Delta y_{t-1} + w_t,$$

допуская, что данные порождаются моделью векторной авторегрессии второго порядка ($p = 2$). Для анализа используем 100 наблюдений.

(I шаг) Исходим из модели $y_t = \alpha + \beta x_t + u_t$.

Оцененная модель:

Dependent Variable: Y

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.006764	0.165007	-0.040992	0.9674
X	1.983373	0.020852	95.11654	0.0000
R-squared	0.989284	Durbin-Watson stat		2.217786

т.е.

$$y_t = -0.006764 + 1.983373 x_t + \hat{u}_t,$$

так что

$$\hat{z}_t = \hat{u}_t = y_t + 0.006764 - 1.983373 x_t.$$

Допустив, что VAR имеет порядок 2, при использовании критерия Дики – Фуллера для проверки рядов y_t и x_t на коинтегрированность в правую часть уравнения включаем одну запаздывающую разность:

$$\Delta \hat{z}_t = \varphi \hat{z}_{t-1} + \theta_1 \Delta \hat{z}_{t-1} + \zeta_t.$$

Оценивая последнее уравнение получаем:

Augmented Dickey-Fuller Test Equation

Dependent Variable: D(Z)

Sample(adjusted): 3 100

Included observations: 98 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
Z(-1)	-1.153515	0.151497	-7.614088	0.0000
D(Z(-1))	0.038156	0.100190	0.380837	0.7042

Полученное значение тестовой статистики $t_\varphi = -7.614$ намного ниже 5% критического уровня -3.396 (см. [Patterson (2000), таблица 8.7]). Гипотеза некоинтегрированности рассматриваемых рядов уверенно отвергается. (Ввиду статистической незначимости коэффициента при запаздывающей разности, можно было бы переоценить модель, не включая запаздывающую разность в правую часть уравнения. Это дало бы значение $t_\varphi = -11.423$, при котором гипотеза некоинтегрированности отвергается еще более уверенно.)

Таким образом, мы принимаем решение о коинтегрированности рядов y_t и x_t , и переходим к построению модели коррекции ошибок.

(Шаг II) Сначала отдельно оцениваем уравнение для Δx_t :

Dependent Variable: D(X)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.028016	0.100847	-0.277810	0.7818
Z(-1)	0.250942	0.176613	1.420858	0.1587
D(X(-1))	0.639967	0.257823	2.482201	0.0148
D(Y(-1))	-0.258740	0.116654	-2.218019	0.0290

Поочередное исключение из правой части уравнения переменных со статистически незначимыми коэффициентами и наибольшим Р-значением приводит к оцененной модели

Dependent Variable: D(X)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(X(-1))	0.115141	0.100249	1.148554	0.2536

и, в конечном счете, к модели

$$\Delta x_t = v_t,$$

которая и была использована при порождении ряда x_t .

Оценивая теперь уравнение для Δy_t , получаем

Dependent Variable: D(Y)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.060101	0.211899	-0.283630	0.7773
Z(-1)	-0.641060	0.371097	-1.727472	0.0874
D(X(-1))	1.313872	0.541733	2.425311	0.0172
D(Y(-1))	-0.482981	0.245111	-1.970459	0.0517

Исключая из правой части оцениваемого уравнения константу, получаем:

Dependent Variable: D(Y)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
Z(-1)	-0.638888	0.369218	-1.730381	0.0868
D(X(-1))	1.317763	0.538932	2.445138	0.0163
D(Y(-1))	-0.483722	0.243908	-1.983217	0.0502

Хотя формально здесь следовало бы начать исключение статистически незначимых переменных с \hat{z}_{t-1} , мы должны принять во внимание уже принятое решение о коинтегрированности рядов y_t и x_t . Но если эти ряды действительно коинтегрированы, то в ЕСМ должно выполняться соотношение $\alpha_1^2 + \alpha_2^2 > 0$. Поскольку же переменная z_{t-1} не вошла в правую часть уравнения для Δx_t , она должна оставаться в правой части уравнения для Δy_t . Если начать исключение с переменной Δy_{t-1} , то в оцененном редуцированном уравнении

Dependent Variable: D(Y)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
Z(-1)	-1.186411	0.248876	-4.767072	0.0000
D(X(-1))	0.331411	0.210732	1.572671	0.1191

статистически незначим коэффициент при Δx_{t-1} , что приводит нас к уравнению $\Delta y_t = \alpha_2 \hat{z}_{t-1} + w_t$, оценивая которое, получаем

Dependent Variable: D(Y)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
Z(-1)	-1.273584	0.247887	-5.137760	0.0000

Проверка гипотезы $H_0: \alpha_2 = -1$ дает:

Null Hypothesis: C(1)= -1

F-statistic	1.218077	Probability	0.272441
Chi-square	1.218077	Probability	0.269738

Поскольку эта гипотеза не отвергается, мы можем остановиться на модели ЕСМ

$$\Delta x_t = \varepsilon_t, \quad \Delta y_t = -\hat{z}_{t-1} + w_t,$$

где

$$\hat{z}_{t-1} = y_{t-1} + 0.006764 - 1.983373 x_{t-1}.$$

Подстановка последнего выражения для \hat{z}_{t-1} в уравнение для Δy_t приводит к соотношению

$$y_t = -0.0068 + 1.983 x_{t-1} + w_t,$$

которое близко к соотношению

$$y_t = 2 x_{t-1} + \eta_t,$$

соответствующему использованному DGP.

Заметим, наконец, что последовательность $w_t = \Delta y_t + \hat{z}_{t-1}$ идентифицируется по наблюдаемой ее реализации как гауссовский белый шум с оцененной дисперсией 4.62 (использованному DGP соответствует значение 5.00), а последовательность $\varepsilon_t = \Delta x_t$ идентифицируется как гауссовский белый шум с оцененной дисперсией 1.04 (использованному DGP соответствует значение 1.00).

Оценив ЕСМ и остановившись на модели

$$\Delta x_t = \varepsilon_t, \quad \Delta y_t = -\hat{z}_{t-1} + w_t,$$

мы тем самым обнаруживаем, что коррекция производится только в отношении ряда y_t : при положительных \hat{z}_{t-1} , т.е. при

$$y_{t-1} - (-0.0068 + 1.983 x_{t-1}) > 0,$$

в правой части уравнения для Δy_t корректирующая составляющая $-\hat{z}_{t-1}$ отрицательна и действует в сторону уменьшения приращения переменной y_t . Напротив, при отрицательных \hat{z}_{t-1} корректирующая составляющая действует в сторону увеличения приращения переменной y_t .

Прошлые значения переменной x_t через посредство \hat{z}_{t-1} помогают в прогнозировании значения y_t , т.е. переменная x_t является причиной по Гренджеру для переменной y_t . В то же время, прошлые значения переменной y_t никак не помогают прогнозированию значения x_t , так что y_t не является причиной по Гренджеру для x_t .

Заметим далее, что даже если в ЕСМ $Cov(v_t, w_t) \neq 0$, оценивание пары уравнений ЕСМ как системы не повышает эффективности оценок, поскольку в правые части обоих уравнений входят одни и те же переменные.

Рассмотренный в нашем примере процесс порождения данных

$$\text{DGP: } x_t = x_{t-1} + \varepsilon_t, \quad y_t = 2x_t + v_t,$$

является частным случаем модели, известной как **треугольная система Филлипса**. В общем случае (для двух рядов) эта система имеет вид

$$y_t = \beta x_t + v_t,$$

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

где $(\varepsilon_t, v_t)^T \sim i.i.d. N_2(0, \Sigma)$ – последовательность независимых, одинаково распределенных случайных векторов, имеющих двумерное нормальное распределение с нулевым математическим ожиданием и ковариационной матрицей Σ . (Такая последовательность называется **двумерным гауссовским белым шумом**.)

Если матрица Σ диагональная, так что $Cov(\varepsilon_t, v_t) = 0$, то тогда x_t является экзогенной переменной в первом уравнении, и никаких проблем с оцениванием коэффициента β в этом случае не возникает.

Если же $Cov(\varepsilon_t, v_t) \neq 0$, то тогда x_t уже не является экзогенной переменной в первом уравнении, т.к. при этом $Cov(x_t, v_t) = Cov(x_{t-1} + \varepsilon_t, v_t) \neq 0$. Поэтому получаемая в первом уравнении оценка наименьших квадратов для β не имеет даже асимптотически нормального распределения.

В дальнейшем мы еще вернемся к проблеме оценивания коинтегрирующего вектора, а сейчас обратимся к вопросу о коинтеграции нескольких временных рядов.

Пусть мы имеем N временных рядов y_{1t}, \dots, y_{Nt} , каждый из которых является интегрированным порядка 1. Если существует такой вектор $\beta = (\beta_1, \dots, \beta_N)^T$, отличный от нулевого, для которого

$$\beta_1 y_{1t} + \dots + \beta_N y_{Nt} \sim I(0) - \text{стационарный ряд,}$$

то говорят, что эти ряды **коинтегрированы (в узком смысле)**; такой вектор β называется **коинтегрирующим** вектором. Если при этом

$$c = E(\beta_1 y_{1t} + \dots + \beta_N y_{Nt}),$$

то тогда можно говорить о **долговременном положении равновесия системы** в виде

$$\beta_1 y_{1t} + \dots + \beta_N y_{Nt} = c.$$

В каждый конкретный момент времени t существует некоторое **отклонение** системы от этого положения равновесия, характеризующееся величиной

$$z_t = \beta_1 y_{1t} + \dots + \beta_N y_{Nt} - c.$$

Ряд z_t , в силу сделанных предположений, является стационарным рядом, имеющим нулевое математическое ожидание, так что он достаточно часто пересекает нулевой уровень, т.е. система колеблется вокруг указанного выше положения равновесия.

Естественной процедурой для проверки коинтегрированности рядов y_{1t}, \dots, y_{Nt} является построение регрессии одного из этих рядов на остальные $N - 1$ рядов и проверка гипотезы наличия единичного корня у ряда z_t на основании исследования ряда остатков от оцененной регрессии. Иначе говоря, мы оцениваем, например, модель

$$y_{1t} = \theta_1 + \theta_2 y_{2t} + \dots + \theta_N y_{Nt} + u_t,$$

и проверяем гипотезу единичного корня на основании исследования ряда остатков

$$\hat{u}_t = y_{1t} - (\hat{\theta}_1 + \hat{\theta}_2 y_{2t} + \dots + \hat{\theta}_N y_{Nt}),$$

опираясь на статистику Дики – Фуллера. Критические значения можно найти, следуя [MacKinnon (1991)] (см. также [Patterson (2000), таблица A8.1]).

Если гипотеза единичного корня отвергается, то вектор

$$\hat{\beta} = (1, -\hat{\theta}_2, \dots, -\hat{\theta}_N)$$

берется в качестве оцененного коинтегрирующего вектора. При этом отклонение системы от положения равновесия оценивается величиной

$$\hat{z}_t = \hat{u}_t.$$

Поясним теперь, что мы имели в виду, оговаривая, что приведенные выше определения коинтеграции соответствуют коинтеграции в узком смысле.

В приведенных определениях ненулевой вектор $\beta = (\beta_1, \dots, \beta_N)^T$ определялся как коинтегрирующий вектор, если $\beta_1 y_{1t} + \dots + \beta_N y_{Nt}$ – стационарный ряд. Это означает, что если ряды y_{1t}, \dots, y_{Nt} (по крайней мере, некоторые из них) содержат, наряду со стохастическим, еще и детерминированные тренды, то тогда коинтегрирующий вектор должен аннулировать оба вида трендов одновременно. И в связи с этим, коинтеграцию в узком смысле называют еще **детерминистской коинтеграцией**.

7.3. Проверка нескольких рядов на коинтегрированность. Критерии Дики – Фуллера

Здесь надо различать несколько случаев.

(1) Коинтегрирующий вектор определяется экономической теорией.

Тогда надо просто проверить на наличие единичного корня соответствующую линейную комбинацию

$$\beta_1 y_{1t} + \dots + \beta_N y_{Nt} .$$

При этом используются те же критические значения, которые рассчитаны на применение к отдельно взятому ряду; эти значения не зависят от количества задействованных рядов N .

Пусть возможный коинтегрирующий вектор не определен заранее.

Тогда отдельно рассматриваются следующие ситуации.

(2) Ряды y_{1t}, \dots, y_{Nt} не имеют детерминированного тренда (точнее, $E(\Delta y_{kt}) = 0$).

(2a) В коинтеграционное соотношение (SM) константа не включается.

В этом случае мы оцениваем

$$SM: y_{1t} = \gamma_2 y_{2t} + \dots + \gamma_N y_{Nt} + u_t ,$$

получаем ряд остатков

$$\hat{u}_t = y_{1t} - (\hat{\gamma}_2 y_{2t} + \dots + \hat{\gamma}_N y_{Nt}) ,$$

оцениваем модель регрессии

$$\Delta \hat{u}_t = \varphi \hat{u}_{t-1} + \zeta_1 \Delta \hat{u}_{t-1} + \dots + \zeta_K \Delta \hat{u}_{t-K} + \varepsilon_t$$

с достаточным количеством запаздывающих разностей и проверяем гипотезу $H_0: \varphi = 0$ против альтернативы $H_0: \varphi < 0$.

На этот раз критические значения для t -статистики t_φ зависят от количества задействованных рядов N . При большом количестве наблюдений можно использовать

критические значения, приведенные в [Hamilton (1994), Table B.9, Case 1]. Однако на практике в правую часть оцениваемого уравнения константа обычно включается.

(2b) В коинтеграционное соотношение (SM) константа включается.

В этом случае мы оцениваем

$$SM: y_{1t} = \alpha + \gamma_2 y_{2t} + \dots + \gamma_N y_{Nt} + u_t,$$

опять получаем ряд остатков – теперь это будет ряд

$$\hat{u}_t = y_{1t} - (\hat{\alpha} + \hat{\gamma}_2 y_{2t} + \dots + \hat{\gamma}_N y_{Nt}),$$

оцениваем модель регрессии

$$\Delta \hat{u}_t = \varphi \hat{u}_{t-1} + \zeta_1 \Delta \hat{u}_{t-1} + \dots + \zeta_K \Delta \hat{u}_{t-K} + \varepsilon_t$$

с достаточным количеством запаздывающих разностей и проверяем гипотезу $H_0: \varphi = 0$ против альтернативы $H_0: \varphi < 0$.

Критические значения в этом случае отличаются от случая (2a). При большом количестве наблюдений можно использовать критические значения, приведенные в [Hamilton (1994), Table B.9, Case 2]. При небольших T критические значения вычисляются по формуле, приведенной в [MacKinnon (1991), таблица 1 (вариант “no trend”)] и воспроизведенной в [Patterson (2000)].

(3) Хотя бы один из рядов y_{2t}, \dots, y_{Nt} имеет линейный тренд, так что $E(\Delta y_{kt}) \neq 0$ хотя бы для одного из регрессоров.

(3a) В коинтеграционное соотношение включается константа.

В этом случае оценивается

$$SM: y_{1t} = \alpha + \gamma_2 y_{2t} + \dots + \gamma_N y_{Nt} + u_t.$$

Далее действуем опять как в (2b), только критические значения другие. При большом количестве наблюдений можно использовать критические значения, приведенные в [Hamilton (1994), Table B.9, Case 3]. При небольших T критические значения вычисляются по формуле, приведенной в работе [MacKinnon (1991), Table 1 (вариант “with trend”)] и воспроизведенной в [Patterson (2000)].

(3b) В коинтеграционное соотношение включается линейный тренд.

В этом случае оценивается

$$SM: y_{1t} = \alpha + \delta t + \gamma_2 y_{2t} + \dots + \gamma_N y_{Nt} + u_t.$$

Действуя так же, как и ранее, используем те же таблицы, что и в (3a), но только не для N , а для $N + 1$ переменных.

Включение тренда в коинтеграционное соотношение приводит к уменьшению мощности критерия из-за необходимости оценивания “мешающего” параметра δ . Однако такой подход вполне уместен в тех случаях, когда нет полной уверенности в том, имеется ли ненулевой тренд хотя бы у одного из рядов $y_{1t}, y_{2t}, \dots, y_{Nt}$.

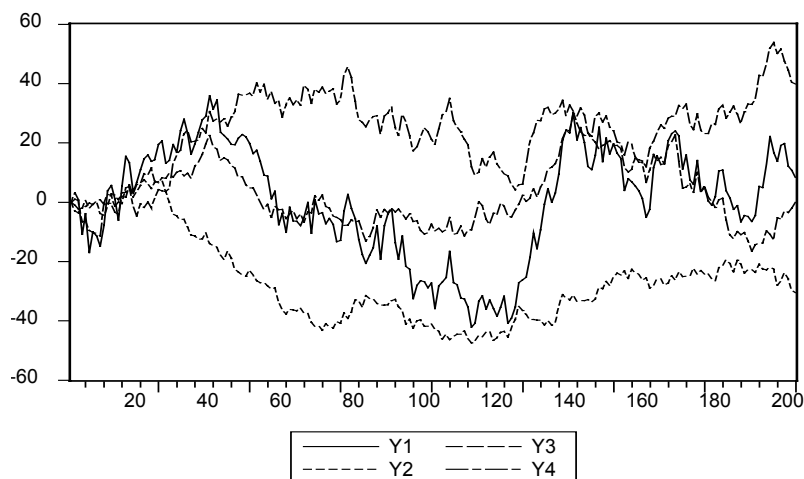
Пример

Смоделируем реализации четырех рядов $y_{1t}, y_{2t}, y_{3t}, y_{4t}$, следуя процессу порождения данных

$$\begin{aligned} DGP: y_{1t} &= y_{2,t} + y_{3,t} + y_{4,t} + \varepsilon_{1t}, \\ y_{2t} &= y_{2,t-1} + \varepsilon_{2t}, \\ y_{3t} &= y_{3,t-1} + \varepsilon_{3t}, \\ y_{4t} &= y_{4,t-1} + \varepsilon_{4t}, \end{aligned}$$

где $\varepsilon_{1t}, \varepsilon_{2t}, \varepsilon_{3t}, \varepsilon_{4t}$ – независимые друг от друга процессы гауссовского белого шума с дисперсиями, равными 1 для $\varepsilon_{2t}, \varepsilon_{3t}, \varepsilon_{4t}$ и 2 для ε_{1t} .

Графики полученных реализаций для $T = 200$ приведены ниже.



Не зная точно процесс порождения данных, мы должны были бы начать с исследования отдельных рядов. У всех четырех рядов не обнаруживается детерминированного тренда. Проверка по критерию Дики – Фуллера дает значения t -статистик, равные $-2.18, -1.78, -0.57, -1.70$, соответственно. Все 4 ряда признаются интегрированными. Продифференцированные ряды идентифицируются как гауссовские белые шумы, так что ряды $y_{1t}, y_{2t}, y_{3t}, y_{4t}$ идентифицируются как AR(1) ряды с единичным корнем, т.е. как интегрированные ряды порядка 1.

Теперь можно приступить к проверке этих четырех рядов на коинтегрированность.

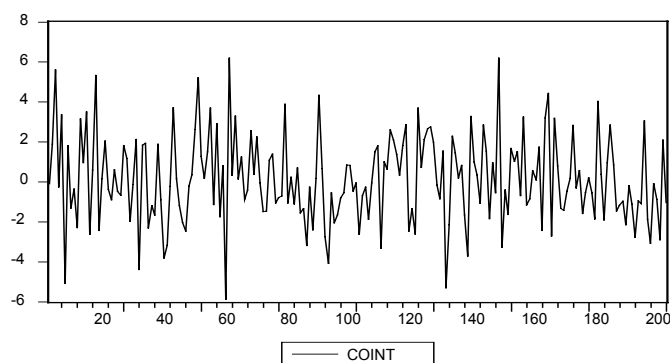
(1) Если “экономическая теория” предполагает теоретическое долговременное соотношение между рассматриваемыми рядами в форме

$$y_{1t} = y_{2,t} + y_{3,t} + y_{4,t},$$

то мы просто проверяем на интегрированность ряд

$$y_{1t} - y_{2,t} - y_{3,t} - y_{4,t}.$$

График этого ряда



вполне похож на график стационарного ряда, что подтверждается проверкой по критерию Дики – Фуллера: вычисленное значение t -статистики критерия равно -15.07 . Гипотеза неинтегрированности рядов отвергается.

Представим теперь, что теория не предлагает нам готового коинтегрирующего вектора.

(2a) Оценивание статистической модели без включения в нее константы дает:

Dependent Variable: Y1

Method: Least Squares

Variable	Coefficient	Std. Error	t-Statistic	Prob.
----------	-------------	------------	-------------	-------

Y2	0.996084	0.009973	99.88161	0.0000
Y3	0.992550	0.009578	103.6296	0.0000
Y4	1.002305	0.012393	80.87922	0.0000

При оценивании тестового уравнения Дики – Фуллера для ряда остатков получаем

Augmented Dickey-Fuller Test Equation

Dependent Variable: D(RESID_2A)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RESID_2A(-1)	-1.075552	0.070892	-15.17178	0.0000

Вычисленное значение t -статистики критерия равно -15.17 , что намного ниже 5% критического значения -3.74 ([Hamilton (1994), Table B.9, Case 1]). Гипотеза неинтегрированности отвергается.

(2b) Оценивание статистической модели с включением константы:

Dependent Variable: Y1

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.332183	0.373542	0.889279	0.3749
Y2	1.002583	0.012369	81.05843	0.0000
Y3	0.987369	0.011215	88.04048	0.0000
Y4	0.999022	0.012937	77.22129	0.0000

При оценивании тестового уравнения Дики – Фуллера для ряда остатков получаем

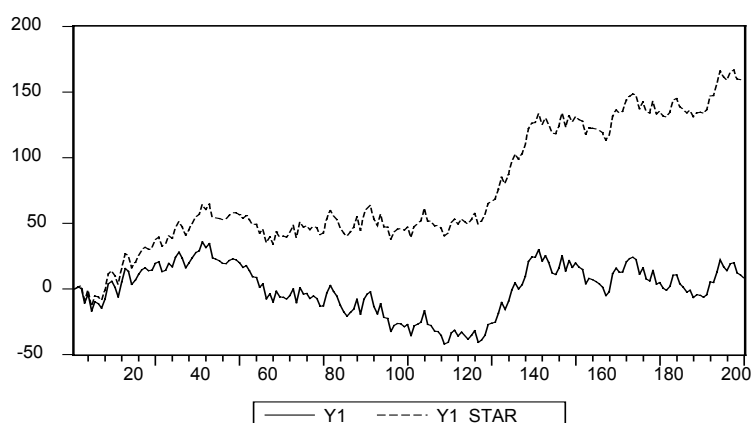
Augmented Dickey-Fuller Test Equation

Dependent Variable: D(RESID_2B)

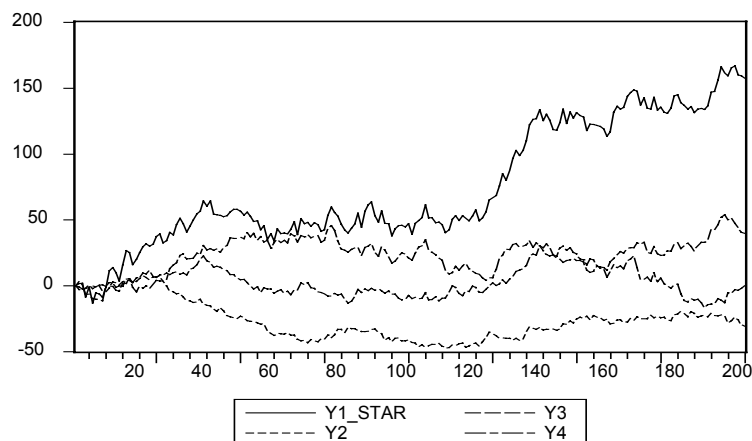
Variable	Coefficient	Std. Error	t-Statistic	Prob.
RESID_2B(-1)	-1.079049	0.070861	-15.22764	0.0000

Вычисленное значение t -статистики -15.23 опять намного ниже 5% критического значения, которое здесь равно -4.11 ([Hamilton (1994), Table B.9, Case 2]). Гипотеза неинтегрированности отвергается.

(3) Модифицируем теперь ряд y_{1t} , переходя к ряду $y^*_{1t} = y_{1t} + 0.75t$, график которого в сравнении с графиком ряда y_{1t} имеет следующий вид:



Картина изменения всех 4 рядов принимает вид

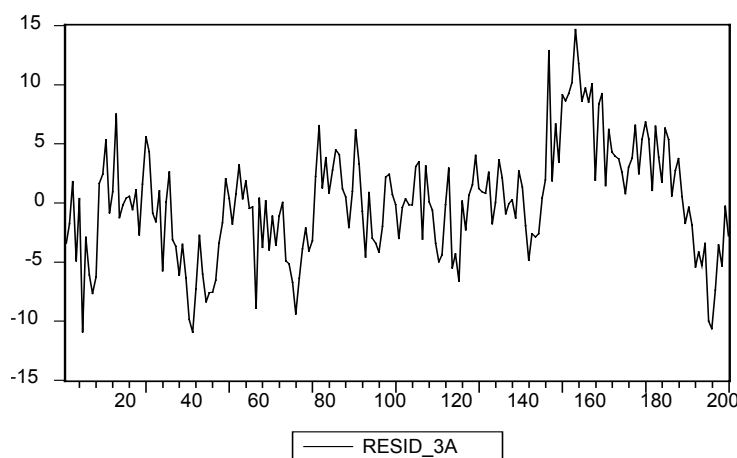


(3a) Оцениваем статистическую модель с константой в правой части:

Dependent Variable: Y1_STAR

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	11.49053	2.704802	4.248195	0.0000
Y2	-1.333762	0.089561	-14.89224	0.0000
Y3	2.856952	0.081207	35.18115	0.0000
Y4	0.072630	0.093677	0.775323	0.4391

В этом случае график остатков имеет несколько отличный вид:



Проверка по Дики – Фуллеру дает следующие результаты:

При оценивании тестового уравнения Дики – Фуллера для ряда остатков получаем Augmented Dickey-Fuller Test Equation

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RESID_3A(-1)	-0.119805	0.033630	-3.562431	0.0005

Вычисленное значение t -статистики -3.56 выше 5% критического значения, которое здесь равно -4.16 ([Hamilton (1994), Table B.9, Case 3]). Гипотеза неинтегрированности не отвергается.

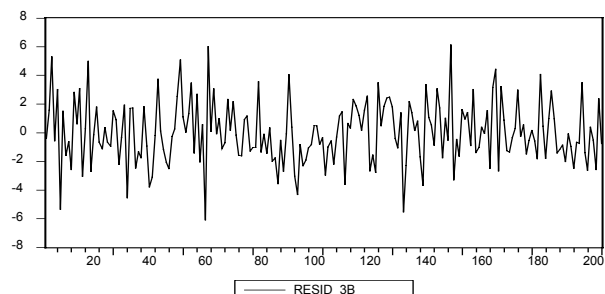
(3b) Включаем в правую часть тренд:

Dependent Variable: Y1_STAR

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.304068	0.390739	0.778187	0.4374

@TREND	0.751890	0.007507	100.1621	0.0000
Y2	1.008470	0.026468	38.10166	0.0000
Y3	0.982658	0.021830	45.01453	0.0000
Y4	1.001356	0.015942	62.81247	0.0000

График остатков:



Последний график похож на график стационарного ряда, что подтверждается проверкой по Дики – Фуллеру:

Augmented Dickey-Fuller Test Equation

Dependent Variable: D(RESID_3B)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RESID_3B(-1)	-1.079492	0.070859	-15.23448	0.0000

Вычисленное значение t -статистики -15.234 намного ниже 5% критического значения, которое здесь равно -4.49 ([Hamilton (1994), Table B.9, Case 3]). Гипотеза неинтегрированности отвергается.

Последние два результата весьма важны для уточнения того, что понимается под коинтеграцией в настоящее время.

Фактически, мы обнаружили следующее. Ряды y_{1t} , y_{2t} , y_{3t} , y_{4t} коинтегрированы в том смысле, который был определен выше (коинтегрированы в узком смысле). Именно в таком виде ввели в обиход понятие коинтеграции Энгл и Гренджер. Ряды y_{1t}^* , y_{2t}^* , y_{3t}^* , y_{4t}^* не являются коинтегрированными в узком смысле. В то же время, включение в правую часть статистической модели трендовой составляющей приводит к стационарным остаткам.

Вспомним в связи с этим, что при включении тренда в правую часть линейного регрессионного уравнения коэффициенты при объясняющих переменных интерпретируются как коэффициенты линейной связи между переменными, очищенными от детерминированного тренда. Последние же действительно были коинтегрированы по построению.

Наблюдаемая ситуация известна теперь под названием **“стохастическая коинтеграция”**. Оно указывает на наличие коинтеграционной связи между стохастическими трендами, входящими в состав рассматриваемых рядов, и не требует согласованности детерминированных трендовых составляющих (если таковые имеются). В этом случае коинтегрирующий вектор аннулирует стохастический тренд, но не обязан одновременно аннулировать и детерминированный тренд. Другими словами, существует линейная комбинация рассматриваемых рядов, которая образует ряд, стационарный относительно детерминированного тренда, но не обязательно стационарный.

В противоположность стохастической коинтеграции, при наличии коинтеграции в узком смысле коинтегрирующий вектор аннулирует и стохастический и

детерминированный тренды, т.е. существует линейная комбинация рассматриваемых рядов, образующая стационарный ряд. В связи с этим, о коинтеграции в узком смысле говорят также как о **детерминистской коинтеграции**.

Замечание

Здесь самое время сделать одно замечание. Мы говорили в разд. 7.1 о том, что при отсутствии коинтеграции между двумя интегрированными рядами непосредственное оценивание модели $y_t = \alpha + \beta x_t + u_t$ бессмысленно, т.к. получаемая оценка $\hat{\beta}_T$ не является оценкой какого-либо теоретического параметра связи между переменными x_t и y_t . Если оба ряда имеют помимо стохастического еще и детерминированный тренд, то оценка $\hat{\beta}_T$ все же не является случайной величиной, а сходится к некоторой постоянной. Соответствующее исследование, проведенное в работе [Entorf (1992)], показало следующее.

Пусть

$$\text{DGP: } x_t = \mu_x + x_{t-1} + \varepsilon_{1t},$$

$$y_t = \mu_y + y_{t-1} + \varepsilon_{2t},$$

где ε_{1t} и ε_{2t} – некоррелированные между собой процессы белого шума, причем $\mu_x, \mu_y \neq 0$. Тогда при оценивании статистической модели

$$\text{SM: } y_t = \alpha + \beta x_t + u_t$$

оценка $\hat{\alpha}_T$ для α , вычисляемая по T наблюдениям, при $T \rightarrow \infty$ расходится, а оценка $\hat{\beta}_T$ для β сходится по вероятности при $T \rightarrow \infty$ к отношению μ_y / μ_x .

Если при тех же условиях оценивать статистическую модель

$$\text{SM: } y_t = \alpha + \beta x_t + \gamma t + u_t,$$

то тогда (при $T \rightarrow \infty$) $\hat{\gamma}_T$ сходится по вероятности к μ_y , а $\hat{\beta}_T$ сходится по распределению к некоторой случайной величине, как и в случае ложной регрессии для случайных блужданий без сносов.

7.4. Оценивание коинтегрированных систем временных рядов

Пусть мы имеем N временных рядов y_{1t}, \dots, y_{Nt} , каждый из которых является интегрированным порядка 1. Если существует такой вектор $\beta = (\beta_1, \dots, \beta_N)^T$, отличный от нулевого, для которого

$$\beta_1 y_{1t} + \dots + \beta_N y_{Nt} \sim I(0) - \text{стационарный ряд,}$$

то ряды **коинтегрированы (в узком смысле)**; такой вектор β называется **коинтегрирующим** вектором. Если при этом

$$c = E(\beta_1 y_{1t} + \dots + \beta_N y_{Nt}),$$

то тогда можно говорить о **долговременном положении равновесия системы** в виде

$$\beta_1 y_{1t} + \dots + \beta_N y_{Nt} = c.$$

В каждый конкретный момент времени t существует некоторое **отклонение** системы от этого положения равновесия, характеризующееся величиной

$$z_t = \beta_1 y_{1t} + \dots + \beta_N y_{Nt} - c.$$

Ряд z_t , в силу сделанных предположений, является стационарным рядом, имеющим нулевое математическое ожидание, так что он достаточно часто пересекает нулевой уровень, т.е. система колеблется вокруг указанного выше положения равновесия.

Положение, однако, осложняется тем, что у коинтегрированной системы $I(1)$ рядов может быть несколько линейно независимых коинтегрирующих векторов. Если максимальное количество линейно независимых коинтегрирующих векторов для заданных рядов y_{1t}, \dots, y_{Nt} равно r , то это число r называется **рангом коинтеграции**. Для коинтегрированной системы, состоящей из N рядов, ранг коинтеграции может принимать значения $r = 1, \dots, N-1$. (Формально, если ряды не коинтегрированы, то $r = 0$. Если же имеется $r = N$ линейно независимых коинтегрирующих векторов, то все N рядов стационарны.) Совокупность всех возможных коинтегрирующих векторов для коинтегрированной системы $I(1)$ рядов образует r -мерное линейное векторное пространство, которое называют **коинтеграционным пространством**. Любой набор r линейно независимых коинтегрирующих векторов образует **базис** этого пространства, и если зафиксировать этот набор в качестве базиса, то тогда любой коинтегрирующий вектор является линейной комбинацией векторов, составляющих базис.

Пусть коинтегрированная система $I(1)$ рядов y_{1t}, \dots, y_{Nt} имеет ранг коинтеграции r и может быть представлена в форме $VAR(p)$ – **векторной авторегрессии** порядка p (VAR – *vector autoregression*)

$$A(L)y_t = \mu + \varepsilon_t,$$

где

$$y_t = (y_{1t}, \dots, y_{Nt})^T,$$

$$\mu = (\mu_1, \dots, \mu_N)^T,$$

$$A(L) = A_0 - A_1L - \dots - A_pL^p,$$

A_0, A_1, \dots, A_p – матрицы размера $(N \times N)$,

$A_0 = I_N$ (единичная матрица),

т.е.

$$y_t = \mu + A_1 y_{t-1} + \dots + A_p y_{t-p} + \varepsilon_t.$$

Тогда ранг матрицы $A(1)$ равен $\text{rank } A(1) = r$ и (по аналогии со случаем $N = 2$) существует представление этой VAR в форме ECM (модели коррекции ошибок)

$$\begin{aligned} \Delta y_{1t} = & \mu_1 + \alpha_{11}z_{1,t-1} + \dots + \alpha_{1r}z_{r,t-1} + \\ & + \sum_{j=1}^{p-1} (\gamma_{11,j}\Delta y_{1,t-j} + \dots + \gamma_{1N,j}\Delta y_{N,t-j}) + \varepsilon_{1t}, \\ & \dots \dots \dots \end{aligned}$$

$$\begin{aligned} \Delta y_{Nt} = & \mu_N + \alpha_{N1}z_{1,t-1} + \dots + \alpha_{Nr}z_{r,t-1} + \\ & + \sum_{j=1}^{p-1} (\gamma_{N1,j}\Delta y_{1,t-j} + \dots + \gamma_{NN,j}\Delta y_{N,t-j}) + \varepsilon_{Nt}, \end{aligned}$$

где

$z_{1,t}, \dots, z_{r,t}$ – стационарные $I(0)$ ряды, соответствующие r линейно независимым коинтегрирующим векторам $\beta_{(1)}, \dots, \beta_{(r)}$,

$(\alpha_{11}, \dots, \alpha_{N1})^T, \dots, (\alpha_{1r}, \dots, \alpha_{Nr})^T$ – линейно независимые векторы корректирующих коэффициентов.

Такую модель коррекции ошибок можно записать в компактном виде как

$$\Delta y_t = \mu + \alpha \beta^T y_{t-1} + \zeta_1 \Delta y_{t-1} \dots + \zeta_{p-1} \Delta y_{t-p+1} + \varepsilon_t,$$

где $\zeta_1, \dots, \zeta_{p-1}$ – матрицы размера $N \times N$, а α и β – $(N \times r)$ -матрицы полного ранга r . При этом столбцы $\beta_{(1)}, \dots, \beta_{(r)}$ матрицы β являются линейно независимыми коинтегрирующими векторами, а элементы α_{ij} матрицы α являются коэффициентами при стационарных линейных комбинациях

$$z_{1,t-1} = \beta_{(1)}^T y_{t-1}, \dots, z_{r,t-1} = \beta_{(r)}^T y_{t-1}$$

(представляющих отклонения в момент $t-1$ от r долговременных соотношений между рядами y_{1t}, \dots, y_{Nt}) в правых частях уравнений для $\Delta y_{1t}, \dots, \Delta y_{Nt}$.

Представление коинтегрированной VAR в форме модели коррекции ошибок не единственно, поскольку в качестве набора $\beta_{(1)}, \dots, \beta_{(r)}$ можно взять любой базис коинтеграционного пространства. Соответственно, неоднозначность имеется и в отношении матрицы α . Один из возможных вариантов выбора базисных коинтегрирующих векторов дается следующим представлением системы коинтегрированных I(1) рядов.

Если ранг коинтеграции равен r , $0 < r < N$, то при соответствующей перенумерации переменных система I(1) рядов y_{1t}, \dots, y_{Nt} допускает представление

$$\begin{pmatrix} y_{1t} \\ \vdots \\ y_{rt} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_r \end{pmatrix} + C \begin{pmatrix} y_{r+1,t} \\ \vdots \\ y_{N,t} \end{pmatrix} + \begin{pmatrix} v_{1t} \\ \vdots \\ v_{rt} \end{pmatrix}, \quad \begin{pmatrix} \Delta y_{r+1,t} \\ \vdots \\ \Delta y_{N,t} \end{pmatrix} = \begin{pmatrix} \delta_{r+1} \\ \vdots \\ \delta_N \end{pmatrix} + \begin{pmatrix} v_{r+1,t} \\ \vdots \\ v_{N,t} \end{pmatrix},$$

(**треугольная система Филлипса**), где

$C = (c_{ij})$ – матрица размера $r \times (N-r)$,

$v_t = (v_{1t}, \dots, v_{Nt})^T$ – стационарный (в широком смысле) векторный случайный процесс с $E(v_t) = 0$,

ряды $y_{r+1,t}, \dots, y_{N,t}$ не коинтегрированы.

Отсюда получаем:

$$y_{1t} - c_{11} y_{r+1,t} - \dots - c_{1,N-r} y_{N,t} = \mu_1 + v_{1t},$$

...

$$y_{rt} - c_{r1} y_{r+1,t} - \dots - c_{r,N-r} y_{N,t} = \mu_r + v_{rt},$$

так что векторы

$$\beta_{(1)} = (1, 0, 0, \dots, 0, -c_{11}, \dots, -c_{1,N-r})^T,$$

$$\beta_{(2)} = (0, 1, 0, \dots, 0, -c_{21}, \dots, -c_{2,N-r})^T,$$

...

$$\beta_{(r)} = (0, 0, 0, \dots, 1, -c_{r1}, \dots, -c_{r,N-r})^T$$

являются линейно независимыми коинтегрирующими векторами. Им соответствуют r стационарных линейных комбинаций рядов $y_{1t}, \dots, y_{N,t}$

$$z_{1,t} = \beta_{(1)}^T y_t = y_{1t} - c_{11} y_{r+1,t} - \dots - c_{1,N-r} y_{N,t},$$

...

$$z_{r,t} = \beta_{(r)}^T y_t = y_{rt} - c_{r1} y_{r+1,t} - \dots - c_{r,N-r} y_{N,t}.$$

Если ряды v_{1t}, \dots, v_{rt} не коррелированы с рядами $v_{r+1,t}, \dots, v_{Nt}$, то переменные $y_{r+1,t}, \dots, y_{N,t}$ являются экзогенными в первой подсистеме, и ее можно оценивать методом наименьших квадратов как многомерную регрессию y_{1t}, \dots, y_{rt} на $y_{r+1,t}, \dots, y_{N,t}$. Полученные оценки \hat{c}_{ij} элементов матрицы C суперсостоятельны, хотя

распределение $T(\hat{c}_{ij} - c_{ij})$ не стремится к нормальному при $T \rightarrow \infty$. В случае $r = 1$ имеем

$$y_{1,t} = \mu_1 + c_{11} y_{2,t} - \dots - c_{1,N-1} y_{N,t} + v_{1,t},$$

условное распределение оценок наименьших квадратов для коэффициентов (при фиксированных значениях $y_{2,t}, \dots, y_{N,t}$) является асимптотически нормальным, и это обеспечивает возможность использования стандартных процедур, основанных на t - и F -статистиках (конечно, в асимптотическом плане), с коррекцией стандартных ошибок коэффициентов в случае, если ряд $v_{1,t}$ не является белым шумом. Коррекция, как и в разд. 7.1, состоит в замене стандартной оценки дисперсии ряда $v_{1,t}$ оценкой долговременной дисперсии этого ряда.

Значения \hat{c}_{ij} можно использовать для построения r линейно независимых $(N \times 1)$ -векторов – оценок коинтегрирующих векторов $\beta_{(1)}, \dots, \beta_{(r)}$:

$$\hat{\beta}_{(1)} = (1, 0, 0, \dots, 0, -\hat{c}_{11}, \dots, -\hat{c}_{1,N-r})^T,$$

...

$$\hat{\beta}_{(r)} = (0, 0, 0, \dots, 1, -\hat{c}_{r1}, \dots, -\hat{c}_{r,N-r})^T.$$

Используя построенные оценки коинтегрирующих векторов $\hat{\beta}_{(1)}, \dots, \hat{\beta}_{(r)}$, получаем оценки искомым стационарных линейных комбинаций в виде

$$\hat{z}_{1,t-1} = \hat{\beta}_{(1)}^T y_{t-1}, \dots, \hat{z}_{r,t-1} = \hat{\beta}_{(r)}^T y_{t-1}.$$

Теперь можно вместо указанной выше “истинной” ЕСМ оценить систему

$$\Delta y_t = \mu + \alpha \hat{z}_{t-1} + \zeta_1 \Delta y_{t-1} \dots + \zeta_{p-1} \Delta y_{t-p+1} + \varepsilon_t,$$

в которой

$$\hat{z}_{t-1} = \begin{pmatrix} \hat{z}_{1,t-1} \\ \vdots \\ \hat{z}_{r,t-1} \end{pmatrix}.$$

При этом оценки наименьших квадратов для коэффициентов последней модели имеют те же самые асимптотические распределения, что и при оценивании истинной ЕСМ.

Заметим, что если мы имеем дело со стохастической (а не с детерминистской) коинтеграцией, то для достижения стационарности рядов $z_{1,t}, \dots, z_{r,t}$ приходится в “остационаривающую” линейную комбинацию рядов $y_{1,t}, \dots, y_{N,t}$ включать еще и дополнительную трендовую составляющую, так что в этом случае речь идет о существовании стационарных линейных комбинаций $(N+1)$ переменных $y_{1,t}, \dots, y_{N,t}$ и t , в которых не все коэффициенты равны нулю.

Если ранг матрицы $A(1)$ равен r , то тогда существует r таких стационарных линейных комбинаций

$$\beta_{11} y_{1t} + \dots + \beta_{1N} y_{1N} + \beta_{1, N+1} t ,$$

.....

$$\beta_{r1} y_{1t} + \dots + \beta_{rN} y_{1N} + \beta_{r, N+1} t .$$

с линейно независимыми $((N+1) \times 1)$ -векторами

$$\beta_{(1)} = (\beta_{11} , \dots , \beta_{1N} , \beta_{1, N+1})^T$$

.....

$$\beta_{(r)} = (\beta_{r1} , \dots , \beta_{rN} , \beta_{r, N+1})^T .$$

При этом последние векторы интерпретируются как линейно независимые коинтегрирующие векторы в системе стохастически коинтегрированных рядов.

Возможность наличия нескольких линейно независимых коинтегрирующих векторов значительно усложняет задачу построения модели коррекции ошибок (ECM), поскольку, как минимум, приходится по реальным статистическим данным оценивать количество таких векторов. Само по себе решение о коинтегрированности нескольких $I(1)$ рядов в результате использования рассмотренных выше процедур Дики – Фуллера отнюдь не дает нам никакой информации о ранге коинтеграции r ; для этого требуются другие статистические процедуры.

Но если мы не знаем ранга коинтеграции, то теряется смысл оценивания уравнения регрессии в уровнях

$$y_{1t} = c + \gamma_2 y_{2t} + \dots + \gamma_N y_{Nt} + u_t$$

(или $y_{1t} = c + \gamma_2 y_{2t} + \dots + \gamma_N y_{Nt} + \gamma_{N+1} t + u_t$). Действительно, если $r > 1$, то вектор $(1, -\hat{\gamma}_2, \dots, -\hat{\gamma}_N)^T$ (или вектор $(1, -\hat{\gamma}_2, \dots, -\hat{\gamma}_N, -\hat{\gamma}_{N+1})^T$) является оценкой всего

лишь одной из возможных линейных комбинаций r линейно независимых коинтегрирующих векторов, которая может и не иметь разумной экономической интерпретации.

Но даже если ранг коинтеграции r по каким-то причинам известен, при $r > 1$ возникает другая проблема. В рассмотренном выше представлении Филлипса линейно независимые коинтегрирующие векторы имели вид

$$\begin{aligned}\beta_{(1)} &= (1, 0, 0, \dots, 0, -c_{11}, \dots, -c_{1, N-r})^T, \\ \beta_{(2)} &= (0, 1, 0, \dots, 0, -c_{21}, \dots, -c_{2, N-r})^T, \\ &\dots \\ \beta_{(r)} &= (0, 0, 0, \dots, 1, -c_{r1}, \dots, -c_{r, N-r})^T.\end{aligned}$$

Любая линейная комбинация этих векторов (не все коэффициенты которой равны нулю) также является коинтегрирующим вектором, а совокупность всех возможных линейных комбинаций этих векторов образует линейное векторное пространство размерности r . Любой вектор из этого пространства (не все коэффициенты которого равны нулю) является коинтегрирующим вектором для y_{1t}, \dots, y_{Nt} , а векторы $\beta_{(1)}, \dots, \beta_{(r)}$ образуют всего лишь один из возможных базисов этого пространства. В практических задачах на первый план (наряду с определением ранга коинтеграции) выходит **идентификация коинтегрирующих векторов**, приводящих к долговременным соотношениям, имеющим разумную экономическую интерпретацию. Мы вернемся к этому вопросу в главе 8.

В настоящее время наиболее распространенной является методика определения ранга коинтеграции, предложенная Йохансенom в работе [Johansen (1988)]. Однако точное описание этой процедуры требует более детального рассмотрения соответствующего математического аппарата. Мы рассмотрим эту процедуру в разд. 8.1, а сейчас сосредоточимся на случае, когда $r = 1$, т.е. (с точностью до пропорциональности) имеется всего один коинтегрирующий вектор.

Наиболее простой является ситуация, когда $N = 2$. В этом случае, если рассматриваемые ряды y_{1t} и y_{2t} коинтегрированы, то ранг коинтеграции может быть равным только единице. Мы уже отмечали ранее, что при построении модели коррекции ошибок на первом шаге процедуры Энгла – Гренджера, вообще говоря, нельзя пользоваться обычными регрессионными критериями (даже в асимптотическом плане). И причина этого в том, что получаемые на первом шаге оценки и статистики в общем случае имеют нестандартные асимптотические распределения.

Об одном исключении из общего случая мы уже говорили – это **треугольная система Филлипса**

$$y_t = \beta x_t + v_t,$$

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

где ε_t и v_t – не коррелированные между собой процессы белого шума.

Вторым исключением является ситуация, исследованная в работе [West (1988)]:

$$y_t = \alpha + \beta x_t + u_t,$$

где

$x_t \sim I(1)$, $E(\Delta x_t) = \mu \neq 0$ (так что ряд x_t содержит и стохастический и детерминированный тренд),

$u_t \sim I(0)$ – стационарный ряд с нулевым средним, не обязательно являющийся процессом белого шума.

В цитированной работе доказывается асимптотическая нормальность соответствующим образом нормированной оценки наименьших квадратов для вектора $(\alpha, \beta)^T$. Если ряд u_t не является процессом белого шума, то для применения этого результата необходимо скорректировать значения t -статистик, вычисляемых по стандартным формулам, соответствующим предположениям классической линейной модели регрессии.

В знаменателях обычных t -статистик для параметров α и β стоят оценки стандартных ошибок оценок $\hat{\alpha}_T$ и $\hat{\beta}_T$ этих параметров, а именно:

$$S \sqrt{(X^T X)^{-1}_{11}} \text{ – для } \hat{\alpha}_T, \quad S \sqrt{(X^T X)^{-1}_{22}} \text{ – для } \hat{\beta}_T.$$

Здесь X – $(T \times 2)$ -матрица значений объясняющих переменных (1 и x_t) в T наблюдениях, а S^2 – несмещенная оценка дисперсии u_t в случае, когда $u_t \sim i.i.d.$,

$$S^2 = \frac{1}{n-2} \sum_{t=1}^T \hat{u}_t^2, \quad \hat{u}_t = y_t - \hat{\alpha} - \hat{\beta} x_t.$$

Поскольку у нас не предполагается, что $u_t \sim i.i.d.$, то для сохранения t -распределения (точнее, $N(0, 1)$ – распределения) t -статистик (хотя бы при больших T), требуется замена S^2 на другую подходящую величину.

Мы предположили, что ряд u_t стационарный. Пусть $\gamma_h = \text{Cov}(u_t, u_{t+h})$ –

последовательность его автоковариаций. West показал, что подходящей является

замена S^2 **долговременной дисперсией** ряда u_t (см. разд. 6.8.1), которая для стационарного ряда вычисляется по формуле

$$\lambda^2 = \sum_{h=-\infty}^{\infty} \gamma_h.$$

Проблема, однако в том, что значение λ^2 не известно, и его приходится оценивать по имеющимся наблюдениям. Для этого, в свою очередь, следовало бы оценить

бесконечное множество автоковариаций γ_h , $h = 0, \pm 1, \pm 2, \dots$, что, конечно, невозможно. Из-за этого, в конечном счете, приходится, так или иначе, делать более определенные предположения о характере автокоррелированности ряда u_t , что дало бы возможность ограничиться при оценивании λ^2 оценкой лишь конечного числа автоковариаций $\gamma_h = \text{Cov}(u_t, u_{t+h})$. В процессе такого оценивания приходится учитывать и то, что автоковариации γ_h с возрастанием h оцениваются все менее точно, и поэтому желательно регулировать (уменьшать) влияние γ_h на оценку долговременной дисперсии λ^2 при возрастании h .

Если исходить из того, что случайный процесс u_t может быть представлен в виде процесса $\text{MA}(q)$ конечного порядка q , то тогда $\gamma_h = 0$ для $|h| > q$, и можно не заниматься получением оценок γ_h для таких h . Это, вместе с предшествующими соображениями, приводит к оценке

$$\hat{\lambda}^2 = \hat{\gamma}_0 + 2 \sum_{h=1}^q \left(1 - \frac{h}{q+1}\right) \hat{\gamma}_h,$$

где $\hat{\gamma}_h = \frac{1}{T} \sum_{t=h+1}^T \hat{u}_t \hat{u}_{t-h}$ – оценки автоковариаций γ_h . При этом можно показать (см., например, [Hamilton (1994), p. 513]), что выбор $q = O(T^{1/5})$ обеспечивает состоятельность такой оценки для λ^2 (*оценка Newey–West*).

В рамках пакета EVIEWS реализация такого метода производится без труда. Следует просто при спецификации уравнения заказать опцию: “вычисление стандартных ошибок методом Newey–West”. (Отметим, однако, что в этой опции используется несколько отличающаяся от приведенной оценка, реализующая еще и поправку на возможную гетероскедастичность ряда.)

Если предположить, что динамика ряда u_t хорошо аппроксимируется моделью авторегрессии $AR(p)$ с конечным p ,

$$u_t = a_1 u_{t-1} + a_2 u_{t-2} + \dots + a_p u_{t-p} + \varepsilon_t,$$

где ε_t – инновационный процесс белого шума с $D(\varepsilon_t) = \sigma_\varepsilon^2$, то тогда

$$\lambda^2 = \sigma_\varepsilon^2 / (1 - a_1 - a_2 - \dots - a_p)^2.$$

Поэтому в такой ситуации в качестве оценки для λ^2 естественно взять величину

$$\hat{\lambda}^2 = \frac{\hat{\sigma}_\varepsilon^2}{(1 - \hat{a}_1 - \hat{a}_2 - \dots - \hat{a}_p)^2},$$

где

$\hat{a}_1, \hat{a}_2, \dots, \hat{a}_p$ – оценки наименьших квадратов для a_1, a_2, \dots, a_p ,

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{T-p} \sum_{t=p+1}^T \hat{\varepsilon}_t^2,$$

$\hat{\varepsilon}_t$ – остатки при оценивании модели авторегрессии для ряда \hat{u}_t .

В любом случае, замена S^2 на $\hat{\lambda}^2$ равносильна умножению значения t -статистики, полученного обычным путем, на $(S/\hat{\lambda})$.

Пример

Смоделируем систему

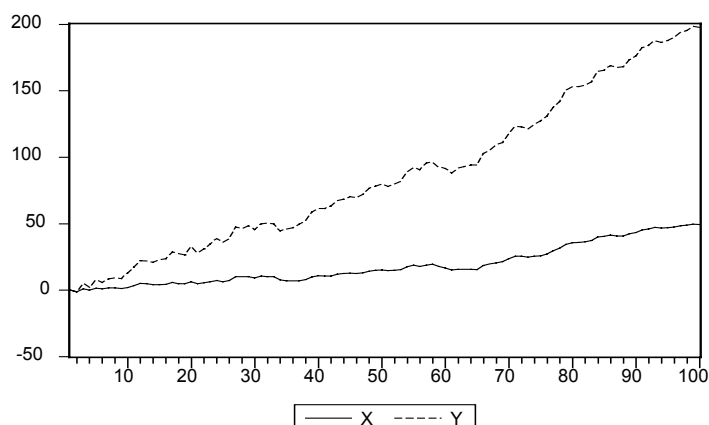
$$\text{DGP: } y_t = 2x_t + u_t,$$

$$x_t = 1 + x_{t-1} + v_t,$$

где

$u_t = 0.4 u_{t-1} + 0.2 u_{t-2} + \varepsilon_t$ – стационарный $AR(2)$ ряд,

ε_t, v_t – гауссовские процессы белого шума, коррелированные в совпадающие моменты времени: $Cov(\varepsilon_t, v_t) = 0.8$:

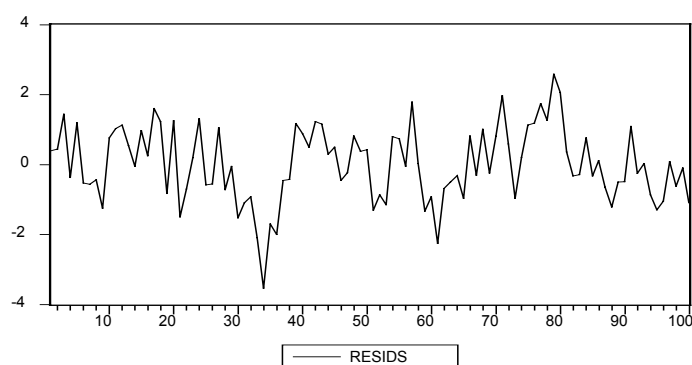


Оценивание статистической модели SM: $y_t = \alpha + \beta x_t + u_t$ обычным методом наименьших квадратов дает следующие результаты:

Dependent Variable: Y

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.398071	0.172093	-2.313111	0.0228
X	2.031938	0.007241	280.6336	0.0000
R-squared	0.998757	Mean dependent var	37.39809	
Adjusted R-squared	0.998745	S.D. dependent var	30.23477	
S.E. of regression	1.071308	Akaike info criterion	2.995436	
Sum squared resid	112.4748	Schwarz criterion	3.047539	
Log likelihood	-147.7718	F-statistic	78755.23	
Durbin-Watson stat	1.080957	Prob(F-statistic)	0.000000	

Если ориентироваться на приведенные значения статистик, то оба параметра оказываются статистически значимыми, хотя в DGP константа в уравнении для y_t отсутствует. Ряд остатков



идентифицируется по коррелограмме как AR(2). Оцененная AR(2) модель:

Dependent Variable: RESIDS

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RESIDS(-1)	0.363522	0.100494	3.617344	0.0005
RESIDS(-2)	0.205074	0.100427	2.042024	0.0439
R-squared	0.240364	Mean dependent var	-0.008547	
Adjusted R-squared	0.232451	S.D. dependent var	1.075097	
S.E. of regression	0.941891	Akaike info criterion	2.738343	
Sum squared resid	85.16727	Schwarz criterion	2.791098	

Отсюда находим оценку для λ :

$$\hat{\lambda} = 0.941891 / (1 - 0.363522 - 0.205074) = 2.183,$$

так что

$$(S / \hat{\lambda}) = 1.071 / 2.183 = 0.491.$$

Это приводит к следующим скорректированным значениям t -статистик и P -значений:

$$t_{\alpha} : -2.313111 \text{ (P-value} = 0.0228) \rightarrow -1.135738 \text{ (P-value} = 0.2588)$$

$$t_{\beta} : 280.6336 \text{ (P-value} = 0.0000) \rightarrow 137.791098 \text{ (P-value} = 0.0000).$$

При использовании скорректированных значений постоянная в оцениваемом уравнении становится статистически незначимой.

Снимем теперь ограничение $N = 2$ и будем интересоваться существующей и **единственной** (по предположению) долговременной связью между N нестационарными $I(1)$ рядами y_{1t}, \dots, y_{Nt} .

Оценивание статистической модели

$$y_{1t} = c + \gamma_2 y_{2t} + \dots + \gamma_N y_{Nt} + u_t$$

приводит в этом случае к суперсостоятельным оценкам, независимо от того, будут ли регрессоры иметь линейный тренд, если только в правую часть уравнения не включается тренд. Однако, как мы уже отмечали выше в разд. 7.2, повышенная скорость сходимости по вероятности оценок коэффициентов к истинным значениям этих коэффициентов вовсе не предотвращает смещения оценок при небольшой длине ряда наблюдений. Многие авторы на основании результатов моделирования отмечали весьма значительное смещение оценок коэффициентов при небольших T .

Как и в случае $N = 2$, особое место в этом отношении занимает **треугольная система Филлипса**

$$y_{1t} = c + \gamma_2 y_{2t} + \dots + \gamma_N y_{Nt} + \varepsilon_{1t}$$

$$y_{2t} = y_{2,t-1} + \varepsilon_{2t},$$

...

$$y_{Nt} = y_{N,t-1} + \varepsilon_{Nt},$$

где $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t}, \dots, \varepsilon_{Nt})^T$ – ***N*-мерный гауссовский белый шум** (т.е. $\varepsilon_1, \varepsilon_2, \dots$ – последовательность независимых, одинаково распределенных случайных векторов, имеющих *N*-мерное нормальное распределение с нулевым математическим ожиданием и ковариационной матрицей $\Sigma = (\sigma_{ij})$), причем случайные величины $\varepsilon_{2t}, \dots, \varepsilon_{Nt}$ могут быть коррелированными между собой, но ε_{1t} не коррелирована ни с одной из них (так что $\sigma_{1j} = 0$ для всех $j = 2, \dots, N$). В этом случае регрессоры y_{2t}, \dots, y_{Nt} не коинтегрированы, и $\beta = (1, -\hat{\gamma}_2, \dots, -\hat{\gamma}_N)^T$ – единственный коинтегрирующий вектор. Условное распределение

$$(\hat{c} - c, \hat{\gamma}_2 - \gamma_2, \dots, \hat{\gamma}_N - \gamma_N)^T \mid \{y_{2t}, \dots, y_{Nt}, t = 1, \dots, T\}$$

является *N*-мерным нормальным, с нулевым средним, так что *F*-критерии для проверки линейных гипотез о значениях коэффициентов $c, \gamma_2, \dots, \gamma_N$ имеют точные *F*-распределения, а *t*-критерии – точные *t*-распределения.

В более общей ситуации, пусть

$$y_{1t} = c + \gamma_2 y_{2t} + \dots + \gamma_N y_{Nt} + u_{1t}$$

$$y_{2t} = y_{2,t-1} + u_{2t},$$

...

$$y_{Nt} = y_{N,t-1} + u_{Nt},$$

где $u_t = (u_{1t}, u_{2t}, \dots, u_{Nt})^T$ – ***N*-мерный гауссовский стационарный векторный ряд** (теперь уже не обязательно *N*-мерный гауссовский белый шум), причем ряды u_{2t}, \dots, u_{Nt} могут быть коррелированными между собой, но ряд u_{1t} не коррелирован с остальными рядами, так что $\text{Cov}(u_{1t}, u_{ks}) = 0$ при $k \neq 1$ для всех t, s . Последнее условие обеспечивает экзогенность переменных в правой части первого уравнения треугольной системы. (Гауссовость ряда u_t означает, что совместное распределение значений ряда в любые *T* различных моментов времени является *NT*-мерным нормальным распределением.)

В такой ситуации для проверки линейных гипотез о коэффициентах можно использовать

скорректированные *F*- и *t*-статистики с асимптотически оправданными *F*- и *t*-распределениями, предварительно заменив обычную оценку S^2 для дисперсии u_{1t} на состоятельную оценку $\hat{\lambda}^2$ “долговременной дисперсии” λ^2 ряда u_{1t} . Последнее соответствует умножению обычной *F*-статистики на $S^2/\hat{\lambda}^2$ и умножению обычной *t*-статистики на $S/\hat{\lambda}$.

Таким образом, проблема нестандартных распределений по-существу связана с возможным нарушением экзогенности регрессоров y_{2t}, \dots, y_{Nt} в первом уравнении треугольной системы.

Сток и Уотсон [Stock, Watson (1993)] и Сайконнен [Saikonnen (1991)] предложили процедуру устранения нежелательной корреляции, состоящую в пополнении правой части первого уравнения треугольной системы *запаздывающими* (“lags”) и *опережающими* (“leads”) значениями приращений регрессоров. (Отсюда наименование метода – “leads” and “lags” .) Именно, вместо первого уравнения системы оценивается его расширенный вариант

$$y_{1t} = c + \gamma_2 y_{2t} + \dots + \gamma_N y_{Nt} + \\ + \sum_{j=-p}^p (\theta_{2j} \Delta y_{2,t-j} + \dots + \theta_{Nj} \Delta y_{N,t-j}) + u_t.$$

Если значение p выбрано правильно (достаточно велико), то тогда статистические выводы в отношении $\gamma_2, \dots, \gamma_N$ можно проводить обычным образом (конечно, имея в виду асимптотическую оправданность соответствующих статистических процедур), но опять с использованием скорректированных значений обычных t - и F -статистик, если u_t не является белым шумом.

Предложенная процедура остается асимптотически оправданной и в случае, когда все или некоторые из рядов y_{2t}, \dots, y_{Nt} имеют детерминированный тренд.

Более того, дополнительных проблем не возникает и в случае, когда в правую часть первого уравнения треугольной системы добавляется линейный тренд и проверяется гипотеза о его значимости. Это позволяет проводить отдельную проверку гипотез о том, что

(a) $y_{1t} - \gamma_2 y_{2t} - \dots - \gamma_N y_{Nt}$ не имеет детерминированного тренда;

(b) $y_{1t} - \gamma_2 y_{2t} - \dots - \gamma_N y_{Nt}$ — стационарный ряд.

(Заметим, что (a) может выполняться при невыполненном (b), если детерминированный тренд устраняется, а стохастический тренд остается.)

Пример

DGP: $y_t = 5 + z_t + u_t$,

$$z_t = z_{t-1} + v_t,$$

где

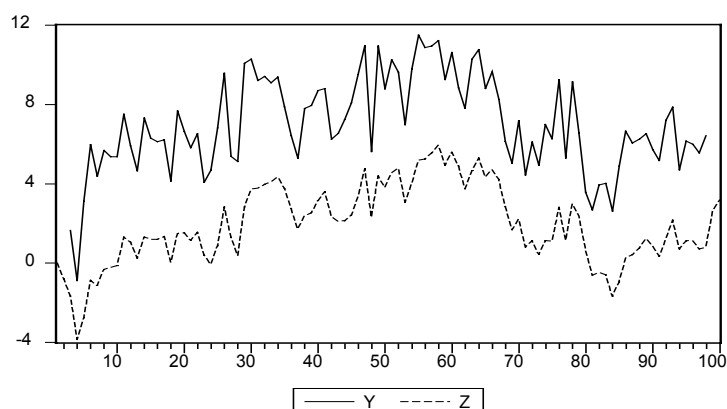
$$z_1 = 0,$$

$$u_t = v_t + 0.25 v_{t-1} + 0.25 v_{t+1} + 0.1 v_{t-2} + 0.1 v_{t+2} + 0.1 \varepsilon_t,$$

ε_t, v_t — не коррелированные между собой гауссовские процессы белого шума.

Здесь случайная величина u_t коррелирована с $v_t, v_{t-1}, v_{t+1}, v_{t-2}, v_{t+2}$, так что непосредственное использование стандартных статистических выводов неоправданно.

Обратимся к смоделированной реализации этого DGP (50 наблюдений):



Оба ряда y_t и z_t идентифицируются по 50 наблюдениям как интегрированные ряды первого порядка. Рассмотрим эту пару в рамках треугольной системы Филлипса. Оценивание методом наименьших квадратов уравнения $y_t = \alpha + \beta z_t + \eta_t$ дает следующий результат:

Dependent Variable: Y

Method: Least Squares

Sample(adjusted): 3 98

Included observations: 96 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.851262	0.133152	36.43410	0.0000
Z	1.088870	0.047570	22.88977	0.0000

Проверять гипотезу $H_0: \beta = 1$, используя обычный t -критерий, нельзя, если $Cov(\eta_t, \Delta z_s) \neq 0$ хотя бы для одной пары значений t, s . Для выяснения вопроса о наличии или отсутствии такой коррелированности обратимся к кросс-коррелограмме, построенной для пары рядов $e_t, \Delta z_t$, где e_t — ряд остатков, полученный при оценивании уравнения

$y_t = \alpha + \beta z_t + \eta_t$. Левый график показывает поведение кросс-корреляций $Cov(e_t, \Delta z_{t-i})$ для $i = 0, 1, 2, \dots$; значения этих кросс-корреляций приведены в столбце “lag”. Правый график показывает поведение кросс-корреляций $Cov(e_t, \Delta z_{t+i})$ для $i = 0, 1, 2, \dots$; значения этих кросс-корреляций приведены в столбце “lead”.

Included observations: 96

Correlations are asymptotically consistent approximations

e, $\Delta Z(-i)$	e, $\Delta Z(+i)$	i	lag	lead
. *****	. *****	0	0.9017	0.9017
.	. *	1	-0.0217	0.0830
.	* .	2	-0.0956	-0.0413
.	. .	3	0.0064	0.0341
* .	. .	4	-0.0510	0.0118
* .	. .	5	-0.0824	-0.0228
. .	. .	6	-0.0171	0.0150
** .	** .	7	-0.1858	-0.1579
. .	. .	8	-0.0292	-0.0272
. *	. *	9	0.0833	0.0701
. .	. .	10	0.0125	0.0216

На основании этой кросс-коррелограммы можно предполагать наличие ненулевых кросс-корреляций в DGP до 7-го порядка. В соответствии с этим, добавим в правую часть оцененного ранее уравнения запаздывающие и опережающие разности переменной z_t вплоть до 7-го порядка.

Dependent Variable: Y

Sample(adjusted): 9 93

Included observations: 85 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.987362	0.020874	238.9236	0.0000
Z	1.000689	0.007818	127.9955	0.0000
D(Z)	1.006216	0.013125	76.66298	0.0000
D(Z(-1))	0.237875	0.012764	18.63643	0.0000
D(Z(-2))	0.089302	0.012810	6.971105	0.0000
D(Z(-3))	-0.008934	0.012368	-0.722323	0.4726
D(Z(-4))	-0.002997	0.012391	-0.241901	0.8096
D(Z(-5))	-0.011646	0.012179	-0.956245	0.3423
D(Z(-6))	-0.010012	0.011925	-0.839615	0.4041
D(Z(-7))	-0.003586	0.011634	-0.308269	0.7588
D(Z(1))	0.262537	0.013373	19.63226	0.0000
D(Z(2))	0.116863	0.013365	8.744236	0.0000
D(Z(3))	-0.010921	0.013219	-0.826184	0.4116
D(Z(4))	0.003903	0.013276	0.294017	0.7696
D(Z(5))	0.021536	0.013232	1.627644	0.1082
D(Z(6))	-0.008452	0.012699	-0.665583	0.5079
D(Z(7))	0.002945	0.012199	0.241376	0.8100

Ряд остатков не обнаруживает автокоррелированности: Р-значения критерия Бройша – Годфри равны 0.252 (при глубине запаздываний $K = 1$) и 0.427 ($K = 2$). Поэтому мы

можем использовать для проверки гипотезы $H_0: \beta = 1$ обычную t -статистику без коррекции стандартной ошибки; ее значение равно

$$t = (1.000689 - 1) / 0.007818 = 0.0081,$$

так что гипотеза $H_0: \beta = 1$ не отвергается.

Пример (продолжение)

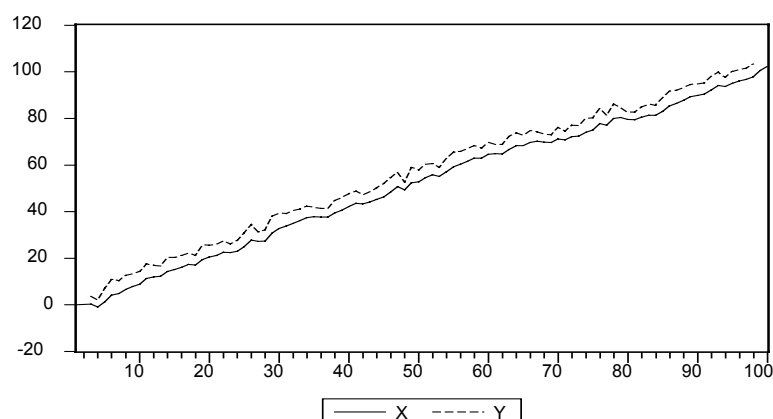
Изменим теперь DGP так, чтобы слева и справа в первом уравнении стояли $I(1)$ ряды с линейным трендом.

$$\text{DGP: } y_t = 5 + x_t + u_t,$$

$$x_t = 1 + x_{t-1} + v_t,$$

где $x_1 = 0$, а u_t, v_t – те же, что и ранее.

Смоделированная реализация этого DGP имеет вид



Оцененное уравнение регрессии $y_t = \alpha + \beta x_t + \eta_t$:

Dependent Variable: Y

Method: Least Squares

Sample(adjusted): 3 98

Included observations: 96 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.172117	0.201812	25.62837	0.0000
X	0.997147	0.003444	289.5306	0.0000

Кросс-коррелограмма ряда остатков от оцененного уравнения и приращений ряда x_t имеет вид, аналогичный предыдущей коррелограмме. Поэтому опять переходим к оцениванию расширенного уравнения, дополненного семью запаздывающими и семью опережающими разностями:

Dependent Variable: Y

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.224675	0.098061	32.88423	0.0000
X	1.000621	0.000520	1923.269	0.0000
D(X)	1.009585	0.012922	78.12803	0.0000
D(X(-1))	0.241440	0.012760	18.92225	0.0000
D(X(-2))	0.093472	0.012938	7.224562	0.0000
D(X(-3))	-0.004479	0.012656	-0.353864	0.7245

D(X(-4))	0.001328	0.012648	0.104986	0.9167
D(X(-5))	-0.007856	0.012272	-0.640121	0.5242
D(X(-6))	-0.006808	0.011934	-0.570501	0.5702
D(X(-7))	-0.001175	0.011513	-0.102067	0.9190
D(X(1))	0.266271	0.013016	20.45699	0.0000
D(X(2))	0.120232	0.012948	9.286042	0.0000
D(X(3))	-0.007953	0.012873	-0.617748	0.5388
D(X(4))	0.006576	0.012866	0.511164	0.6109
D(X(5))	0.023956	0.012814	1.869580	0.0658
D(X(6))	-0.007153	0.012294	-0.581813	0.5626
D(X(7))	0.003367	0.011821	0.284813	0.7767

В ряде остатков и здесь не обнаруживается автокоррелированности, так что можно использовать для проверки гипотезы $H_0: \beta = 1$ обычную t -статистику без коррекции стандартной ошибки; ее значение равно

$$t = (1.000621 - 1) / 0.00520 = 1.194,$$

гипотеза $H_0: \beta = 1$ не отвергается.

Включим в правую часть оцениваемого уравнения еще и тренд. При этом получаем:

Dependent Variable: Y

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.223725	0.101508	31.75848	0.0000
@TREND	-0.000317	0.007802	-0.040686	0.9677
X	1.000938	0.007798	128.3544	0.0000
D(X)	1.009461	0.013372	75.49241	0.0000
D(X(-1))	0.241345	0.013061	18.47841	0.0000
D(X(-2))	0.093380	0.013231	7.057897	0.0000
D(X(-3))	-0.004552	0.012877	-0.353495	0.7248
D(X(-4))	0.001255	0.012867	0.097531	0.9226
D(X(-5))	-0.007941	0.012541	-0.633231	0.5287
D(X(-6))	-0.006888	0.012181	-0.565473	0.5736
D(X(-7))	-0.001256	0.011767	-0.106726	0.9153
D(X(1))	0.266438	0.013736	19.39718	0.0000
D(X(2))	0.120397	0.013657	8.815884	0.0000
D(X(3))	-0.007809	0.013441	-0.580980	0.5632
D(X(4))	0.006723	0.013451	0.499805	0.6189
D(X(5))	0.024098	0.013370	1.802394	0.0760
D(X(6))	-0.007035	0.012719	-0.553156	0.5820
D(X(7))	0.003469	0.012172	0.285016	0.7765

Гипотеза $H_0: \beta = 1$ не отвергается и для переменных, очищенных от тренда. Коэффициент при трендовой переменной статистически незначим. Полученные результаты указывают на то, что мы имеем дело с детерминистской коинтеграцией.

Пример

Рассмотрим теперь следующий DGP:

$$W_t = 5 + t + rw_t,$$

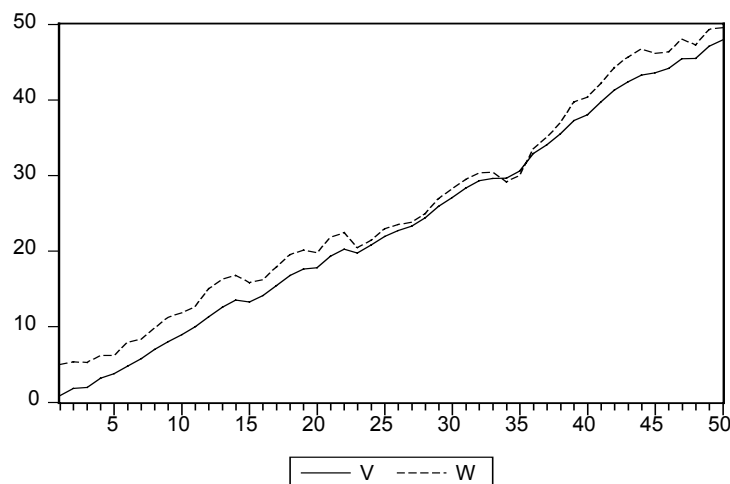
$$V_t = 1 + t + 0.5 * rw_t + 0.1 * n2_t,$$

где

$$rw_t = rw_{t-1} + 0.5 * n3_t - \text{случайное блуждание без сноса},$$

$n2_t$, $n3_t$ – некоррелированные гауссовские процессы белого шума с единичной дисперсией.

Смоделированная реализация длины 50 имеет вид:



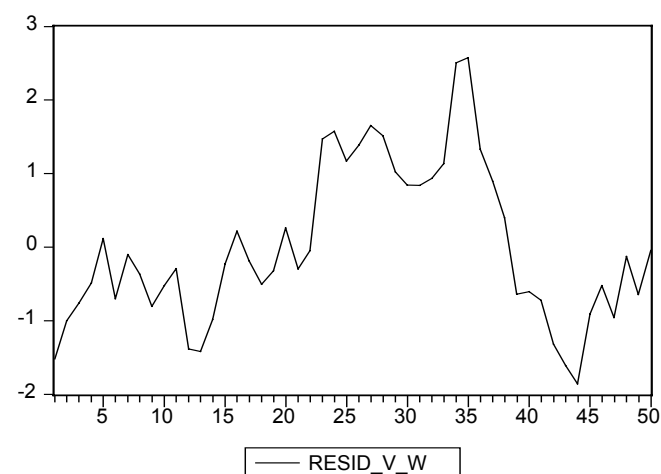
Оцениваем статистическую модель $V_t = \alpha + \beta W_t + \eta_t$:

Dependent Variable: V

Included observations: 50

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2.657128	0.325116	-8.172855	0.0000
W	1.021898	0.011165	91.52908	0.0000
Durbin-Watson stat	0.290480	Prob(F-statistic)		0.000000

Ряд остатков



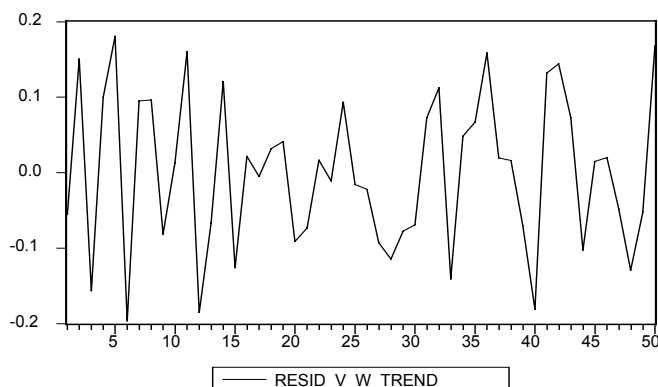
идентифицируется как интегрированный (статистика Дики – Фуллера равна -2.22 при 5% критическом значении -3.46), так что ряды V_t и W_t не являются детерминистски коинтегрированными. Близость к 1 оценки коэффициента β соответствует равенству угловых коэффициентов детерминированных трендов, входящих в состав в рядов V_t и W_t . Ряд $V_t - W_t$ не имеет выраженного детерминированного тренда и его график отличается от только что представленного практически только сдвигом.

Добавим в правую часть оцениваемого уравнения трендовую составляющую:

Dependent Variable: V

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.493506	0.035792	-41.72787	0.0000
@TREND	0.503277	0.007131	70.57274	0.0000
W	0.496897	0.007519	66.08810	0.0000
Durbin-Watson stat	2.275079	Prob(F-statistic)		0.000000

Теперь ряд остатков имеет вид



и идентифицируется как стационарный (статистика Дики – Фуллера равна -7.09). Кросс-коррелограмма ряда остатков и приращений ряда W_t

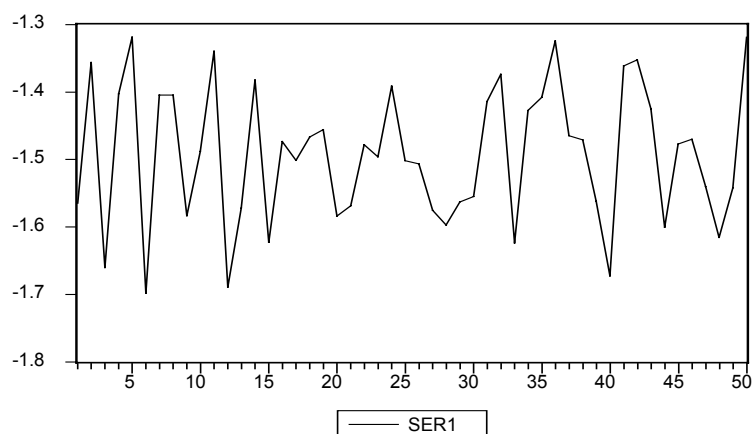
e ,W_DIF(-i)	e ,W_DIF(+i)	i	lag	lead
. .	. .	0	0.0353	0.0353
. .	. *	1	-0.0237	0.1217
* .	. .	2	-0.0846	0.0115
. .	* .	3	0.0052	-0.1083
* .	. *	4	-0.0776	0.1174
. *	. *	5	0.1352	0.1018
. **	. .	6	0.1986	0.0347
. *	. **	7	0.1093	0.1669
** .	. *	8	-0.1751	0.0614
** .	**** .	9	-0.2456	-0.3565
. *	* .	10	0.1177	-0.0421

указывает на то, что здесь для пополнения оцениваемого уравнения достаточно ограничиться включением в правую часть 9 запаздывающих и опережающих разностей ряда W_t .

Оценивая пополненное уравнение, получаем следующий результаты для коэффициентов при тренде и W_t :

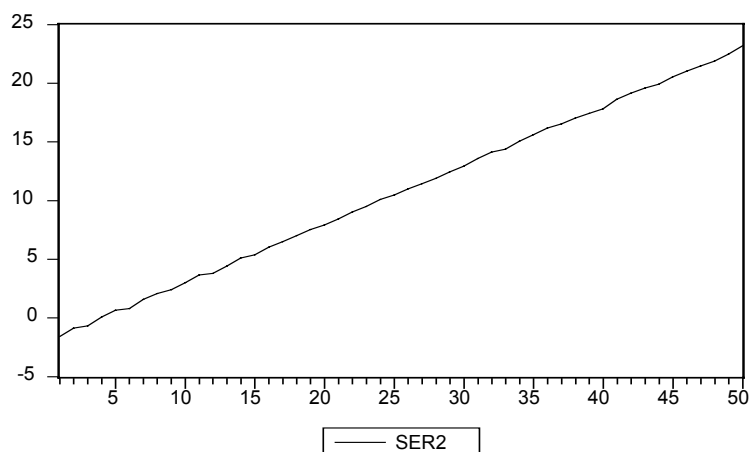
@TREND	0.497313	0.023984	20.73533	0.0000
W	0.505495	0.025512	19.81437	0.0000

При этом гипотеза гауссовского белого шума для ряда остатков не отвергается. Это означает, что мы имеем здесь дело со стохастической коинтеграцией. В рамках расширенной модели не отвергается гипотеза о равенстве 0.5 коэффициентов при тренде и W_t . График ряда $V_t - 0.5t - 0.5W_t$ имеет вид



и этот ряд идентифицируется как стационарный.

Подведем итог. Ряд $V_t - W_t$ не имеет выраженного детерминированного тренда, но имеет стохастический тренд. Ряд $V_t - 0.5W_t$ имеет выраженный линейный тренд, но не имеет стохастического тренда:



Наконец, ряд $V_t - 0.5t - 0.5W_t$ идентифицируется как стационарный, со средним значением -1.493 и стандартным отклонением 0.104 . И это находится в полном соответствии с использованным при моделировании реализаций процессом порождения данных. Действительно, в соответствии с этим DGP,

$$\begin{aligned} V_t - 0.5t - 0.5W_t &= (1 + t + 0.5 * rw_t + 0.1 * n2_t) - 0.5t - 0.5(5 + t + rw_t) = \\ &= -1.5 + 0.1 * n2_t. \end{aligned}$$

Глава 8. Процедура Йохансена

8.1. Оценивание ранга коинтеграции

Пусть $I(1)$ ряды y_{1t}, \dots, y_{Nt} в совокупности образуют векторный ряд

$$y_t = (y_{1t}, \dots, y_{Nt})^T,$$

следующий модели векторной авторегрессии $VAR(p)$

$$A(L)y_t = \mu + \varepsilon_t,$$

где

$$A(L) = A_0 - A_1L - \dots - A_pL^p,$$

A_0, A_1, \dots, A_p – матрицы размера $(N \times N)$,

$A_0 = I_N$ (единичная матрица),

т.е.

$$y_t = \mu + A_1 y_{t-1} + \dots + A_p y_{t-p} + \varepsilon_t.$$

Путем алгебраических преобразований эту модель можно представить также в виде

$$\Delta y_t = \mu + \zeta_0 y_{t-1} + \zeta_1 \Delta y_{t-1} + \dots + \zeta_{p-1} \Delta y_{t-p+1} + \varepsilon_t,$$

где

$$\zeta_0 = A_1 + \dots + A_p - I_N,$$

$$\zeta_k = -(A_{k+1} + \dots + A_p), \quad k = 1, 2, \dots, p-1.$$

Заметим, что

$$\zeta_0 = A_1 + \dots + A_p - I_N = -A(1),$$

так что $\text{rank } \zeta_0 = \text{rank } A(1)$.

Мы уже отмечали выше (разд.7.4), что если ряды y_{1t}, \dots, y_{Nt} коинтегрированы, то матрица $A(1)$ имеет пониженный ранг ($\text{rank } A(1) < N$). Этот же пониженный ранг будет иметь в этом случае и матрица ζ_0 . В общем случае, ранг матрицы ζ_0 может принимать значения $r = \text{rank } \zeta_0 = 0, 1, \dots, N$.

- Значения $r = 1, \dots, N-1$ соответствуют коинтегрированной VAR (ряды $y_{1t}, \dots, y_{Nt} \sim I(1)$ коинтегрированы).
- Если $r = 0$, то ряды y_{1t}, \dots, y_{Nt} не коинтегрированы.
- Если $r = N$, то тогда любой N -мерный вектор является коинтегрирующим, так что коинтегрирующими будут, например, векторы $(1, 0, \dots, 0)^T, (0, 1, \dots, 0)^T, \dots, (0, 0, \dots, 1)^T$. Но это означает, что все ряды y_{1t}, \dots, y_{Nt} являются стационарными.

Ранг матрицы ζ_0 , $r = \text{rank } \zeta_0$, обычно называют рангом коинтеграции рассматриваемой системы рядов y_{1t}, \dots, y_{Nt} , вне зависимости от того, имеет ли место действительная коинтеграция этих рядов.

Выяснение ранга коинтеграции является ключевым моментом в построении **ЕСМ – модели коррекции ошибок** – по наблюдаемым статистическим данным. Один из возможных путей решения этой задачи был предложен Йохансеном ([Johansen (1988)], [Johansen (1991)], [Johansen (1992)], [Johansen (1994)], [Johansen, Juselius (1990)]). Изложение этого метода требует перехода к гораздо более высокому математическому уровню. Поэтому мы, не выходя слишком далеко за принятую планку строгости и

детальности изложения, дадим здесь только самое общее представление об этом методе.

Мы уже говорили о том, что если коинтегрированная система $I(1)$ рядов y_{1t}, \dots, y_{Nt} может быть представлена в форме VAR с $\text{rank } A(1) = r$, то существует соответствующее представление VAR в форме ECM. Исходя из этого, Йохансен в качестве отправной точки берет представление¹

$$\Delta y_t = \mu + \zeta_0 y_{t-1} + \zeta_1 \Delta y_{t-1} \dots + \zeta_{p-1} \Delta y_{t-p+1} + \varepsilon_t$$

с матрицей

$$\zeta_0 = \alpha \beta^T,$$

где α и β – $(N \times r)$ -матрицы полного ранга r . При этом столбцы $\beta_{(1)}, \dots, \beta_{(r)}$ матрицы β являются линейно независимыми коинтегрирующими векторами, а элементы матрицы α являются коэффициентами при стационарных линейных комбинациях

$$z_{1,t-1} = \beta_{(1)}^T y_{t-1}, \dots, z_{r,t-1} = \beta_{(r)}^T y_{t-1}$$

(представляющих отклонения от r долговременных соотношений между рядами y_{1t}, \dots, y_{Nt}) в правых частях уравнений для $\Delta y_{1t}, \dots, \Delta y_{Nt}$.

В процедуре Йохансена предполагается, что ε_t – N -мерный гауссовский белый шум, так что случайный вектор $\varepsilon_t = (\varepsilon_{1t}, \dots, \varepsilon_{Nt})^T$ имеет N -мерное нормальное распределение с нулевым средним и ковариационной матрицей $\text{Cov}(\varepsilon_t) = \Omega$, и $\text{Cov}(\varepsilon_{kt}, \varepsilon_{js}) = 0$ при $t \neq s$ для всех $k, j = 1, \dots, N$.

Прежде, чем применять процедуру Йохансена, следует определиться с порядком p векторной авторегрессии, которой следует векторный ряд. Для этой цели можно использовать стандартные t - и F -критерии (с асимптотическим $N(0, 1)$ распределением для t -статистик и асимптотическими χ^2 распределениями для qF) и, применяя их к VAR в уровнях, порядок которой взят “с запасом”, понизить, по возможности, порядок этой “избыточной” VAR. Заметим в этой связи, что процедура Йохансена достаточно чувствительна к выбору порядка VAR, в рамках которой эта процедура реализуется.

Сама процедура начинается с того, что по имеющимся наблюдениям значений y_{1t}, \dots, y_{Nt} , $t = 1, \dots, T$, вычисляются максимумы логарифмических функций правдоподобия $L(\Omega, \mu, \zeta_0, \zeta_1, \dots, \zeta_{p-1})$ для неизвестных параметров $\Omega, \mu, \zeta_0, \zeta_1, \dots, \zeta_{p-1}$ при различных предположениях о ранге коинтеграции r . С точностью до слагаемых, одинаковых при различных r , эти максимумы равны

$$L_{\max}(r) = -(T/2) \sum_{i=1}^r \ln(1 - \hat{\lambda}_i), \quad r = 1, \dots, N,$$

где $\hat{\lambda}_1, \dots, \hat{\lambda}_N$ – некоторые величины, вычисляемые на основании одних только статистических данных без всяких предположений о ранге коинтеграции, $1 > \hat{\lambda}_1 > \dots > \hat{\lambda}_N > 0$. Сравнивая значения $L_{\max}(r)$, полученные при различных r , можно отдать в итоге предпочтение той или иной гипотезе об истинном ранге коинтеграции. Для формализации соответствующего решения в виде некоторой статистической процедуры можно использовать известный из математической статистики критерий отношения правдоподобий для различения двух гипотез.

¹ Мы ограничиваемся в этой книге системами $I(1)$ рядов. Йохансен рассматривал также и системы, включающие ряды типа $I(2)$, см. например, [Johansen (1994a)], [Johansen (1994b)], [Johansen (1995b)].

Пусть в качестве исходной (“нулевой”) выступает гипотеза $H_0: r = r^*$, а в качестве альтернативной – гипотеза $H_A: r = r^* + 1$. Для различения этих гипотез сравниваются значения

$$L_{\max}(r^*) = -(T/2) \sum_{i=1}^{r^*} \ln(1 - \hat{\lambda}_i)$$

и

$$L_{\max}(r^* + 1) = -(T/2) \sum_{i=1}^{r^*+1} \ln(1 - \hat{\lambda}_i).$$

Критерий основывается на статистике

$$\lambda_{\max}(r^*) = 2(L_{\max}(r^* + 1) - L_{\max}(r^*)) = -(T/2) \ln(1 - \hat{\lambda}_{r^*+1}).$$

Асимптотическое (при $T \rightarrow \infty$) распределение этой статистики при гипотезе H_0 зависит от r^* и N ; для него рассчитаны соответствующие таблицы (см., например, [Patterson (2000), таблицы 14.3 – 14.7], [Enders (1995), таблица B] или [Hamilton (1994), таблица B.11]).

Если гипотеза $H_0: r = r^*$ верна, то значения $\hat{\lambda}_{r^*+1}, \dots, \hat{\lambda}_N$ близки к нулю. Если верна альтернативная гипотеза, то значение $\hat{\lambda}_{r^*+1}$ отделено от нуля, и значения статистики $\lambda_{\max}(r^*)$ смещаются в сторону больших положительных значений. Поэтому гипотезу $H_0: r = r^*$ следует отвергать в пользу гипотезы $H_A: r = r^* + 1$ при больших положительных значениях статистики $\lambda_{\max}(r^*)$, превышающих соответствующий критический уровень. Пусть теперь в качестве исходной (“нулевой”) опять выступает гипотеза $H_0: r = r^*$, но в качестве альтернативной берется гипотеза $H_A: r > r^*$. Для различения этих гипотез сравниваются значения

$$L_{\max}(r^*) = -(T/2) \sum_{i=1}^{r^*} \ln(1 - \hat{\lambda}_i)$$

и

$$L_{\max}(N) = -(T/2) \sum_{i=1}^N \ln(1 - \hat{\lambda}_i).$$

Критерий основывается на статистике

$$\lambda_{\text{trace}}(r^*) = 2(L_{\max}(N) - L_{\max}(r^*)) = -(T/2) \sum_{i=r^*+1}^N \ln(1 - \hat{\lambda}_i).$$

Асимптотическое (при $T \rightarrow \infty$) распределение этой статистики при гипотезе H_0 зависит от r^* и N ; для него также рассчитаны соответствующие таблицы (см., например, [Patterson (2000), таблицы 14.3 – 14.7], [Enders (1995), таблица B] или [Hamilton (1994), таблица B.10]).

Если гипотеза $H_0: r = r^*$ верна, то значения $\hat{\lambda}_{r^*+1}, \dots, \hat{\lambda}_N$ близки к нулю. Если верна альтернативная гипотеза, то эти значения отделены от нуля, и значения статистики $\lambda_{\text{trace}}(r^*)$ смещаются в сторону больших положительных значений. Поэтому гипотезу $H_0: r = r^*$ следует отвергать в пользу гипотезы $H_A: r > 1$ при больших положительных значениях статистики $\lambda_{\text{trace}}(r^*)$, превышающих соответствующий критический уровень. Проблема, однако, в том, что заранее обычно не известно, на какое значение r следует рассчитывать. В таком случае возникает целое множество альтернативных пар гипотез, при проверке которых можно получить несогласующиеся результаты. Йохансен предложил последовательную процедуру проверки гипотез, с помощью которой можно получить состоятельную оценку истинного ранга коинтеграции.

Именно, зададимся некоторым уровнем значимости α , скажем, 0.05, и начнем с проверки гипотезы $H_0: r = 0$ против альтернативы $H_A: r > 0$. Если эта гипотеза не отвергается, то полагаем $\hat{r} = 0$. В противном случае проверяем гипотезу $H_0: r = 1$ против альтернативы $H_A: r > 1$. Если гипотеза $H_0: r = 1$ не отвергается, то полагаем $\hat{r} = 1$; в противном случае проверяем гипотезу $H_0: r = 2$ против $H_A: r > 2$ и т.д.

Полученная оценка \hat{r} состоятельна в следующем смысле. Если в действительности $r = r_0$, то при $T \rightarrow \infty$

$$P\{\hat{r} = k\} \rightarrow 0, \quad k = 0, 1, \dots, r_0 - 1,$$

$$P\{\hat{r} = r_0\} \rightarrow 1 - \alpha.$$

Таким образом, оценить ранг коинтеграции можно, по крайней мере, теоретически.

Однако при обсуждении процедуры Йохансена мы не упомянули еще об одной серьезной проблеме, стоящей на пути к оцениванию истинного ранга коинтеграции. Дело в том, что критические значения статистик критериев отношения правдоподобий зависят не только от r^* и N . Они зависят также от того, имеют ли ряды детерминированные тренды, включается ли константа и/или тренд в коинтеграционное соотношение (*коинтеграционное уравнение*, *CE* – *cointegrating equation*). В связи с этим, при каждом значении r ранга коинтеграции можно рассмотреть следующие 5 ситуаций (именно эти ситуации учитываются, например, в пакете EVIEWS).

- $H_2(r)$: в данных нет детерминированных трендов;
в CE не включаются ни константа ни тренд.
- $H_1^*(r)$: в данных нет детерминированных трендов;
в CE включается константа, но не включается тренд.
- $H_1(r)$: в данных есть детерминированный линейный тренд;
в CE включается константа, но не включается тренд.
- $H^*(r)$: в данных есть детерминированный линейный тренд;
в CE включаются константа и линейный тренд.
- $H(r)$: в данных есть детерминированный квадратичный тренд;
в CE включаются константа и линейный тренд.

При фиксированном ранге r перечисленные 5 ситуаций образуют цепочку вложенных гипотез:

$$H_2(r) \subset H_1^*(r) \subset H_1(r) \subset H^*(r) \subset H(r).$$

Это дает возможность, опять используя критерий отношения правдоподобий, проверять выполнение гипотезы, стоящей левее в этой цепочке, в рамках гипотезы, расположенной непосредственно справа. Во всех случаях асимптотическое распределение статистики критерия является распределением хи-квадрат. Что касается степеней свободы у этого асимптотического распределения, то оно равно

r – для пар $H_2(r) \subset H_1^*(r)$ и $H_1(r) \subset H^*(r)$,

$(N - r)$ – для пар $H_1^*(r) \subset H_1(r)$ и $H^*(r) \subset H(r)$.

Заметим, что для каждой из 5 ситуаций, в свою очередь, образуются цепочки вложенных гипотез:

$$H(0) \subset \dots \subset H(r) \subset \dots \subset H(N)$$

$$H^*(0) \subset \dots \subset H^*(r) \subset \dots \subset H^*(N)$$

$$H_1(0) \subset \dots \subset H_1(r) \subset \dots \subset H_1(N)$$

$$H_1^*(0) \subset \dots \subset H_1^*(r) \subset \dots \subset H_1^*(N)$$

$$H_2(0) \subset \dots \subset H_2(r) \subset \dots \subset H_2(N).$$

Критические значения статистик λ_{\max} и λ_{trace} , используемые при решении вопроса о ранге коинтеграции, различны для этих 5 цепочек, и это осложняет задачу оценивания ранга коинтеграции, поскольку приходится предварительно выбирать цепочку, в рамках которой будет производиться оценивание.

Некоторым подспорьем в этом отношении является сводка значений информационных критериев Акаике (AIC) и Шварца для всех упомянутых $5(N+1)$ вариантов. Как и обычно, наилучшая модель выбирается по минимуму значений критерия Акаике или критерия Шварца. Впрочем, практика показывает, что более доверять в этом отношении стоит критерию Шварца. При анализе смоделированных данных выбор по критерию Акаике часто приводит к результатам, совершенно не соответствующим процессу порождения данных.

Для лучшего уяснения возникающих вариантов, рассмотрим 5 ситуаций, перечисленных выше, на простейшем примере треугольной системы Филлипса для двух I(1) рядов.

$$\mathbf{DGP}_1 : y_t = \beta x_t + \varepsilon_t , \\ x_t = x_{t-1} + v_t .$$

На основании этих двух уравнений находим:

$$y_t - y_{t-1} = -y_{t-1} + \beta(x_{t-1} + v_t) + \varepsilon_t = -(y_{t-1} - \beta x_{t-1}) + u_t ,$$

где

$$u_t = \beta v_t + \varepsilon_t ,$$

так что получаем ЕСМ в виде

$$\Delta y_t = -z_{t-1} + u_t ,$$

$$\Delta x_t = v_t ,$$

где

$$z_t = y_t - \beta x_t \quad (\text{“константа и тренд не включаются в СЕ”}).$$

Поскольку ряды y_t и x_t не содержат детерминированного тренда (“тренда в данных нет”), то все это соответствует ситуации $\mathbf{H}_2(r)$.

$$\mathbf{DGP}_2 : y_t = \rho_0 + \beta x_t + \varepsilon_t , \\ x_t = x_{t-1} + v_t .$$

Эту систему можно записать в виде

$$\Delta y_t = -z_{t-1} + u_t ,$$

$$\Delta x_t = v_t ,$$

где

$$z_t = y_t - \rho_0 - \beta x_t \quad (\text{“константа включается в СЕ”}).$$

Поскольку ряды y_t и x_t не содержат детерминированного тренда (“тренда в данных нет”), то все это соответствует ситуации $\mathbf{H}_1^*(r)$.

$$\mathbf{DGP}_3 : y_t = \rho_0 + \beta x_t + \varepsilon_t , \\ x_t = \gamma_0 + x_{t-1} + v_t .$$

В этом случае

$$y_t - y_{t-1} = -y_{t-1} + \rho_0 + \beta(\gamma_0 + x_{t-1} + v_t) + \varepsilon_t = \\ = -(y_{t-1} - \rho_0 - \beta x_{t-1}) + \beta \gamma_0 + u_t ,$$

так что получаем

$$\Delta y_t = -z_{t-1} + \beta \gamma_0 + u_t,$$

$$\Delta x_t = \gamma_0 + v_t,$$

где

$$z_t = y_t - \rho_0 - \beta x_t \quad (\text{“константа включается в СЕ”}).$$

Ряд x_t содержит линейный тренд (“тренд в данных”). Все это соответствует ситуации $\mathbf{H}_1(r)$.

$$\mathbf{DGP}_4 : y_t = \rho_0 + \rho_1 t + \beta x_t + \varepsilon_t,$$

$$x_t = \gamma_0 + x_{t-1} + v_t.$$

Здесь

$$\begin{aligned} y_t - y_{t-1} &= -y_{t-1} + \rho_0 + \rho_1 t + \beta (\gamma_0 + x_{t-1} + v_t) + \varepsilon_t = \\ &= -(y_{t-1} - \rho_0 - \rho_1 t - \beta x_{t-1}) + \beta \gamma_0 + u_t, \end{aligned}$$

или

$$\Delta y_t = -z_{t-1} + \beta \gamma_0 + u_t,$$

$$\Delta x_t = \gamma_0 + v_t,$$

где

$$z_t = y_t - \rho_0 - \rho_1 t - \beta x_t \quad (\text{“константа и тренд включаются в СЕ”}).$$

Ряд x_t содержит линейный тренд (“тренд в данных”). Все это соответствует ситуации $\mathbf{H}^*(r)$.

$$\mathbf{DGP}_5 : y_t = \rho_0 + \rho_1 t + \beta x_t + \varepsilon_t,$$

$$x_t = \gamma_0 + \gamma_1 t + x_{t-1} + v_t.$$

В этом случае

$$\begin{aligned} y_t - y_{t-1} &= -y_{t-1} + \rho_0 + \rho_1 t + \beta (\gamma_0 + \gamma_1 t + x_{t-1} + v_t) + \varepsilon_t = \\ &= -(y_{t-1} - \rho_0 - \rho_1 t - \beta x_{t-1}) + \beta \gamma_0 + \beta \gamma_1 t + u_t, \end{aligned}$$

или

$$\Delta y_t = -z_{t-1} + \beta \gamma_0 + \beta \gamma_1 t + u_t,$$

$$\Delta x_t = \gamma_0 + \gamma_1 t + v_t,$$

где

$$z_t = y_t - \rho_0 - \rho_1 t - \beta x_t \quad (\text{“константа и тренд включаются в СЕ”}).$$

Ряд x_t содержит квадратичный тренд (“квадратичный тренд в данных”). Все это соответствует ситуации $\mathbf{H}(r)$.

Пример

В качестве примера мы проведем анализ смоделированных данных, реализующих 5 только что рассмотренных вариантов DGP. При этом использовались следующие значения параметров:

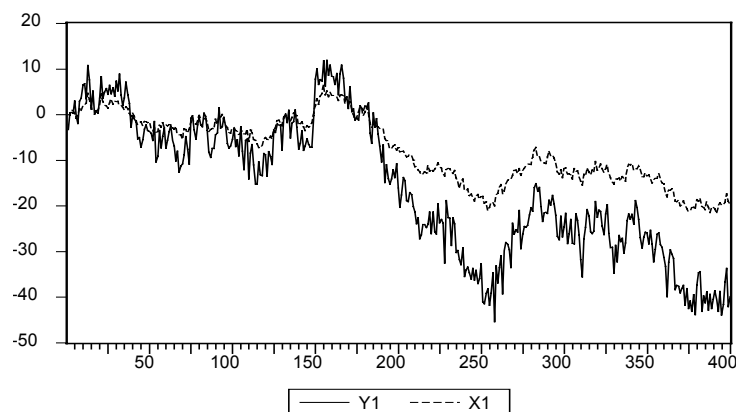
$$\beta = 2, \rho_0 = 5, \rho_1 = 0.2, \gamma_0 = 0.2, \gamma_1 = 0.01.$$

В качестве рядов ε_t и v_t брались имитации длины $T = 400$ независимых между собой гауссовских белых шумов, имеющих дисперсии, равные 4 и 1, соответственно.

$$\mathbf{DGP}_1 : y_t = 2 x_t + \varepsilon_t,$$

$$x_t = x_{t-1} + v_t.$$

Смоделированная реализация имеет вид



Сводка статистик для определения ранга коинтеграции по этим смоделированным данным, получаемая в пакете EVIEWS (в предположении, что VAR в уровнях имеет порядок 2):

Sample: 1 400

Included observations: 398

Series: Y1 X1

Lags interval: 1 to 1

Data Trend:	None	None	Linear	Linear	Quadratic
Rank or	No Intercept	Intercept	Intercept	Intercept	Intercept
No. of CEs	No Trend	No Trend	No Trend	Trend	Trend
Log Likelihood					
0	-1526.582	-1526.582	-1526.015	-1526.015	-1525.928
1	-1434.108	-1433.925	-1433.357	-1433.324	-1433.242
2	-1434.100	-1432.926	-1432.926	-1430.264	-1430.264
AIC					
0	7.691369	7.691369	7.698565	7.698565	7.708180
1	7.246775	7.250880	7.253051	7.257911	7.262521
2	7.266836	7.270985	7.270985	7.267658	7.267658
Schwarz Criteria					
0	7.731434	7.731434	7.758663	7.758663	7.788309
1	7.326905	7.341026	7.353213	7.368089	7.382716
2	7.387030	7.411212	7.411212	7.427918	7.427918
L.R. Test:	Rank = 1	Rank = 1	Rank = 1	Rank = 1	Rank = 2

В таблице для каждой из 5 указанных выше возможных ситуаций при 3 возможных рангах коинтеграции ($r = 0, 1, 2$) приведены:

- Значение $L_{\max}(r)$ максимума логарифма функции правдоподобия (Log Likelihood), соответствующее выделенному варианту.
- Значение информационного критерия Акаике (AIC – Akaike Information Criteria), соответствующее выделенному варианту.
- Значение информационного критерия Шварца (Schwarz Criteria), соответствующее выделенному варианту.

В первом столбце указывается “испытываемый” ранг коинтеграции (Rank or No. of CEs). Следующие 5 столбцов соответствуют 5 ситуациям, перечисленным выше. (В

порядке слева направо это ситуации $H_2(r)$, $H_1^*(r)$, $H_1(r)$, $H^*(r)$, $H(r)$.) В конце каждого столбца указан результат оценивания ранга коинтеграции в рамках цепочки, соответствующей данному столбцу. Таким образом, в рамках первых четырех ситуаций ранг коинтеграции оценивается как $r = 1$, а в рамках пятой ситуации ($H(r)$) ранг коинтеграции оценивается как $r = 2$.

Ориентируясь на критерий Акаике, наилучшей следует признать модель $H_2(1)$ (нет тренда в данных, в СЕ не включаются ни константа ни тренд; ранг коинтеграции равен 1) – для нее значение критерия минимально (равно 7.246775). Та же модель выбирается и критерием Шварца (для нее значение критерия равно 7.326905).

Следующая таблица расшифровывает процесс получения оценки ранга коинтеграции в ситуации, соответствующей первому столбцу.

Sample: 1 400

Included observations: 398

Test assumption: No deterministic trend in the data

Series: Y1 X1

Lags interval: 1 to 1

Eigenvalue	Likelihood Ratio	5 Percent Critical Value	1 Percent Critical Value	Hypothesized No. of CE(s)
0.371673	184.9642	12.53	16.31	None **
4.00E-05	0.015911	3.84	6.51	At most 1

*(**) denotes rejection of the hypothesis at 5%(1%) significance level

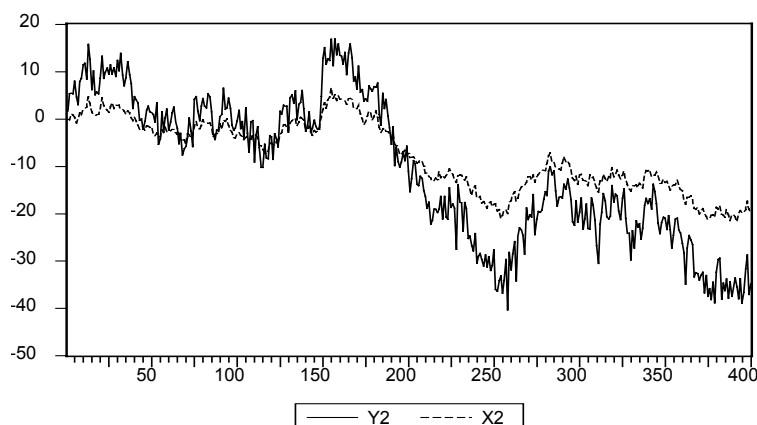
L.R. test indicates 1 cointegrating equation(s) at 5% significance level

В первом столбце (Eigenvalue) указаны значения $\hat{\lambda}_1, \hat{\lambda}_2$, используемые в критерии отношения правдоподобий. В первой строке 184.9642 – наблюдаемое значение статистики $\lambda_{\text{trace}}(0)$, используемой при проверке гипотезы $H_0: r = 0$ против альтернативы $H_A: r > 0$. Рядом с ним приводятся 5% и 1% критические значения статистики $\lambda_{\text{trace}}(0)$ в рассматриваемой ситуации. Поскольку наблюдаемое значение значительно превосходит оба критических значения, гипотеза $H_0: r = 0$ отвергается в пользу альтернативы $H_A: r > 0$. Во второй строке 0.015911 – наблюдаемое значение статистики $\lambda_{\text{trace}}(1)$, используемой при проверке гипотезы $H_0: r = 1$ против альтернативы $H_A: r > 1$. Рядом приводятся 5% и 1% критические значения статистики $\lambda_{\text{trace}}(1)$. Наблюдаемое значение статистики $\lambda_{\text{trace}}(1)$ намного меньше критических, так что гипотеза $H_0: r = 1$ не отвергается. Как итог оцененное значение ранга коинтеграции принимается равным 1, что соответствует истинному положению вещей.

$$\text{DGP}_2 : y_t = 5 + 2x_t + \varepsilon_t,$$

$$x_t = x_{t-1} + v_t.$$

Смоделированная реализация имеет вид



Сводка статистик для определения ранга коинтеграции по этим смоделированным данным:

Sample: 1 400

Included observations: 398

Series: Y2 X2

Lags interval: 1 to 1

Data Trend:	None	None	Linear	Linear	Quadratic
Rank or	No Intercept	Intercept	Intercept	Intercept	Intercept
No. of CEs	No Trend	No Trend	No Trend	Trend	Trend
Log Likelihood					
0	-1526.582	-1526.582	-1526.015	-1526.015	-1525.928
1	-1513.785	-1433.925	-1433.357	-1433.324	-1433.242
2	-1513.777	-1432.926	-1432.926	-1430.264	-1430.264
AIC					
0	7.691369	7.691369	7.698565	7.698565	7.708180
1	7.647159	7.250880	7.253051	7.257911	7.262521
2	7.667223	7.270985	7.270985	7.267658	7.267658
Schwarz Criteria					
0	7.731434	7.731434	7.758663	7.758663	7.788309
1	7.727289	7.341026	7.353213	7.368089	7.382716
2	7.787417	7.411212	7.411212	7.427918	7.427918
L.R. Test:	Rank = 1	Rank = 1	Rank = 1	Rank = 1	Rank = 2

Ориентируясь на критерий Акаике, наилучшей следует признать модель $H_1^*(1)$ (нет тренда в данных, в СЕ включается константа; ранг коинтеграции равен 1) – для нее значение критерия минимально (равно 7.250880). Та же модель выбирается и критерием Шварца (для нее значение критерия равно 7.341026).

Следующая таблица расшифровывает процесс получения оценки ранга коинтеграции в ситуации, соответствующей второму столбцу.

Test assumption: No deterministic trend in the data

Series: Y2 X2

Lags interval: 1 to 1

	Likelihood	5 Percent	1 Percent	Hypothesized
Eigenvalue	Ratio	Critical Value	Critical Value	No. of CE(s)

0.372251	187.3130	19.96	24.60	None **
0.005008	1.998159	9.24	12.97	At most 1

*(**) denotes rejection of the hypothesis at 5%(1%) significance level

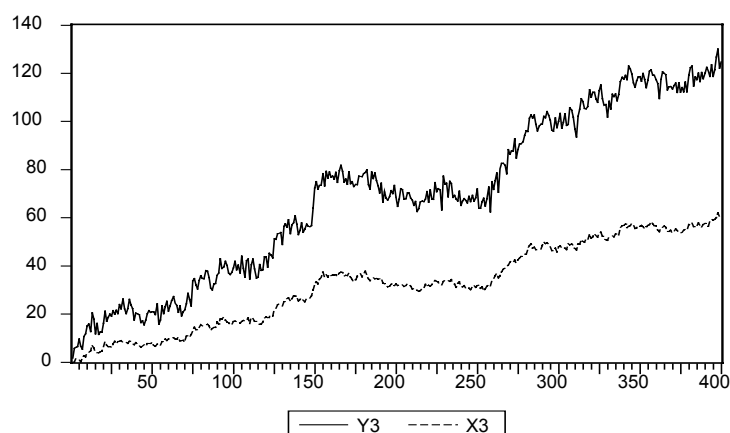
L.R. test indicates 1 cointegrating equation(s) at 5% significance level

Наблюдаемое значение 187.310 статистики $\lambda_{\text{trace}}(0)$, используемой при проверке гипотезы $H_0: r = 0$ против альтернативы $H_A: r > 0$, значительно превосходит оба критических значения. Гипотеза $H_0: r = 0$ отвергается в пользу альтернативы $H_A: r > 0$. Наблюдаемое значение 1.998159 статистики $\lambda_{\text{trace}}(1)$, используемой при проверке гипотезы $H_0: r = 1$ против альтернативы $H_A: r > 1$, намного меньше критических значений. Гипотеза $H_0: r = 1$ не отвергается. Как итог оцененное значение ранга коинтеграции принимается равным 1, что соответствует истинному положению вещей.

$$\text{DGP}_3 : y_t = 5 + 2x_t + \varepsilon_t,$$

$$x_t = 0.2 + x_{t-1} + v_t.$$

Смоделированная реализация имеет вид



Сводка статистик для определения ранга коинтеграции по этим смоделированным данным:

Sample: 1 400

Included observations: 398

Series: Y3 X3

Lags interval: 1 to 1

Data Trend:	None	None	Linear	Linear	Quadratic
Rank or	No Intercept	Intercept	Intercept	Intercept	Intercept
No. of CEs	No Trend	No Trend	No Trend	Trend	Trend
	Log Likelihood				
0	-1531.268	-1531.268	-1526.015	-1526.015	-1525.928
1	-1506.877	-1438.576	-1433.325	-1433.324	-1433.242
2	-1504.370	-1432.880	-1432.880	-1430.264	-1430.264
	AIC				
0	7.714915	7.714915	7.698565	7.698565	7.708180
1	7.612447	7.274253	7.252889	7.257911	7.262521
2	7.619950	7.270753	7.270753	7.267658	7.267658
	Schwarz Criteria				

0	7.754980	7.754980	7.758663	7.758663	7.788309
1	7.692577	7.364399	7.353051	7.368089	7.382716
2	7.740144	7.410980	7.410980	7.427918	7.427918
L.R. Test:	Rank = 2	Rank = 2	Rank = 1	Rank = 1	Rank = 2

Ориентируясь на критерий Акаике, наилучшей следует признать модель $H_1(1)$ (тренд в данных, в СЕ включается константа; ранг коинтеграции равен 1) – для нее значение критерия минимально (равно 7.252889). Та же модель выбирается и критерием Шварца (для нее значение критерия равно 7.353051).

Расшифровка процесса получения оценки ранга коинтеграции в ситуации, соответствующей третьему столбцу:

Sample: 1 400

Included observations: 398

Test assumption: Linear deterministic trend in the data

Series: Y3 X3

Lags interval: 1 to 1

	Likelihood	5 Percent	1 Percent	Hypothesized
Eigenvalue	Ratio	Critical Value	Critical Value	No. of CE(s)
0.372353	186.2692	15.41	20.04	None **
0.002234	0.890114	3.76	6.65	At most 1

*(**) denotes rejection of the hypothesis at 5%(1%) significance level

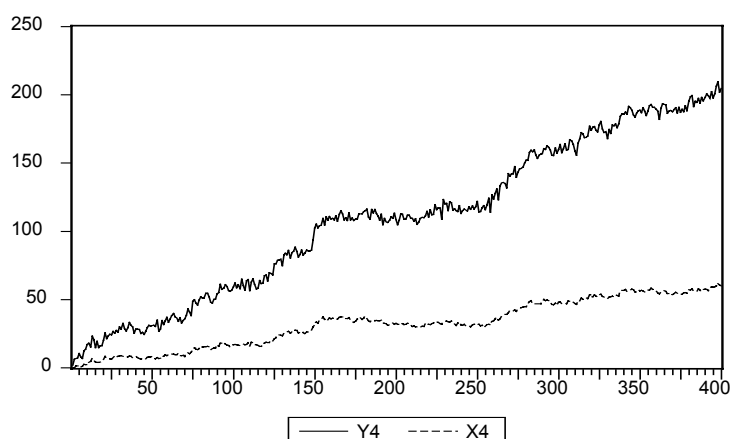
L.R. test indicates 1 cointegrating equation(s) at 5% significance level

Гипотеза $H_0: r = 0$ отвергается в пользу альтернативы $H_A: r > 0$. Гипотеза $H_0: r = 1$ не отвергается в пользу альтернативы $H_A: r > 1$. Как итог оцененное значение ранга коинтеграции принимается равным 1, что соответствует истинному положению вещей.

$$\text{DGP}_4 : y_t = 5 + 0.2 t + 2 x_t + \varepsilon_t,$$

$$x_t = 0.2 + x_{t-1} + v_t.$$

Смоделированная реализация имеет вид



Сводка статистик для определения ранга коинтеграции по этим смоделированным данным:

Sample: 1 400

Included observations: 398

Series: Y4 X4

Lags interval: 1 to 1

Data Trend:	None	None	Linear	Linear	Quadratic
Rank or	No Intercept	Intercept	Intercept	Intercept	Intercept
No. of CEs	No Trend	No Trend	No Trend	Trend	Trend
Log Likelihood					
0	-1533.049	-1533.049	-1526.015	-1526.015	-1525.928
1	-1525.311	-1525.279	-1518.361	-1433.324	-1433.242
2	-1521.201	-1518.005	-1518.005	-1430.264	-1430.264
AIC					
0	7.723863	7.723863	7.698565	7.698565	7.708180
1	7.705079	7.709944	7.680208	7.257911	7.262521
2	7.704525	7.698520	7.698520	7.267658	7.267658
Schwarz Criteria					
0	7.763928	7.763928	7.758663	7.758663	7.788309
1	7.785208	7.800090	7.780370	7.368089	7.382716
2	7.824720	7.838747	7.838747	7.427918	7.427918
L.R. Test:	Rank = 2	Rank = 2	Rank = 1	Rank = 1	Rank = 2

Ориентируясь на критерий Акаике, наилучшей следует признать модель $H^*(1)$ (тренд в данных, в СЕ включаются константа и линейный тренд; ранг коинтеграции равен 1) – для нее значение критерия минимально (равно 7.257911). Та же модель выбирается и критерием Шварца (для нее значение критерия равно 7.368089).

Расшифровка процесса получения оценки ранга коинтеграции в ситуации, соответствующей четвертому столбцу:

Test assumption: Linear deterministic trend in the data

Series: Y4 X4

Lags interval: 1 to 1

	Likelihood	5 Percent	1 Percent	Hypothesized
Eigenvalue	Ratio	Critical Value	Critical Value	No. of CE(s)
0.372355	191.5010	25.32	30.45	None **
0.015260	6.120353	12.25	16.26	At most 1

*(**) denotes rejection of the hypothesis at 5%(1%) significance level

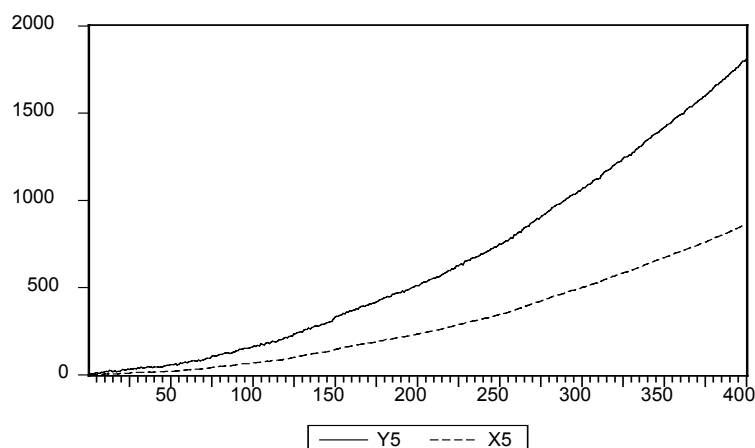
L.R. test indicates 1 cointegrating equation(s) at 5% significance level

Гипотеза $H_0: r = 0$ отвергается в пользу альтернативы $H_A: r > 0$. Гипотеза $H_0: r = 1$ не отвергается в пользу альтернативы $H_A: r > 1$. Оцененное значение ранга коинтеграции принимается равным 1, что соответствует истинному положению вещей.

$$DGP_5 : y_t = 5 + 0.2 t + 2 x_t + \varepsilon_t,$$

$$x_t = 0.2 + 0.01 t + x_{t-1} + v_t.$$

Смоделированная реализация имеет вид



Сводка статистик для определения ранга коинтеграции по этим смоделированным данным:

Series: Y5 X5

Lags interval: 1 to 1

Data Trend:	None	None	Linear	Linear	Quadratic
Rank or	No Intercept	Intercept	Intercept	Intercept	Intercept
No. of CEs	No Trend	No Trend	No Trend	Trend	Trend
Log Likelihood					
0	-1672.222	-1672.222	-1630.634	-1630.634	-1525.928
1	-1527.340	-1527.331	-1525.738	-1525.724	-1432.667
2	-1527.280	-1520.795	-1520.795	-1432.659	-1432.659
AIC					
0	8.423224	8.423224	8.224289	8.224289	7.708180
1	7.715279	7.720258	7.717274	7.722231	7.259633
2	7.735077	7.712538	7.712538	7.279694	7.279694
Schwarz Criteria					
0	8.463289	8.463289	8.284386	8.284386	7.788309
1	7.795408	7.810404	7.817436	7.832409	7.379827
2	7.855272	7.852765	7.852765	7.439953	7.439953
L.R. Test:	Rank = 1	Rank = 2	Rank = 2	Rank = 2	Rank = 1

Ориентируясь на критерий Акаике, наилучшей следует признать модель $H(1)$ (квадратичный тренд в данных, в СЕ включаются константа и линейный тренд; ранг коинтеграции равен 1) – для нее значение критерия минимально (равно 7.259633). Та же модель выбирается и критерием Шварца (для нее значение критерия равно 7.379827).

Расшифровка процесса получения оценки ранга коинтеграции в ситуации, соответствующей пятому столбцу:

Test assumption: Quadratic deterministic trend in the data

Series: Y5 X5

Lags interval: 1 to 1

Likelihood	5 Percent	1 Percent	Hypothesized
------------	-----------	-----------	--------------

Eigenvalue	Ratio	Critical Value	Critical Value	No. of CE(s)
0.374152	186.5374	18.17	23.46	None **
3.97E-05	0.015819	3.74	6.40	At most 1

*(**) denotes rejection of the hypothesis at 5%(1%) significance level

L.R. test indicates 1 cointegrating equation(s) at 5% significance level

Гипотеза $H_0: r = 0$ отвергается в пользу альтернативы $H_A: r > 0$. Гипотеза $H_0: r = 1$ не отвергается в пользу альтернативы $H_A: r > 1$. Оцененное значение ранга коинтеграции принимается равным 1, что соответствует истинному положению вещей.

Следующий пример, в котором процесс порождения данных представляет треугольную систему, иллюстрирует процедуру определения ранга коинтеграции при большем количестве рядов.

Пример

Рассмотрим процесс порождения данных, образующий треугольную систему Филлипса.

$$\text{DGP: } L234_t = 0.5W_{2t} + W_{3t} + 2W_{4t} + \varepsilon_{1t},$$

$$L23_t = W_{2t} + 0.5W_{3t} + \varepsilon_{2t},$$

$$W_{2t} = W_{2,t-1} + \varepsilon_{3t},$$

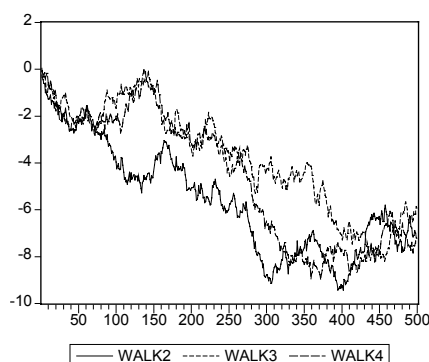
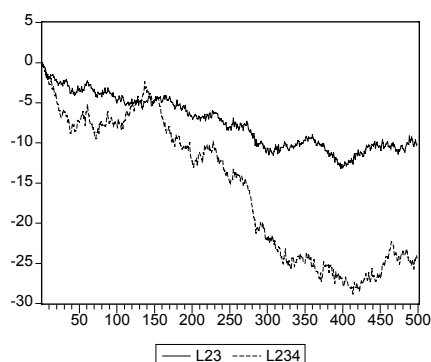
$$W_{3t} = W_{3,t-1} + \varepsilon_{4t},$$

$$W_{4t} = W_{4,t-1} + \varepsilon_{5t},$$

где $\varepsilon_{1t}, \varepsilon_{2t}, \varepsilon_{3t}, \varepsilon_{4t}, \varepsilon_{5t}$ — независимые между собой гауссовские процессы белого шума, имеющие нулевые математические ожидания и дисперсии, равные 0.04.

Ряды W_{2t}, W_{3t}, W_{4t} являются случайными блужданиями и интерпретируются в рамках этой системы как **общие тренды ("common trends")**, в том смысле, что вся (стохастическая) нестационарность системы управляется этими тремя рядами. Нестационарное поведение ряда $L23_t$ регулируется рядами W_{2t} и W_{3t} , а нестационарное поведение ряда $L234_t$ регулируется рядами W_{2t}, W_{3t} и W_{4t} .

Смоделированные данные содержат 501 значение для каждого из 5 входящих в DGP рядов $L234_t, L23_t, W_{2t}, W_{3t}$ и W_{4t} . Следующие графики показывают поведение смоделированных реализаций.



Оценивая по смоделированной реализации 5 рядов модель VAR(1) в уровнях (без ограничений на ранг коинтеграции) и анализируя коррелограммы полученных рядов остатков, мы не обнаруживаем автокоррелированности у всех 5 рядов остатков. Например, коррелограмма ряда остатков от ряда $L234_t$ имеет вид

Included observations: 498

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
-----------------	---------------------	----	-----	--------	------

. .	. .	1	-0.044	-0.044	0.9530	0.329
. .	. .	2	-0.037	-0.039	1.6395	0.441
. .	. .	3	0.031	0.028	2.1276	0.546
. .	. .	4	-0.020	-0.019	2.3355	0.674
. .	. .	5	0.031	0.031	2.8073	0.730
. .	. .	6	0.036	0.037	3.4726	0.748
. .	. .	7	0.008	0.015	3.5055	0.835
. .	. .	8	0.039	0.041	4.2674	0.832
. .	. .	9	0.003	0.007	4.2734	0.893
. .	. .	10	0.001	0.005	4.2743	0.934

Поэтому мы можем остановиться на статистической модели в виде VAR(1) для уровней. Соответствующая ей модель коррекции ошибок не содержит в правых частях уравнений запаздывающих разностей, и это следует учитывать при оценивании ранга коинтеграции.

Сводка статистик для определения ранга коинтеграции по смоделированным данным:

Series: L234 L23 WALK2 WALK3 WALK4

Lags interval: No lags

Data Trend:	None	None	Linear	Linear	Quadratic
Rank or	No Intercept	Intercept	Intercept	Intercept	Intercept
No. of CEs	No Trend	No Trend	No Trend	Trend	Trend
Log Likelihood					
0	99.73833	99.73833	103.2311	103.2311	104.7734
1	279.2178	279.4377	282.9301	283.2575	284.7997
2	457.0419	457.6821	461.1675	462.6611	464.2006
3	461.1754	463.9622	466.3818	469.2587	470.4661
4	464.5438	468.0578	469.2034	472.1756	473.0952
5	464.7799	470.6760	470.6760	473.8191	473.8191
Akaike					
0	-0.400556	-0.400556	-0.394502	-0.394502	-0.380616
1	-1.081196	-1.078063	-1.076025	-1.073323	-1.063453
2	-1.755188	-1.749727	-1.751677	-1.749643	-1.743777
3	-1.731628	-1.730772	-1.732457	-1.731963	-1.728780
4	-1.704995	-1.703044	-1.703628	-1.699501	-1.699178
5	-1.665783	-1.669382	-1.669382	-1.661924	-1.661924
Schwarz					
0	-0.400556	-0.400556	-0.352227	-0.352227	-0.296066
1	-0.996646	-0.985058	-0.949199	-0.938043	-0.894352
2	-1.586088	-1.563717	-1.540301	-1.521357	-1.490127
3	-1.477977	-1.451756	-1.436531	-1.410672	-1.390579
4	-1.366794	-1.331023	-1.323152	-1.285205	-1.276427
5	-1.243032	-1.204355	-1.204355	-1.154623	-1.154623
L.R. Test:	Rank = 2	Rank = 2	Rank = 2	Rank = 2	Rank = 2

В рамках каждого столбца цепочки критериев выводят на ранг 2. Оба информационных критерия (Акаике и Шварца) указывают на вариант “нет тренда в данных, в

коинтеграционное соотношение не включаются ни константа ни тренд; ранг коинтеграции равен 2”.

Приведем теперь сводку статистик для определения ранга коинтеграции тройки рядов W_{2t} , W_{3t} и W_{4t} .

Series: WALK2 WALK3 WALK4

Lags interval: No lags

Data Trend:	None	None	Linear	Linear	Quadratic
Rank or	No Intercept	Intercept	Intercept	Intercept	Intercept
No. of CEs	No Trend	No Trend	No Trend	Trend	Trend

Akaike					
0	-1.209358	-1.209358	-1.211316	-1.211316	-1.205445
1	-1.201837	-1.206417	-1.208105	-1.209616	-1.206429
2	-1.191311	-1.194747	-1.195348	-1.193234	-1.192934
3	-1.168162	-1.177195	-1.177195	-1.171753	-1.171753

Schwarz					
0	-1.209358	-1.209358	-1.185951	-1.185951	-1.154715
1	-1.151107	-1.147232	-1.132009	-1.125065	-1.104968
2	-1.089850	-1.076376	-1.068523	-1.049499	-1.040743
3	-1.015971	-0.999639	-0.999639	-0.968832	-0.968832

L.R. Test:	Rank = 0	Rank = 0	Rank = 0	Rank = 0	Rank = 0
------------	----------	----------	----------	----------	----------

Здесь в рамках каждого столбца цепочки критериев выводят на ранг 0, что соответствует DGP. Критерий Акаике указывает на варианты с трендом в данных, тогда как критерий Шварца останавливается на вариантах без тренда в данных, что и соответствует DGP.

8.2. Оценивание модели коррекции ошибок

После оценивания ранга коинтеграции в рамках процедуры Йохансена имеется возможность получения (при выбранном ранге коинтеграции r) оценок максимального правдоподобия для r линейно независимых коинтегрирующих векторов. Реализация такого оценивания в пакете EVIEWS для группы из 5 рядов, рассмотренной в предыдущем примере ($r = 2$), дает следующие результаты.

Test assumption: No deterministic trend in the data

Series: L234 L23 WALK2 WALK3 WALK4

Lags interval: No lags

Unnormalized Cointegrating Coefficients:					
L234	L23	WALK2	WALK3	WALK4	
-0.079261	-0.198108	0.236127	0.178603	0.159704	
-0.202709	0.079211	0.022787	0.161363	0.406370	
0.001194	0.000453	-0.014625	-0.034465	0.037834	
-0.002101	0.001543	0.019423	-0.024621	0.007077	
-0.000206	0.000771	0.011197	-0.009764	-0.012244	

Каждая строка этой таблицы содержит компоненты одного из возможных коинтегрирующих векторов. Всего, таким образом, предлагается к рассмотрению 5 вариантов коинтегрирующих векторов, причем эти пять векторов являются линейно

независимыми. Первая из 5 строк содержит коэффициенты линейной комбинации указанных рядов, “наиболее похожей на стационарную”. Вторая строка соответствует линейной комбинации, занимающей в этом отношении второе место, и т.д.

Если бы мы оценили ранг коинтеграции как $r = 1$, то тогда в качестве оценки коинтегрирующего вектора можно было взять вектор с компонентами, приведенными в первой строке, т.е. вектор

$$(-0.079261, -0.198108, 0.236127, 0.178603, 0.159704)^T,$$

или любой пропорциональный ему вектор. Выбирая из этого множества вектор, нормализованный на первую компоненту, т.е. вектор, полученный из указанного делением всех его компонент на первую компоненту, мы получили бы вектор

$$(1, 2.499451, -2.979119, -2.25363, -2.014916)^T.$$

Поскольку ранг коинтеграции был оценен как $r = 2$, то в качестве оценок двух линейно независимых коинтегрирующих векторов можно взять векторы с компонентами, приведенными в первых двух строках, т.е. векторы

$$\beta_{(1)}^* = (-0.079261, -0.198108, 0.236127, 0.178603, 0.159704)^T$$

и

$$\beta_{(2)}^* = (-0.202709, 0.079211, 0.022787, 0.161363, 0.406370)^T.$$

Дело, однако, в том, что помимо этих двух векторов в качестве коинтегрирующих векторов с тем же успехом могут выступать и любые линейные комбинации этих двух векторов. И в реальных экономических задачах важно, чтобы выбранная в итоге из этого множества пара векторов выражала осмысленные с экономической точки зрения (экономической теории) долговременные связи между рассматриваемыми переменными (например, паритет покупательной способности, спрос на деньги и т.п.). Это, в свою очередь, требует наложения на коинтегрирующие векторы соответствующих идентифицирующих ограничений, позволяющих различать эти векторы, выделяя их из всего множества линейных комбинаций базисных векторов.

Если ранг коинтеграции равен $r > 1$, то для различения коинтегрирующих векторов достаточно наложить на каждый из коинтегрирующих векторов $q = r - 1$ линейных ограничений (причем эти линейные ограничения сами должны быть линейно независимыми – иначе различения не получится). Это дает возможность определить каждый из коинтегрирующих векторов с точностью до коэффициента пропорциональности, а затем получить единственный набор коинтегрирующих векторов, нормируя компоненты каждого вектора на какую-либо из его (ненулевых) компонент.

В нашем примере $r = 2$, так что на каждый из двух коинтегрирующих векторов достаточно наложить по одному линейному ограничению, например, приравнять одну из компонент коинтегрирующего вектора нулю. При этом зануляемые компоненты в двух векторах должны быть различными. Выбор зануляемых компонент на практике определяется, исходя из представлений той или иной экономической теории.

Имитируя такой выбор, мы будем исходить из наличия информации о том, что “в соответствии с некоторой экономической теорией”, между переменными $L234_t$, $L23_t$, W_{2t} , W_{3t} и W_{4t} должны существовать две долговременные связи, одна из которых связывает переменные $L234_t$, W_{2t} , W_{3t} , W_{4t} и не включает переменную $L23_t$, а другая связывает переменные $L23_t$, W_{2t} , W_{3t} , W_{4t} и не включает переменную $L234_t$. Если при этом из той же “экономической теории” следует также, что в первой долговременной связи “объясняемой” переменной является $L234_t$, а во второй – переменная $L23_t$, то, нормируя первый коинтегрирующий вектор на первую компоненту, а второй коинтегрирующий вектор – на вторую компоненту, мы представляем эти коинтегрирующие векторы в виде

$$\beta_{(1)} = (1, 0, \beta_{13}, \beta_{14}, \beta_{15})^T$$

$$\beta_{(2)} = (0, 1, \beta_{23}, \beta_{24}, \beta_{25})^T.$$

Таким образом, после получения двух произвольных линейно независимых оценок двух коинтегрирующих векторов,

$$(-0.079261, -0.198108, 0.236127, 0.178603, 0.159704)^T,$$

$$(-0.202709, 0.079211, 0.022787, 0.161363, 0.406370)^T,$$

задача состоит в отыскании линейных комбинаций этих оцененных векторов, имеющих вид

$$(1, 0, \beta_{13}, \beta_{14}, \beta_{15})^T \text{ и } (0, 1, \beta_{23}, \beta_{24}, \beta_{25})^T.$$

Решение этой задачи в пакете EVIEWS приводит к следующему результату:

Normalized Cointegrating Coefficients: 2 Cointegrating Equation(s)

L234	L23	WALK2	WALK3	WALK4
1.000000	0.000000	-0.499995	-0.993065	-2.006077
		(0.00541)	(0.00868)	(0.00852)
0.000000	1.000000	-0.991867	-0.504230	-0.003537
		(0.00549)	(0.00881)	(0.00865)

В соответствии с этой таблицей,

$$\beta_{(1)} = (1, 0, -0.499995, -0.993065, -2.006077)^T$$

$$\beta_{(2)} = (0, 1, -0.991867, -0.504230, -0.003537)^T.$$

Это соответствует двум долговременным соотношениям

$$L234_t = 0.499995W_{2t} + 0.993065W_{3t} + 2.006077W_{4t},$$

$$L23_t = 0.991867W_{2t} + 0.504230W_{3t} + 0.003537W_{4t},$$

которые близки к теоретическим долговременным соотношениям, определяемым использованным DGP, а именно,

$$L234_t = 0.5W_{2t} + W_{3t} + 2W_{4t},$$

$$L23_t = W_{2t} + 0.5W_{3t}.$$

Нетрудно заметить, что оцененные векторы $\beta_{(1)}^*$ и $\beta_{(2)}^*$ являются линейными комбинациями векторов $\beta_{(1)}$ и $\beta_{(2)}$:

$$\beta_{(1)}^* = -0.079261\beta_{(1)} - 0.198108\beta_{(2)},$$

$$\beta_{(2)}^* = -0.202709\beta_{(1)} + 0.079211\beta_{(2)}.$$

Приведенные в таблице под оценками коэффициентов оцененные значения их стандартных ошибок дают некоторую ориентацию в отношении того, какими в действительности могут быть компоненты “истинных” коинтегрирующих векторов. Следует только учитывать, что оценки компонент коинтегрирующих векторов не являются нормально распределенными (даже асимптотически).

После получения оценок подходящих коинтегрирующих векторов можно приступить к оцениванию коэффициентов ЕСМ обычными методами. Только и здесь следует учитывать, что асимптотически нормальными являются лишь оценки кратковременной динамики, т.е. коэффициенты при запаздывающих разностях переменных.

В нашем примере получаем:

	D(L234)	D(L23)	D(WALK2)	D(WALK3)	D(WALK4)
ecm1	-0.971220	-0.051197	-0.040351	0.005400	0.031996
	(0.11006)	(0.06585)	(0.04248)	(0.04468)	(0.04231)
	(-8.82480)	(-0.77748)	(-0.94983)	(0.12086)	(0.75621)

<i>ecm2</i>	-0.033474	-1.015363	-0.001894	0.010505	-0.024292
	(0.10788)	(0.06455)	(0.04164)	(0.04380)	(0.04148)
	(-0.31028)	(-15.7298)	(-0.04549)	(0.23985)	(-0.58571)

Согласно этой таблице оцененная ЕСМ имеет вид

$$\Delta(L234)_t = -0.971220 (ecm1)_{t-1} - 0.033474 (ecm2)_{t-1} + e_{1t},$$

$$\Delta(L23)_t = -0.051197 (ecm1)_{t-1} - 1.015363 (ecm2)_{t-1} + e_{2t},$$

$$\Delta(W2)_t = -0.040351 (ecm1)_{t-1} - 0.001894 (ecm2)_{t-1} + e_{3t},$$

$$\Delta(W3)_t = 0.005400 (ecm1)_{t-1} + 0.010505 (ecm2)_{t-1} + e_{4t},$$

$$\Delta(W4)_t = 0.031996 (ecm1)_{t-1} - 0.024292 (ecm2)_{t-1} + e_{5t},$$

где

$$(ecm1)_t = L234_t - 0.499995 W_{2t} - 0.993065 W_{3t} - 2.006077 W_{4t},$$

$$(ecm2)_t = L23_t - 0.991867 W_{2t} - 0.504230 W_{3t} - 0.003537 W_{4t}.$$

В рамках процедуры Йохансена имеется также возможность проверки гипотез о выполнении дополнительных (“сверхидентифицирующих”) ограничений на коинтегрирующие векторы, например, гипотезы

$$H_0: \beta_{25} = 0$$

о занулении последней компоненты второго коинтегрирующего вектора, или гипотезы

$$H_0: \alpha_{12} = \alpha_{21} = \alpha_{31} = \alpha_{32} = \alpha_{41} = \alpha_{42} = \alpha_{51} = \alpha_{52} = 0,$$

означающей отсутствие составляющей $(ecm1)_{t-1}$ во всех уравнениях кроме первого и отсутствие составляющей $(ecm2)_{t-1}$ во всех уравнениях кроме второго (что и соответствует использованному процессу порождения данных).

Однако процедура проверки выполнения таких ограничений на коинтегрирующие векторы, как и процедура проверки выполнения тех или иных линейных ограничений на элементы матрицы α корректирующих коэффициентов, не встроена в пакет EVIEWS и требует привлечения для ее реализации других специализированных пакетов статистического анализа данных. (Напомним, что распределения оценок коэффициентов коинтегрирующих векторов и элементов матрицы α не являются асимптотически нормальными.) Проведение детального коинтеграционного анализа нестационарных рядов возможно с использованием макропакета CATS (*Cointegration Analysis of Time Series*), оформленного в виде процедуры для пакета RATS (*Regression Analysis of Time Series*). Краткое описание соответствующих процедур с подробными примерами анализа экономических данных можно найти, например, в книге [Patterson (2000)].

Завершая описание процедуры Йохансена, следует обратить особое внимание на следующие обстоятельства.

- Процедура Йохансена исходит из предположения о гауссовости процесса белого шума в VAR модели.
- Процедура Йохансена чувствительна к выбору порядка p модели VAR.
- Используемые критические значения статистик λ_{\max} и λ_{trace} — асимптотические, так что при малом количестве наблюдений к полученным выводам следует относиться достаточно осторожно.

В связи с последним обстоятельством, при работе с умеренным количеством наблюдений рекомендуется корректировать наблюдаемые значения статистик λ_{\max} и λ_{trace} , умножая их на $(T - Np)/T$ (“коррекция на число степеней свободы”).

Все эти замечания означают, что при коинтеграционном анализе реальных экономических (а не смоделированных) данных интерпретация полученных результатов может оказаться довольно затруднительной.

Заключение

Как было отмечено во Введении, данное учебное пособие является лишь кратким введением в современные методы эконометрического анализа статистических данных, представленных в виде временных рядов. Естественно, что при этом за рамками пособия осталось достаточное количество вопросов, имеющих важное значение.

Что касается собственно временных рядов, то в пособии не были затронуты проблемы, связанные с прогнозированием и спектральным анализом временных рядов, моделями дробно-интегрированных рядов (ARFIMA) и рядов с авторегрессионной условной гетероскедастичностью (ARCH, GARCH и т.п.), моделями с переключениями, широко используемыми при анализе высокочастотных финансовых временных рядов.

В отношении моделей связи между несколькими интегрированными рядами читателю полезно ознакомиться с проблемой построения структурных моделей коррекции ошибок (структурных ЕСМ) и связанной с ней проблемой более точного определения понятия экзогенности (слабая экзогенность, строгая экзогенность). Можно упомянуть также обобщение процедуры Йохансена на системы, включающие $I(2)$ переменные, сезонную коинтеграцию, построение моделей связи при наличии структурных изменений, байесовский подход к анализу связей между временными рядами.

Подготовленный читатель может обратиться по этим вопросам, например, к уже цитировавшимся книгам [Patterson (2000)], [Maddala, Kim (1998)], [Hamilton (1994)], [Mills (1993)], к работам [Гренджер и Хатанака (1972)] и [Favero (2001)], а также к монографиям [Clements, Hendry (1998)] и [Clements, Hendry (1999)].

Список литературы

1. **Бокс Дж., Дженкинс Г. (1974)** *Анализ временных рядов. Прогноз и управление.* Вып. 1, 2. – М., Мир.
2. **Гренджер К., М. Хатанака (1972)** *Спектральный анализ временных рядов в экономике.* (пер. с англ.) – М., Статистика.
3. **Носко В.П. (2000)** *Эконометрика для начинающих.* М., ИЭПП.
4. **Хеннан Э. (1974)** *Многомерные временные ряды.* (пер. с англ.) – М., Мир.
5. **Эконометрический анализ динамических рядов основных макроэкономических показателей (2001).** Научные труды ИЭПП №34Р. М., ИЭПП.
6. **Akaike (1973)** “Information Theory and an Extension of the Maximum Likelihood Principle”, in Petrov B.N. and Csáki F. (Eds), *Proceedings, 2nd International Symposium on Information Theory*, 267-281. Akadémia Kiado, Budapest.
7. **Andrews D.W.K. (1991)** “Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation,” *Econometrica*, 59, 817–858.
8. **Ardeni P.G., D. Lubian (1991)** “Is There Trend Reversion in Purchasing Power Parity”, *European Economic Review*, 35, 1035-1055.
9. **Bartlett M.S. (1946)** “On the Theoretical Specification of sampling properties of Autocorrelated Time Series”, *Journal of the Royal Statistical Society, Series B*, 8, 27-41.
10. **Bierens H.J. (1997)** “Testing the Unit Root with Drift Hypothesis Against Nonlinear Trend Stationarity, with an Application to the US Price Level and Interest Rate”, *Journal of Econometrics*, 81, 29-64.
11. **Box G.E.P., Pierce D.A. (1970)** “Distribution of Residual Autocorrelations in Autoregressive Integrated Moving Average Time Series Models”, *Journal of the American Statistical Association*, 65, 1509-1526.
12. **Chan K.H, J.C.Hayya, J.K.Ord (1977)** “A Note on Trend Removal Methods: The Case of polynomial versus variate differencing”, *Econometrica*, 45, 737-744.
13. **Cheung Y.-W., M.D. Chinn (1996)** “Deterministic, Stochastic, and Segmented Trends in Aggregate Output: a Cross-country Analysis”, *Oxford Economic Papers*, 48, №1, 134-162.
14. **Cheung Y.-W., K.S. Lay (1995)** “Lag Order and Critical Values of a Modified Dickey-Fuller Test”, *Oxford Bulletin of Economics and Statistics*, 57, №3, 411-419.
15. **Christiano L.J., M. Eichenbaum (1990)** “Unit Roots in Real GDP: Do We Know, and Do We Care?”, *Carnegie-Rochester Conference Series on Public Policy*, 32, 7-62.
16. **Clark P.K. (1989)** “Trend Reversion in Real Output and Unemployment”, *Journal of Econometrics*, 40, 15-32.

17. **Clements M. P., D. F. Hendry (1998)** *Forecasting Economic Time series*, Cambridge University Press, Cambridge.
18. **Clements M. P., D. F. Hendry (1999)** *Forecasting Non-stationary Economic Time series*, MIT Press, Cambridge, Massachusetts, London.
19. **Cochrane J.H. (1998)** "How Big is the Random Walk in GNP?", *Journal of Political Economy*, 96, 893-920.
20. **Cogley T. (1990)** "International Evidence on the Size of the Random Walk in Output", *Journal of Political Economy*, 98, 501-518.
21. **Copeland L.S. (1991)** "Cointegration Tests with Daily Exchange Rate Data", *Oxford Bulletin of Economics and Statistics*, 53, 185-198.
22. **Davidson R., J.G. MacKinnon (1993)** *Estimation and Inference in Econometrics*, Oxford University Press
23. **Dickey D.A. (1976)** "Estimation and Hypothesis Testing for Nonstationary Time Series", Ph.D. dissertation, Iowa State University.
24. **Dickey D.A., W.R.Bell, R.B. Miller (1986)** "Unit Roots in Time Series Models: Tests and Implications", *American Statistician*, 40, 12-26.
25. **Dickey D.A. , W.A. Fuller (1979)** "Distribution of the Estimators for Autoregressive Time Series with a Unit Root", *Journal of the American Statistical Association*, 74, 427-431.
26. **Dickey, D.A., W.A. Fuller (1981)** "Likelihood Ratio Statistics for Autoregressive Time Series With a Unit Root", *Econometrica*, 49, 1057-1072.
27. **Dickey D.A., S. Pantula (1987)** "Determining the Order of Differencing in Autoregressive Processes", *Journal of Business and Economic Statistics*, 15, 455-461.
28. **Dolado H., T. Jenkinson, S. Sosvilla-Rivero (1990)** "Cointegration and Unit Roots", *Journal of Economic Surveys*, 4, 243-273.
29. **Dutt S.D. (1998)** "Purchasing Power Parity Revisited: Null of Cointegration Approach", *Applied Economic Letters*, 5, 573-576.
30. **Dutt S.D., D. Ghosh (1999)** "An Empirical Examination of Exchange Market Efficiency", *Applied Economic Letters*, 6, №2, 89-91.
31. **Dweyer G.P., Wallace M.S. (1992)** "Cointegration and Market Efficiency", *Journal of International Money and Finance*, 11 318-327.
32. **Elliott G., T.J. Rothenberg, J.H. Stock (1996)** "Efficient Tests for an Autoregressive Unit Root", *Econometrica*, 64, 813-836.
33. **Enders W. (1995)** *Applied Econometric Time Series*, Wiley, New York

34. **Engle R.F., C.W.J. Granger (1987)** "Co-integration and Error Correction: Representation, Estimation, and Testing", *Econometrica*, 55, 251-276.
35. **Engle R.F., C.W.J. Granger (1991)** "Cointegrated Economic Time Series: An Overview with New Results", in R.F. Engle and C.W.J. Granger (eds.), *Long-Run Economic Relationships, Readings in Cointegration*, Oxford University Press, 237-266.
36. **Entorf H. (1992)** "Random Walk with Drift, Simultaneous Errors, and Small Samples: Simulating the Bird's Eye View", Institut National de la Statistique et des Etudes Economiques.
37. **Fama E.F., French K.R. (1988)** "Permanent and Temporary Components of Stock Prices", *Journal of Political Economy*, 96, 246-273.
38. **Favero C. A. (2001)** *Applied Macroeconometrics*, Oxford University Press Inc., New-York.
39. **Fuller W.A. (1976)** *Introduction to Statistical Time Series*, Wiley, New York.
40. **Fuller W.A. (1996)** *Introduction to Statistical Time Series*, 2nd ed, Wiley, New York
41. **Funke N., J. Thornton (1999)** "The Demand for Money in Italy, 1861-1988", *Applied Economic Letters*, 6, №5, 299-301.
42. **Ghysels E., Perron P. (1992)** "The Effect of Seasonal Adjustment Filters on Tests for a Unit Root", *Journal of Econometrics*, 55, 57-98.
43. **Granger C.W.J. (1983)** UCSD Discussion Paper, 83-13a.
44. **Green W.H. (1997)** "Econometric Analysis". 3rd edition, Prentice-Hall.
45. **den Haan W.J. (2000)** "The Comovement Between Output and Prices", *Journal of Monetary Economics*, 46, №1, 3-30.
46. **Hafer R.W., D.W. Jansen (1991)** "The Demand for Money in the United States: Evidence from Cointegration Tests", *Journal of Money, Credit, and Banking*, 23 (1991), 155-168.
47. **Hall A. (1994)** "Testing for a Unit Root in Time Series with Pretest Data-Based Model Selection", *Journal of Business and Economic Statistics*, 12, 451-470.
48. **Hamilton, James D. (1994)** *Time Series Analysis*, Princeton University Press, Prinseton.
49. **Hannan E.J., Quinn B.G. (1979)** "The Determination of the Order of an Autoregression", *Journal of the Royal Statistical Society, Series B*, 41, 190-195.
50. **Hasan M.S. (1998)** "The Choice of Appropriate Monetary Aggregate in the United Kindom", *Applied Economic Letters*, 5, №9, 563-568.
51. **Hatanaka M. (1996)** *Time Series-Based Econometrics: Unit Roots and Cointegration*, Oxford University Press.

52. **Holden D., Perman R. (1994)** "Unit Roots and Cointegration for Economists", в сборнике *Cointegration for the Applied Economists* (редактор Rao B.B.), Macmillan.
53. **Jarque C., A. Bera (1980)** "Efficient Tests for Normality, Homoskedasticity, and Serial Independence of Regression Residuals," *Economics Letters*, 6, 255–259.
54. **Johansen S. (1988)** "Statistical Analysis of Cointegration Vectors", *Journal of Economic Dynamics and Control*, 12, 231-254.
55. **Johansen S. (1991)** "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models", *Econometrica*, 59, 1551-1580.
56. **Johansen S. (1992)** "Determination of Cointegration Rank in the Presence of a Linear Trend", *Oxford Bulletin of Economics and Statistics*, 54, 383-397.
57. **Johansen S. (1994a)** "The Role of the Constant Term in Cointegration Analysis of Nonstationary Variables", *Econometric Reviews*, 13, 205-219.
58. **Johansen S. (1994b)** "A Likelihood Analysis of the I(2) model", *Scandinavian Journal of Statistics* **Johansen S. (1995a)** *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press, Oxford.
59. **Johansen S. (1995)** "A Statistical Analysis of Cointegration for I(2) Variables", *Econometric Theory*, 11, 25-29.
60. **Johansen S., K. Juselius (1990)** "Maximum Likelihood Estimation and Inferences on Cointegration—with applications to the demand for money," *Oxford Bulletin of Economics and Statistics*, 52, 169–210.
61. **Kavalieris (1991)** "A Note on Estimating Autoregressive-Moving average Order", *Biometrika*, 78, 920-922.
62. **Kim B. J.C., Soowon Mo (1995)** "Cointegration and the Long-run Forecast of Exchange Rates", *Economics Letters*, 48, №№ 3-4, 353-359.
63. **Kwan A.C.C. (1996a)** "On the Finite-sample Distribution of Modified Portmanteau Tests for Randomness of a Gaussian Time Series", *Biometrika*, 83, №4, 938-943.
64. **Kwan A.C.C. (1996b)** "A Comparative Study of the Finite-sample Distribution of some Portmanteau Tests for Univariate Time Series Models", *Commun. Statist.-Simula*, 25, №4, 867-904.
65. **Kwiatkowski D., P.C.B. Phillips, P. Schmidt, Y. Shin (1992)** "Testing of the Null Hypothesis of Stationary against the Alternative of a Unit Root", *Journal of Econometrics*, 54, 159-178.
66. **Kwiatkowski D., P. Schmidt (1990)** "Dickey–Fuller Tests with Trend", *Commun. Statist.-Theory Meth.*, 19, №10, 3645-3656.
67. **Leybourne S.J. (1995)** "Testing for Unit Roots Using Forward and Reverse Dickey-Fuller Regressions", *Oxford Bulletin of Economics and Statistics*, 57, 559-571.

68. **Leybourne S., T. Mills, P. Newbold (1998)** "Spurious Rejections by Dickey-Fuller Tests in the Presence of a Break Under Null", *Journal of Econometrics*, 87, 191-203.
69. **Ljung G., G.E.P. Box (1979)** "On a Measure of Lack of Fit in Time Series Models", *Biometrika*, 66, 255-270.
70. **Lomnicki Z.A. (1961)** "Tests for Departure from Normality in the Case of Linear Stochastic Processes", *Metrika*, 4, 37-62.
71. **Lumsdaine R.L., Kim I.M. (1997)** "Structural Change and Unit Roots", *The Review of Economics and Statistics*, 79, 212-218.
72. **MacKinnon, J.G. (1991)** "Critical Values for Cointegration Tests," Глава 13 в *Long-run Economic Relationships: Readings in Cointegration*, edited by R.F.Engle and C.W.J. Granger, Oxford University Press.
73. **Maddala G.S., In-Moo Kim (1998)** *Unit Roots, Cointegration, and Structural Change*. Cambridge University Press, Cambridge.
74. **Mann H.B., A. Wald (1943)** "On Stochastic Limit and Order Relationships", *Annals of Mathematical Statistics*, 14, 217-277.
75. **Metin K. (1995)** "An Integrated Analysis of Turkish Inflation", *Oxford Bulletin of Economics and Statistics*, 57, №4, 513-532.
76. **Milas C. (1998)** "Demand for Greek Imports Using Multivariate Cointegration Technique", *Applied Economics*, 30, №11, 1483-1492.
77. **Mills T.C. (1993)** *The Econometric Modelling of Financial Time Series*. Cambridge University Press, Cambridge.
78. **Molana H. (1994)** "Consumption and Fiscal Theory. UK Evidence from a Cointegration Approach", *Dundee Discussion Papers*, University of Dundee, Dundee, Scotland.
79. **Murray C.J., C.R. Nelson (2000)** "The Uncertain Trend in U.S. GDP", *Journal of Monetary Economics*, 46, 79-95.
80. **Nadal-De Simone F., W.A. Razzak (1999)** "Nominal Exchange Rates and Nominal Interest Rate Differentials", *IMF Working Paper* WP/99/141.
81. **Nelson C.R., H. Kang (1981)** "Spurious Periodicity in Inappropriately Detrended Time Series", *Journal of Monetary Economics*, 10, 139-162.
82. **Nelson C.R., C.I. Plosser (1982)** "Trends and Random Walks in Macroeconomic Time Series", *Journal of Monetary Economics*, 10, 139-162.
83. **Newey W., K. West (1987)** "A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703-708

84. **Newey W., K. West (1994)** "Automatic Lag Selection in Covariance Matrix Estimation," *Review of Economic Studies*, 61, 631–653.
85. **Ng S., P. Perron (1995)** "Unit Root Tests in ARMA models With Data-Dependent Methods for the Selection of the Truncation Lag", *Journal of American Statistical Association*, 90, 268-281.
86. **Nunes L.S., Newbold P., C.-M. Kuan (1997)** "Testing for Unit Roots With Breaks. Evidence on the Great Crash and the Unit Root Hypothesis Reconsidered", *Oxford Bulletin of Economics and Statistics*, 59, №4, 435-448.
87. **Patterson K. (2000)** *An Introduction to Applied Econometrics: A Time Series Approach*. New York: St's Martin Press.
88. **Perron P. (1988)** "Trends and Random Walks in Macroeconomic Time Series: Furter Evidence from a New Approach", *Journal of Economic Dynamic and Control*, 12, 297-332.
89. **Perron P. (1989a)** "The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis, *Econometrica*, 577, 1361-1401.
90. **Perron P. (1989b)** "Testing for a Random Walk: A Simulation Experiment When the Sampling Interval Is Varied" – в сборнике *Advances in Econometrics and Modelling* (редактор B.Ray), Kluwer Academic Publishers, Dordrecht and Boston.
91. **Perron P. (1997)** "Further evidence on breaking trend functions in macroeconomic variables, *Journal of Econometrics*, 80, №2, 355-385.
92. **Perron P., Vogelsang T.J. (1993)** "Erratum", *Econometrica*, 61, №1, 248-249.
93. **Phillips P.C.B. (1987)** "Time Series Regression with a Unit Root", *Econometrica*, 55, 277-301.
94. **Phillips P.C.B., P. Perron (1988)** "Testing for a Unit Root in Time Series Regression," *Biometrika*, 75, 335–346.
95. **Saikonnen P. (1991)** "Asymptotically Efficient Estimation of Cointegrated Regressions", *Econometric Theory*, 7, 1-21.
96. **Said E., D.A. Dickey (1984)** "Testing for Unit Roots in Autoregressive Moving Average Models of Unknown Order," *Biometrika*, 71, 599–607.
97. **Sargan J.D., Bhargava A. (1983)** "Testing Residuals from Least Squares Regression for Being Generated by the Gaussian Random Walk", *Econometrica*, 51, N1 153-174
98. **Shiller R.J., Perron P. (1985)** "Testing the Random Walk Hypothesis: Power versus Frequency of Observation", *Economic Letters*, 18, 381-386.
99. **Schmidt P., Phillips P.C.B. (1992)** "LM Tests for a Unit Root in the Presence of Deterministic Trends", *Oxford Bulletin of Economics and Statistics*, 54, 257-287.

100. **Schwarz G. (1978)** "Estimating the Dimension of a Model", *The Annals of Statistics*, 16, 461-464.
101. **Schwert G.W. (1989)** "Tests for Unit Roots: A Monte Carlo Investigation", *Journal of Business and Economic Statistics*, 7, 147-159.
102. **Sims C.A., J.H. Stock, M.W. Watson (1990)** "Inference in Linear Time Series Models with Some Unit Roots", *Econometrica* 58, 113-144.
103. **Stock Watson (1993)** "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems", *Econometrica*, 61, 783-820.
104. **Taylor A.M.R. (2000)** "The Finite Sample Effects of Deterministic Variables on Conventional Methods of Lag-Selection in Unit-Root Tests", *Oxford Bulletin of Economics and Statistics*, 62, 293-304.
105. **West K.D. (1988)** "Asymptotic Normality, When Regressors Have a Unit Root", *Econometrica*, 56, 1397-1417.
106. **White J. S. (1958)** "The Limiting Distribution of the Serial Correlation Coefficient in the Explosive Case", *Annals of Mathematical Statistics*, 29, 1188-1197.
107. **Wirjanto T. S., R.A. Amano (1996)** "Nonstationary Regression Models with a Lagged Dependent Variable", *Commun. Statist.-Theory Meth.*, 25, №7, 1489-1503.
108. **Woodward G., R. Pillarisetti (1999)** "Empirical Evidence on Alternative Theories of Inflation and Unemployment: a Re-Evaluation for the Scandinavian Countries", *Applied Economic Letters*, 6, №1, 55-58.
109. **Zivot E., Andrews D.W.K. (1992)** "Further Evidence on the Great Crash, the Oil Price Shock and the Unit Root Hypothesis", *Journal of Business and Economic Statistics*, 10, 251-270.

Указатель

Q

Q-статистики, 42, 43

Бокса – Пирса, 54

Люнга – Бокса, 55

A

Автокоррелированность

критерий Бройша – Годфри, 69, 86

критерий Дарбина – Уотсона, 69

смещение, 69

Авторегрессионные ошибки

преобразование Кохрейна – Оркатта, 83

Б

Белый шум, 16

N-мерный гауссовский, 230, 242

векторный, 203

гауссовский, 17

двумерный гауссовский, 212

Броуновское движение, 133

В

Векторная авторегрессия, 73

замкнутая модель, 77

модель коррекции ошибок (ECM), 204, 242

оценивание, 258

открытая модель, 77

матрица долгосрочных

мультипликаторов, 78

пониженного ранга, 204

порядок, 74

ранг коинтеграции, 241

оценивание, 242, 255

условие стабильности, 76

Временной ряд, 14

гауссовский, 15

векторный, 231

дифференцирование, 114

долговременная дисперсия, 160

оценивание, 160

интегрированный порядка k , 116

остационаривание, 127

передифференцированный, 127

разностно стационарный, 117

стационарный в широком смысле, 15

стационарный векторный, 66

гауссовский, 231

кросс-корреляционная функция, 66

кросс-корреляция, 66

разложение Вольда, 203

стационарный относительно
детерминированного тренда, 63, 116,
133

строго стационарный, 14, 15

типа $ARIMA(p, k, q)$, 117

Временные ряды

коинтегрированные, 195, 203

в узком смысле, 203, 212, 221

векторное ARMA представление, 204

долговременная связь, 206

коинтеграционное пространство, 221

базис, 221

коинтегрирующий вектор, 203, 221

кратковременная динамика, 206

общие тренды, 256

оценивание

метод *leads and lags*, 232

ряды с линейным трендом, 226

треугольная система Филлипса, 211, 226, 230

ранг коинтеграции, 221

треугольная система Филлипса, 223

некоинтегрированные, 195

Д

Динамические модели

авторегрессионные ошибки, 83

модель коррекции ошибок, 83

модель опережающего показателя, 81

модель распределенных запаздываний, 82

модель скорости роста, 82

модель частичной корректировки, 82

приведенная форма, 82

процесс авторегрессии, 81

Динамические модели, типы моделей, 81

Долговременная дисперсия, 160, 197, 227

оценка Newey – West, 160, 227

ширина окна, 160

Долговременная связь, 78

Долговременное положение равновесия системы, 221

отклонение от положения равновесия, 221

И

Инновация, 23, 73

вектор инноваций, 74

Итерационные методы, 47

выбор стартового значения, 47, 48

К

Коинтеграция

детерминистская, 203, 213, 220, 236

проверка гипотезы коинтегрированности, 212

заданный коинтегрирующий вектор, 213

мощность критериев, 215

неизвестный коинтегрирующий вектор

ряды без тренда, 213

ряды с трендом, 214

стохастическая, 220, 239

Коинтегрированная VAR, 222

модель коррекции ошибок (ECM), 222

оценивание коэффициентов ECM, 261

сверхидентифицирующие ограничения, 261

ранг коинтеграции, 222
Коинтегрирующие векторы
оценивание
 идентифицирующие ограничения, 259
 сверхидентифицирующие ограничения, 261
Коинтегрирующий вектор
 нормализованный, 259
 оценивание, 206, 213
 двухшаговая процедура, 206
 метод Йохансена, 258
Коррелограмма, 16, 55
Коэффициент автокорреляции, 16
Критерий DF-GLS, 164
Критерий Дики – Фуллера, 134, 140, 147
 для процессов $ARMA(p, q)$, 146
 мощность критериев, 142, 148
 расширенный, 145
 таблицы, 150
Критерий информационный
 Акаике, 44
 Хеннана – Куинна, 45
 Шварца, 44
Критерий Квятковского – Филлипса – Шмидта – Шина (KPSS), 164
Критерий Лейбурна, 163
 мощность, 163
Критерий Перрона, 172
 датировка точки излома, 181
 обобщенный, 182
Критерий Филлипса – Перрона, 159
 мощность, 161

Критерий Шмидта – Филлипса, 163
Куртозис, 182

Л

Линейная модель наблюдений
 классическая нормальная, 4, 8
Ложная периодичность, 119, 127
Ложная регрессия, 186, 193

М

Метод максимального правдоподобия, 47
Модели
 ARX
 стабильность, 66
 динамические (ADL), 67
 мультипликаторы
 долгосрочные, 67, 68
 импульсные, 69
 коррекции ошибок (ECM), 204
Мультипликатор
 импульсный, 124

О

Обратимости условие, 31, 47, 48, 49, 54
Объясняющие переменные
 стохастические, 9
Оператор запаздывания, 23
Оценка наименьших квадратов
 обобщенная, 13
 суперсостоятельная, 207

П

Паразитная связь, 196
Передаточная функция, 78
Подбор стационарной модели ARMA

диагностика модели, 35, 53
 проверка нормальности, 58
идентификация модели, 35
оценивание модели, 35, 47
 метод максимального
 правдоподобия, 47

Применимость стандартных
статистических выводов

 ситуация А, 9
 ситуация А', 11
 ситуация В, 11
 ситуация С, 12
 ситуация D, 63
 ситуация Е, 64
 ситуация F, 65

Принцип экономности модели, 34
 экономная модель, 32

Причинность по Гренджеру, 205

Проверка нормальности
 критерий Jarque – Bera, 58

Процесс авторегрессии, 18, 24, 25, 28,
37, 99, 139, 144
 векторный, 208
 взрывной, 111
 порядка p , 23

Процесс порождения данных, 84

Р

Различение TS и DS рядов, 128
 TS-гипотеза, 129
 близкие альтернативы, 143
 влияние протяженности ряда, 167
 гипотеза единичного корня, 129
 количество единичных корней, 169

коррекция сезонности, 167
многовариантная процедура, 153
нулевая TS гипотеза, 143
 критерий KPSS, 164
процедура Кохрейна, 165
согласованность статистических
выводов, 168
структурные изменения модели, 172
 аддитивный выброс, 174
 модель с инновационным
 выбросом, 175

С

Сезонные модели

 авторегрессии, 32
 аддитивные, 34
 мультипликативные, 34
 скользящего среднего, 32

Скользящее среднее порядка q
 оценивание, 26, 48
 backcasting, 49
 условие обратимости, 48

Случайное блуждание, 111
 со сносом, 114

Смешанный процесс авторегрессии –
скользящего среднего (ARMA), 30
 сезонные модели, 32
 условие стационарности, 30

Стабильности условие, 21, 67, 84, 94

Статистика

 Бокса – Пирса, 42
 Люнга – Бокса, 42
 отношения дисперсий, 166

Статистическая модель, 84

Статическая регрессия, 81
Стационарные временные ряды
 белый шум, 16
Структурная форма, 83

Т

Тренд
 детерминированный, 113
 имеющий излом, 128
 квадратичный, 152, 171
детрендирование, 114
изменение наклона, 176
изменение наклона и уровня, 176
изменение уровня, 176
сегментированный, 177
стохастический, 113, 220

Ф

Фиктивная линейная связь, 187

Функция

автокорреляционная, 16, 36, 38, 48,
127, 159
 выборочная, 37
частная автокорреляционная, 36
 выборочная, 38, 46

Э

Экзогенные переменные, 77
Эндогенные переменные, 77

Ю

Юла – Уокера уравнения, 25
 использование
 при выборе стартовых значений, 48
 при вычислении частных
 автокорреляций, 26

при идентификации модели ARMA,
36