Временные ряды

Эконометрика. Openedu. Неделя 12

Временные ряды:

- Одномерные

Один показатель для каждого момента времени.

- Многомерные

Несколько показателей для каждого момента времени.

Пример многомерного временного ряда

Год	Население (тыс. чел.)	ВВП (млрд. руб.)
2002	145649	10830
2003	144964	13208
2004	144168	17027
2005	143474	21610
2006	142754	26917
2007	142220	33248
2008	141980	41277
2009	141900	38807
2010	142962	46308
2011	142914	59698
2012	143103	66927
2013	143395	71055

Одномерный временной ряд

Временной ряд — последовательность случайных величин

$$y_1, y_2, y_3, y_4 \ldots$$

Без предположений невозможно прогнозировать

1, 2, 3, 4, 5, ?

Какое число будет следующим?

Без предположений невозможно прогнозировать

1, 2, 3, 4, 5, 42

Базовое предположение — стационарность

Временной ряд называется стационарным, если:

- $E(y_1) = E(y_2) = E(y_3) = \ldots$
- $Var(y_1) = Var(y_2) = Var(y_3) = \ldots = \gamma_0$
- $Cov(y_1, y_2) = Cov(y_2, y_3) = Cov(y_3, y_4) = \ldots = \gamma_1$
- $Cov(y_1, y_3) = Cov(y_2, y_4) = Cov(y_3, y_5) = \ldots = \gamma_2$
- $\ldots$

Предпосылки коротко:

Временной ряд называется стационарным, если:

- $E(y_t) = const$
- $Cov(y_t, y_{t-k}) = \gamma_k$, т.е. не зависит от t

Автоковариационная функция

$\gamma_k = Cov(y_t, y_{t-k})$ — (авто)-ковариационная функция процесса

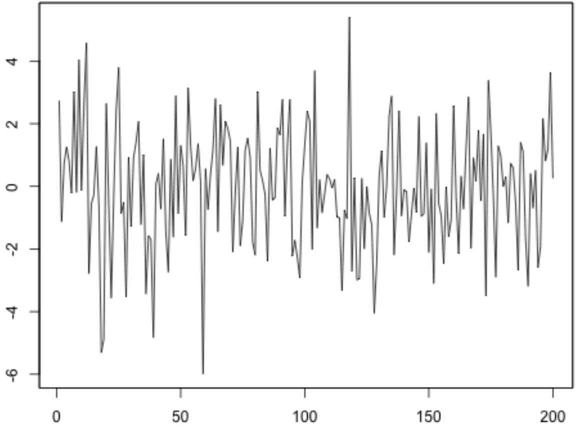
Самый простой пример — белый шум

Ряд ε_t — белый шум, если:

- $E(\varepsilon_t) = 0$
- $Var(\varepsilon_t) = \sigma^2$
- $Cov(\varepsilon_t, \varepsilon_{t-k}) = 0$

Пример белого шума

$\varepsilon_t \sim N(0, 4)$ и независимы



Конвенция об обозначениях

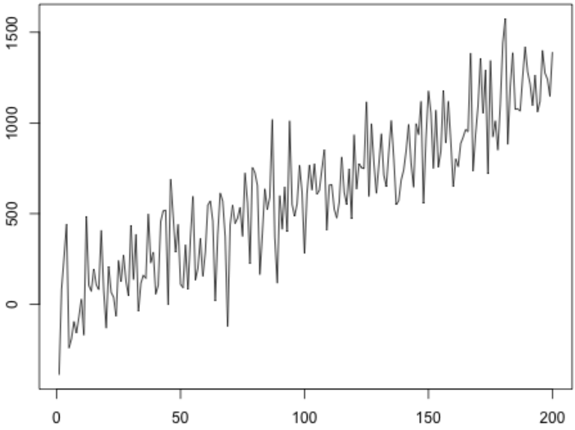
На эту лекцию ε_t всегда обозначает белый шум!

Примеры нестационарных процессов

- Процесс с детерминистическим трендом
- Случайное блуждание

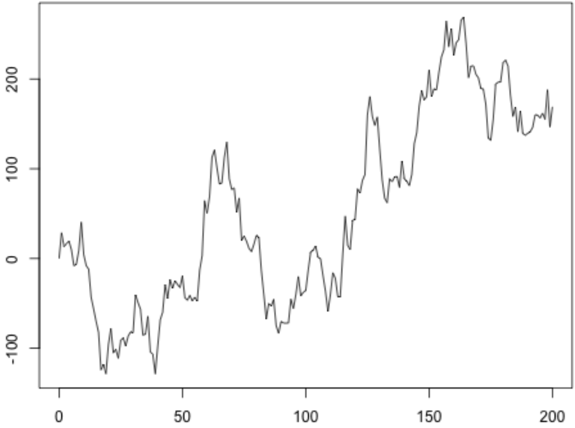
Процесс с детерминистическим трендом

- $y_t = 5 + 6t + \varepsilon_t$. Нестационарность: $E(y_t) = 5 + 6t \neq const$



Случайное блуждание

- $\begin{cases} y_0 = 0 \\ y_t = y_{t-1} + 2 + \varepsilon_t \end{cases}$. Нестационарность: $Var(y_t) = t\sigma^2 \neq const$



Процесс скользящего среднего

Процесс представимый в виде

$$y_t = \mu + \varepsilon_t + a_1\varepsilon_{t-1} + \ldots + a_q\varepsilon_{t-q}$$

Обозначение процесса скользящего среднего

$y_t \sim MA(q)$, Moving Average

Пример MA процесса [у доски]

$$y_t = 5 + \varepsilon_t + 3\varepsilon_{t-1} - 2\varepsilon_{t-2}$$

Найдите $E(y_t)$, $Var(y_t)$, $Cov(y_t, y_{t-k})$

Запись с помощью оператора лага

L — оператор лага:

- $Ly_t = y_{t-1}$
- $L \cdot L \cdot y_t = L^2 y_t = y_{t-2}$
- $\ldots$

Пример записи с помощью оператора лага

$MA(2)$:

$$y_t = 2 + \varepsilon_t + 3\varepsilon_{t-1} - 2\varepsilon_{t-2}$$

$$y_t = 2 + (1 + 3L - 2L^2)\varepsilon_t$$

Интерпретация:

Коэффициенты плохо интерпретируемы.
У стационарного процесса есть (авто)-корреляционная функция:

$$\rho_k = Corr(y_t, y_{t-k}) = \frac{Cov(y_t, y_{t-k})}{\sqrt{Var(y_t)Var(y_{t-k})}} = \frac{\gamma_k}{\gamma_0}$$

Интерпретация

Если y_t — стационарный процесс и $y_t \sim N(\mu_y, \sigma_y^2)$, то:
ρ_k — на сколько в среднем изменится y_t при росте y_{t-k} на единицу

Автокорреляционная функция MA процесса [у доски]

$$y_t = 5 + \varepsilon_t + 3\varepsilon_{t-1} - 2\varepsilon_{t-2}$$

Найдите ρ_k

Частная автокорреляционная функция-идея

ρ_k — автокорреляционная функция. Измеряет совокупный эффект воздействия y_{t-k} на y_t как напрямую, так и через y_{t-k+1}, y_{t-k+2}, $\ldots$, y_{t-1}

ϕ_k — частная автокорреляционная функция. Измеряет прямой эффект воздействия y_{t-k} на y_t, устранив сквозное воздействие через y_{t-k+1}, y_{t-k+2}, $\ldots$, y_{t-1}.

$$\begin{tikzcd}
y_1 \arrow[r] \arrow[rr, bend left] \arrow[rrr, bend left, "\phi_3"] & y_2 \arrow[r] \arrow[rr, bend left] & y_3 \arrow[r] & y_4
\end{tikzcd}$$

Частная автокорреляционная функция-интерпретация

Если y_t — стационарный процесс и $y_t \sim N(\mu_y, \sigma_y^2)$, то:
ϕ_k — на сколько в среднем изменится y_t при росте y_{t-k} на единицу при фиксированных y_{t-1}, y_{t-2}, ..., y_{t-k+1}

Частная автокорреляционная функция-определение

$$\phi_k = Cor(y_t - P(y_t), y_{t-k} - P(y_{t-k}))$$

где $P(y_t)$ — проекция случайной величины y_t на линейную оболочку величин y_{t-1}, y_{t-2}, $\ldots$, y_{t-k+1}.

Частная автокорреляция алгоритм подсчёта

$$\gamma_0\phi_1 = \gamma_1$$

$$\begin{cases} \gamma_0 *_1 + \gamma_1\phi_2 = \gamma_1 \\ \gamma_1 *_1 + \gamma_0\phi_2 = \gamma_2 \end{cases}$$

$$\begin{cases} \gamma_0 *_1 + \gamma_1 *_2 + \gamma_2\phi_3 = \gamma_1 \\ \gamma_1 *_1 + \gamma_0 *_2 + \gamma_1\phi_3 = \gamma_2 \\ \gamma_2 *_1 + \gamma_1 *_2 + \gamma_0\phi_3 = \gamma_3 \end{cases}$$

...

Частная автокорреляционная функция MA процесса [у доски]

$$y_t = 5 + \varepsilon_t + 3\varepsilon_{t-1} - 2\varepsilon_{t-2}$$

Найдите ϕ_1, ϕ_2, ϕ_3

Процесс авторегрессии

- Стационарный процесс вида

$$y_t = c + b_1 y_{t-1} + b_2 y_{t-2} + \ldots + b_p y_{t-p} + \varepsilon_t$$

Обозначение процесса авторегрессии

$y_t \sim AR(p)$, AutoRegression

Частная и обычная автокорреляционные функции для AR процесса [у доски]

$$y_t = 2 + 0.5y_{t-1} + \varepsilon_t \ \varepsilon_t \sim N(0, \sigma^2)$$

Найдите ρ_k, ϕ_k

Альтернативная форма записи

$$y_t = 2 + 0.5y_{t-1} + \varepsilon_t$$

Или

$$(y_t - 4) = 0.5(y_{t-1} - 4) + \varepsilon_t$$

Важное предупреждение

Из одного уравнения $y_t = 2 + 0.5y_{t-1} + \varepsilon_t$ не следует автоматически стационарность (!)

Пример множества решений [у доски]

$$y_t = 2 + 0.5y_{t-1} + \varepsilon_t \; \varepsilon_t \sim N(0, 1)$$

- $y_0 = 0$, $y_1 \sim N(2, 1)$, $y_2 \sim N(3, 1.25)$, ...
- $y_0 \sim N(3, 4/3)$, $y_1 \sim N(3, 4/3)$, $y_2 \sim N(3, 4/3)$, ...

Подразумеваем стационарное решение

Пишем:

$$y_t = 2 + 0.5y_{t-1} + \varepsilon_t \ \varepsilon_t \sim N(0,1)$$

Подразумеваем:

- $y_0 \sim N(3, 4/3)$, $y_1 \sim N(3, 4/3)$, $y_2 \sim N(3, 4/3)$, ...

AR процесс можно записать с помощью лага

$$y_t = 2 + 0.5y_{t-1} - 0.06y_{t-2} + \varepsilon_t$$

$$(1 - 0.5L + 0.06L^2)y_t = 2 + \varepsilon_t$$

Характеристический многочлен

$$(1 - 0.5L + 0.06L^2)y_t = 2 + \varepsilon_t$$

$$f(L)y_t = 2 + \varepsilon_t$$

$f(L)$ — характеристический многочлен

Когда у есть стационарное решение?

$$f(L)y_t = c + \varepsilon_t$$

Если корни характеристического уравнения AR процесса, $f(z) = 0$, по модулю больше единицы, то существует единственное стационарное решение, в котором y_t выражается через прошлые шумы, то есть через ε_t, ε_{t-1}, ε_{t-2}, $\ldots$

Упражнение [у доски]

Пример 1. $y_t = 7 + 0.5y_{t-1} - 0.06y_{t-2} + \varepsilon_t$
Пример 2. $y_t = -3 + 1.2y_{t-1} - 0.2y_{t-2} + \varepsilon_t$
Есть ли у этих уравнений стационарные решения?

Прогнозирование

Прогноз на h шагов вперед: $E(y_{t+h}|y_t, y_{t-1}, y_{t-2}, \ldots)$
Часто кратко обозначают: $\hat{y}_{t+h}$

Упражнение на прогнозирование [у доски]

$$y_t = 2 + 0.5y_{t-1} - 0.06y_{t-2} + \varepsilon_t \ \varepsilon_t \sim N(0;4)$$

$y_{100} = 4$, $y_{99} = 3$.
Постройте точечный и интервальный прогноз на 1 и 2 шага вперед

Модель авторегрессии и скользящего среднего

- Стационарный процесс вида

$$y_t = c + b_1 y_{t-1} + b_2 y_{t-2} + \ldots + b_p y_{t-p} + \\ + \varepsilon_t + a_1 \varepsilon_{t-1} + \ldots + a_q \varepsilon_{t-q}$$

где сумма $p + q$ минимально возможна

Обозначение

- $y_t \sim ARMA(p, q)$

Сумма $p + q$ минимально возможная

- $y_t = \varepsilon_t$
- $y_t - y_{t-1} = \varepsilon_t - \varepsilon_{t-1}$

В этом примере $y_t \sim ARMA(0, 0)$

ARMA — это наше всё!

Теорема.
Любой стационарный процесс можно представить в виде $MA(\infty)$
Практический вывод. С помощью $ARMA(p, q)$ можно компактно и сколь угодно точно описать любой стационарный процесс

Итого про ARMA(p, q)

- коэффициенты не интерпретируемы
- используются для прогнозирования

Оценивание коэффициентов

Есть T наблюдений: $y_1, y_2, y_3, \ldots, y_T$
Чаще всего используется метод максимального правдоподобия

Подробности метода максимального правдоподобия

- Как правило, предполагается нормальность $\varepsilon_t \sim N(0;\sigma^2)$
- Стационарность y_t

$$y_t = c + b_1 y_{t-1} + b_2 y_{t-2} + \ldots + b_p y_{t-p} + \\ + \varepsilon_t + a_1 \varepsilon_{t-1} + \ldots + a_q \varepsilon_{t-q}$$

Результат метода максимального правдоподобия

На выходе получаем оценки

$$\hat{\theta} = (\hat{c}, \hat{a}_1, \ldots, \hat{a}_q, \hat{b}_1, \ldots, \hat{b}_q, \hat{\sigma}^2)$$

И оценку их ковариационной матрицы $\widehat{Var}(\hat{\theta})$

Проверка гипотез и доверительные интервалы

$$\frac{\hat{a}_j - a_j}{se(\hat{a}_j)} \to N(0;1)$$

Выборочная автокорреляционная функция

ACF — autocorrelation function

$$\hat{\rho}_k = \frac{\sum_{t=k+1}^{T}(y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{T}(y_t - \bar{y})^2}$$

Выборочная частная автокорреляционная функция

PACF — partial autocorrelation function
Получим $\hat{\phi}_k$ из оценки регрессии

$$\hat{y}_t = * + * \cdot y_{t-1} + * \cdot y_{t-2} + \ldots + * \cdot y_{t-k+1} + \phi_k y_{t-k} + u_t$$

Примечания к расчету автокорреляционной функции

- Для оценки каждого $\hat{\phi}_k$ строится отдельная регрессия
- Из каждой регрессии нужен только последний коэффициент

Алгоритм на практике

1. Графики ряда, ACF, PACF
2. Если ряд нестационарный, то преобразуем
3. Выбираем p и q
4. Оцениваем $ARMA(p, q)$
5. Прогнозируем

Основное преобразование

Взятие разности: переход от y_t к Δy_t

Обозначение

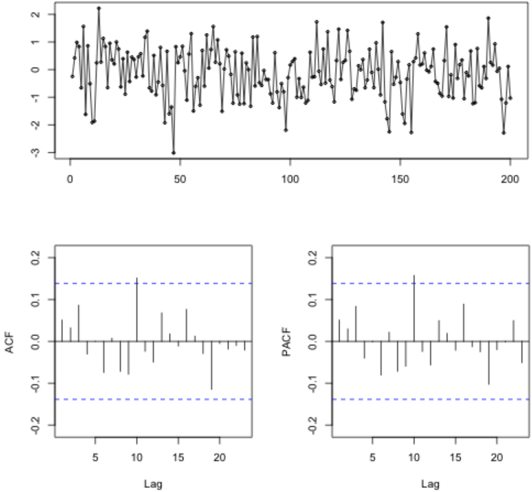
- $y_t \sim ARIMA(p, 1, q)$ равносильно $\Delta y_t \sim ARMA(p, q)$
- $y_t \sim ARIMA(p, 0, q)$ равносильно $y_t \sim ARMA(p, q)$

Выбор p и q по графикам

График выборочной корреляционной функции есть даже у нестационарного процесса!

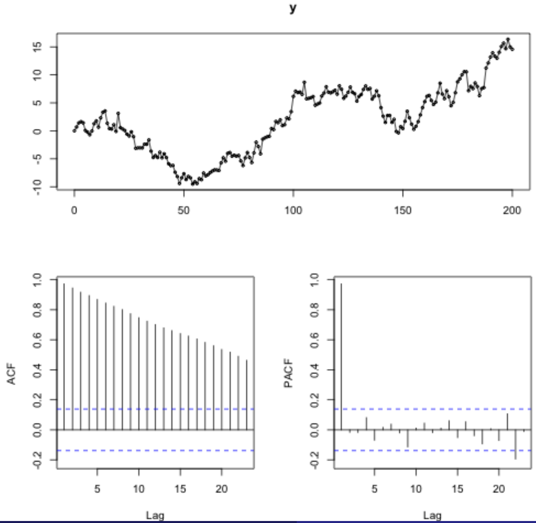
У нестационарного процесса ρ_k, ϕ_k не существуют, однако компьютер всегда может построить график выборочной автокорреляционной и выборочной частной автокорреляционной функции!

Белый шум, $y_t = \varepsilon_t$



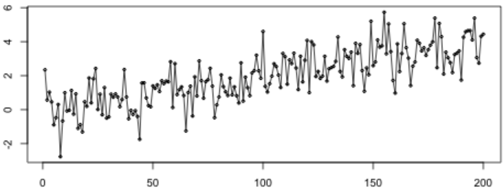
Случайное блуждание (нестационарный процесс!)

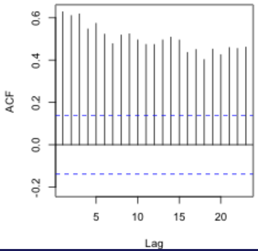
Случайное блуждание, $y_t = y_{t-1} + \varepsilon_t$. Истинные ρ_k и ϕ_k НЕ существуют!

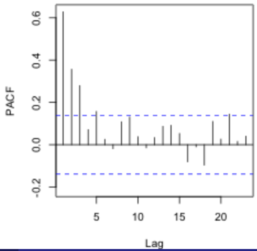


Процесс с трендом (нестационарный процесс!)

Процесс с трендом, $y_t = 0.02 \cdot t + \varepsilon_t$. Истинные ρ_k и ϕ_k НЕ существуют!

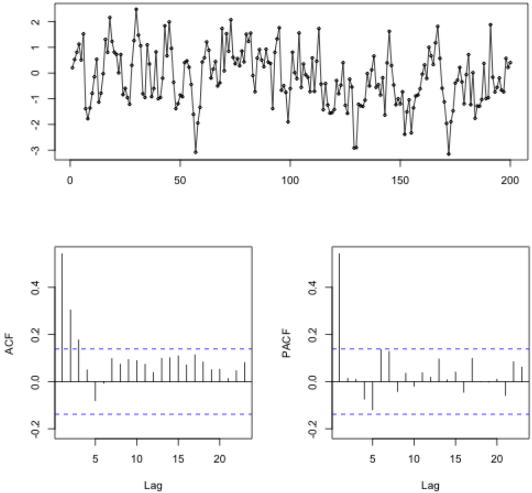






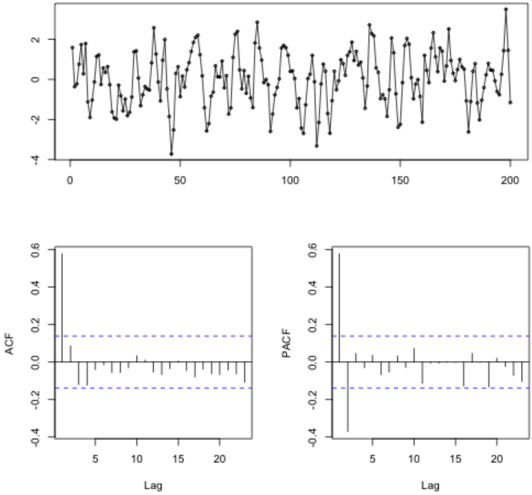
AR(1)

AR(1), $y_t = 0.7 y_{t-1} + \varepsilon_t$



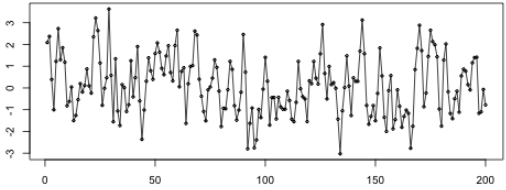
AR(2)

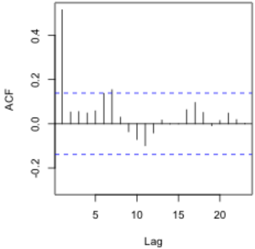
AR(2), $y_t = 0.9y_{t-1} - 0.5y_{t-2} + \varepsilon_t$

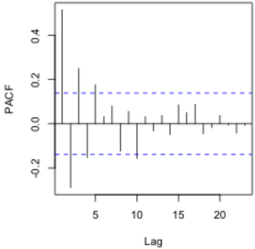


MA(1)

MA(1), $y_t = 0.7\varepsilon_{t-1} + \varepsilon_t$

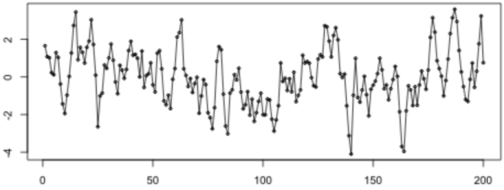


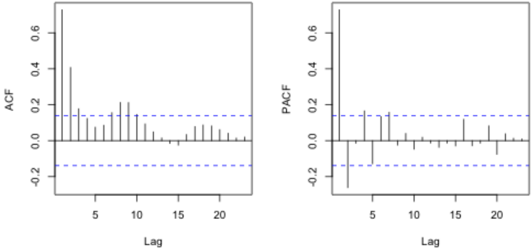




MA(2)

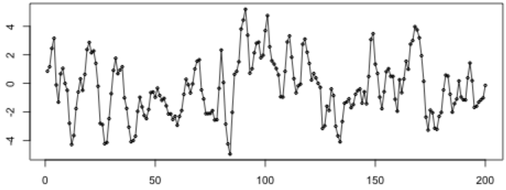
MA(2), $y_t = 0.9\varepsilon_{t-1} + 0.5\varepsilon_{t-1} + \varepsilon_t$

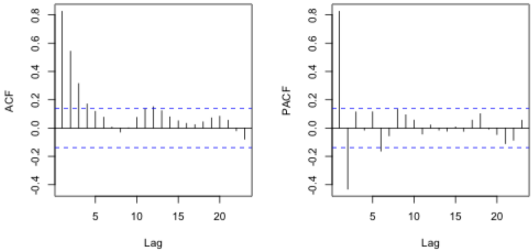




ARMA(1,1)

ARMA(1,1), $y_t = 0.7y_{t-1} + 0.5\varepsilon_{t-1} + \varepsilon_t$





Мораль

- Временные ряды: стационарные и нет
- Для стационарных — модель ARMA

Источники мудрости:

- Борзых Д.А., Демешев Б.Б. Эконометрика в задачах и упражнениях: глава 12
- Катышев П.К., Пересецкий А. А. Эконометрика. Начальный курс: главы 11.3, 11.4
- Носко В.П., Эконометрика. Введение в регрессионный анализ временных рядов: глава 2