Лекция 13
Анализ временных рядов

Габдуллин Р.А., Макаренко В.А.

МГУ им. М.В. Ломоносова

31 марта 2021

Временной ряд

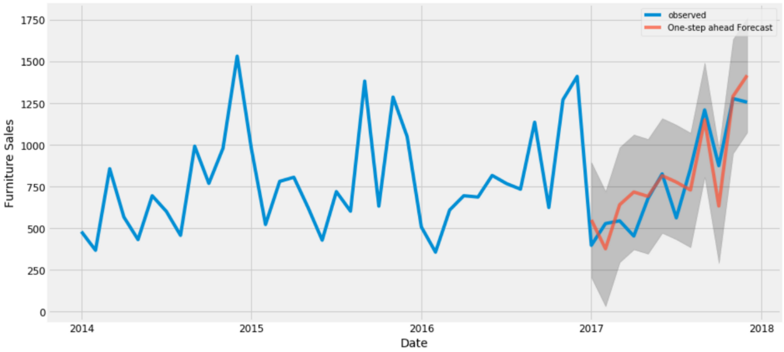

Рис.: Источник: becominghuman.ai

- Последовательность случайных величин (случайный процесс с дискретным временем) $y_1, y_2, \dots, y_t, \dots$ будем называть временным рядом.

Стационарные временные ряды

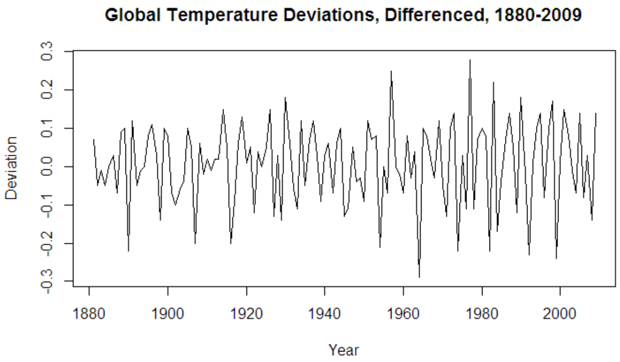

Рис.: Источник: datascienceplus.com

- Будем называть временной ряд стационарным (в широком смысле), если математическое ожидание и дисперсия постоянны во времени.

Нестационарные временные ряды. Тренд

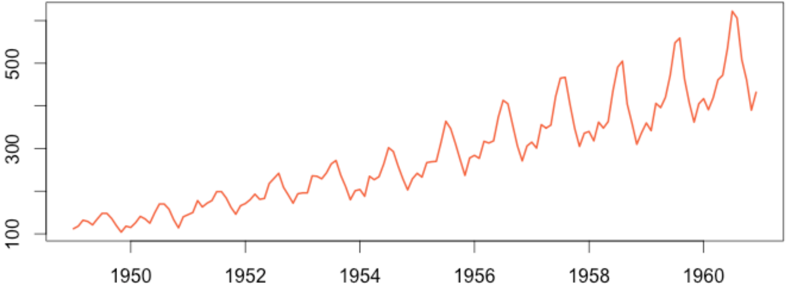

Рис.: Источник: anomaly.io

- Среднее значение меняется во времени.

Нестационарные временные ряды. Гетероскедастичность

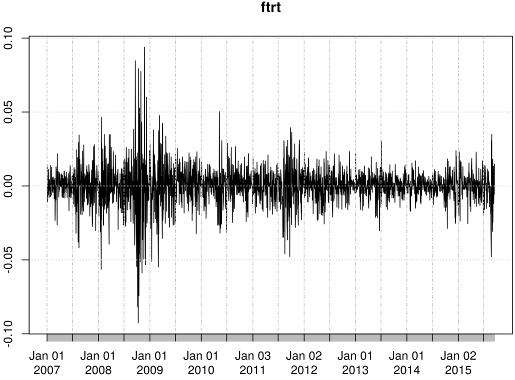

Рис.: Источник: quantstart.com

- Дисперсия меняется во времени.

Модель авторегрессии (AR)

- Процесс AR(p):

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \ldots + \alpha_p y_{t-p} + \varepsilon_t,$$

где $\{\varepsilon_t\}$ – н.о.р.с.в. с $\mathbb{E}\varepsilon_t = 0$ и $\mathbb{D}\varepsilon_t = \sigma_\varepsilon^2$.

- Альтернативная запись:

$$\phi(B) y_t = \varepsilon_t, \quad \phi(B) = 1 - \alpha_1 B - \alpha_2 B^2 - \ldots - \alpha_p B^p,$$

где $B^k y_t = y_{t-k}$.

Модель авторегрессии (AR)

Теорема

Любой процесс AR(p) может быть представлен в следующем виде:

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \ldots + \alpha_p y_{t-p} + \varepsilon_t = \sum_{k=0}^{+\infty} \beta_k \varepsilon_{t-k}.$$

Теорема

Если все корни характеристического уравнения

$$\phi(B) = 1 - \alpha_1 B - \alpha_2 B^2 - \ldots - \alpha_p B^p = 0$$

по модулю больше единицы (лежат вне единичного круга), то процесс AR(p) является стационарным.

Модель авторегрессии (AR). Числовые характеристики

- Процесс AR(p):

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \ldots + \alpha_p y_{t-p} + \varepsilon_t,$$

- Математическое ожидание:

$$\mathbb{E} y_t = 0$$

- Дисперсия:

$$\mathbb{D} y_t = \sigma_\varepsilon^2 + \sum_{k=1}^{p} \alpha_k \operatorname{cov}(y_t, y_{t+k}).$$

- Автокорреляция (ACF) определяется из системы уравнений Юла-Уокера:

$$\rho_k = \alpha_1 \rho_{k-1} + \ldots + \alpha_p \rho_{k-p}.$$

Модель авторегрессии (AR). Частная автокорреляция (PACF)

- Частная автокорреляция (PACF):

$$\pi_k = \begin{cases} \operatorname{corr}(y_t, y_{t+1}), & k = 1, \\ \operatorname{corr}(y_{t+k} - P_{t,k}(y_{t+k}), y_t - P_{t,k}(y_t)), & k > 1 \end{cases},$$

  где $P_{t,k}(x)$ – регрессия (наилучшая линейная оценка) случайной величины x на случайные величины $y_{t+1}, \dots, y_{t+k-1}$.

- В модели AR(p):

$$\begin{cases} \pi_k > 0, & 1 \leqslant k \leqslant p, \\ 0, & k > p. \end{cases}$$

Модель авторегрессии (AR)

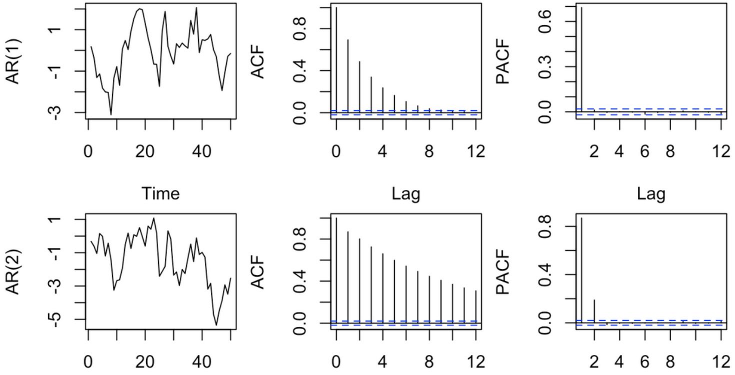

Рис.: Источник: nwfsc-timeseries.github.io

Модель авторегрессии (AR)

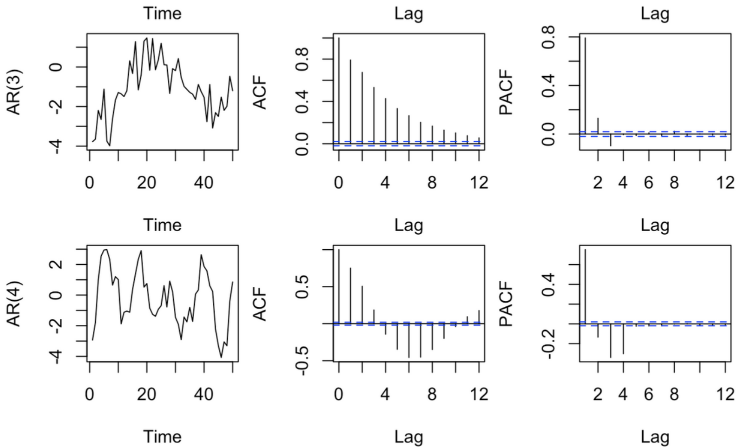

Рис.: Источник: nwfsc-timeseries.github.io

Модель скользящего среднего (MA)

- Процесс MA(q):

$$y_t = \theta_0\varepsilon_t + \theta_1\varepsilon_{t-1} + \ldots + \theta_q\varepsilon_{t-q},$$

где $\{\varepsilon_t\}$ – н.о.р.с.в. с $\mathbb{E}\varepsilon_t = 0$ и $\mathbb{D}\varepsilon_t = \sigma_\varepsilon^2$.

- Математическое ожидание:

$$\mathbb{E}y_t = 0.$$

- Дисперсия:

$$\mathbb{D}y_t = \sigma_\varepsilon^2(\theta_0^2 + \theta_1^2 + \ldots + \theta_q^2).$$

- Автокорреляция (ACF):

$$\rho_k = \operatorname{corr}(y_t, y_{t+k}) = \begin{cases} 1, & \text{если} \quad k = 0, \\ \frac{\sum_{j=0}^{q-k}\theta_j\theta_{j+k}}{\sum_{j=0}^{q}\theta_j^2}, & \text{если} \quad 1 \leqslant k \leqslant q, \\ 0, & \text{если} \quad k > q. \end{cases}$$

Модель скользящего среднего (MA)

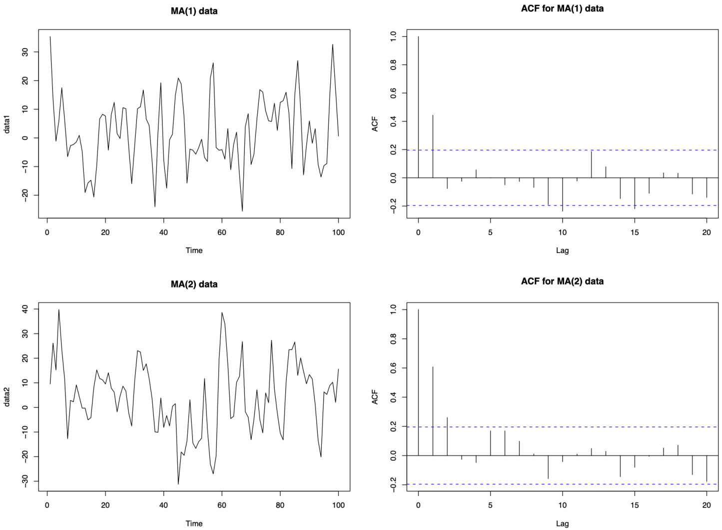

Рис.: Источник: bookdown.org

Модель ARMA

- Процесс ARMA(p,q):

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \ldots + \alpha_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q} =$$

$$= \sum_{k=1}^{p} \alpha_k y_{t-k} + \sum_{k=1}^{q} \theta_k \varepsilon_{t-k} + \varepsilon_t.$$

- Альтернативная запись

$$\phi(B) y_t = \psi(B) \varepsilon_t,$$

$$\phi(B) = 1 - \alpha_1 B - \alpha_2 B^2 - \ldots - \alpha_p B^p,$$

$$\psi(B) = 1 + \theta_1 B + \theta_2 B^2 + \ldots + \theta_q B^q,$$

$$B^k x_t = x_{t-k}.$$

Модель ARMA

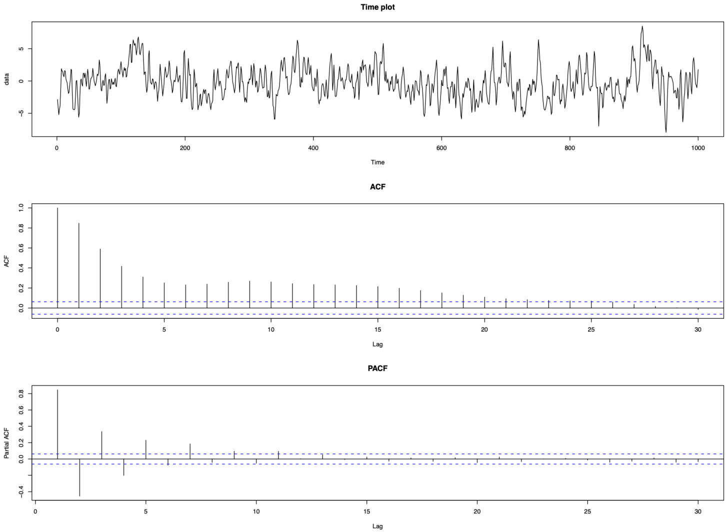

Рис.: Источник: bookdown.org

Модель ARMA

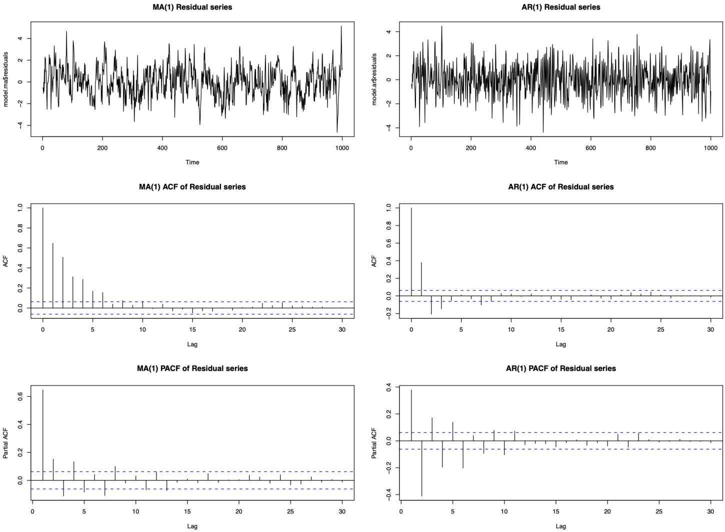

Рис.: Источник: bookdown.org

Модель ARIMA

- Дифференцирование временного ряда:

$$\nabla y_t = (1 - B)y_t = y_t - y_{t-1},$$

$$\nabla^d y_t = (1 - B)^d y_t.$$

- y_t – процесс ARIMA(p,d,q), если процесс $\nabla^d y_t$ является ARMA(p,q) процессом:

$$\phi(B)(1 - B)^d y_t = \psi(B)\varepsilon_t,$$

где

$$\phi(B) = 1 - \alpha_1 B - \alpha_2 B^2 - \ldots - \alpha_p B^p,$$

$$\psi(B) = 1 + \theta_1 B + \theta_2 B^2 + \ldots + \theta_q B^q,$$

модель ARIMA

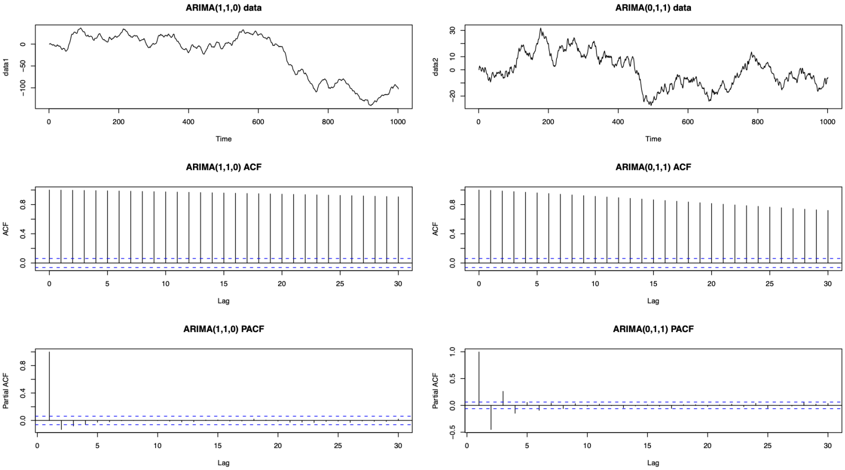

Рис.: Источник: bookdown.org

Модель SARIMA

- Процесс SARIMA(p,d,q,P,D,Q,s):

$$\Phi(B^s)\phi(B)\nabla_s^D\nabla^d y_t = \Psi(B^s)\psi(B)\varepsilon_t,$$

где

$$\nabla^d = (1-B)^d y_t, \quad \nabla_s^D = (1-B^s)^D,$$

$$\phi(B) = 1 - \alpha_1 B - \alpha_2 B^2 - \ldots - \alpha_p B^p,$$

$$\psi(B) = 1 + \theta_1 B + \theta_2 B^2 + \ldots + \theta_q B^q,$$

$$\Phi(B^s) = 1 - \widetilde{\alpha}_1 B^s - \widetilde{\alpha}_2 B^{2s} - \ldots - \widetilde{\alpha}_P B^{Ps},$$

$$\Psi(B^s) = 1 + \widetilde{\theta}_1 B^s + \widetilde{\theta}_2 B^{2s} + \ldots + \widetilde{\theta}_Q B^{Qs},$$

- В модель добавлены сезонные циклы.
- Модель SARIMAX. Добавляются внешние признаки X_t.

Другие методы прогнозирования временных рядов

- Адаптивные методы краткосрочного прогнозирования (модели Хольта-Уинтерса, Тейла-Вейджа, Тригга-Лича, …)
- Сведение задачи пронозирования временного ряда к задаче регрессии по внешним признакам.
- Нейросетевые подходы (rnn, transformers, …).

Кросс-валидация для временных рядов

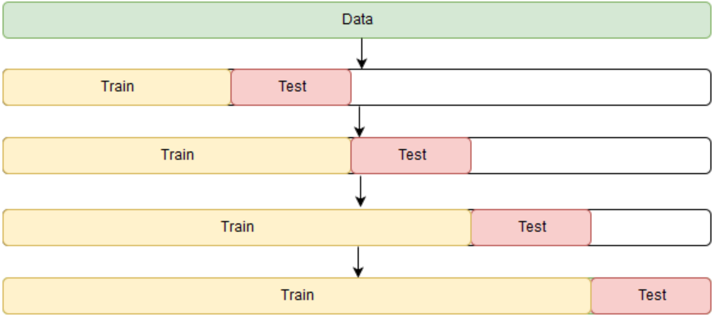

Рис.: Источник: stats.stackexchange.com

- В обучение попадают примеры из прошлого, а в валидацию – из будущего.