

Time Series

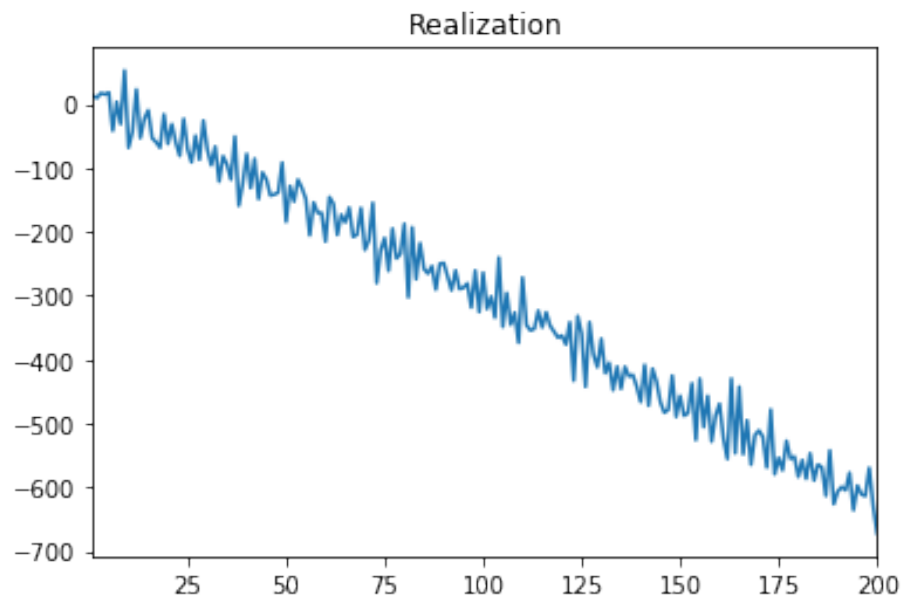
Исаев Андрей

March 2019

1 Введение

Временной ряд: в файле “input.csv”. Количество наблюдений $N = 200$

Нарисуем реализацию



Модель ВР

$$\begin{aligned}
X_t &= \mu + bt + \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} + \varepsilon_t, \\
X_t &= \mu + bt + \sum_{j=1}^p \alpha_j X_{t-1} + (X_{t-1} - X_{t-2})(-\alpha_2 - \dots - \alpha_p) \\
&= \mu + bt + \sum_{j=1}^p \alpha_j X_{t-1} + \theta_1 \Delta X_{t-1} + \dots + \theta_p \Delta X_{t-p} + \varepsilon_t, \\
\theta &= -\alpha_{j+1} - \dots - \alpha_p, \\
\Delta X_t &= X_t - X_{t-1} = \mu + bt + \left(\sum_{j=1}^p \alpha_j - 1\right) X_{t-1} + \sum_{j=1}^p \theta_j \Delta X_{t-j} + \varepsilon_t \\
\Delta X_t &= \mu + bt + \gamma X_{t-1} + \sum_{j=1}^p \theta_j \Delta X_{t-j} + \varepsilon_t \tag{1}
\end{aligned}$$

Где

$$\begin{aligned}
\varepsilon_t &\sim WN(0, \sigma^2) \\
p &\leq N^{1/3}
\end{aligned}$$

Нужно вычислить какое количество запаздываний являются значимыми и уточнить формулу (1). Количество запаздываний $p \leq 5.8$. Для этого посчитаем $p - value$ на уровне значимости $\alpha = 0.05$ для коэффициентов $\alpha_1, \dots, \alpha_5$ модели с $p = 5$ запаздываниями.

$$\begin{aligned}
\alpha_1 &= 0.000, \\
\alpha_2 &= 0.005, \\
\alpha_3 &= 0.421, \\
\alpha_4 &= 0.558, \\
\alpha_5 &= 0.855
\end{aligned}$$

Таким образом, значимые только α_1, α_2 . Модели принимает вид:

$$\Delta X_t = \mu + bt + \gamma X_{t-1} + \theta_1 \Delta X_{t-1} + \varepsilon_t \tag{2}$$

Используем Процедуру Dolado-Jenkinson-Sosvilla-Rivero для того, чтобы определить к какому типу относится ВР - TSP или DSP . Процедура состоит из пяти шагов.

Шаг1. Проверить гипотезу $H_0 : \sum_{j=1}^p \alpha_j - 1 = \gamma = 0$ против альтернативной гипотезы $H_A : \gamma < 0$, с помощью ADF-статистики $DF_{tr}(\alpha = 0.05)$, если H_0 отвергается, то $\gamma < 0$ и проверяемый ВР можно отнести к типу TSP.

$$\hat{DF}_{tr} = \frac{\hat{\gamma}}{S(\hat{\gamma})}$$

Гипотеза отвергается на уровне значимости $\alpha = 0.05$, если $\hat{DF}_{tr} < DF_{tr}(\alpha = 0.05)$

ADF Statistic: -10.485569

p-value: 0.000000

usedlags: 2

Critical Values:

1%: -4.005

5%: -3.433

10%: -3.140

Поскольку $-10.485569 < -3.433$ то необходимо отвергнуть H_0 . И временной ряд - типа TSP.

2 Оценка тренда

Рассмотрим аддитивную модель ВР: $x_t = d_t + \varepsilon_t$, $t = 1, \dots, n$

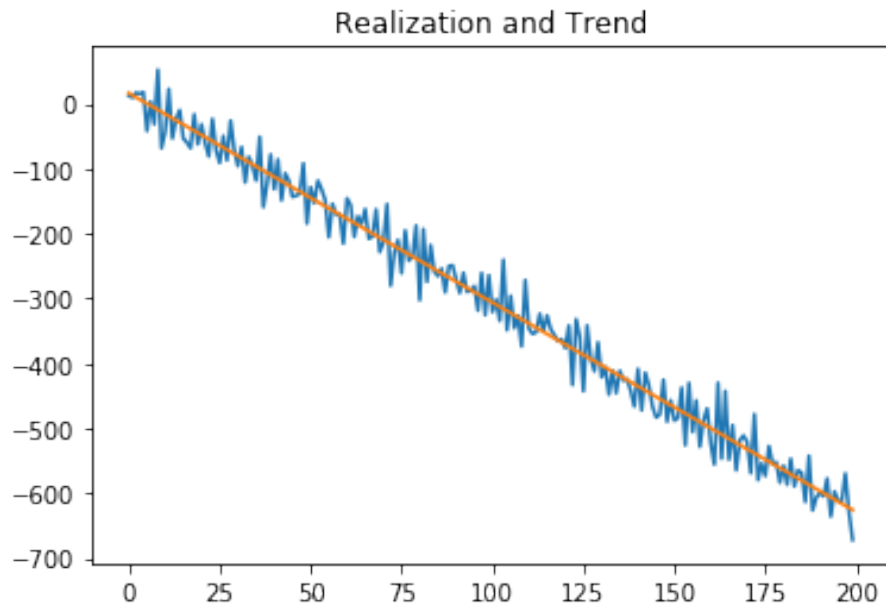
Необходимо оценить детерминированную составляющую ряда. Для этого нужно использовать Метод Наименьших Квадратов (МНК).

Пусть тренд $f(t) = \theta_0 + \theta_1 t$ - линейный тренд.

И тогда

$$\hat{\theta}_0 = \bar{y} - \hat{\theta}_1 \bar{x},$$
$$\hat{\theta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

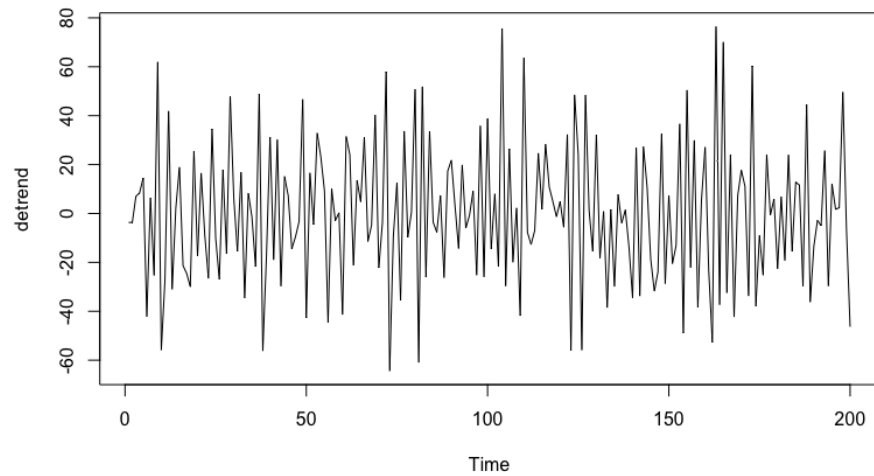
По расчетам $\hat{\theta}_0 = 19.70$, $\hat{\theta}_1 = -3.227509$



3 Детрендрование ряда

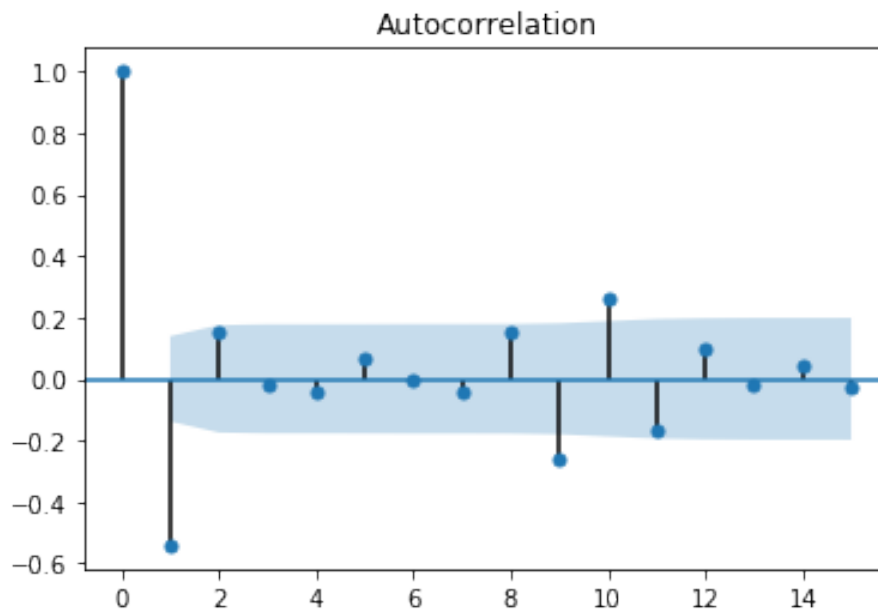
Для того чтобы детрендровать ВР, необходимо посчитать $X_i - (\hat{\theta}_0 + \hat{\theta}_1 i)$
Детрендрованный ряд находится в файле “output.csv”

Нарисуем

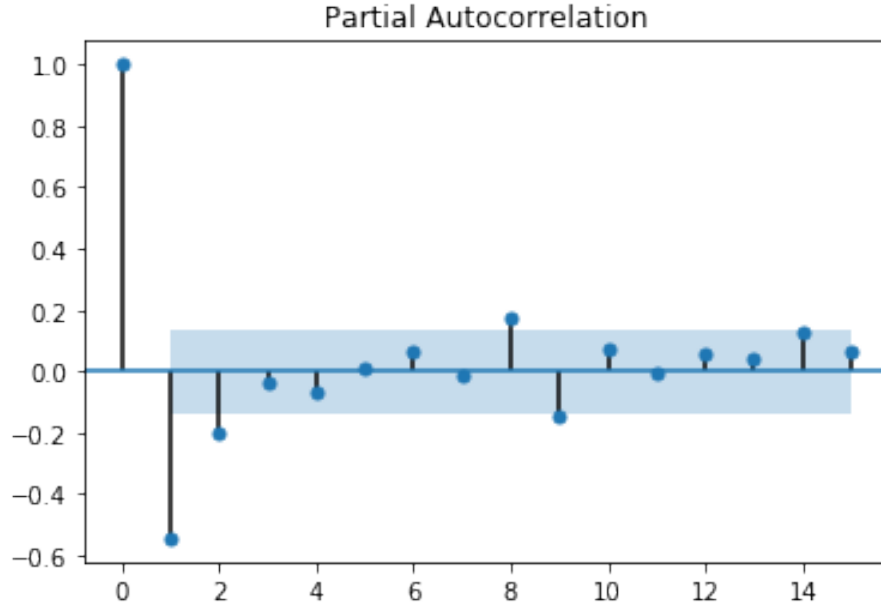


4 Идентификация случайно составляющей ряда

Необходимо построить выборочные Автокорреляционную и Частную Автокорреляционную функции для $X_{notrend}$, и опираясь на характерные свойства этих функций, выбрать наиболее вероятную модель.



$\rho = 1.000, -0.541, 0.154, -0.021, -0.046, 0.064, -0.000, -0.045, 0.151, -0.258, 0.261, -0.166, 0.100, -0.019, 0.045, -0.029, 0.030, -0.058, 0.114, -0.187, 0.132, -0.014, -0.021, -0.005, -0.023, 0.031, 0.040, -0.102, 0.118, -0.062, 0.005, -0.067, 0.016, 0.089, -0.139, 0.134, -0.072, -0.062, 0.129, -0.104, 0.119$



$\phi = 1.000, -0.544, -0.199, -0.040, -0.071, 0.007, 0.066, -0.012, 0.170, -0.144, 0.074, -0.004, 0.053, 0.038, 0.129, 0.068, 0.048, -0.002, 0.049, -0.118, -0.099, 0.050, 0.001, -0.025, -0.090, 0.004, 0.052, -0.051, -0.026, 0.066, 0.012, -0.162, -0.135, 0.119, -0.082, 0.062, 0.021, -0.069, 0.042, -0.018, 0.135$

Асимптотическое распределение выборочной ЧАКФ $\hat{\phi}(h)$ модели $AR(p)$ имеет гауссовское распределение начиная с $p + 1$ лага ($h > p$), с дисперсией

$$Var(\hat{\phi}(h)) \sim \frac{1}{n}$$

Т.е. с вероятностью 0.95 попадает в доверительный интервал $[-\frac{1.96}{\sqrt{n}}; \frac{1.96}{\sqrt{n}}]$

Если $\hat{\phi}(h)$ для $h > p$ незначимы, то предполагаем, что модель - $AR(p)$

Асимптотическое распределение выборочной АКФ $\hat{\rho}(h)$ модели $MA(q)$ имеет гауссовское распределение начиная с $q + 1$ лага ($h > q$), с дисперсией

$$Var(\hat{\rho}(h)) = \frac{1}{n} (1 + 2 \sum_{k=1}^q \rho^2(k))$$

В общем случае, если $\hat{\rho}(h)$ для $h > p$ незначимы, то предполагаем, что модель - $MA(p)$

Рассмотрев график выборочной ЧАКФ, можно сделать предположение, что искомая модель - $AR(2)$, судя по количеству значимых $\phi(h)$ - значения лагов 1,2,8,9. Однако, следует рассмотреть также модель $ARMA(1,1)$.

5.1 AR(2)

$$x_n = \varepsilon_n + \sum_{j=1}^p \alpha_j x_{n-j}$$
[illegible]

Для $p = 2$:

$$\begin{aligned} x_n &= \varepsilon_n + \alpha_1 x_{n-1} + \alpha_2 x_{n-2} \\ \alpha &= R_2^{-1} \rho_2, \\ R_2^{-1} &= \frac{1}{1 - \rho^2(1)} \begin{bmatrix} 1 & -\rho(1) \\ -\rho(1) & 1 \end{bmatrix}, \\ \rho_2 &= \begin{pmatrix} \rho(1) \\ \rho(2) \end{pmatrix}, \end{aligned}$$

$$\hat{\alpha} = \begin{pmatrix} -0.647 \\ -0.196 \end{pmatrix}$$

Модель авторегрессии-скользящего среднего порядка (p, q) представляет собой

$$x_n - \sum_{j=1}^p \alpha_j x_{n-j} = \varepsilon_n + \sum_{j=1}^q \beta_j \varepsilon_{n-j}$$

$$x_n - \alpha_1 x_{n-1} = \varepsilon_n + \beta_1 \varepsilon_{n-1}$$

Параметры для оценки: α_1, β_1

Для оценки параметров α и β модели ARMA необходимо решить систему уравнений:

$$\begin{cases} \rho_X(1) = \frac{(\alpha + \beta)(1 + \alpha\beta)}{1 + \beta^2 + 2\alpha\beta}, \\ \rho_X(2) = \frac{\alpha(\alpha + \beta)(1 + \alpha\beta)}{1 + \beta^2 + 2\alpha\beta}. \end{cases}$$

По подсчетам:

$$\hat{\alpha} = -0.285$$

$$\hat{\beta} = 0.3764$$

6 Оценка адекватности модели

6.1 Контроль качества

Для оценки качества модели будут использоваться информационные критерии Акаике (AIC) и Шварца (BIC). Информационные критерии необходимо для того, чтобы предотвратить переобучение с помощью функции потерь для каждого дополнительного параметра в модели.

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

Для $AR(2)$ модели:

$$AIC = 1843.4$$
$$BIC = 1856.59$$

Для $ARMA(1, 1)$ модели:

$$AIC = 1843.19$$
$$BIC = 1856.38$$

Согласно критериям Акаике и Шварца, лучшая модель - $AR(2)$

6.2 Проверка остатков на Белый Шум

Для проверки нормальности остатков на Белый Шум будет использоваться тест Бокса-Пирса. Проверить нулевую гипотезу $H_0: \rho_\varepsilon(1) = \dots = \rho_\varepsilon(m) = 0$ с тестовой статистикой

$$Q^* = n \sum_{j=1}^m \hat{\rho}_\varepsilon^2(j)$$

Q^* аппроксимируется распределением χ^2 . Адекватность модели отвергается на уровне значимости α если

$$Q^* > \chi_{1-\alpha}^2(m-p-q)$$

Для $AR(2)$ модели

$$\hat{\varepsilon}_t = X_t - \hat{X}_t = X_t - \alpha_1 X_{t-1} - \alpha_2 X_{t-2}, t = 2, \dots, n$$
$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{t=3}^n \hat{\varepsilon}_t^2$$

X-squared = 0.019629, df = 1, p-value = 0.8886

Нулевая гипотеза не отвергается, следовательно остатки - белый шум. Для $ARMA(1, 1)$ модели:

$$\hat{\varepsilon}_t = \frac{1 - \hat{\alpha}_1 L}{1 + \hat{\beta}_1 L} X_t$$
$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{t=3}^n \hat{\varepsilon}_t^2$$

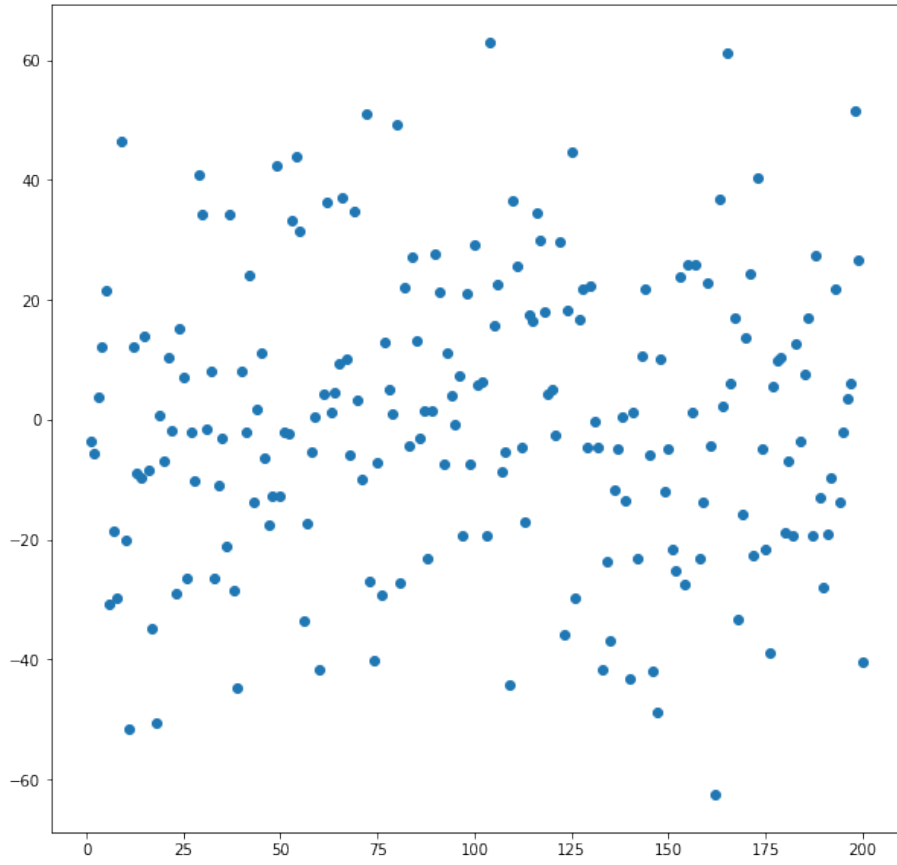
X-squared = 0.0085845, df = 1, p-value = 0.9262

Нулевая гипотеза не отвергается, следовательно остатки - белый шум.

7 Диагностика остатков

$$\hat{\varepsilon}_t = X_t - \hat{X}_t$$

$$\hat{\sigma}^2 = \frac{1}{n-p} \sum_{t=p+1}^n \hat{\varepsilon}_t^2$$



Для оценка нормальности остатков необходимо выполнить тест Харке-Бера.

$$JB = (n - p - q - 1) \left(\frac{\hat{S}^2}{6} - \frac{\hat{K}^2}{24} \right)$$

$$\hat{K} = \frac{\frac{1}{n} \sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^4}{\left(\frac{1}{n-1} \sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^2 \right)^2} - 3$$

$$\hat{S} = \frac{\frac{1}{n} \sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^3}{\left(\frac{1}{n-1} \sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^2 \right)^{\frac{3}{2}}}$$

Где \hat{S} - выборочная асимметричность, и \hat{K} - выборочный эксцесс.
Проверить гипотезу $H_0 : \{\varepsilon_t\}$ - i.i.d. ($K = 3, S = 0$) против альтернативной гипотезы $H_A : \{\varepsilon_t\}$ - не i.i.d. ($K \neq 3, S \neq 0$), с помощью χ^2 -статистики ($df = 2, \alpha = 0.05$), если H_0 не отвергается, то остатки нормальны.

sample skew of the residuals is 0.066

sample kurtosis of the residuals is 2.739

Jarque-Bera statistic for AR(2) residuals: 0.7115

p-value: 0.7

Т.е. на уровне значимости $\alpha = 0.05$ гипотеза о нормальности остатков не отвергается. Отметим, что тест на Харке-Бера не необходим для прогнозирования, но для тестов Дики-Фулера используется предположение о нормальности остатков.

8 Прогноз на один шаг

$$\begin{aligned}X_t &= \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \varepsilon_t \\X_{t+1} &= E(X_{t+1}|X_t) = E(\alpha_1 X_t + \varepsilon_t|X_t) + E(\alpha_2 X_{t-1} + \varepsilon_{t-1}|X_{t-1}) \\&= \alpha_1 X_t + \alpha_2 X_{t-1}\end{aligned}$$

По расчетам при $AR(2)$ модели:

$$\begin{aligned}\hat{X}_{200} &= -5.951; \text{ real } X_{200} = -46.334 \\ \hat{X}_{199} &= -33.203; \text{ real } X_{199} = -6.493 \\ \hat{X}_{198} &= -1.893; \text{ real } X_{198} = 49.608 \\ \hat{X}_{197} &= -3.620; \text{ real } X_{197} = 2.337 \\ \hat{X}_{196} &= -1.880; \text{ real } X_{196} = 1.702\end{aligned}$$