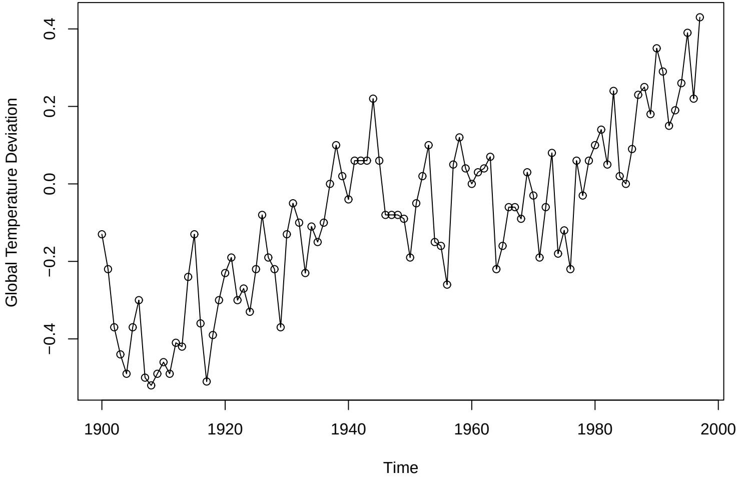
Прикладная статистика. Занятие 9. Анализ временных рядов.

26 апреля 2011 г.

Исходные данные

Отклонение от среднегодовой температуры в градусах Цельсия

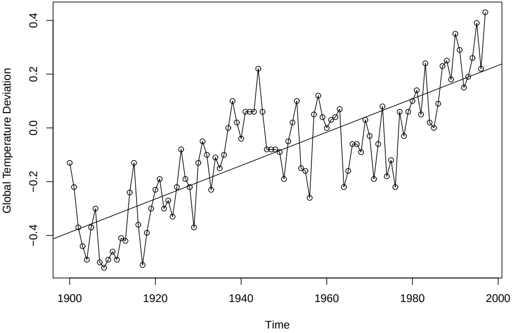


Линейный тренд: регрессия

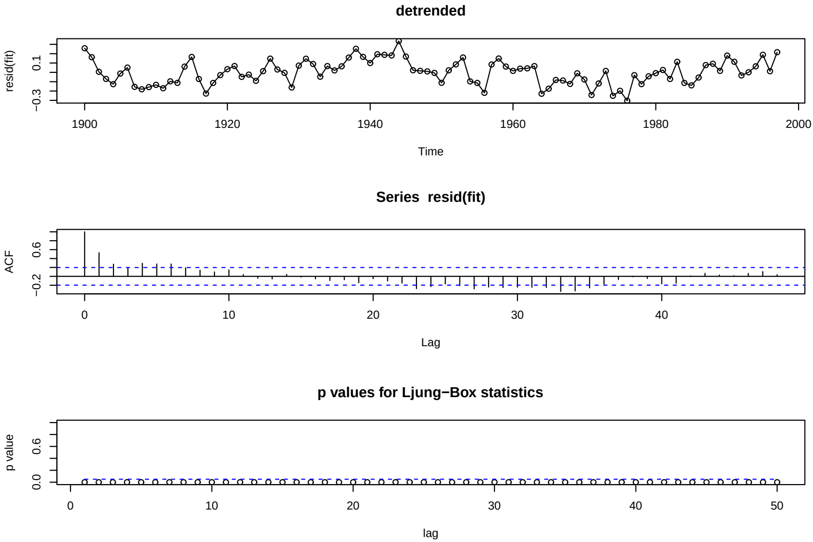
Построим зависимость отклонения температуры от года:

$$x_t = \beta_1 + \beta_2 t + \varepsilon_t, \quad t = 1900, \dots, 2000.$$

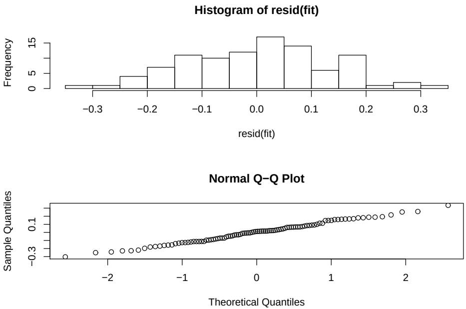
$\beta_1 = -12.186, \quad \beta_2 = 0.006;$
$SE(\beta_1) = 0.9, \quad SE(\beta_2) = 0.005;$
$R^2 = 0.6515, \quad R_A^2 = 0.6479;$
$F = 179.5, \quad p < 2.2 \times 10^{-16}.$



Остатки


detrended
resid(fit)
0.1
-0.3
1900
1920
1940
1960
1980
2000
Time
Series resid(fit)
ACF
0.6
-0.2
0
10
20
30
40
Lag
p values for Ljung-Box statistics
p value
0.6
0.0
0
10
20
30
40
50
lag

Остатки



Критерий нормальности Шапиро-Уилка: $p = 0.8618$.
Критерий гомоскедастичности Бройша-Пагана: $p = 0.9335$.
Критерий стационарности KPSS: $p > 0.1$.
Критерий независимости Бокса-Пирса: $p = 1.52 \times 10^{-7}$.
Критерий автокоррелированности Дарбина-Уотсона: $p = 1.78 \times 10^{-10}$.

Авторегрессия

$$AR\left(p\right):\quad x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots \phi_p x_{t-p} + \omega_t,$$

где x_t — стационарный ряд с нулевым средним, $\phi_1, \ldots, \phi_p$ — константы ($\phi_p \neq 0$), ω_t — гауссов белый шум с нулевым средним и постоянной дисперсией σ_ω^2.

Если среднее равно μ, модель принимает вид

$$x_t = \alpha + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots \phi_p x_{t-p} + \omega_t,$$

где $\alpha = \mu\left(1 - \phi_1 - \ldots - \phi_p\right)$.

Другой способ записи:

$$\phi\left(B\right) x_t = \left(1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p\right) x_t = \omega_t,$$

где B — разностный оператор ($B x_t = x_{t-1}$).

Линейная комбинация p подряд идущих членов ряда x_t даёт белый шум.

Скользящее среднее

$$MA\,(q): \quad x_t = \omega_t + \theta_1 \omega_{t-1} + \theta_2 \omega_{t-2} + \ldots \theta_q \omega_{t-q},$$

где x_t — стационарный ряд с нулевым средним, $\theta_1, \ldots, \theta_q$ — константы $(\theta_q \neq 0)$, ω_t — гауссов белый шум с нулевым средним и постоянной дисперсией σ_ω^2.

Если среднее равно μ, модель принимает вид

$$x_t = \mu + \omega_t + \theta_1 \omega_{t-1} + \theta_2 \omega_{t-2} + \ldots \theta_q \omega_{t-q}.$$

Другой способ записи:

$$x_t = \theta\,(B)\,\omega_t = \left(1 + \theta_1 B + \theta_2 B^2 + \ldots + \theta_q B^q\right)\omega_t,$$

где B — разностный оператор.

Линейная комбинация q компонент белого шума ω_t даёт элемент ряда.

Автокорреляции

В моделях $MA\,(q)$ автокорреляция ряда равна нулю при лаге, большем q, и строго больше нуля при лаге q.

Частичная автокорреляция стационарного ряда x_t:

$$\phi_{hh} = \begin{cases} corr(x_1, x_0), & h = 1, \\ corr(x_h - x_h^{h-1}, x_0 - x_0^{h-1}), & h \geq 2, \end{cases}$$

где x_h^{h-1} — регрессия x_h на $\{x_{h-1}, x_{h-2}, \dots, x_1\}$:

$$x_h^{h-1} = \beta_1 x_{h-1} + \beta_2 x_{h-2} + \ldots + \beta_{h-1} x_1,$$

$$x_0^{h-1} = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_{h-1} x_{h-1}.$$

В моделях $AR\,(p)$ частичная автокорреляция ряда равна нулю при лаге, большем p, и строго больше нуля при лаге p.

ARMA (Autoregressive moving average)

$$ARMA\,(p,q): \quad x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots \phi_p x_{t-p} + \omega_t + \theta_1 \omega_{t-1} + \theta_2 \omega_{t-2} + \ldots \theta_q \omega_{t-q},$$

где x_t — стационарный ряд с нулевым средним, $\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q$ — константы ($\phi_p \neq 0, \theta_q \neq 0$), ω_t — гауссов белый шум с нулевым средним и постоянной дисперсией σ_ω^2.

Если среднее равно μ, модель принимает вид

$$x_t = \alpha + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots \phi_p x_{t-p} + \omega_t + \theta_1 \omega_{t-1} + \theta_2 \omega_{t-2} + \ldots \theta_q \omega_{t-q},$$

где $\alpha = \mu\,(1 - \phi_1 - \ldots - \phi_p)\,.$

Другой способ записи:

$$\phi\,(B)\,x_t = \theta\,(B)\,\omega_t.$$

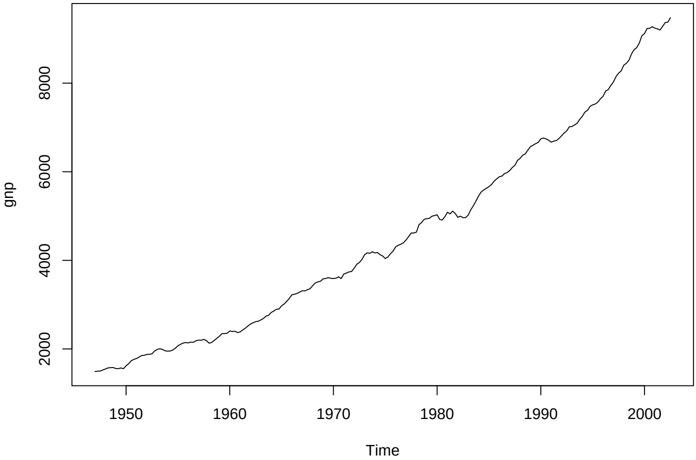
ARIMA (Autoregressive integrated moving average)

Для нестационарного ряда стационарным может оказаться ряд его разностей.

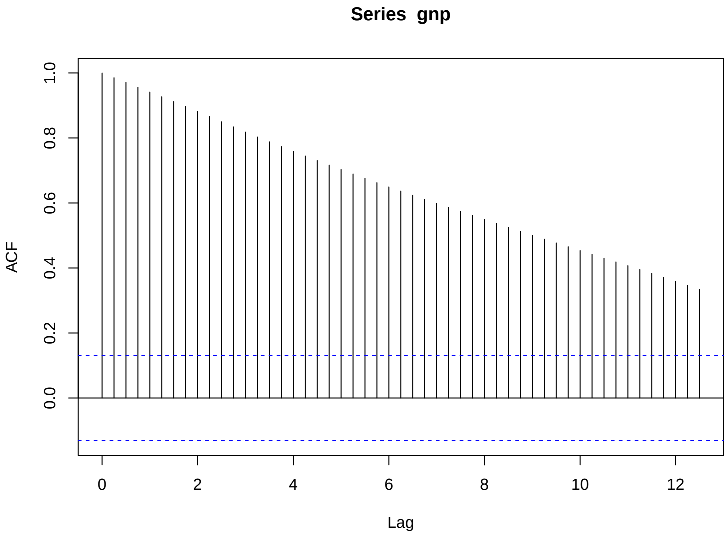
Ряд описывается моделью $ARIMA(p,d,q)$, если ряд его разностей $\nabla^d x_t = (1-B)^d$ описывается моделью $ARMA(p,q)$.

Исходные данные

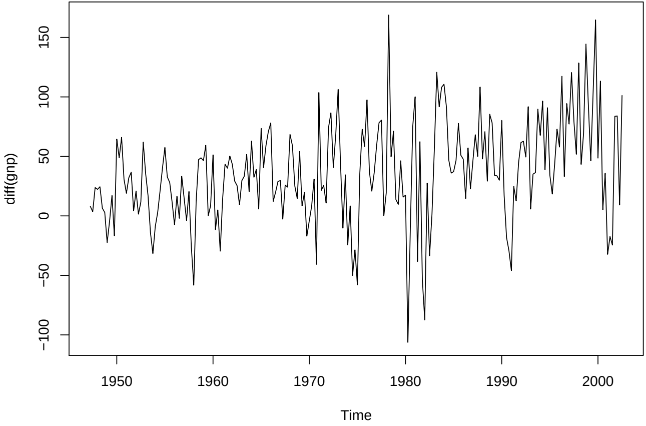
Поквартальные очищенные от сезонности данные о ВВП США в миллиардах долларов 1996 года.



Автокорреляция

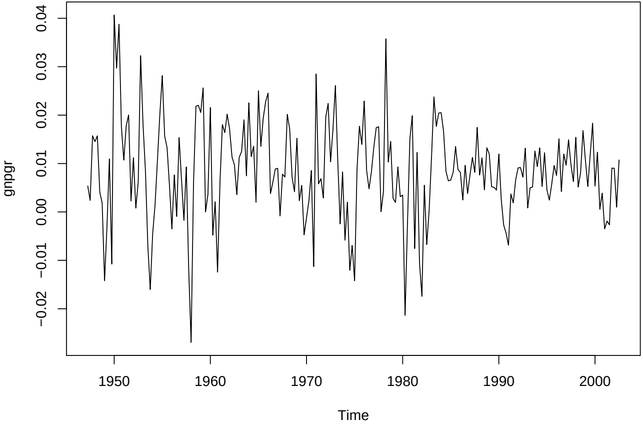


Ряд первых разностей



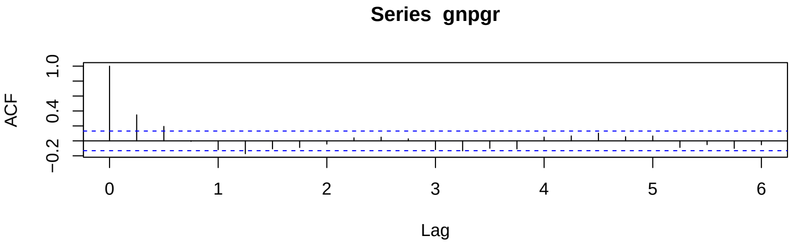
Нестационарен, вариация данных выше во второй половине ряда.

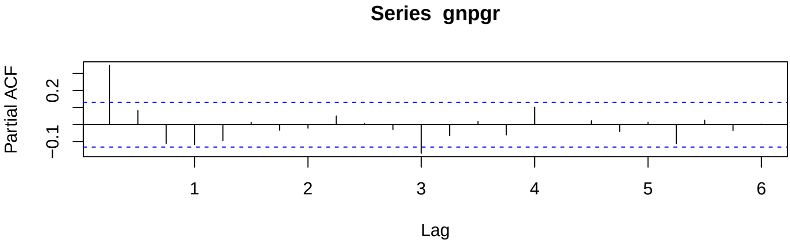
Ряд разностей логарифмов ряда



Стационарен, интерпретируется как прирост ВВП в процентах.

Автокорреляция и частная автокорреляция ряда прироста





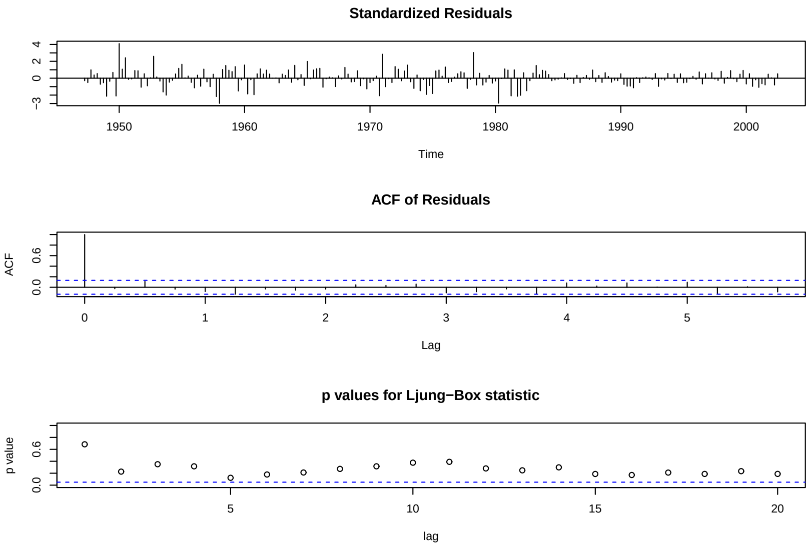
Модели

Варианты интерпретации графиков:

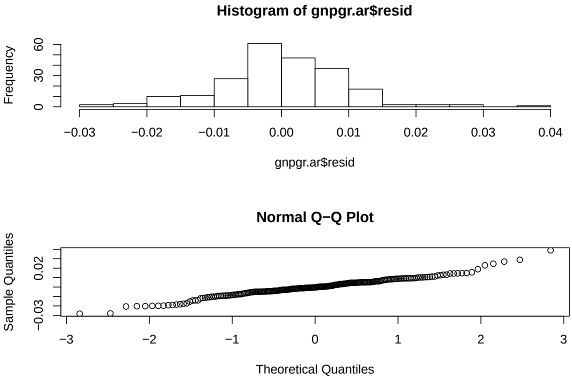
- AC равна нулю после лага 2, PAC убывает — модель $MA(2)$;
- PAC равна нулю после лага 1, AC убывает — модель $AR(1)$;
- модель $ARMA(1,2)$.

$AR(1)\colon\quad x_t = 0.005 + 0.347x_{t-1} + \hat{\omega}_t,\ \hat{\sigma}_\omega = 0.0095.$
$MA(2)\colon\quad x_t = 0.008 + 0.303\hat{\omega}_{t-1} + 0.204\hat{\omega}_{t-2} + \hat{\omega}_t,\ \hat{\sigma}_\omega = 0.0094.$
$ARMA(1,2)\colon\quad x_t = 0.008 + 0.241x_{t-1} + 0.076\hat{\omega}_{t-1} + 0.162\hat{\omega}_{t-2} + \hat{\omega}_t,\ \hat{\sigma}_\omega = 0.0089.$

Диагностика $AR(1)$

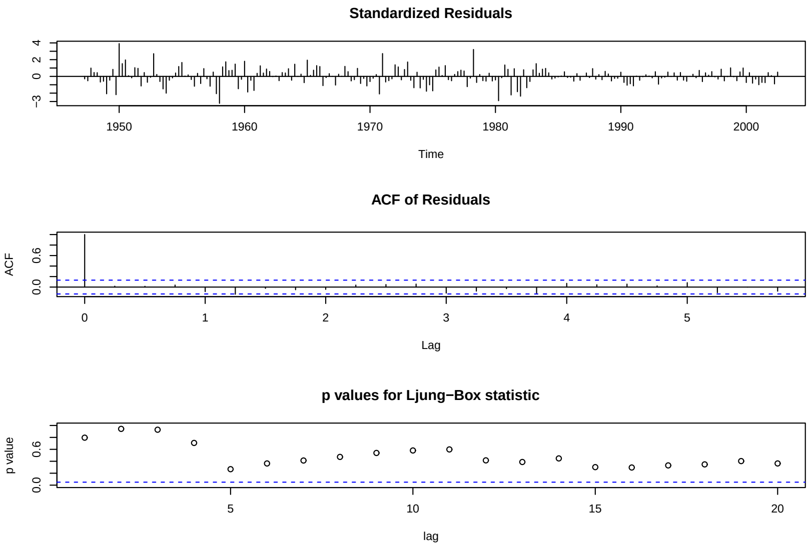


Диагностика $AR(1)$

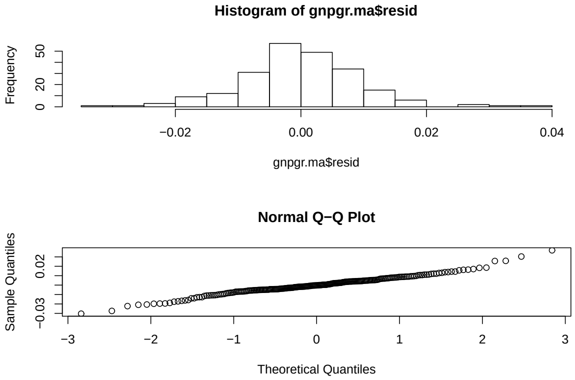


Критерий нормальности Шапиро-Уилка: $p = 0.0006886$.
Критерий стационарности KPSS: $p > 0.1$.
Критерий независимости Бокса-Пирса: $p = 0.6867$.

Диагностика $MA(2)$

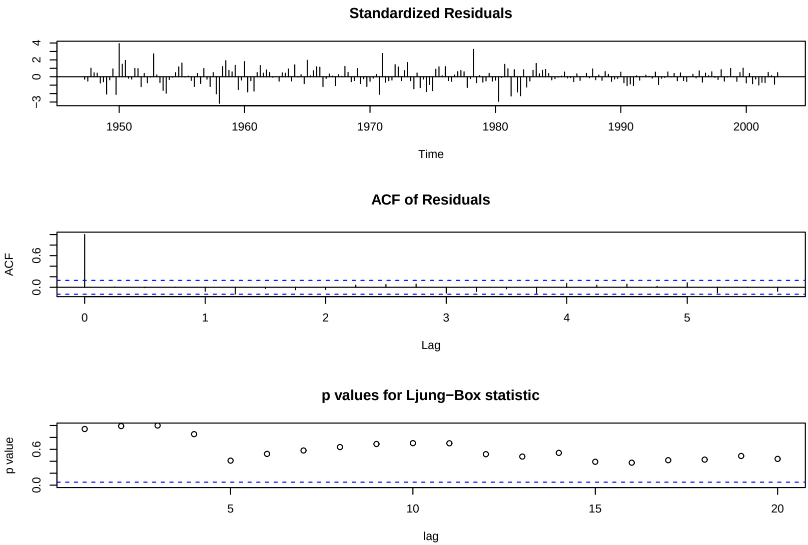


Диагностика $MA(2)$

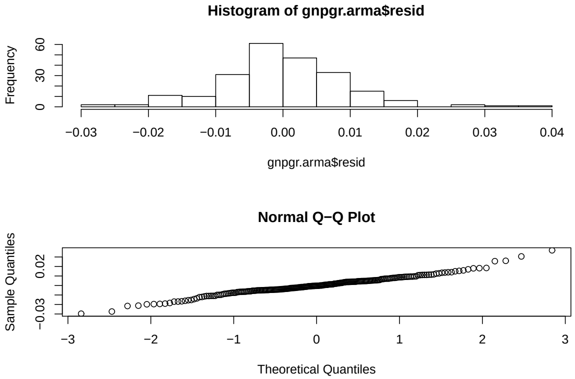


Критерий нормальности Шапиро-Уилка: $p = 0.003416$.
Критерий стационарности KPSS: $p > 0.1$.
Критерий независимости Бокса-Пирса: $p = 0.797$.

Диагностика $ARMA(1,2)$



Диагностика $ARMA(1,2)$



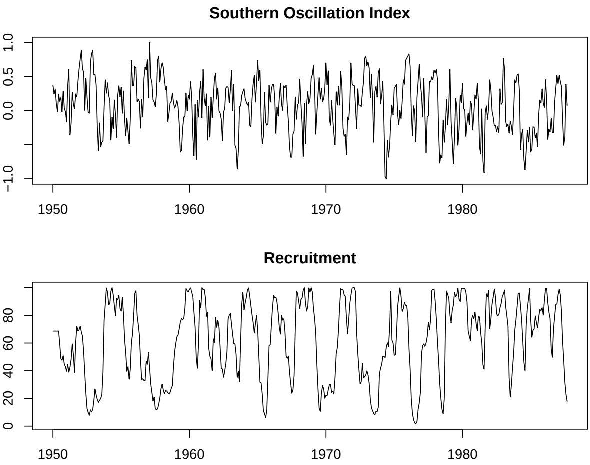
Критерий нормальности Шапиро-Уилка: $p = 0.003497$.
Критерий стационарности KPSS: $p > 0.1$.
Критерий независимости Бокса-Пирса: $p = 0.9411$.

Сравнение моделей

AIC — информационный критерий Акаике;
AICc — он же с поправкой на случай небольшого размера выборки;
BIC (SIC) — байесовский (Шварца) информационный критерий.

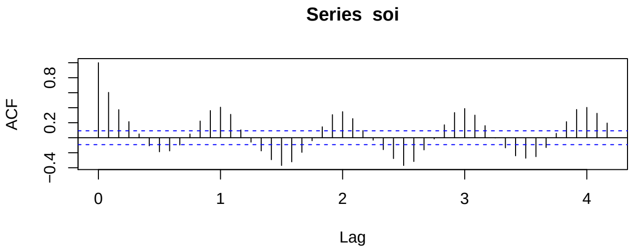
	AIC	AICc	BIC (SIC)
$AR(1)$	-1431.221	-8.284898	-9.263748
$MA(2)$	-1431.929	-8.297199	-9.276049
$ARMA(1,2)$	**-1430.948**	**-8.301886**	**-9.280737**

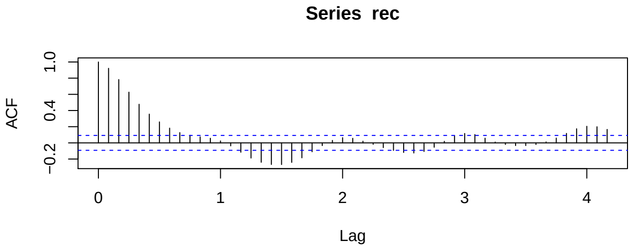
Исходные данные



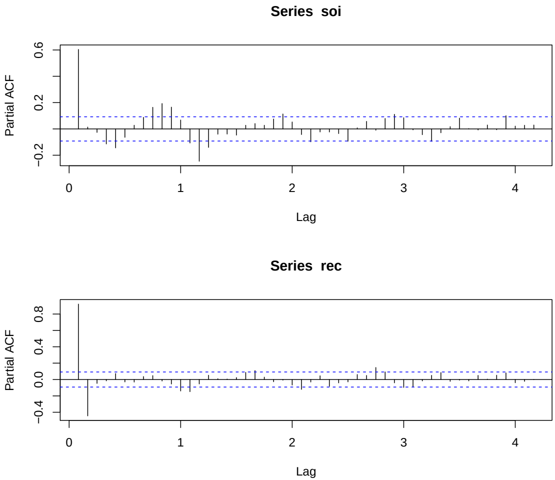
Southern Oscillation Index — изменения атмосферного давления, связанные с колебаниями температуры поверхности воды в Тихом океане.
Recruitment — число новых особей рыбы.

Автокорреляции





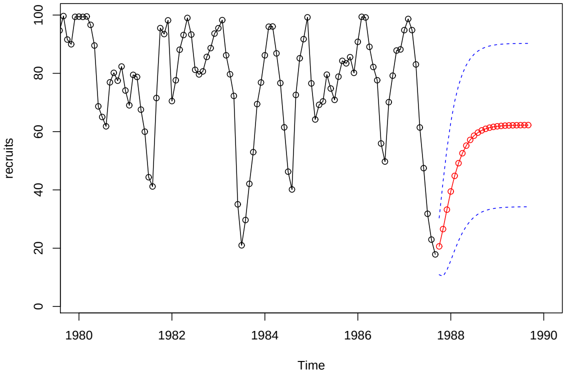
Частичные автокорреляции



Прогнозирование ряда Recruitment

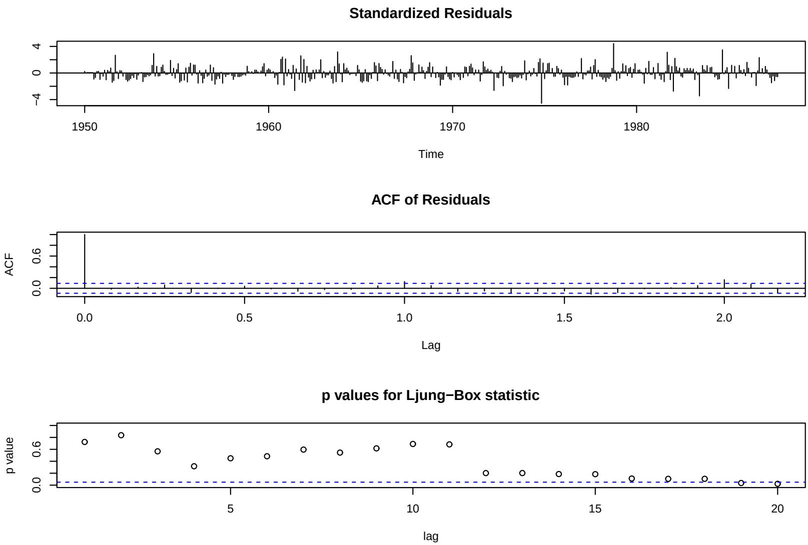
Выбор — модель $AR(2)$:

$$x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \omega_t.$$

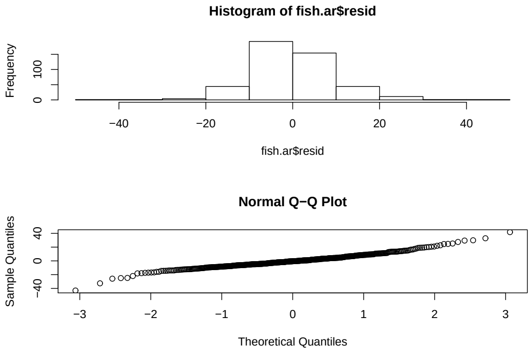


В моделях $ARMA(p, q)$ с увеличением горизонта прогноз всё больше похож на константу.

Остатки


Standardized Residuals
4
0
−4
1950
1960
1970
1980
Time
ACF of Residuals
ACF
0.0
0.6
0.0
0.5
1.0
1.5
2.0
Lag
p values for Ljung−Box statistic
p value
0.0
0.6
5
10
15
20
lag

Остатки



Критерий нормальности Шапиро-Уилка: $p = 2.72 \times 10^{-7}$.
Критерий стационарности KPSS: $p > 0.1$.
Критерий независимости Бокса-Пирса: $p = 0.7248$.

Seasonal multiplicative ARMA/ARIMA

$$ARMA(p,q) \times (P,Q)_s: \quad \Phi_P(B^s)\phi(B)x_t = \Theta_Q(B^s)\theta(b)\omega_t,$$

где

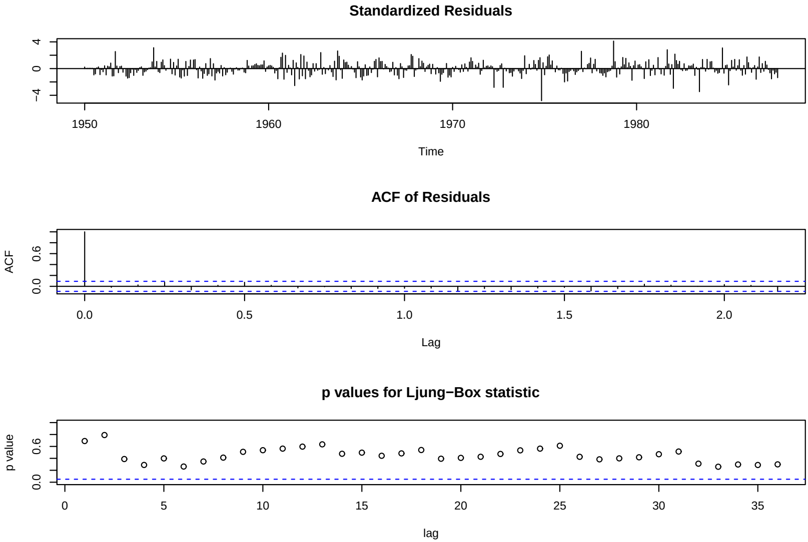
$$\Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \ldots - \Phi_P B^{Ps},$$
$$\Theta_Q(B^s) = 1 + \Theta_1 B^s + \Theta_2 B^{2s} + \ldots + \Theta_Q B^{Qs}.$$

$$SARIMA: \quad \Phi_P(B^s)\phi(B)\nabla_s^D \nabla^d x_t = \alpha + \Theta_Q(B^s)\theta(b)\omega_t,$$

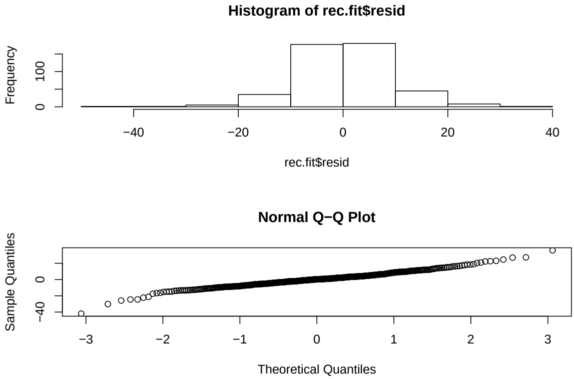
Прогнозирование ряда Recruitment

	AIC
$ARMA(2,0) \times (3,0)_{12}$	3308.515
$ARMA(2,0) \times (2,0)_{12}$	3316.283
$ARMA(2,0) \times (1,0)_{12}$	3325.706
$ARMA(2,0) \times (0,1)_{12}$	3327.352
$ARMA(2,0) \times (0,2)_{12}$	3321.88
$ARMA(2,0) \times (0,3)_{12}$	3314.787
$ARMA(2,0) \times (1,1)_{12}$	3283.717

Диагностика $ARMA(2,0) \times (1,1)_{12}$

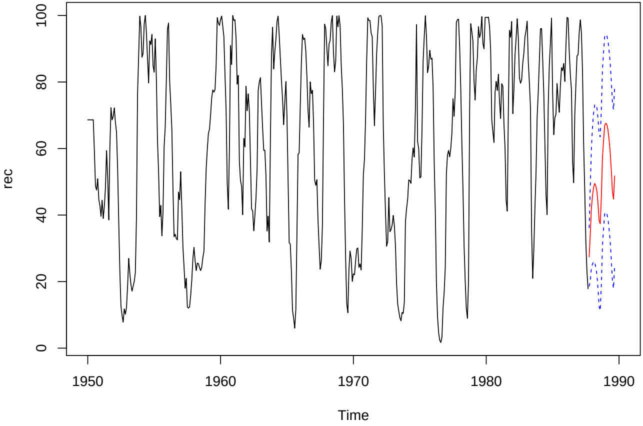


Диагностика $ARMA(2,0) \times (1,1)_{12}$

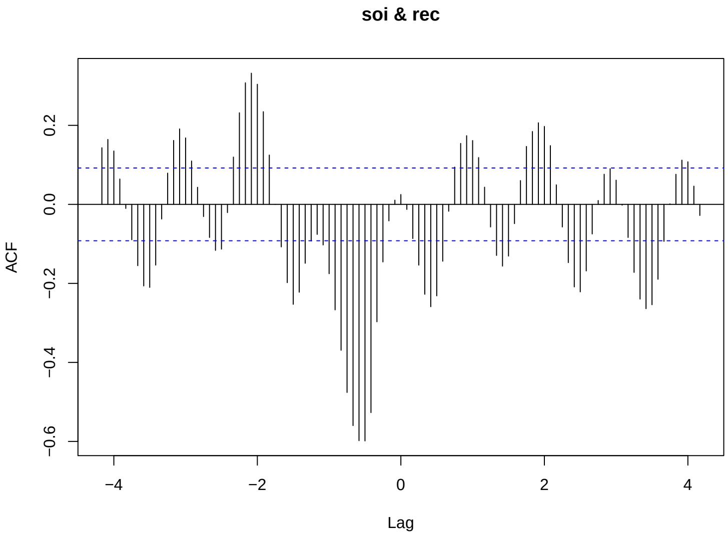


Критерий нормальности Шапиро-Уилка: $p = 6.781 \times 10^{-6}$.
Критерий стационарности KPSS: $p > 0.1$.
Критерий независимости Бокса-Пирса: $p = 0.6915$.

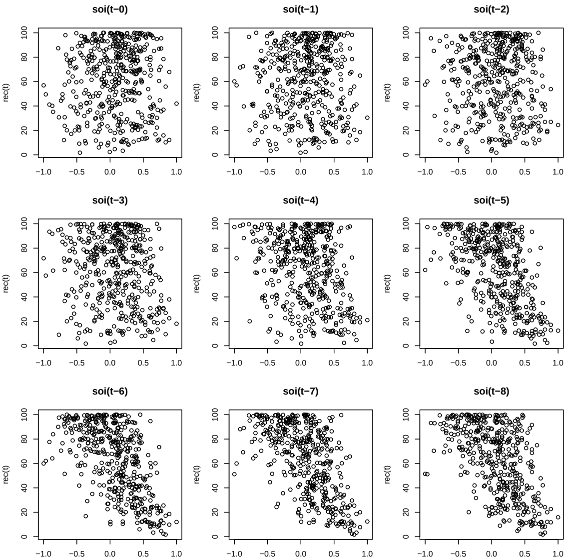
Прогнозирование ряда Recruitment



Кросс-корреляция



Кросс-корреляция



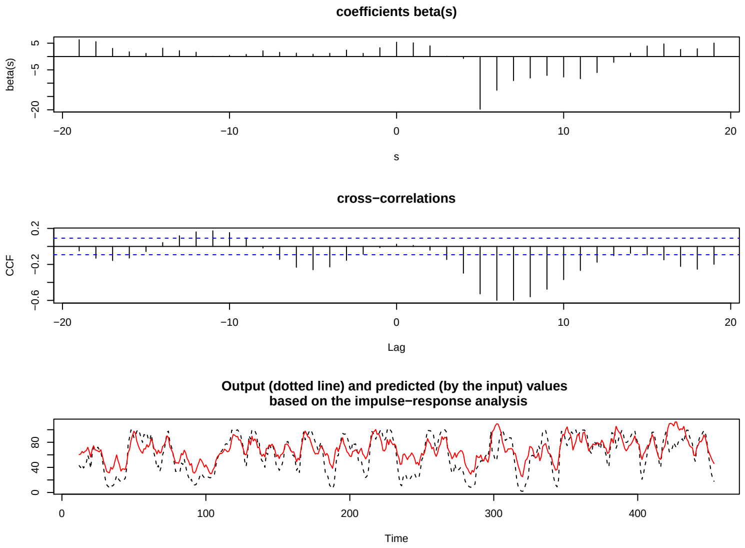
Lagged regression models

$$y_t = \sum_{r=-\infty}^{\infty} \beta_r x_{t-r} + v_t,$$

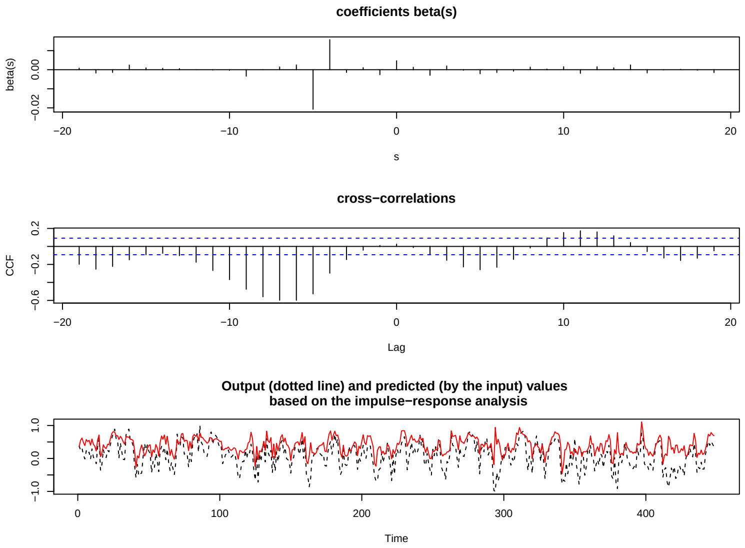
где v_t — стационарный шум.

ДАЛЕЕ ИСПОЛЬЗУЕТСЯ СПЕКТРАЛЬНЫЙ АНАЛИЗ
(Impulse-response analysis)
(изучайте сами, если надо)

Прогнозирование ряда Recruitment по SOI



Прогнозирование ряда SOI по Recruitment



VARIMA

Рассматривается и прогнозируется векторный ряд $[SOI, Recrutment]$, можно задавать любые варианты зависимости между рядами.

...

Прикладная статистика
Семинар 9. Анализ временных рядов.

Рябенко Евгений
riabenko.e@gmail.com