

Прикладная статистика. Занятие 10. Анализ временных рядов.

25 апреля 2012 г.

Линейный тренд: регрессия

Построим зависимость отклонения температуры от года:

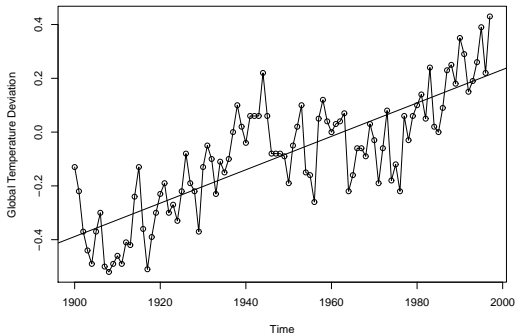
$$x_t = \beta_1 + \beta_2 t + \varepsilon_t, \quad t = 1900, \dots, 2000.$$

$$\beta_1 = -12.186, \quad \beta_2 = 0.006;$$

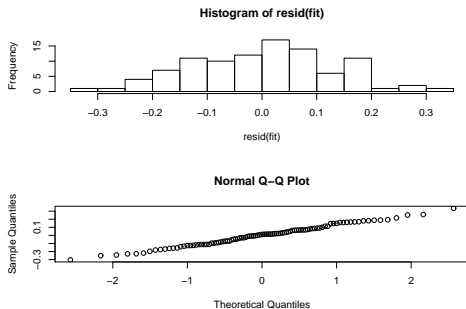
$$SE(\beta_1) = 0.9, \quad SE(\beta_2) = 0.005;$$

$$R^2 = 0.6515, \quad R_A^2 = 0.6479;$$

$$F = 179.5, \quad p < 2.2 \times 10^{-16}.$$



Остатки



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

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

где x_t — стационарный ряд с нулевым средним, ϕ_1, \dots, ϕ_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

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

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

Другой способ записи:

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

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

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

Скользящее среднее

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

где x_t — стационарный ряд с нулевым средним, $\theta_1, \dots, \theta_q$ — 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

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

Другой способ записи:

$$x_t = \theta(B)\omega_t = (1 + \theta_1B + \theta_2B^2 + \dots + \theta_qB^q)\omega_t,$$

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 $\{x_{h-1}, x_{h-2}, \dots, x_1\}$:

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

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

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

ARMA (Autoregressive moving average)

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

где x_t — стационарный ряд с нулевым средним, $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$ — 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

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

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

Другой способ записи:

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

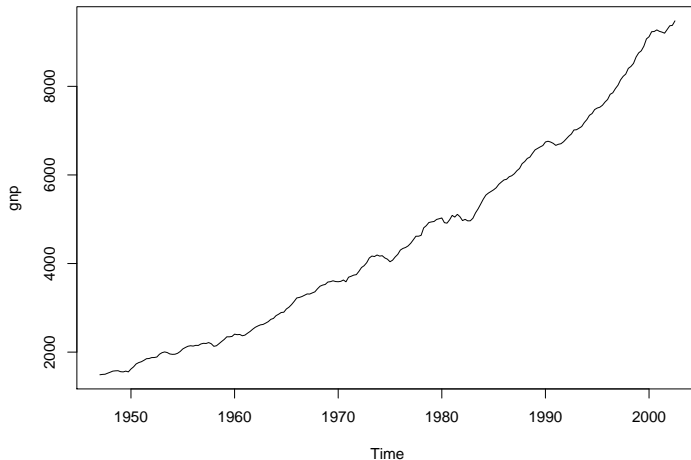
ARIMA (Autoregressive integrated moving average)

Для нестационарного ряда стационарным может оказаться ряд его разностей.

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

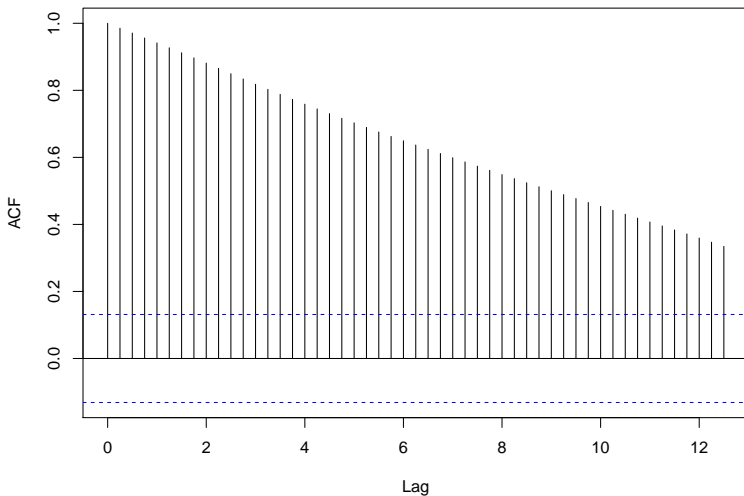
Исходные данные

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

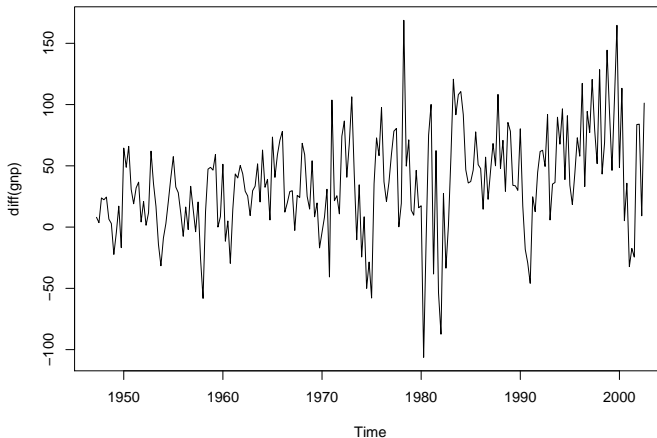


Автокорреляция

Series gnp

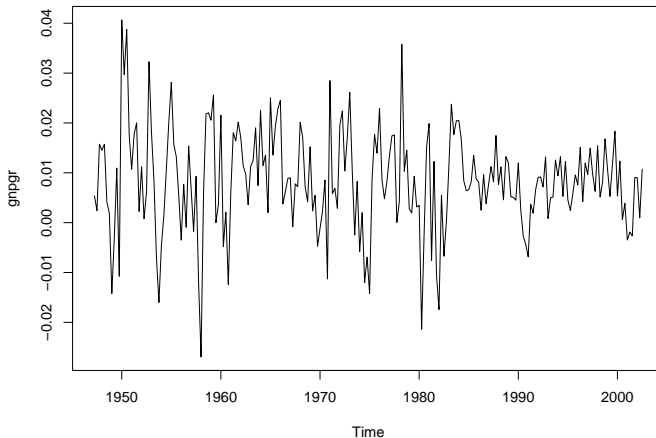


Ряд первых разностей



Нестационарен, вариация данных выше во второй половине ряда (KPSS $p < 0.01$).

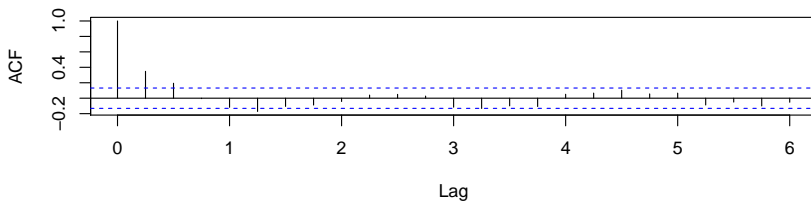
Ряд разностей логарифмов ряда



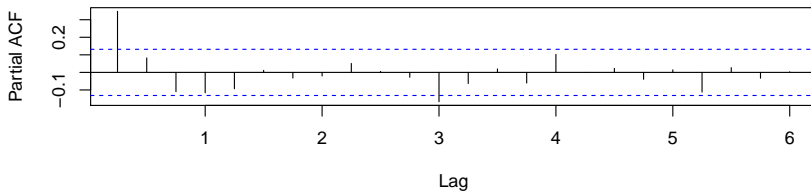
Стационарен (KPSS $p > 0.1$), интерпретируется как прирост ВВП в процентах.

Автокорреляция и частная автокорреляция ряда прироста

Series gnpgr



Series gnpgr



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 $ARMA(1, 2)$.

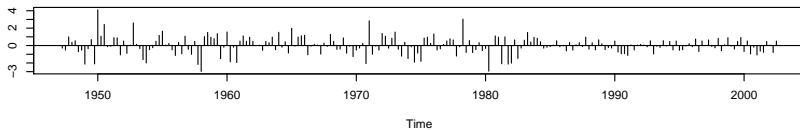
$$AR(1): x_t = 0.005 + 0.347x_{t-1} + \hat{\omega}_t, \hat{\sigma}_\omega = 0.0095.$$

$$MA(2): x_t = 0.008 + 0.303\hat{\omega}_{t-1} + 0.204\hat{\omega}_{t-2} + \hat{\omega}_t, \hat{\sigma}_\omega = 0.0094.$$

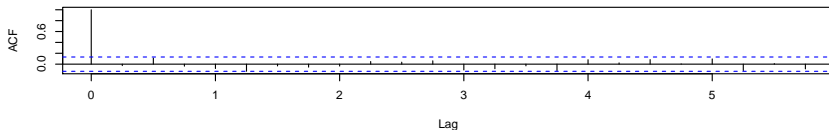
$$ARMA(1, 2): x_t = 0.008 + 0.241x_{t-1} + 0.076\hat{\omega}_{t-1} + 0.162\hat{\omega}_{t-2} + \hat{\omega}_t, \hat{\sigma}_\omega = 0.0089.$$

Диагностика $AR(1)$

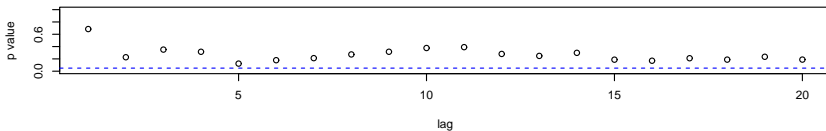
Standardized Residuals

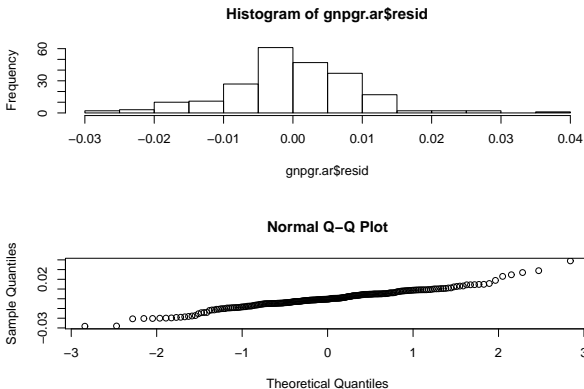


ACF of Residuals



p values for Ljung-Box statistic



Диагностика $AR(1)$ 

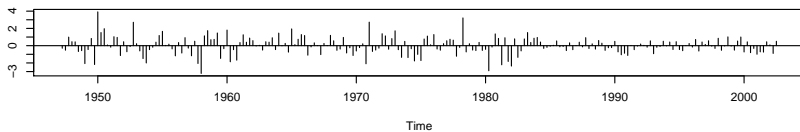
Критерий нормальности Шапиро-Уилка: $p = 0.0006886$.

Критерий стационарности KPSS: $p > 0.1$.

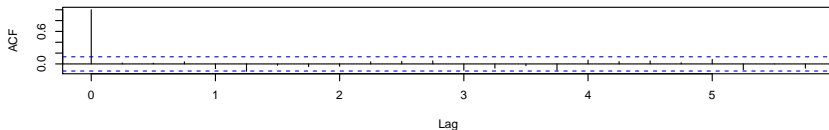
Критерий независимости Бокса-Пирса: $p = 0.6867$.

Диагностика $MA(2)$

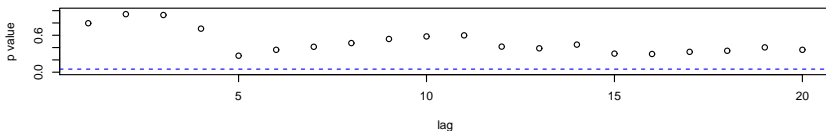
Standardized Residuals

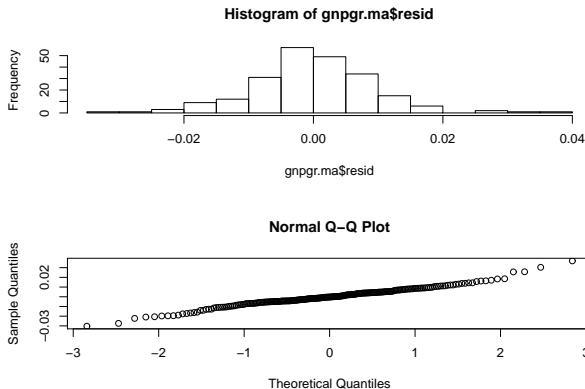


ACF of Residuals



p values for Ljung-Box statistic



Диагностика $MA(2)$ 

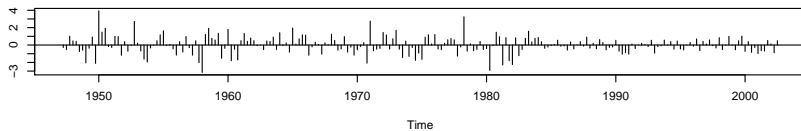
Критерий нормальности Шапиро-Уилка: $p = 0.003416$.

Критерий стационарности KPSS: $p > 0.1$.

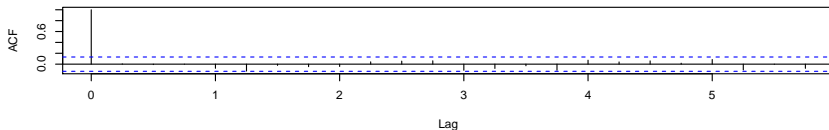
Критерий независимости Бокса-Пирса: $p = 0.797$.

Диагностика $ARMA(1,2)$

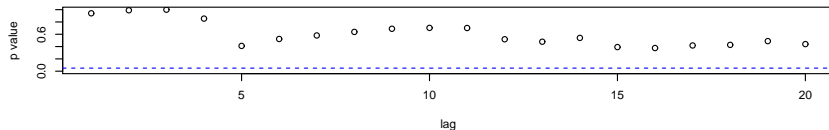
Standardized Residuals

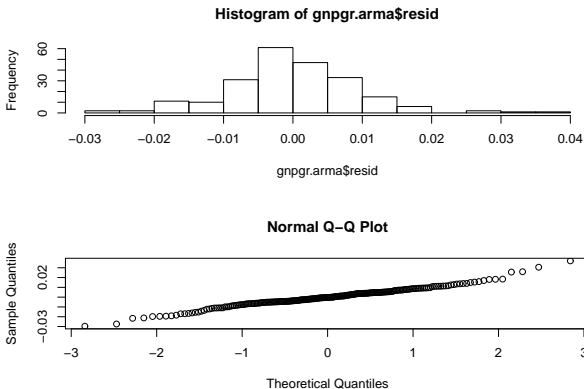


ACF of Residuals



p values for Ljung-Box statistic



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

$\hat{\sigma}^2 = \frac{SSE_k}{n}$, k — число параметров модели.

Информационный критерий Акаике:

$$AIC = \log \hat{\sigma}^2 + \frac{n + 2k}{n}$$

Он же с поправкой на случай небольшого размера выборки:

$$AIC_c = \log \hat{\sigma}^2 + \frac{n + k}{n - k - 2}$$

Байесовский (Шварца) информационный критерий:

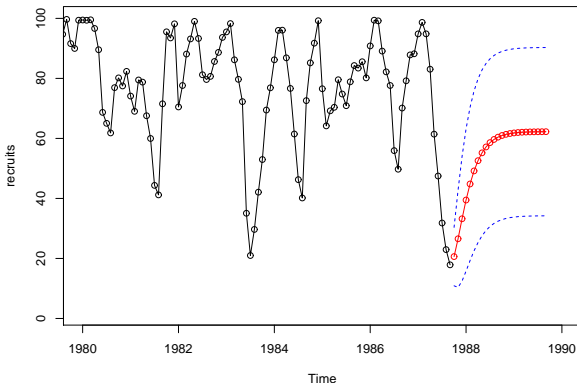
$$BIC = \log \hat{\sigma}^2 + \frac{k \log n}{n}$$

	AIC	AIC _c	BIC (SIC)
<i>AR</i> (1)	-1431.221	-8.284898	-9.263748
<i>MA</i> (2)	-1431.929	-8.297199	-9.276049
<i>ARMA</i> (1, 2)	-1430.948	-8.301886	-9.280737

Прогнозирование ряда

Выбор — модель $AR(2)$:

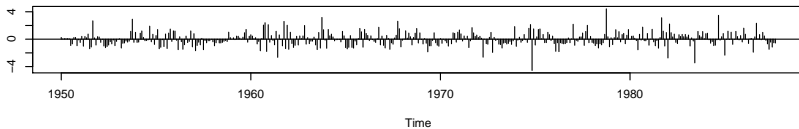
$$x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \omega_t.$$



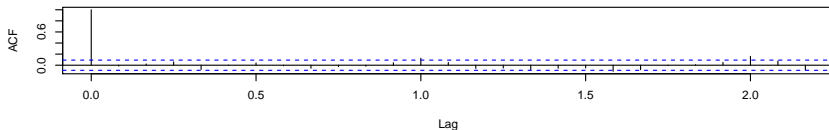
В моделях $ARMA(p, q)$ с увеличением горизонта прогноз всё больше похож на константу.

Остатки

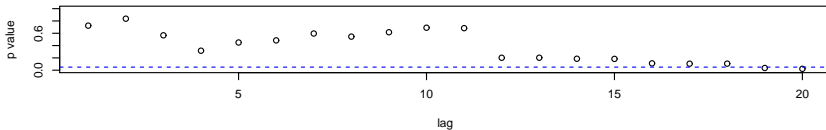
Standardized Residuals



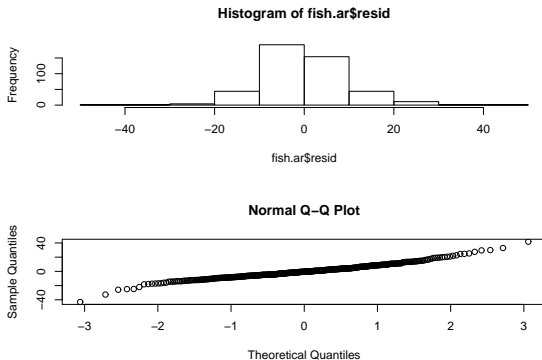
ACF of Residuals



p values for Ljung-Box statistic



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

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

где

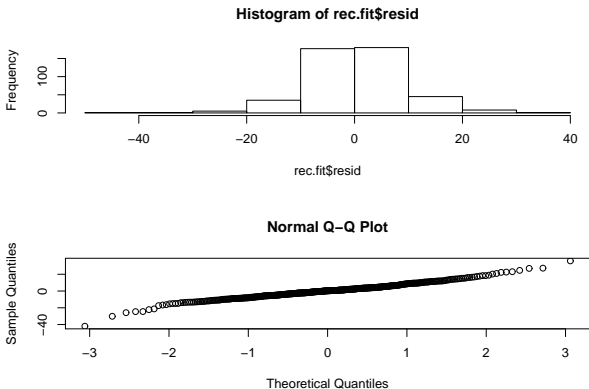
$$\Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_P B^{Ps},$$

$$\Theta_Q(B^s) = 1 + \Theta_1 B^s + \Theta_2 B^{2s} + \dots + \Theta_Q B^{Qs}.$$

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

Сравнение моделей

	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

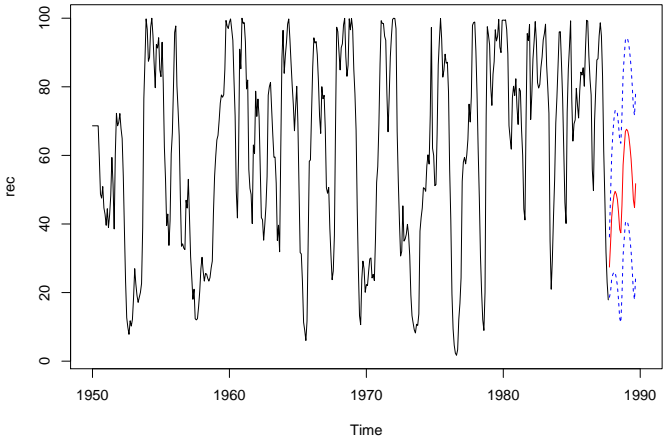
Диагностика $ARMA(2, 0) \times (1, 1)_{12}$ 

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



Прикладная статистика
Семинар 10. Анализ временных рядов.

Рябенко Евгений
riabenko.e@gmail.com