

Прогнозирование временных рядов



Содержание

- Примеры задач
- Модель авторегрессии
- Модель скользящего среднего
- Модель ARMA
- ARIMA - интегрированная ARMA
- Подбор параметров модели.
Авторегрессионный спектр

Примеры задач

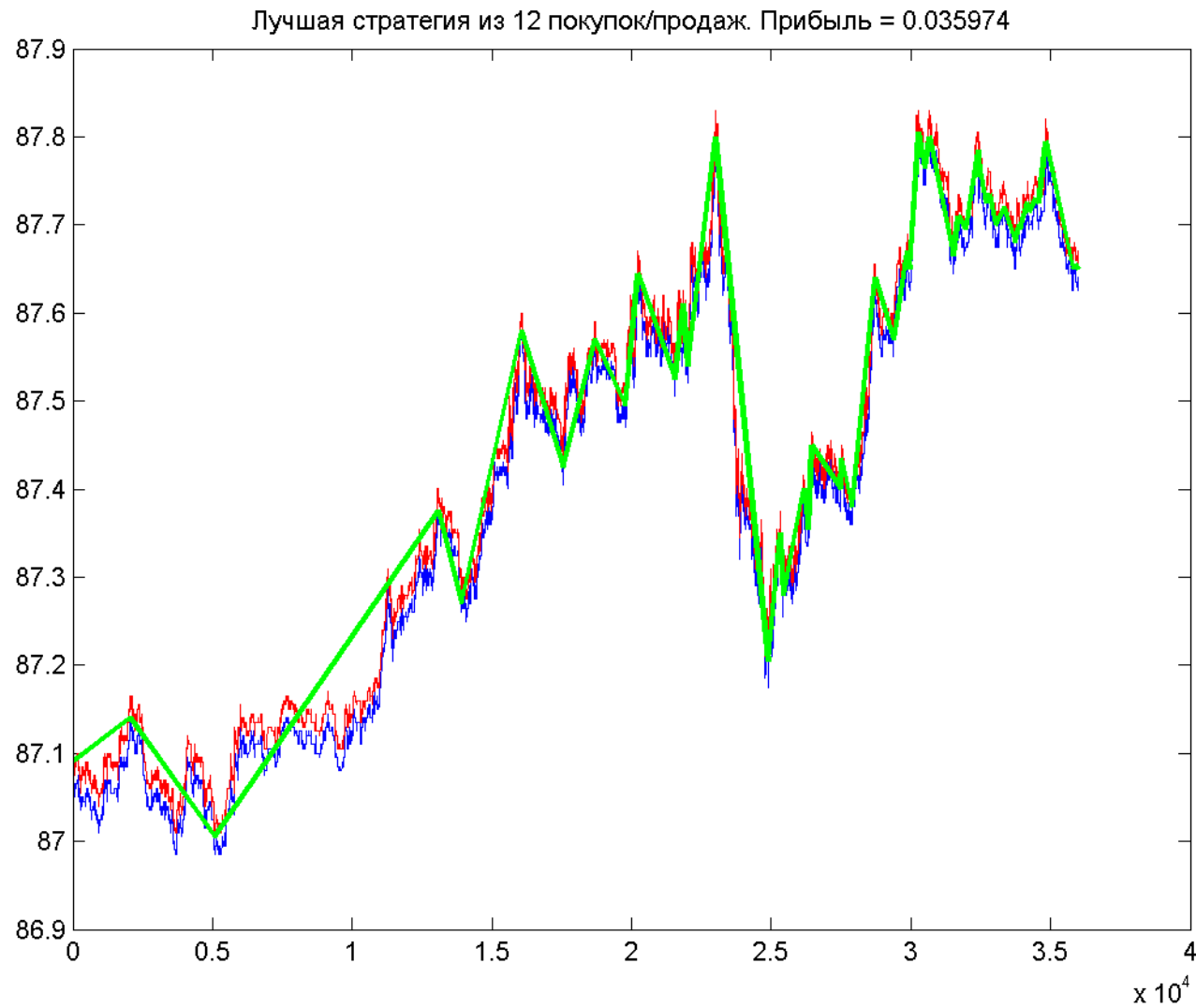
Динамика цен на нефть Brent (ICE.Brent, USD за баррель)



Динамика кросс-курса евро к доллару США (EUR/USD)

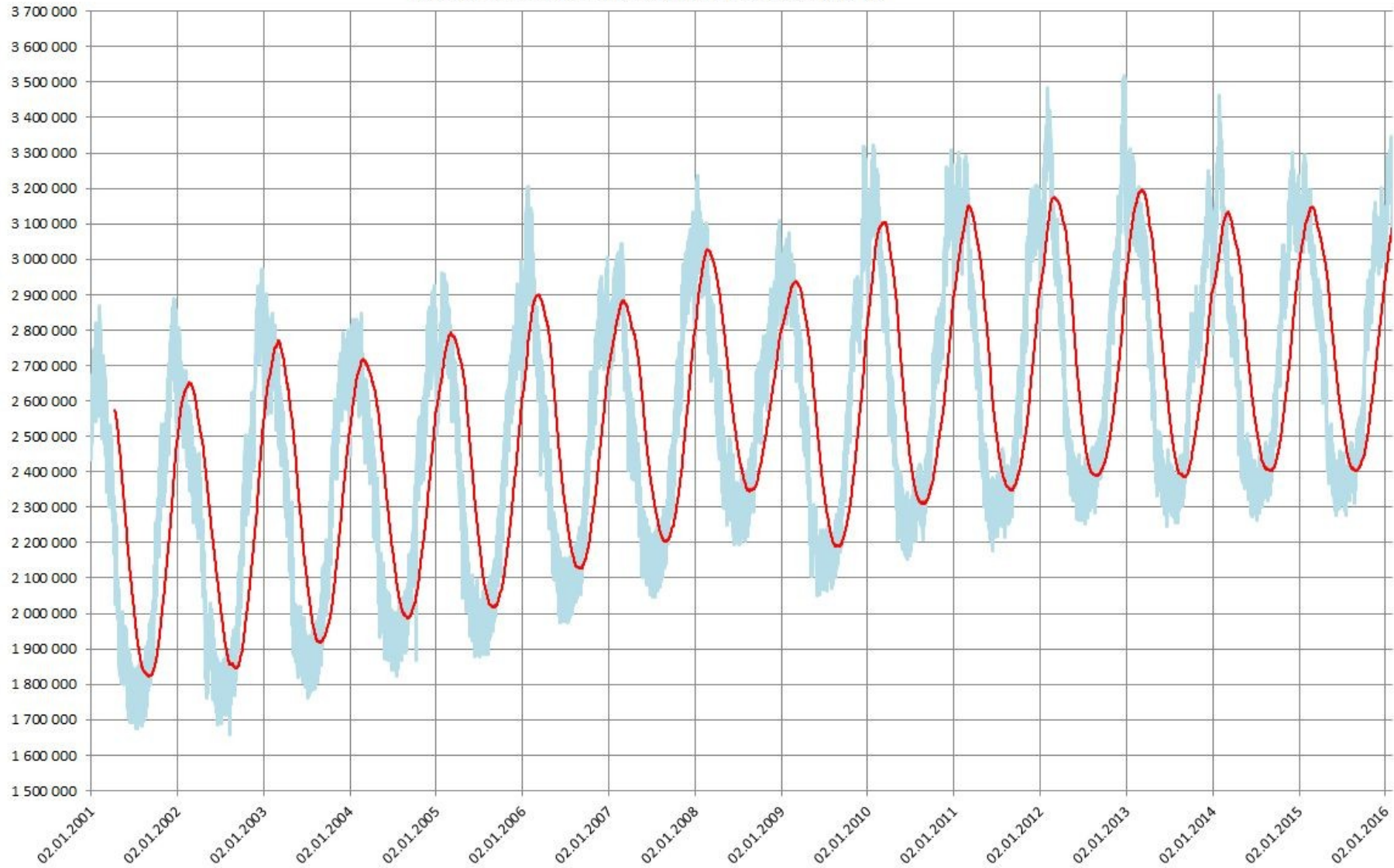


Тиковые данные

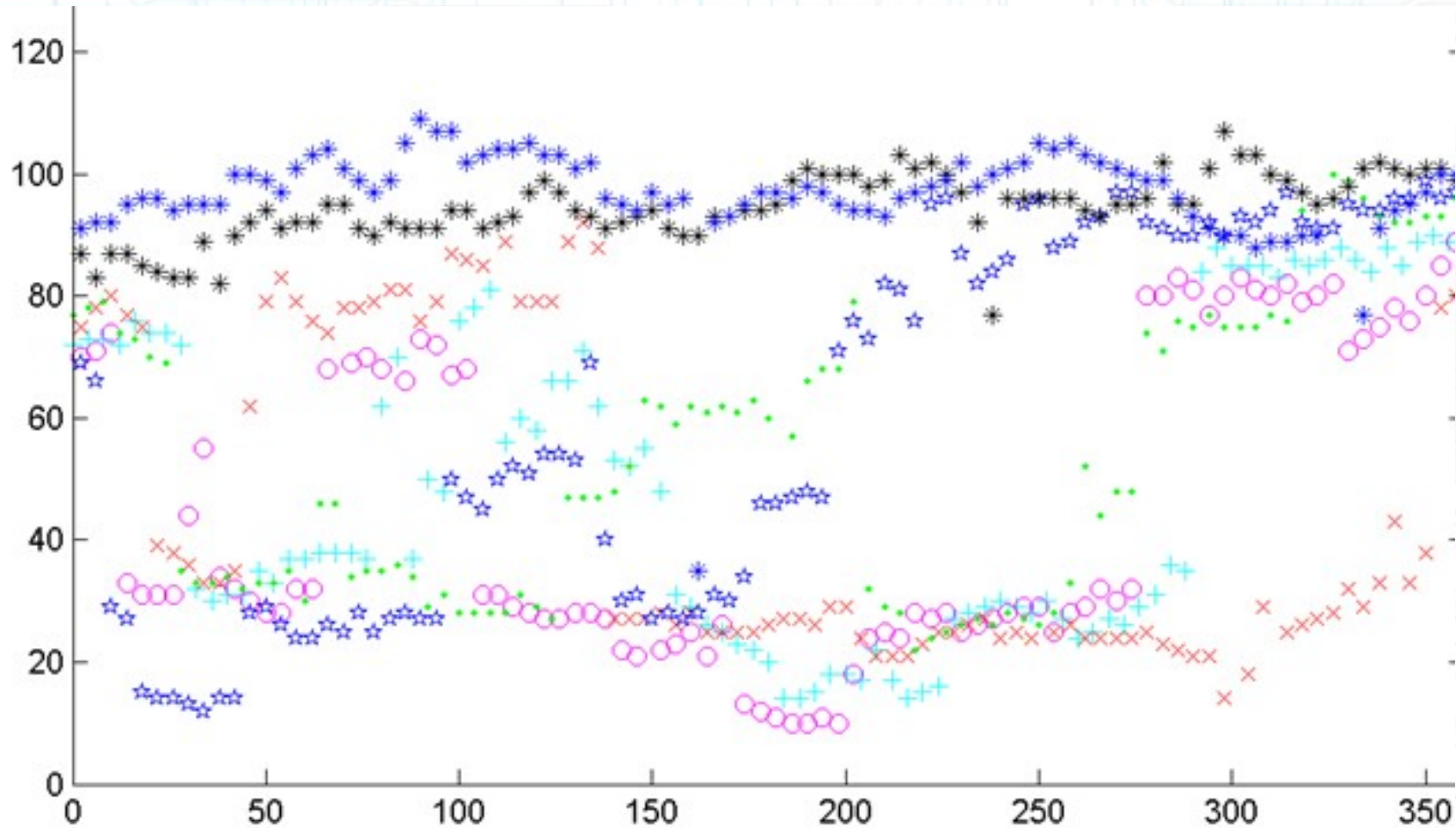


Потребление электроэнергии

Потребление электроэнергии по данным ЕЭС России (МВт*ч)



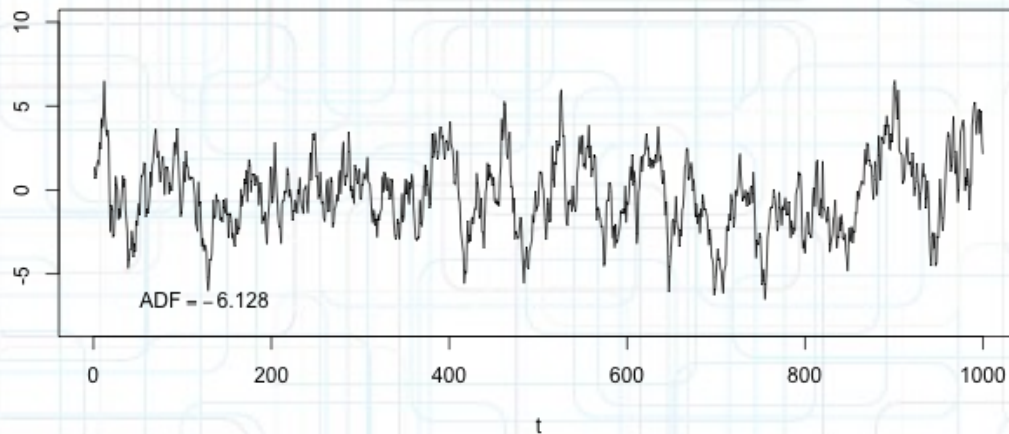
Скорость движения автомобилей в разные дни недели с 15:00 до 21:00



Авторегрессионная модель

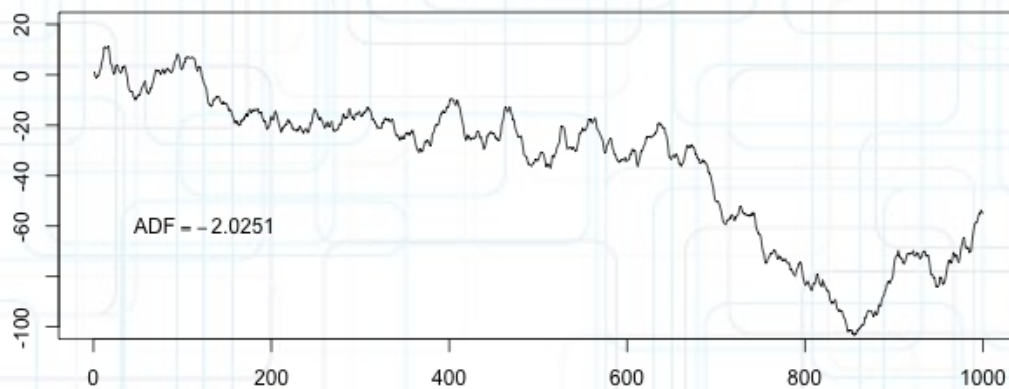
- Случайный процесс называется стационарным, если случайное распределение значений функции зависит только от предыдущих значений и расстояния по времени до них, но не от самих значений времени

Stationary Time Series



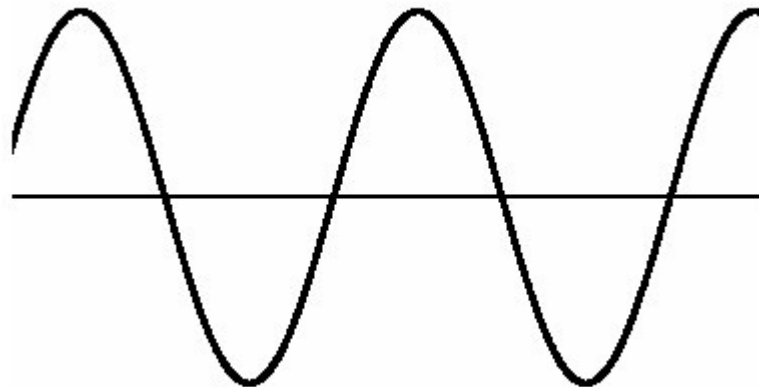
$$X_t = c + \sum_{i=1}^p a_i X_{t-i} + \varepsilon_t,$$

Non-stationary Time Series



Авторегрессионная модель

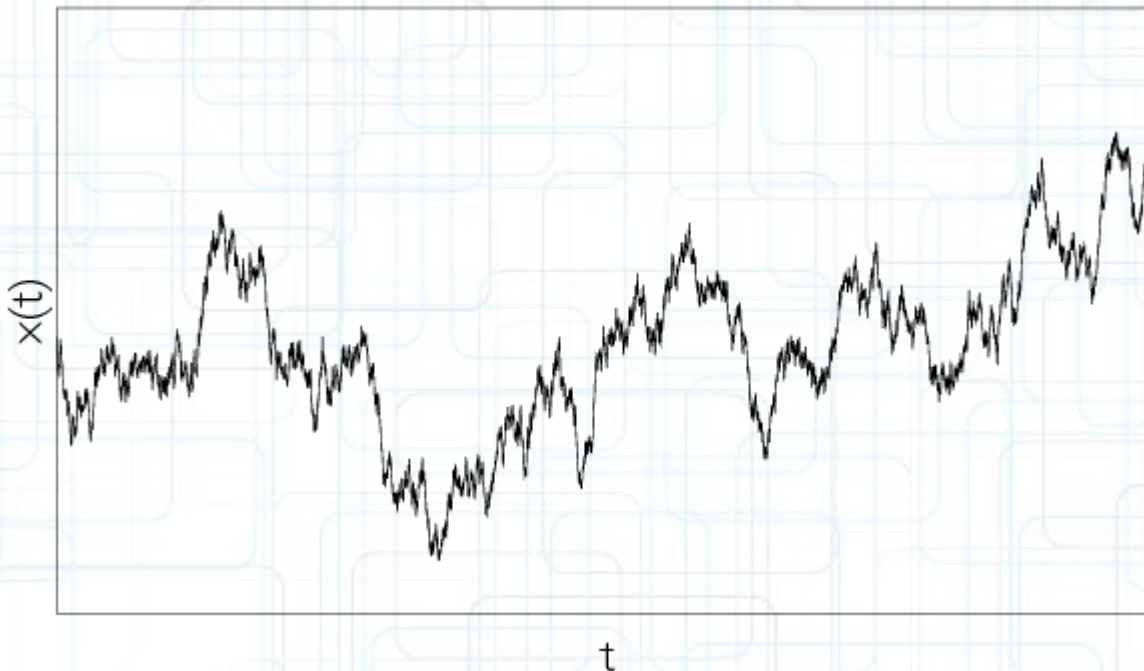
- Реализации стационарного случайного процесса могут быть периодическими
Пример: процесс $X(t)=\sin(t+s)$, где s – равномерно распределенная на $[0;2\pi]$ случайная величина
- $X(t) = X(t-2\pi)$



Авторегрессионный процесс первого порядка AR(1)

- Стационарный процесс – марковский, если значение зависит только от ближайшего предыдущего значения
- Пример: случайное блуждание

$$X_t = c + rX_{t-1} + \varepsilon_t$$



Поиск коэффициентов авторегрессии

- Метод наименьших квадратов:

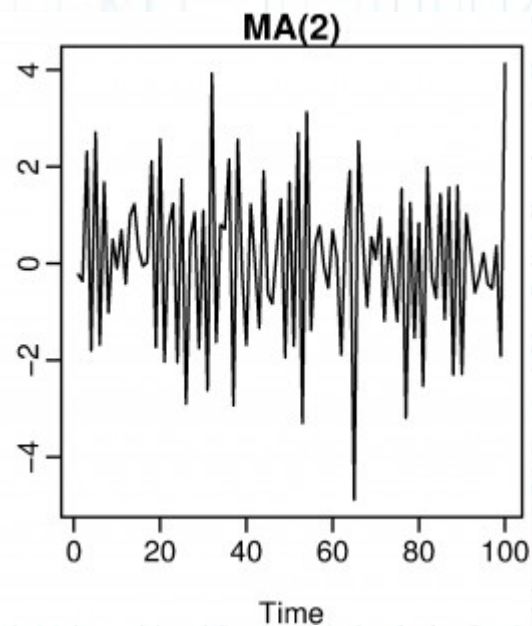
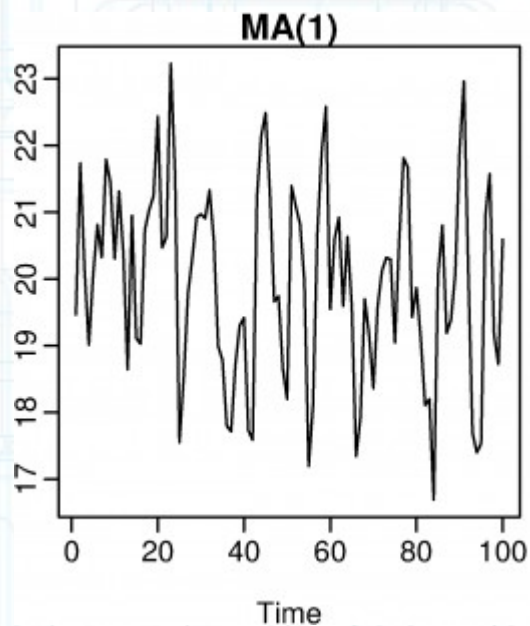
$$\hat{y}_{t+1}(w) = \sum_{j=1}^n w_j y_{t-j+1},$$

$$F_{\ell \times n} = \begin{pmatrix} y_{t-1} & y_{t-2} & y_{t-3} & \dots & y_{t-n} \\ y_{t-2} & y_{t-3} & y_{t-4} & \dots & y_{t-n-1} \\ \dots & \dots & \dots & \dots & \dots \\ y_n & y_{n-1} & y_{n-2} & \dots & y_1 \\ y_{n-1} & y_{n-2} & y_{n-3} & \dots & y_0 \end{pmatrix}, \quad y_{\ell \times 1} = \begin{pmatrix} y_t \\ y_{t-1} \\ \dots \\ y_{n+1} \\ y_n \end{pmatrix}$$

$$Q_t(w, X^\ell) = \sum_{i=n}^t (\hat{y}_i(w) - y_i)^2 = \|Fw - y\|^2 \rightarrow \min_w$$

Модель скользящего среднего (Moving Average - MA)

$$X_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$



Модель ARMA (autoregressive moving average)

$$X_t = c + \varepsilon_t + \sum_{i=1}^p \alpha_i X_{t-i} + \sum_{i=1}^q \beta_i \varepsilon_{t-i}.$$



ARIMA - интегрированная модель авторегрессии скользящего среднего (Box-Jenkins model)

Временной ряд называется интегрированным порядка k , если разности ряда порядка k являются стационарными

$$\Delta^d X_t = c + \sum_{i=1}^p a_i \Delta^d X_{t-i} + \sum_{j=1}^q b_j \varepsilon_{t-j} + \varepsilon_t$$



ARIMA (2,0,1) $y_t = a_1 y_{t-1} + a_2 y_{t-2} + b_1 \varepsilon_{t-1}$

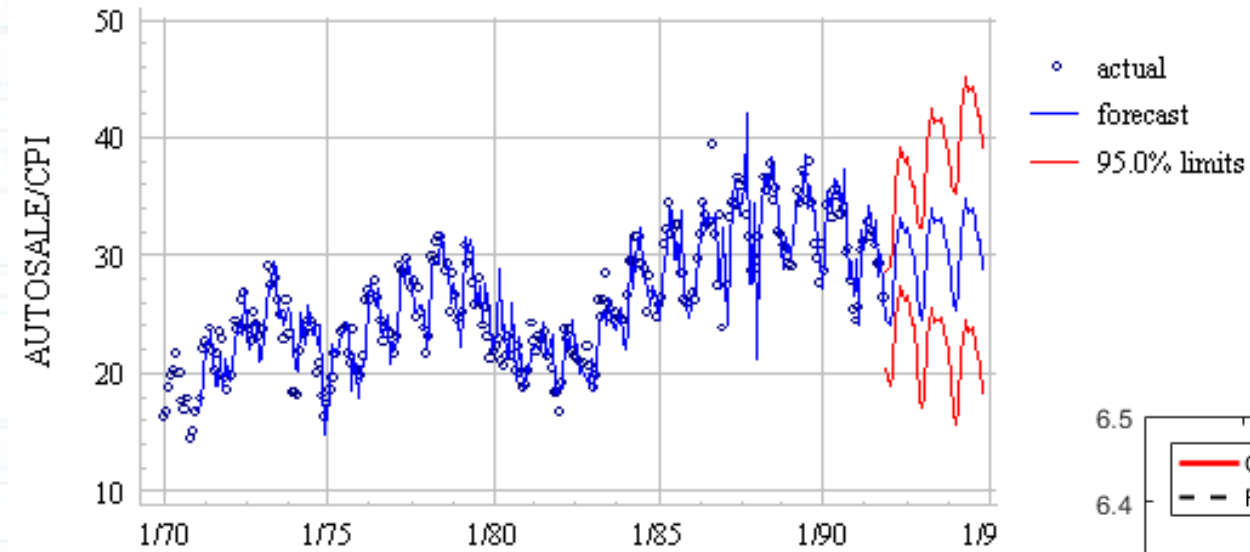
ARIMA (3,0,1) $y_t = a_1 y_{t-1} + a_2 y_{t-2} + a_3 y_{t-3} + b_1 \varepsilon_{t-1}$

ARIMA (1,1,0) $\Delta y_t = a_1 \Delta y_{t-1} + \varepsilon_t$, where $\Delta y_t = y_t - y_{t-1}$

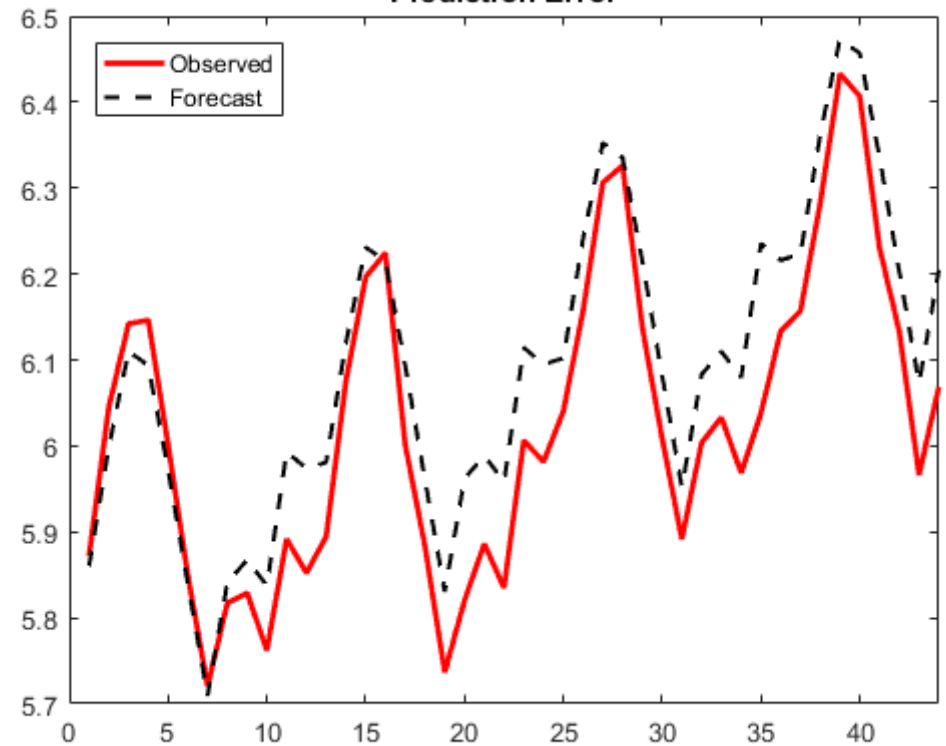
ARIMA (2,1,0) $\Delta y_t = a_1 \Delta y_{t-1} + a_2 \Delta y_{t-2} + \varepsilon_t$ where $\Delta y_t = y_t - y_{t-1}$

ARIMA

Forecasts for AUTOSALE/CPI from November '91
ARIMA(1,0,0)x(0,1,0)12 with constant



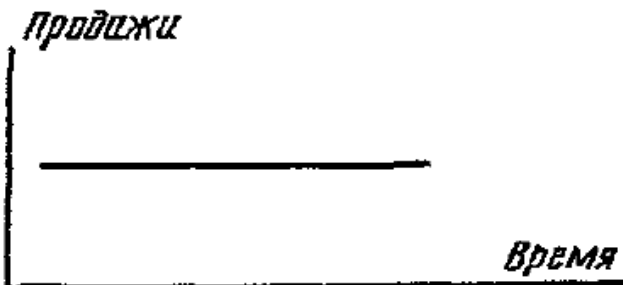
Prediction Error



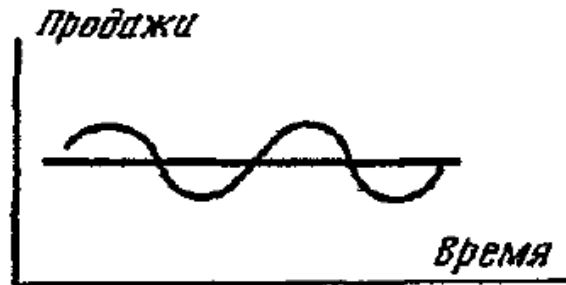
Общий план решения задачи предсказания

- Подготовка данных – сведение к стационарному случайному процессу
- Определение типа модели
- Оценка параметров
- Предсказание

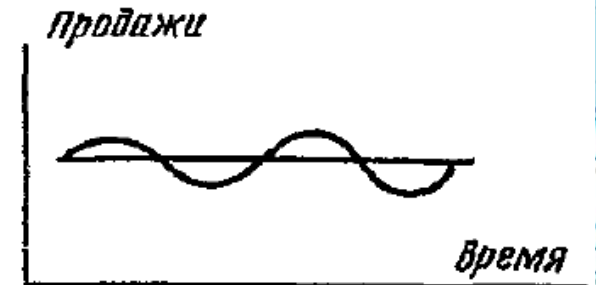
Модели с трендом и сезонным эффектом



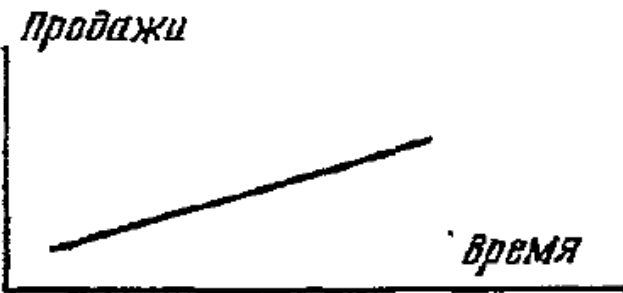
Модель 1-А



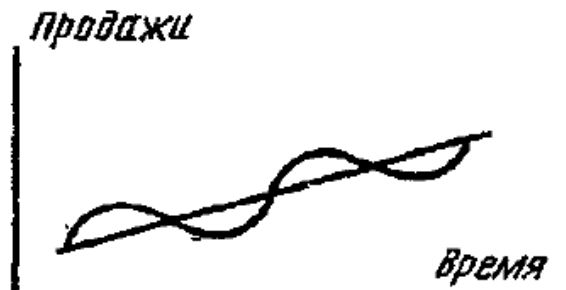
Модель 2-А



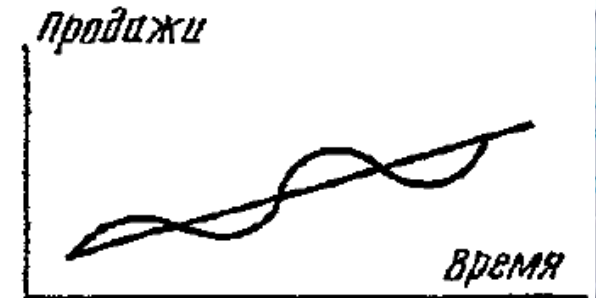
Модель 3-А



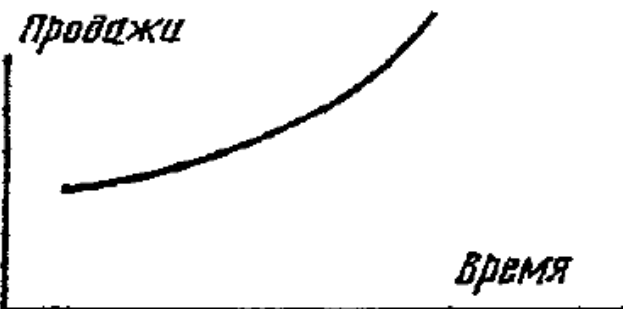
Модель 1-В



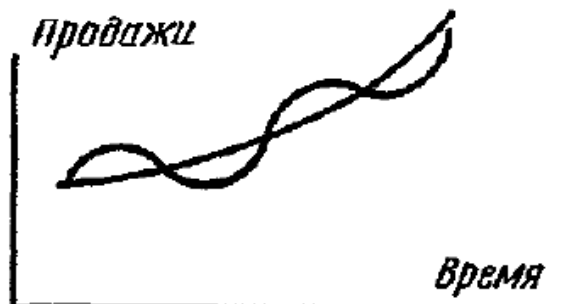
Модель 2-В



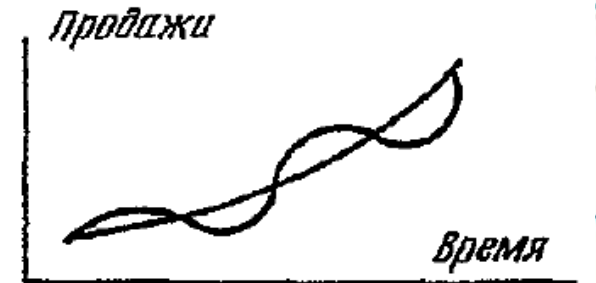
Модель 3-В



Модель 1-С



Модель 2-С



Модель 3-С

а)

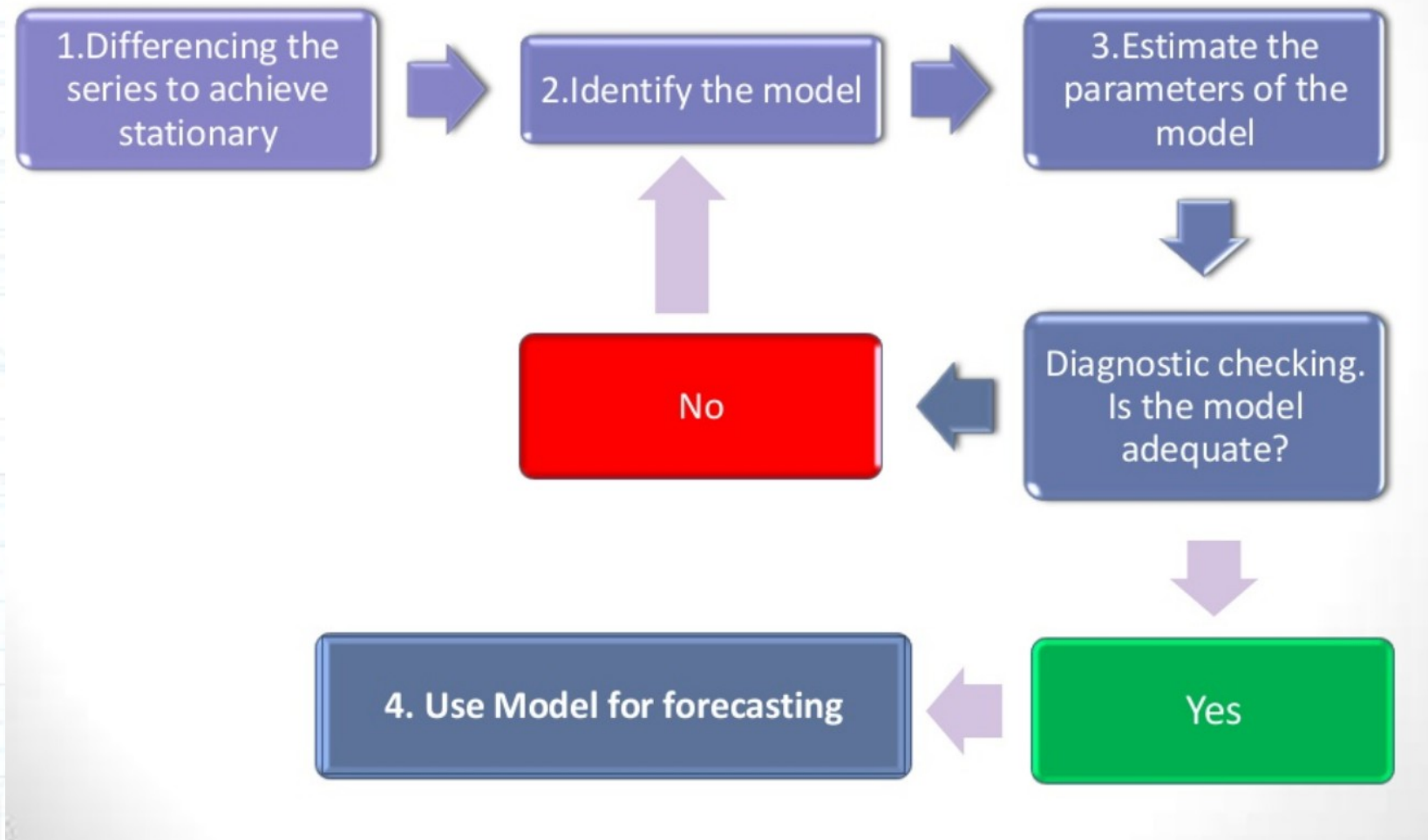
б)

в)

Моделирование с помощью ARIMA(p,d,q)

- Стационарность – определение правильного d , исключение сезонности
- Подбор p и q , используя ACF, PACF и unit root тесты
- Проверка – расчет оценки качества
- Оценка невязки – является ли она белым шумом?
- Предсказание

Авторы Бокс и Дженкинс предлагают схему:



Стационарность

- Процесс из разностей какого порядка является стационарным?
- Исключить сезонность, используя
 - сезонные добавки/множители к среднему значению за этот сезон
 - сезонную $ARIMA(p,d,q) \times (P,D,Q)$ модель, например $ARIMA(0,0,0) \times (0,1,0)$: $\hat{Y}_t = Y_{t-12} + \mu$

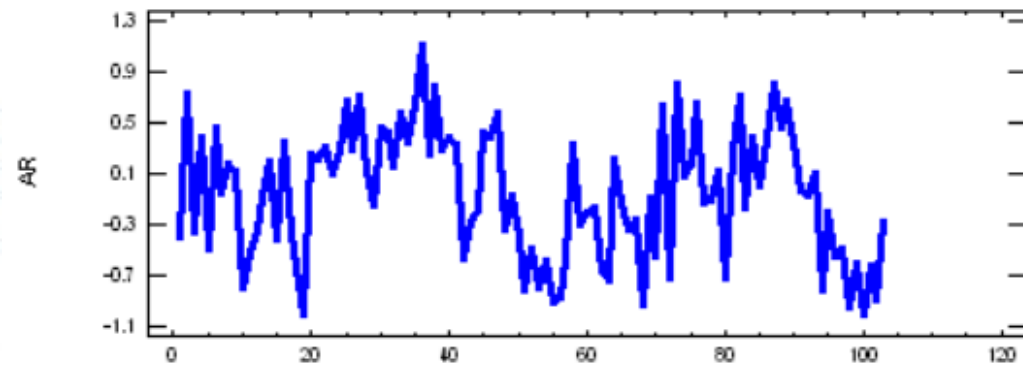
Автокорреляция (ACF)

- Корреляция между значениями процесса, с зафиксированным расстоянием по времени между ними
- Частичная автокорреляция (PACF) - “часть корреляции между Y_t и Y_{t-k} ”, которая не объясняется промежуточными корреляциями”. Коэффициент в AR-модели

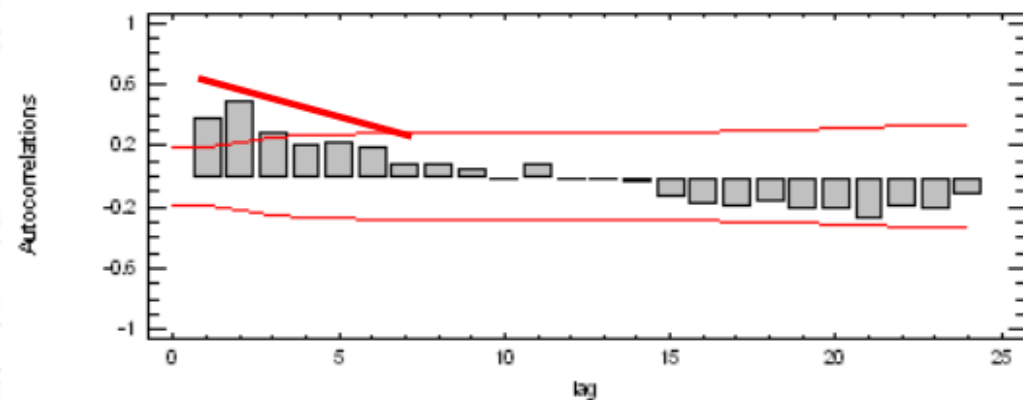
Признаки AR модели

- Процесс стремится вернуться к некоторому среднему значению
- АСФ убывает плавно, РАСФ - резко

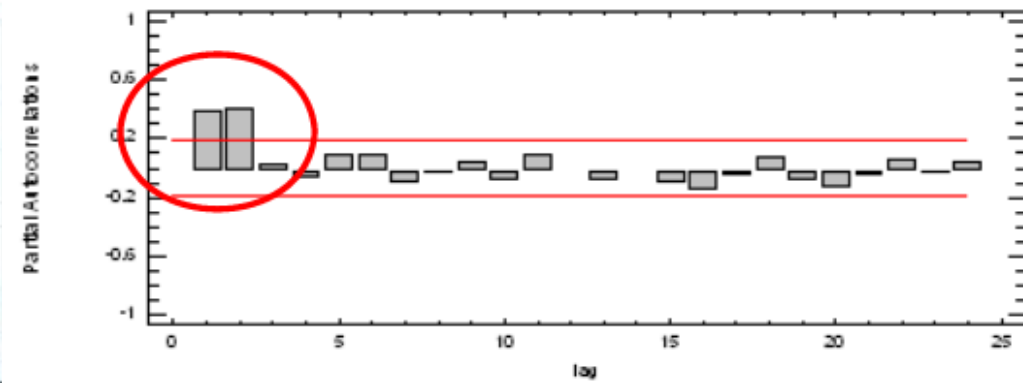
Time Series Plot for AR



Estimated Autocorrelations for AR



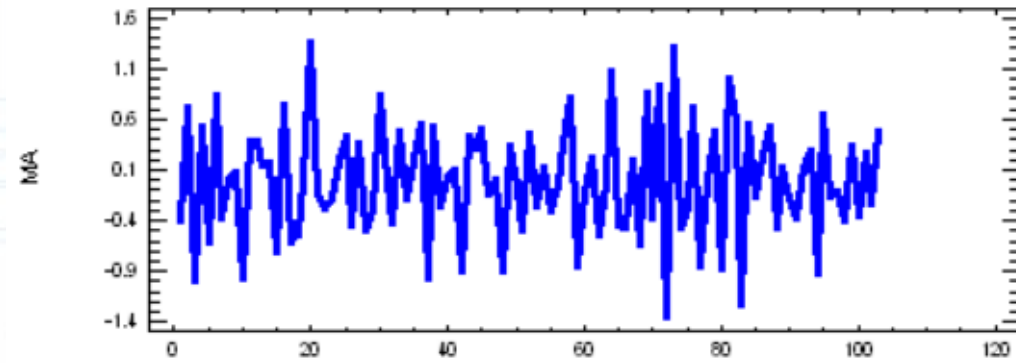
Estimated Partial Autocorrelations for AR



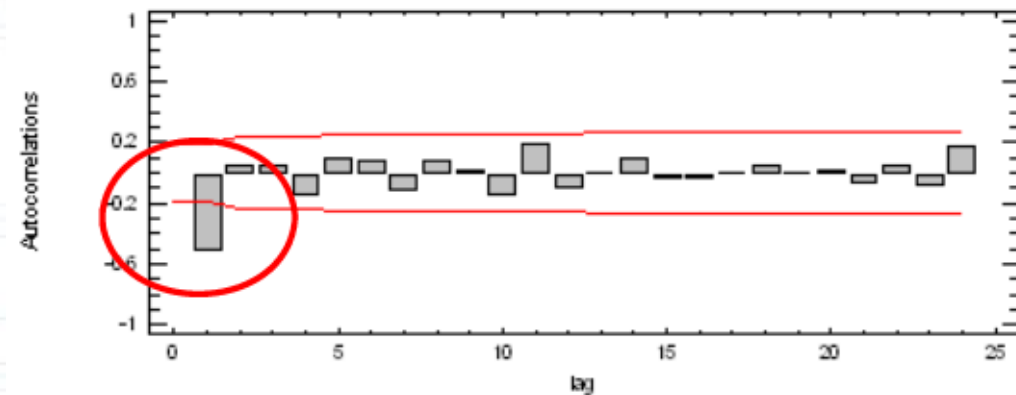
Признаки MA модели

- Похожа на белый шум
- АСФ убывает резко, РАСФ - постепенно

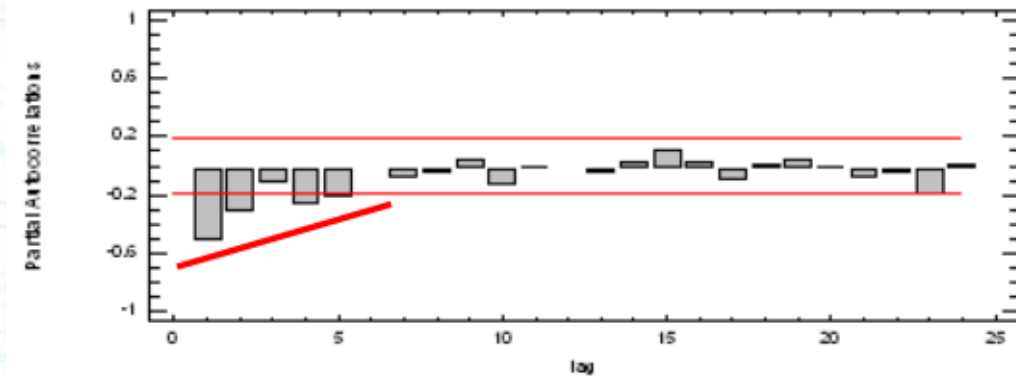
Time Series Plot for MA



Estimated Autocorrelations for MA



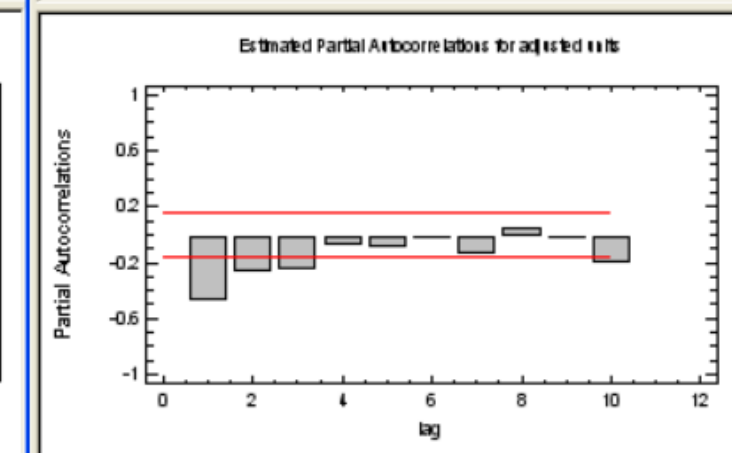
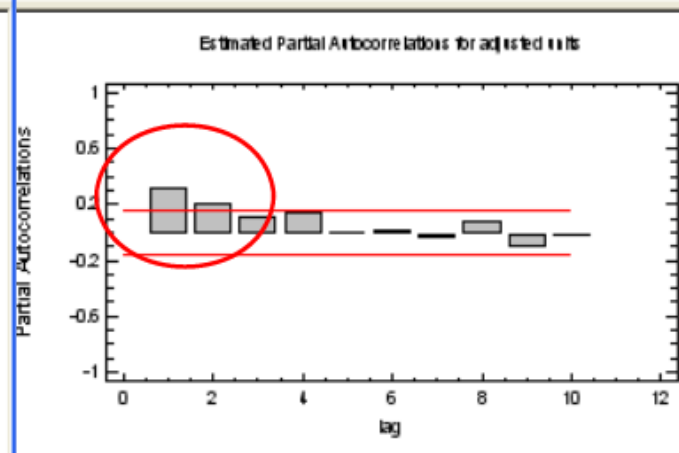
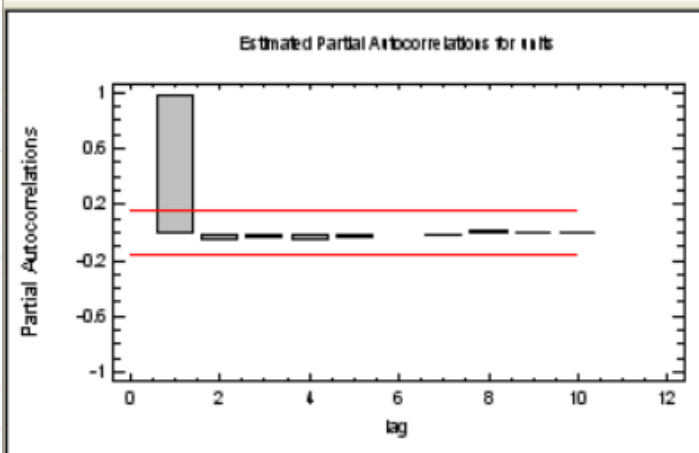
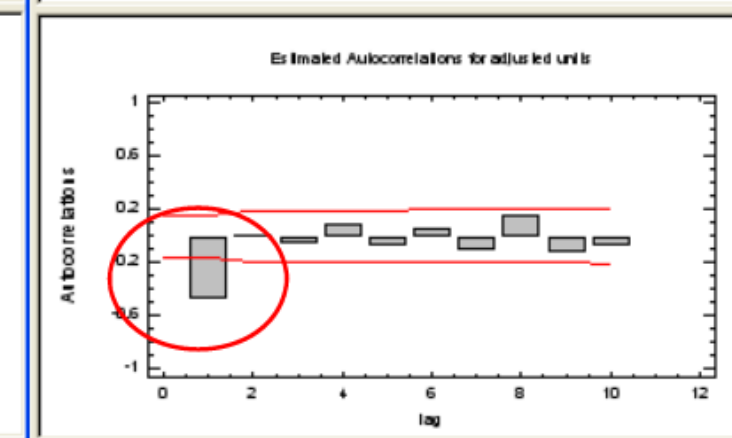
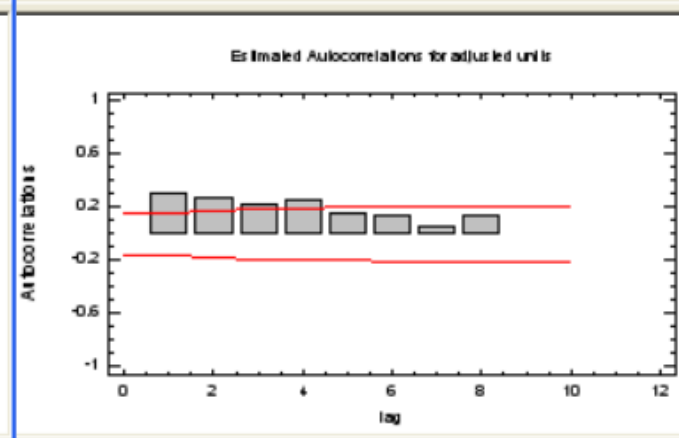
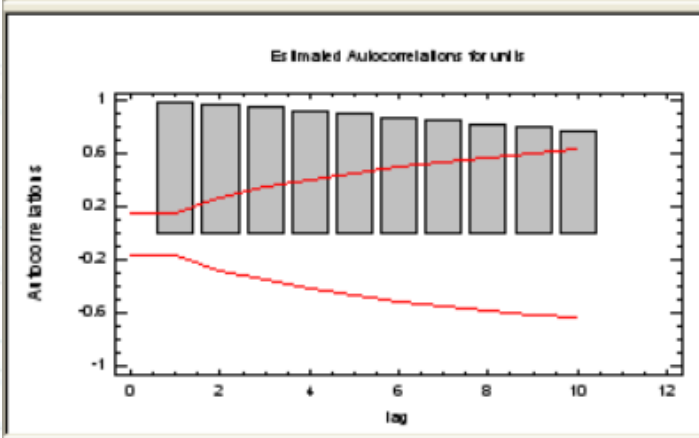
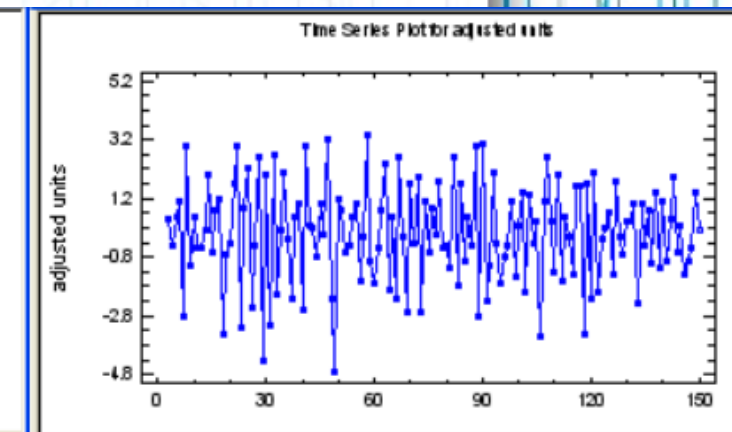
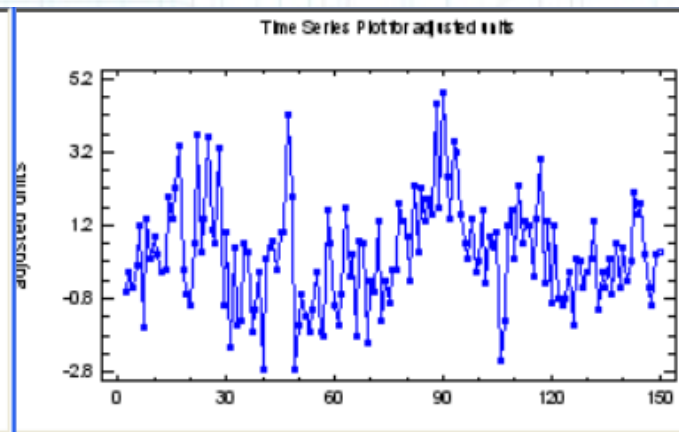
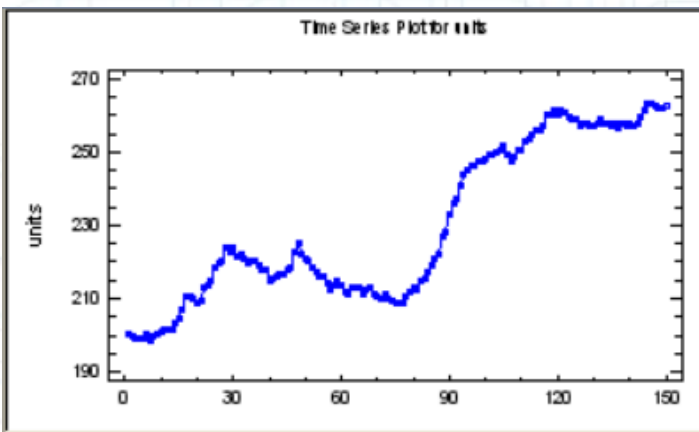
Estimated Partial Autocorrelations for MA



AR или MA

- Все зависит от порядка d дифференцирования процесса
- Исходный процесс обычно похож на AR
- После вычисления нескольких разностей он превращается в MA-процесс
- Не нужно дифференцировать слишком много раз – это переобучение

Пример

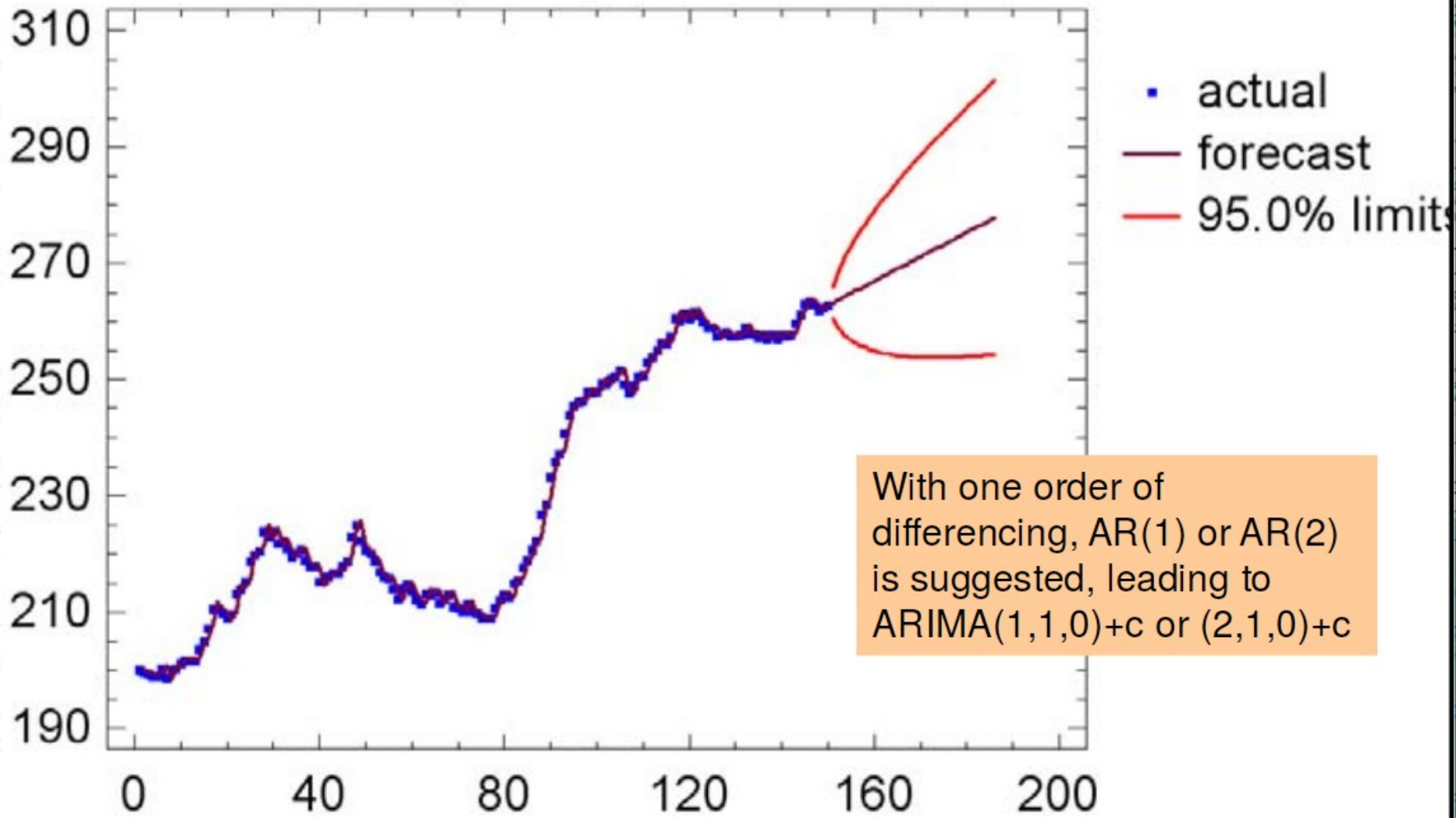


Original series: nonstationary

1st difference: AR signature

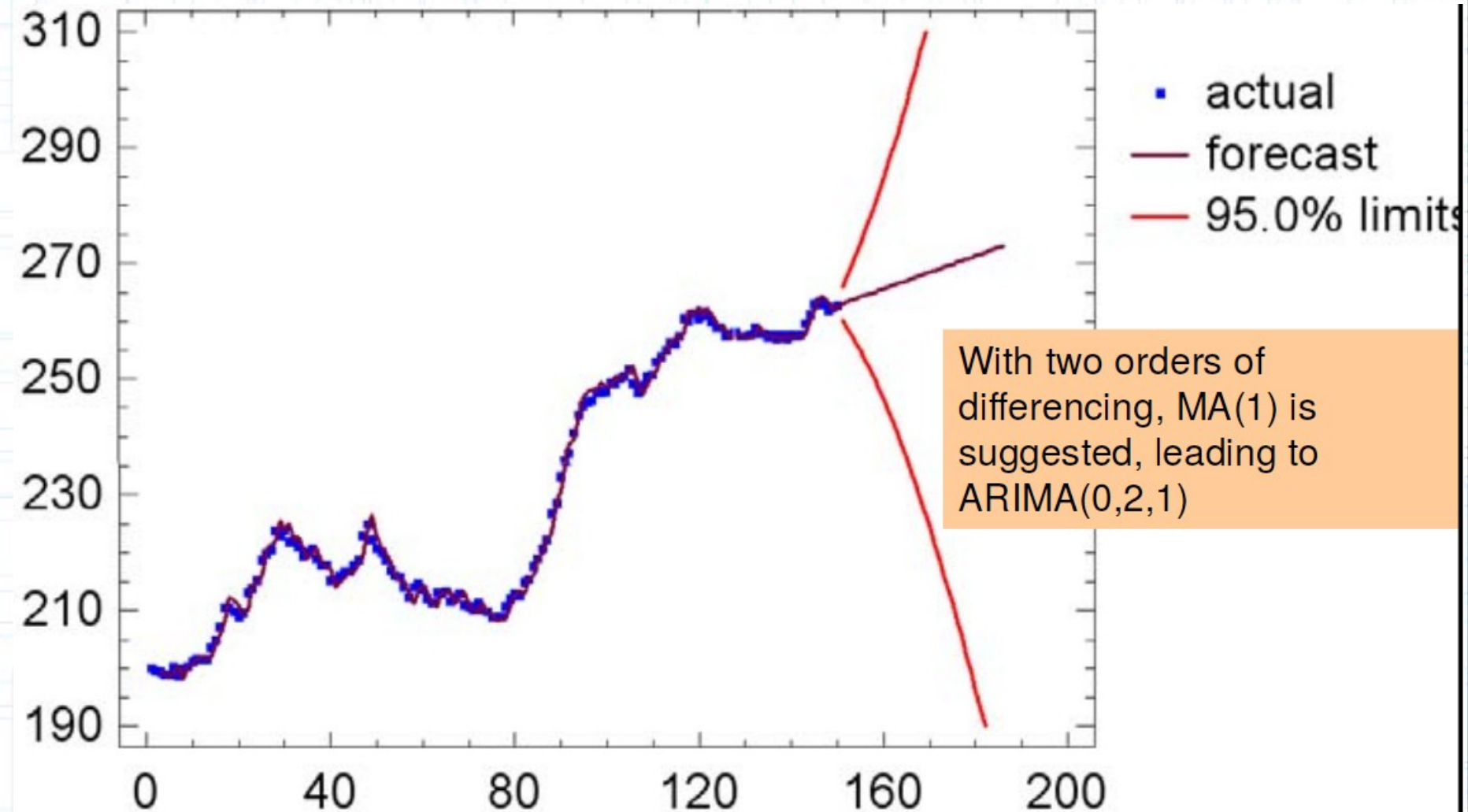
2nd difference: MA signature

Пример ARIMA(1,1,0)



With one order of differencing, AR(1) or AR(2) is suggested, leading to ARIMA(1,1,0)+c or (2,1,0)+c

Пример ARIMA(0,2,1)



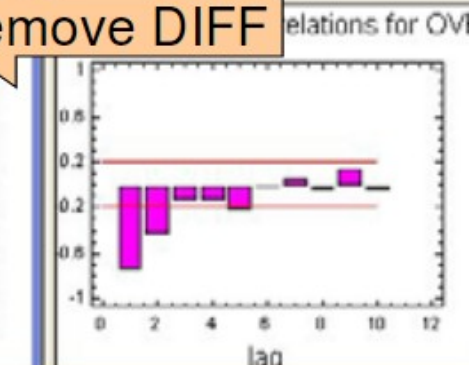
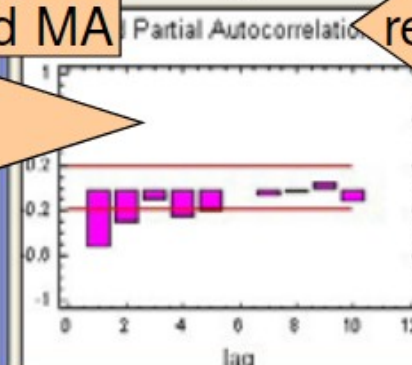
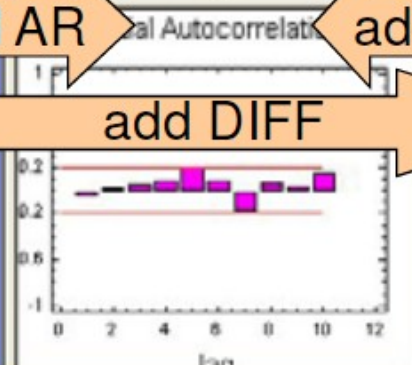
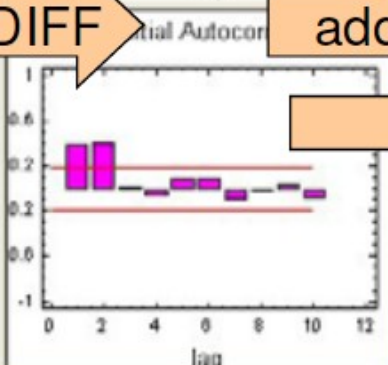
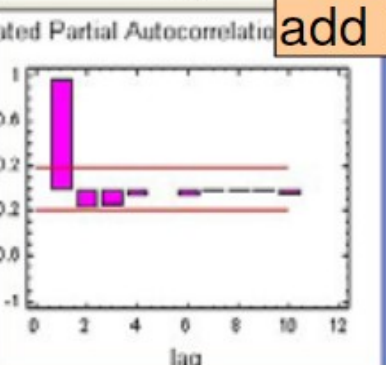
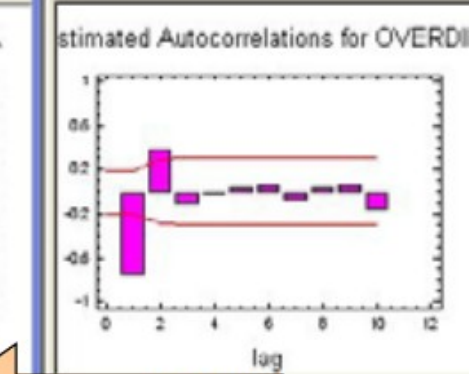
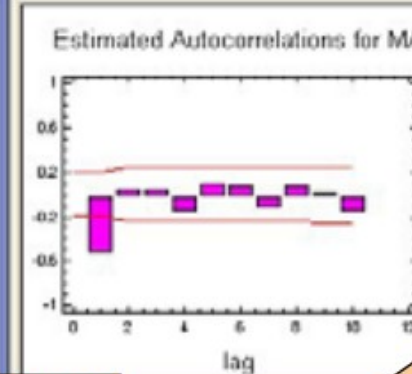
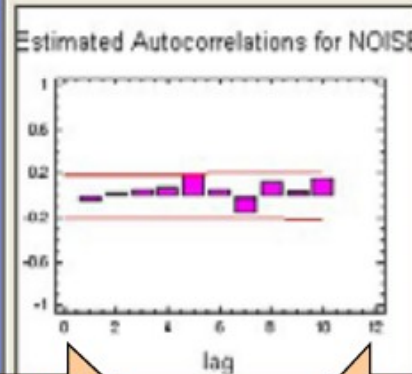
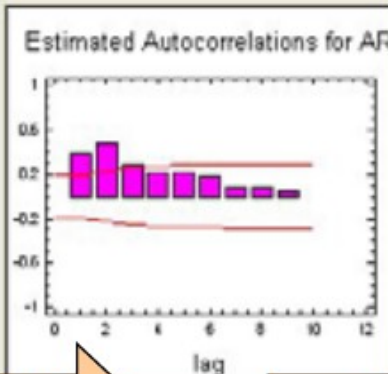
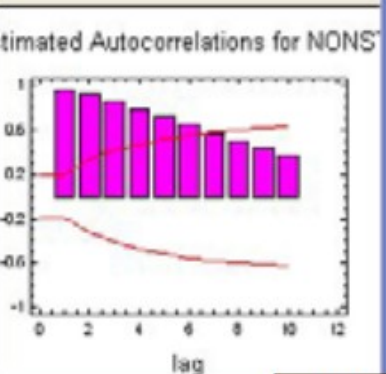
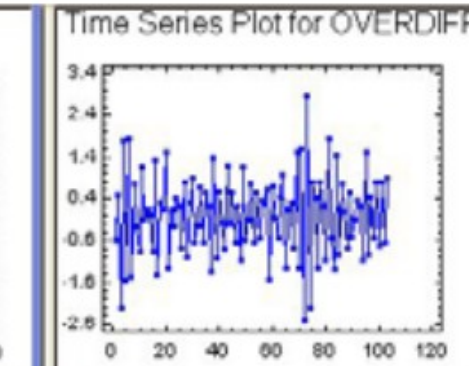
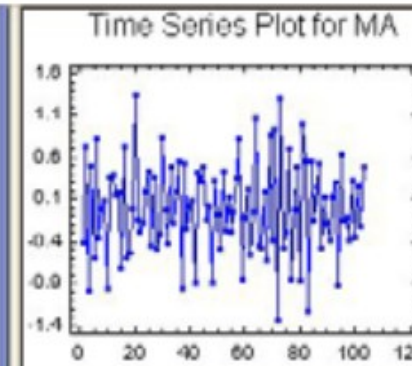
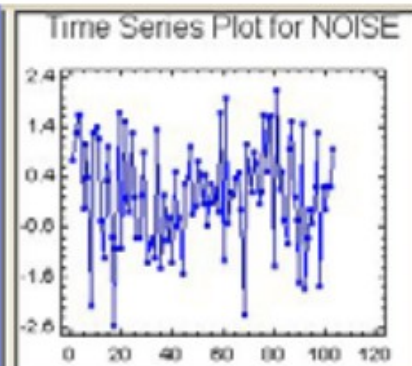
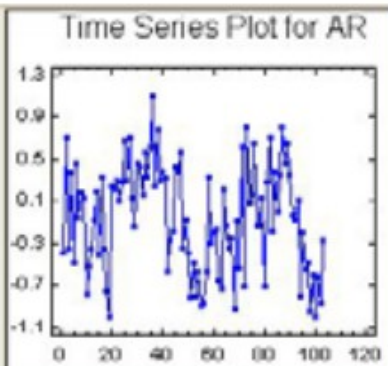
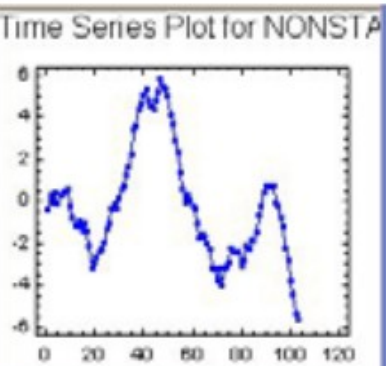
With two orders of differencing, MA(1) is suggested, leading to ARIMA(0,2,1)

Подбор параметров модели

← Positive autocorrelation

No autocorrelation

Negative autocorrelation →



add DIFF

add AR

add MA

remove DIFF

add DIFF

Nonstationary

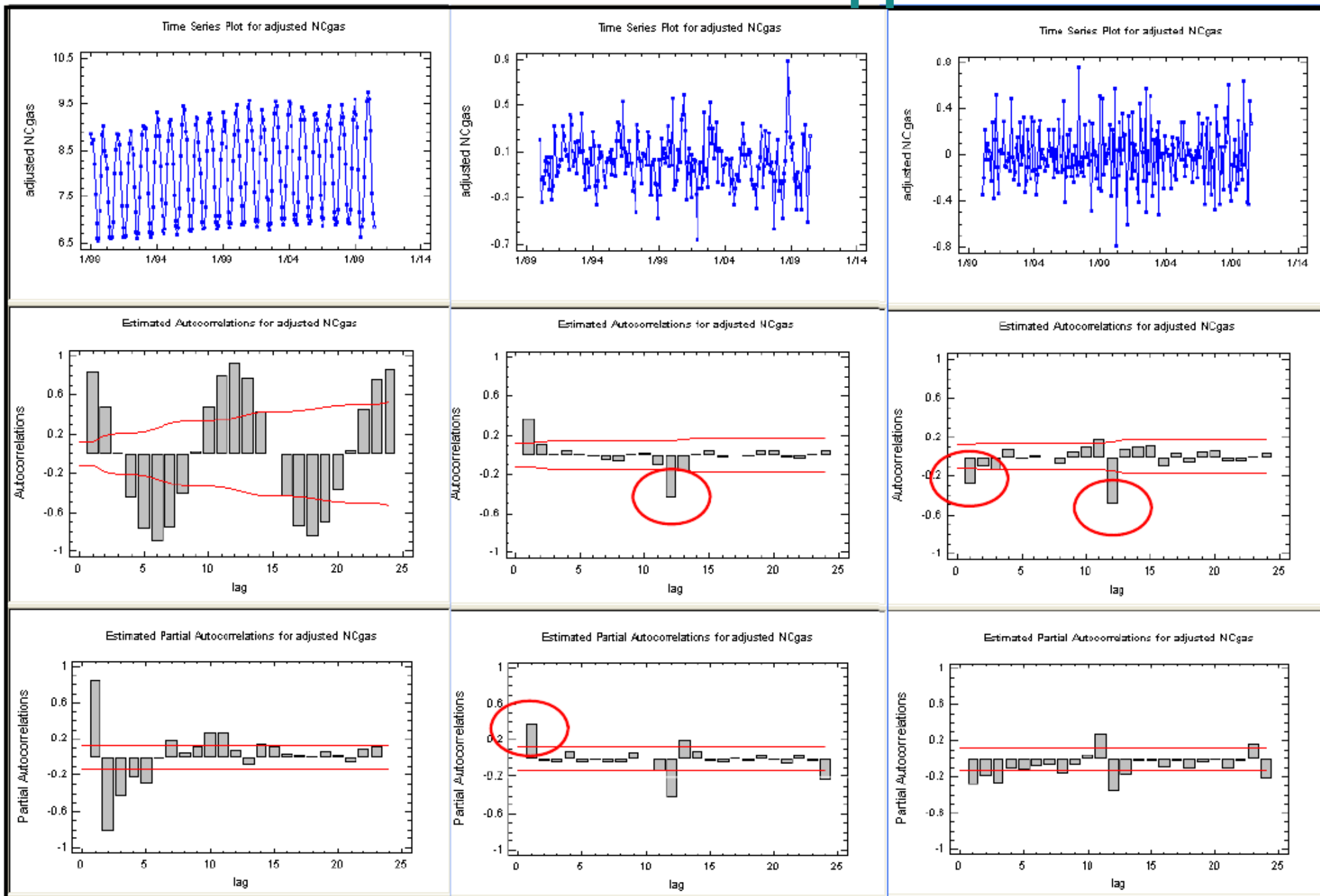
Auto-Regressive

White Noise

Moving-Average

Overdifferenced

Сезонная ARIMA-модель



Original series: nonstationary

Seas. diff: need AR(1) & SMA(1)

Both diff: need MA(1) & SMA(1)