ВЫБОР
ARIMA И ПРОГНОЗИРОВАНИЕ

ПОДБОР ПАРАМЕТРОВ

- α, ϕ, θ
- d, D
- q, Q
- p, P

α, ϕ, θ

- Если все остальные параметры фиксированы, коэффициенты регрессии подбираются методом наименьших квадратов
- Чтобы найти коэффициенты θ, шумовая компонента предварительно оценивается с помощью остатков авторегрессии
- Если шум белый (независимый одинаково распределённый гауссовский), то МНК даёт оценки максимального правдоподобия

d, D

- Порядки дифференцирования подбираются так, чтобы ряд стал стационарным
- Ещё раз: если ряд сезонный, рекомендуется начинать с сезонного дифференцирования
- Чем меньше раз мы продифференцируем, тем меньше будет дисперсия итогового прогноза

q, Q, p, P

- Гиперпараметры нельзя выбирать из принципа максимума правдоподобия: L всегда увеличивается с их ростом
- Для сравнения моделей с разными q, Q, p, P можно использовать критерий Акаике:
$$AIC = -2\ln L + 2k,$$
$k = P + Q + p + q + 1$ — число параметров в модели
- Начальные приближения можно выбрать с помощью автокорреляций

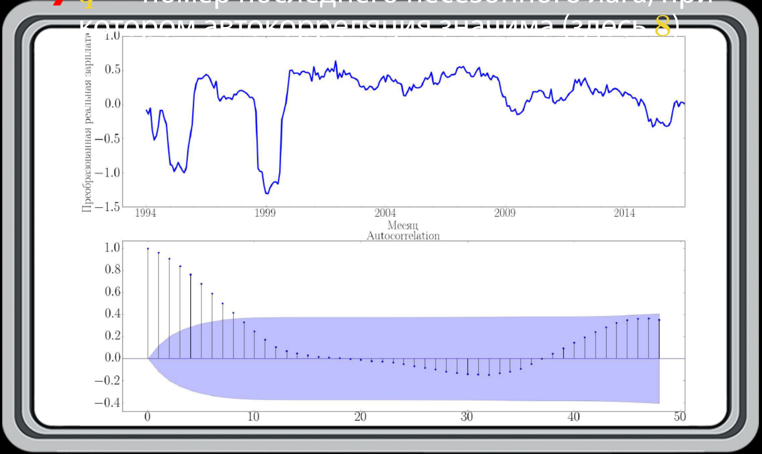
q, Q

- $Q * S$ — номер последнего сезонного лага, при котором автокорреляция значима (здесь 0)
- q — номер последнего несезонного лага, при котором автокорреляция значима (здесь 8)

q, Q

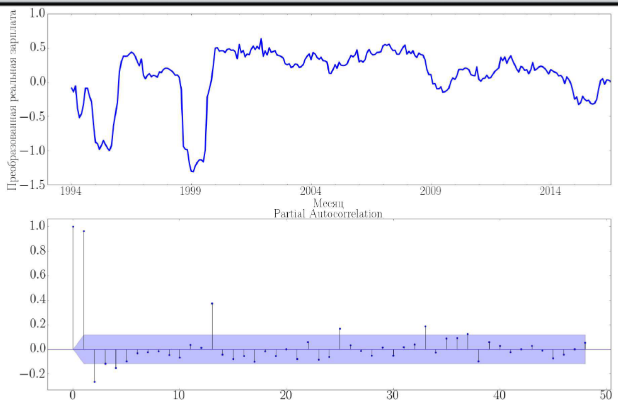
- $Q * S$ — номер последнего сезонного лага, при котором автокорреляция значима (здесь 0)

- q — номер последнего несезонного лага, при котором автокорреляция значима (здесь 8)



p, P

- Частичная автокорреляция — автокорреляция после снятия авторегрессии предыдущего порядка

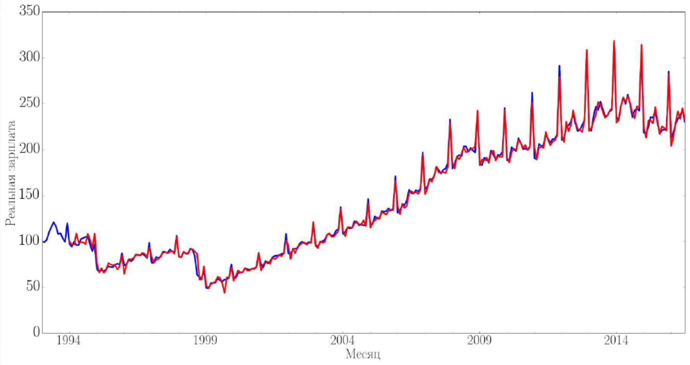


p, P

- $P * S$ — номер последнего сезонного лага, при котором частичная автокорреляция значима (здесь 2)
- p — номер последнего несезонного лага, при котором частичная автокорреляция значима (здесь 1)

РЕАЛЬНАЯ ЗАРАБОТНАЯ ПЛАТА

- Перебирая модели с $D = 1$, $d = 0$ и преобразованием Бокса-Кокса, получаем наименьший AIC на $ARIMA(2,0,1) \times (2,1,2)$:



ПОДБОР ARIMA

- Смотрим на ряд
- При необходимости стабилизируем дисперсию
- Если ряд нестационарен, подбираем порядок дифференцирования
- Анализируем ACF/PACF, определяем примерные p, q, P, Q
- Обучаем модели-кандидаты, сравниваем их по AIC, выбираем победителя
- Смотрим на остатки полученной модели, если они плохие, пробуем что-то поменять

ПРОГНОЗИРОВАНИЕ

МФТИ

$$y_t = \hat{\alpha} + \hat{\phi}_1 y_{t-1} + \cdots + \hat{\phi}_p y_{t-p} +$$
$$+ \varepsilon_t + \hat{\theta}_1 \varepsilon_{t-1} + \cdots + \hat{\theta}_q \varepsilon_{t-q}$$

ПРОГНОЗИРОВАНИЕ

- Заменяем t на $T+1$

$$\hat{y}_{T+1|T} = \hat{\alpha} + \hat{\phi}_1 y_T + \cdots + \hat{\phi}_p y_{T+1-p} +$$
$$+\varepsilon_{T+1} + \hat{\theta}_1 \varepsilon_T + \cdots + \hat{\theta}_q \varepsilon_{T+1-q}$$

ПРОГНОЗИРОВАНИЕ

$$\hat{y}_{T+1|T} = \hat{\alpha} + \hat{\phi}_1 y_T + \cdots + \hat{\phi}_p y_{T+1-p} +$$
$$+ \varepsilon_{T+1} + \hat{\theta}_1 \varepsilon_T + \cdots + \hat{\theta}_q \varepsilon_{T+1-q}$$

Прогнозирование

- Заменяем будущие ошибки на нули:

$$\hat{y}_{T+1|T} = \hat{\alpha} + \hat{\phi}_1 y_T + \dots + \hat{\phi}_p y_{T+1-p} +$$
$$+ \hat{\theta}_1 \varepsilon_T + \dots + \hat{\theta}_q \varepsilon_{T+1-q}$$

ПРОГНОЗИРОВАНИЕ

- Заменяем прошлые ошибки на остатки:

$$\hat{y}_{T+1|T} = \hat{\alpha} + \hat{\phi}_1 y_T + \cdots + \hat{\phi}_p y_{T+1-p} +$$
$$+ \hat{\theta}_1 \hat{\varepsilon}_T + \cdots + \hat{\theta}_q \hat{\varepsilon}_{T+1-q}$$

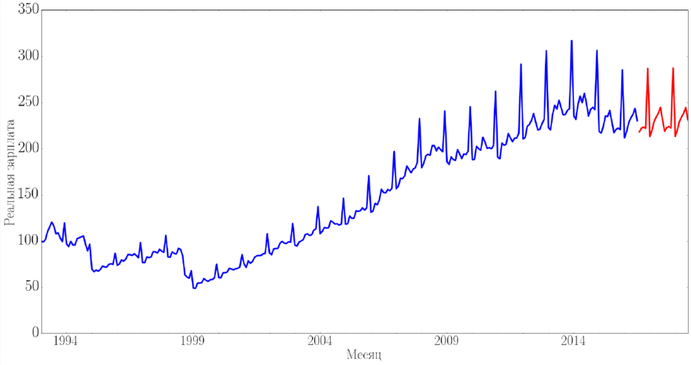
ПРОГНОЗИРОВАНИЕ

- Если мы прогнозируем на момент времени $T+2$, в формуле появляется значение ряда из будущего:

$$\hat{y}_{T+2|T} = \hat{\alpha} + \hat{\phi}_1 y_{T+1} + \dots + \hat{\phi}_p y_{T+2-p} +$$
$$+ \hat{\theta}_1 \hat{\varepsilon}_{T+1} + \dots + \hat{\theta}_q \hat{\varepsilon}_{T+2-q}$$

- Заменяем его на прогноз $\hat{y}_{T+1|T}$

ПРОГНОЗИРОВАНИЕ



РЕЗЮМЕ

- Подбор параметров модели
- Прогнозирование

ДАЛЕЕ В ПРОГРАММЕ

- Как понять, что модель достаточно хороша?
- Анализ остатков