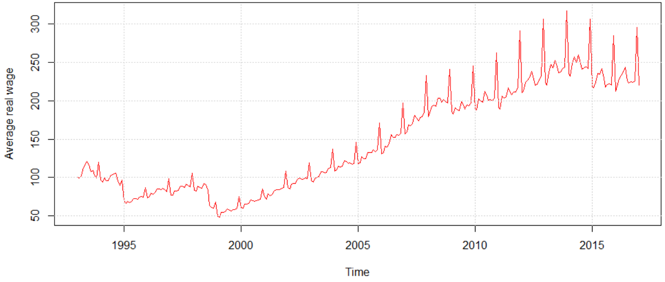
Прикладной статистический анализ данных
Анализ временных рядов

Андрей Грабовой

Прогнозирование временного ряда

Временной ряд: $y_1, \dots, y_T, \dots, \quad y_t \in \mathbb{R}$, — значения признака, измеренные через постоянные временные интервалы.



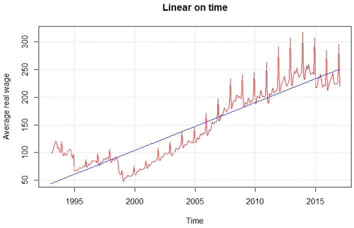
Задача прогнозирования — найти функцию f_T:

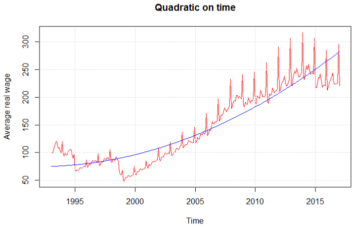
$$y_{T+d} \approx f_T\left(y_T, \dots, y_1, d\right) \equiv \hat{y}_{T+d|T},$$

где $d \in \{1, \dots, D\}$ — отсрочка прогноза, D — горизонт прогнозирования.

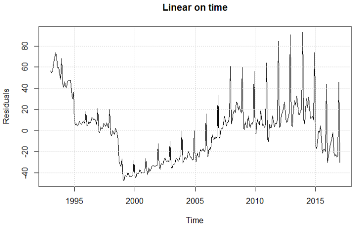
Регрессия

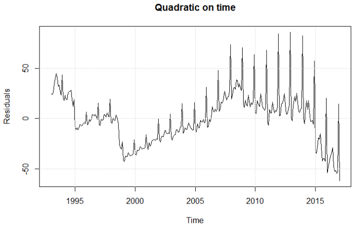
Простейшая идея: сделать регрессию на время.



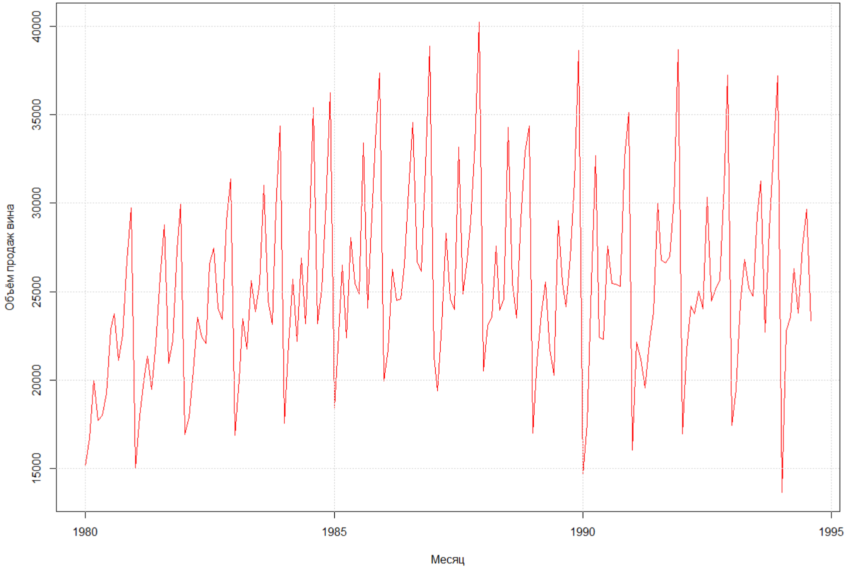


Остатки не выглядят как шум:

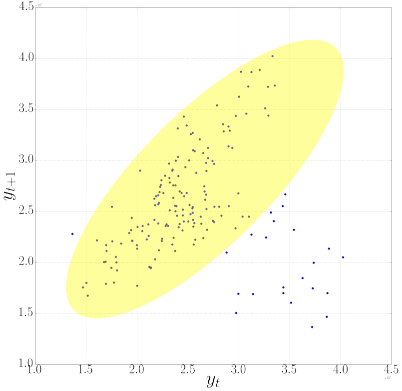




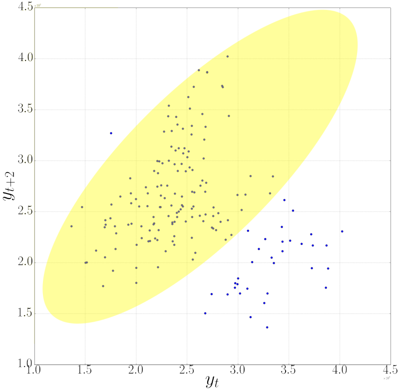
Продажи вина в Австралии



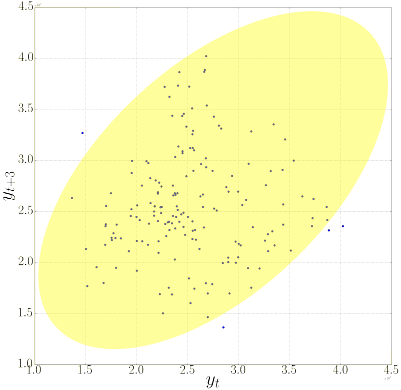
Продажи в соседние месяцы



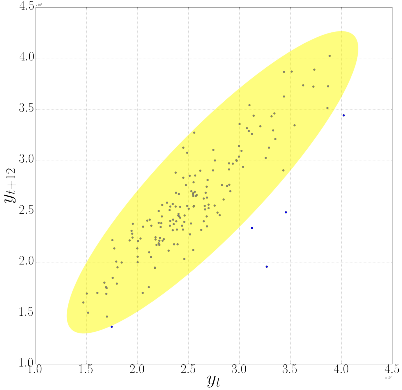
Продажи через 1 месяц



Продажи через 2 месяца



Продажи через год



Автокорреляционная функция (ACF)

Наблюдения временного ряда автокоррелированы.

Автокорреляция:

$$r_\tau = r_{y_t y_{t+\tau}} = \frac{\sum\limits_{t=1}^{T-\tau} (y_t - \bar{y})(y_{t+\tau} - \bar{y})}{\sum\limits_{t=1}^{T} (y_t - \bar{y})^2}, \quad \bar{y} = \frac{1}{T} \sum_{t=1}^{T} y_t.$$

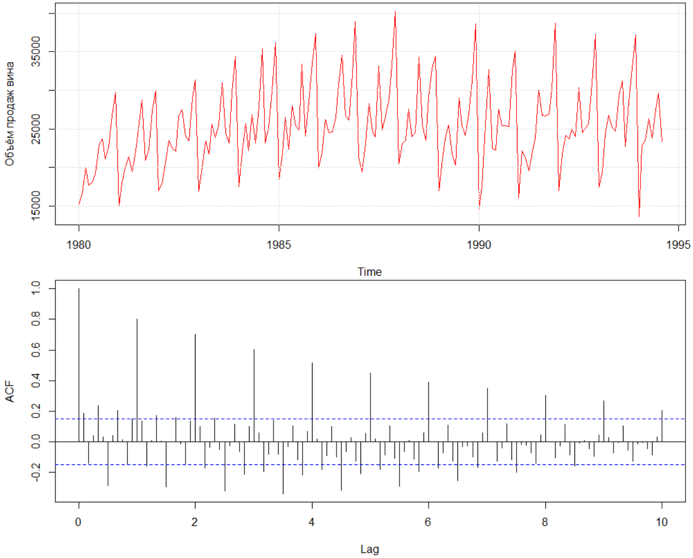
$r_\tau \in [-1, 1]$, τ — лаг автокорреляции.

Проверка значимости отличия автокорреляции от нуля:

временной ряд:	$Y^T = Y_1, \dots, Y_T$;
нулевая гипотеза:	$H_0 \colon r_\tau = 0$;
альтернатива:	$H_1 \colon r_\tau \neq 0$;
статистика:	$T\left(Y^T\right) = \frac{r_\tau \sqrt{T-\tau-2}}{\sqrt{1-r_\tau^2}}$;
нулевое распределение:	$St\left(T - \tau - 2\right)$.

Автокорреляционная функция (ACF)

Коррелограмма:



Компоненты временных рядов

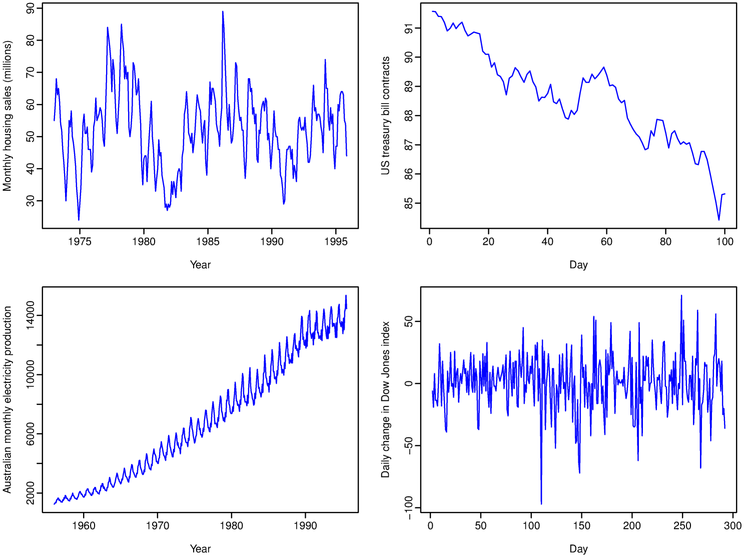
Тренд — плавное долгосрочное изменение уровня ряда.

Сезонность — циклические изменения уровня ряда с постоянным периодом.

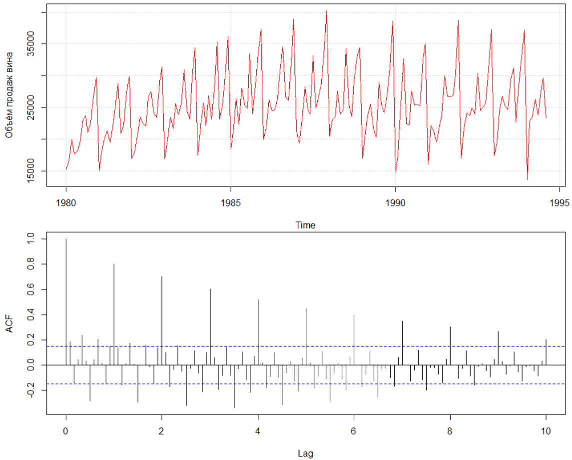
Цикл — изменения уровня ряда с переменным периодом (цикл жизни товара, экономические волны, периоды солнечной активности).

Ошибка — непрогнозируемая случайная компонента ряда.

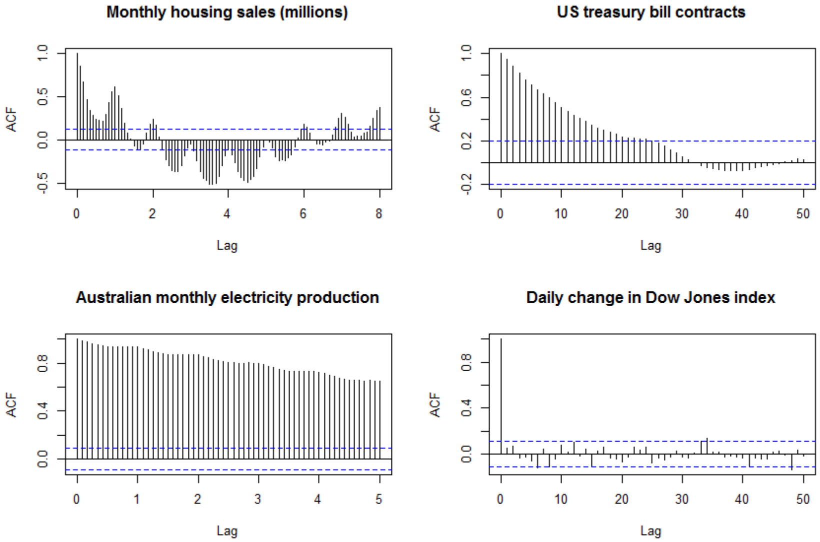
Компоненты временных рядов



Компоненты временных рядов

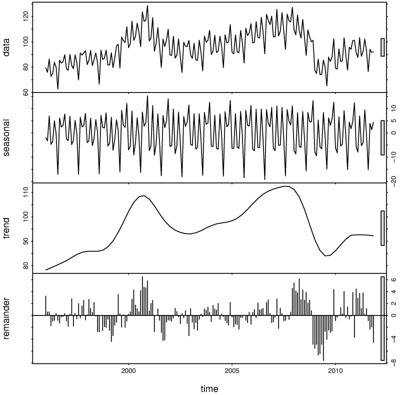


Компоненты временных рядов



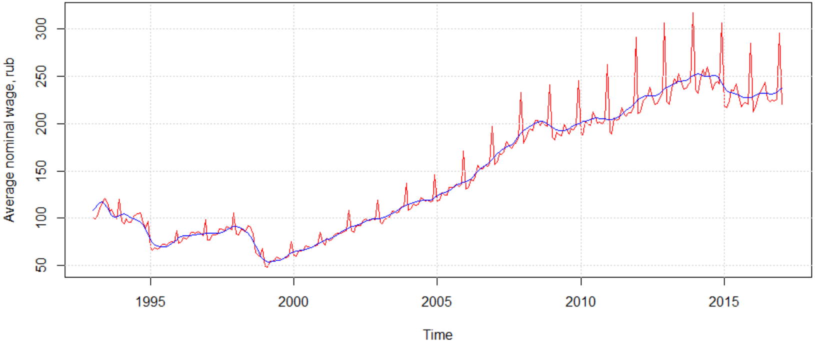
Компоненты временных рядов

STL-декомпозиция:



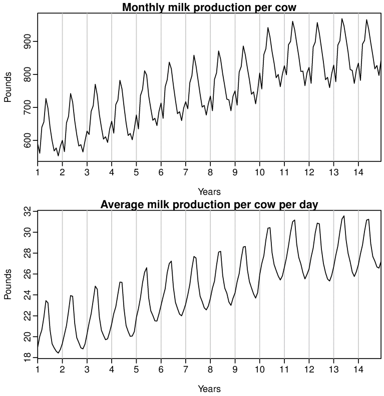
Снятие сезонности

Часто для удобства интерпретации ряда сезонная компонента вычитается:



Календарные эффекты

Иногда упростить структуру временного ряда можно за счёт учёта неравномерности отсчётов:



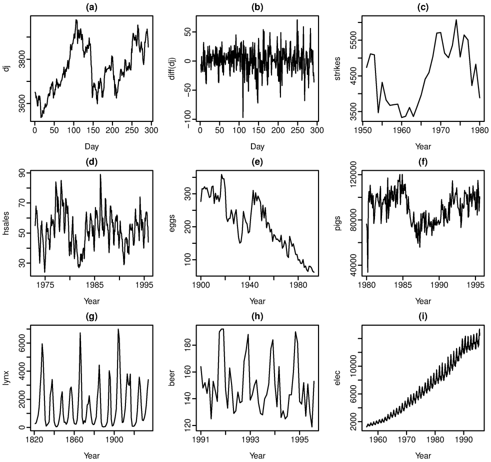
Стационарность

Ряд $y_1, \ldots, y_T$ **стационарен**, если $\forall s$ распределение $y_t, \ldots, y_{t+s}$ не зависит от t, т. е. его свойства не зависят от времени.

Ряды с трендом или сезонностью нестационарны.

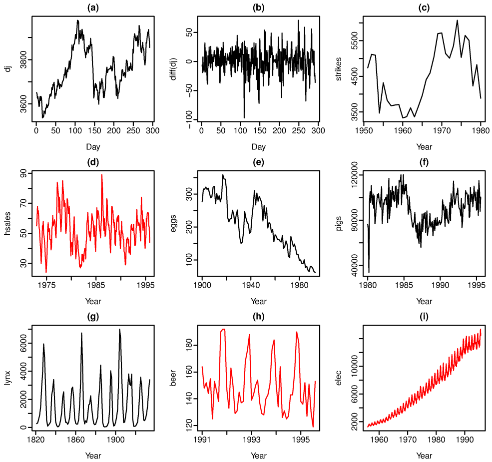
Ряды с непериодическими циклами стационарны, поскольку нельзя предсказать заранее, где будут находится максимумы и минимумы.

Стационарность



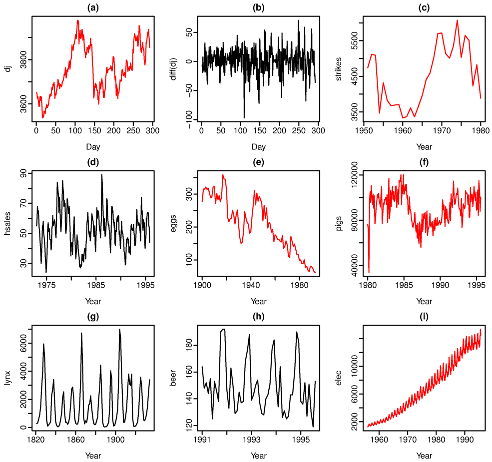
Стационарность

Нестационарны из-за сезонности:



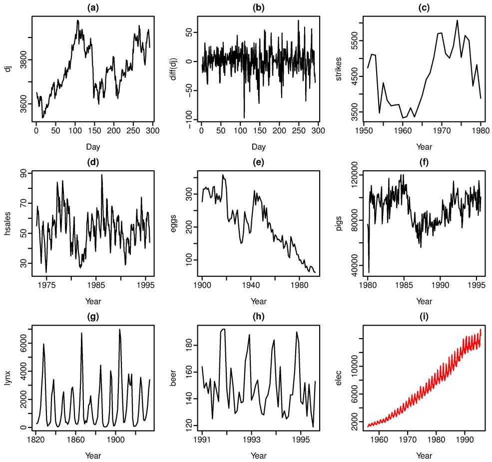
Стационарность

Нестационарны из-за тренда:



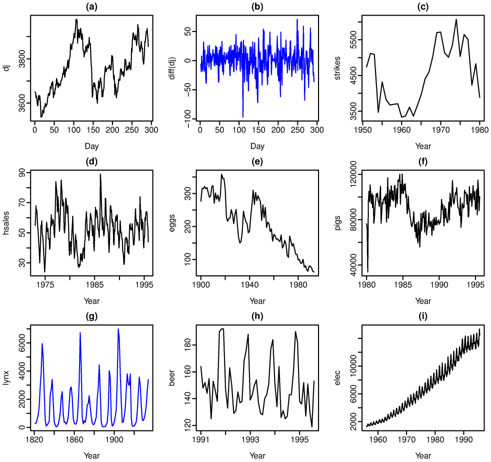
Стационарность

Нестационарны из-за меняющейся дисперсии:



Стационарность

Стационарны:



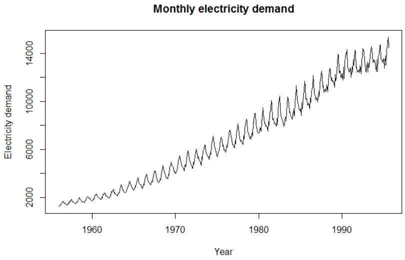
Критерий KPSS (Kwiatkowski-Philips-Schmidt-Shin)

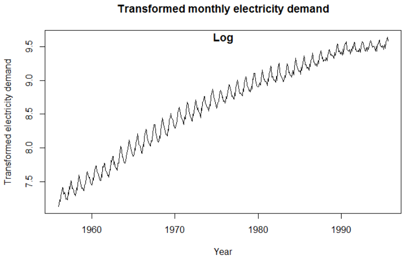
ряд ошибок прогноза:	$\varepsilon^T = \varepsilon_1, \dots, \varepsilon_T$;
нулевая гипотеза:	H_0: ряд ε^T стационарен;
альтернатива:	H_1: ряд ε^T описывается моделью вида $\varepsilon_t = \alpha\varepsilon_{t-1}$;
статистика:	$KPSS\left(\varepsilon^T\right) = \frac{1}{T^2}\sum\limits_{i=1}^{T}\left(\sum\limits_{t=1}^{i}\varepsilon_t\right)^2 \Big/ \lambda^2$, λ^2—оценка дисперсии ошибок;
нулевое распределение:	табличное.

Другие критерии для проверки стационарности: Дики-Фуллера, Филлипса-Перрона и их многочисленные модификации (см. Patterson K. *Unit root tests in time series, volume 1: key concepts and problems*. — Palgrave Macmillan, 2011).

Для рядов с монотонно меняющейся дисперсией можно использовать стабилизирующие преобразования.

Часто используют логарифмирование:



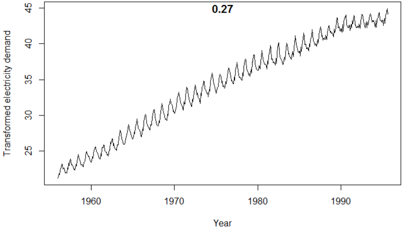


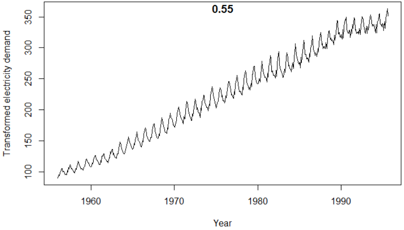
Преобразования Бокса-Кокса

Параметрическое семейство стабилизирующих дисперсию преобразований:

$$y'_t = \begin{cases} \ln y_t, & \lambda = 0, \\ \left(y_t^\lambda - 1\right)/\lambda, & \lambda \neq 0. \end{cases}$$

Параметр λ выбирается так, чтобы минимизировать дисперсию или максимизировать правдоподобие модели.





Преобразования Бокса-Кокса

После построения прогноза для трансформированного ряда его нужно преобразовать в прогноз исходного:

$$\hat{y}_t = \begin{cases} \exp\left(\hat{y}'_t\right), & \lambda = 0, \\ \left(\lambda \hat{y}'_t + 1\right)^{1/\lambda}, & \lambda \neq 0. \end{cases}$$

- если некоторые $y_t \leqslant 0$, преобразования Бокса-Кокса невозможны (нужно прибавить к ряду константу)
- часто оказывается, что преобразование вообще не нужно
- можно округлять значение λ, чтобы упростить интерпретацию
- как правило, стабилизирующее преобразование слабо влияет на прогноз и сильно — на предсказательный интервал

Дифференцирование

Дифференцирование ряда — переход к попарным разностям его соседних значений:

$$y_1, \dots, y_T \;\longrightarrow\; y'_2, \dots, y'_T,$$

$$y'_t = y_t - y_{t-1}.$$

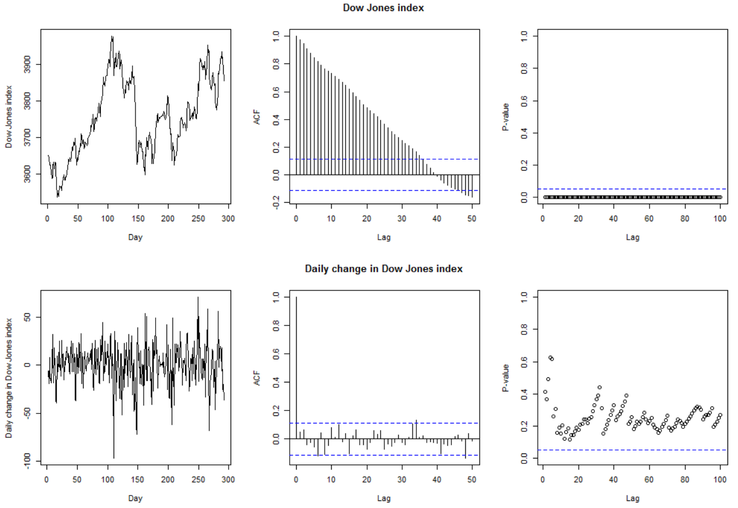
Дифференцированием можно стабилизировать среднее значение ряда и избавиться от тренда и сезонности.

Может применяться неоднократное дифференцирование; например, для второго порядка:

$$y_1, \dots, y_T \longrightarrow y'_2, \dots, y'_T \longrightarrow y''_3, \dots, y''_T,$$

$$y''_t = y'_t - y'_{t-1} = y_t - 2y_{t-1} + y_{t-2}.$$

Дифференцирование



Критерий KPSS: для исходного ряда $p < 0.01$, для ряда первых разностей — $p > 0.1$.

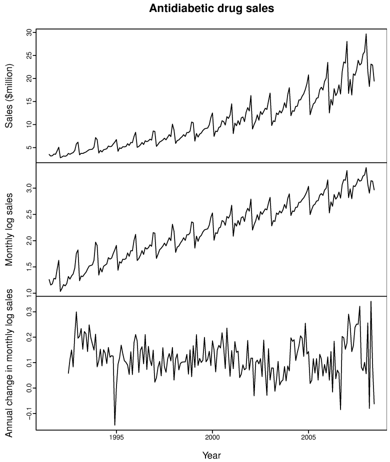
Сезонное дифференцирование

Сезонное дифференцирование ряда — переход к попарным разностям его значений в соседних сезонах:

$$y_1, \ldots, y_T \;\longrightarrow\; y'_{s+1}, \ldots, y'_T,$$

$$y'_t = y_t - y_{t-s}.$$

Сезонное дифференцирование



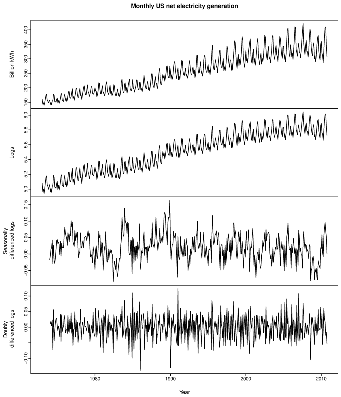
Критерий KPSS: для исходного ряда $p < 0.01$, для логарифмированного — $p < 0.01$, после сезонного дифференцирования — $p > 0.1$.

Комбинированное дифференцирование

Сезонное и обычное дифференцирование может применяться к одному ряду в любом порядке.

Если ряд имеет выраженный сезонный профиль, рекомендуется начинать с сезонного дифференцирования — после него ряд уже может оказаться стационарным.

Комбинированное дифференцирование



Критерий KPSS: для исходного ряда $p < 0.01$, для логарифмированного — $p < 0.01$, после сезонного дифференцирования — $p = 0.0355$, после ещё одного дифференцирования — $p > 0.1$.

Остатки

Остатки — разность между фактом и прогнозом:

$$\hat{\varepsilon}_t = y_t - \hat{y}_t.$$

Прогнозы $\hat{y}_t$ могут быть построены с фиксированной отсрочкой:

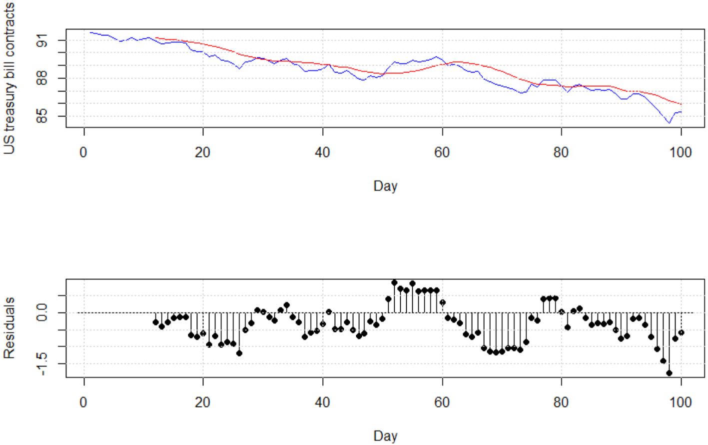
$$\hat{y}_{R+d|R}, \ldots, \hat{y}_{T|T-d},$$

или с фиксированным концом истории при разных отсрочках:

$$\hat{y}_{T-D+1|T-D}, \ldots, \hat{y}_{T|T-D}.$$

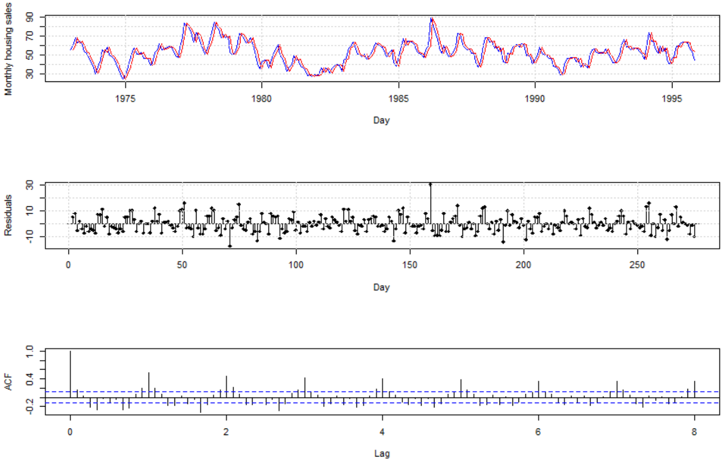
Необходимые свойства остатков прогноза

- Несмещённость — равенство среднего значения нулю:



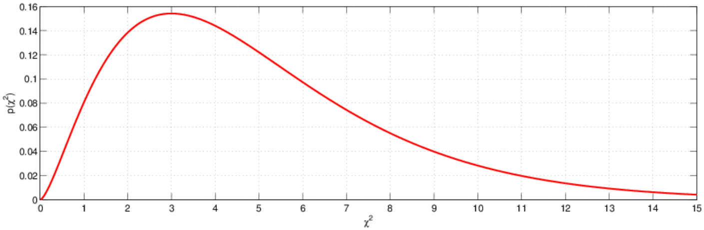
Необходимые свойства остатков прогноза

- Неавтокоррелированность — отсутствие неучтённой зависимости от предыдущих наблюдений:

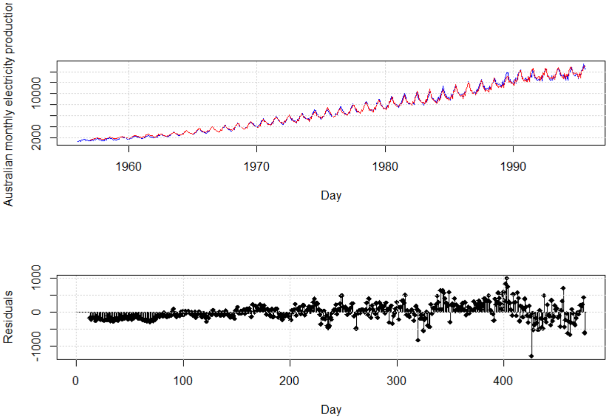


Q-критерий Льюнга-Бокса

ряд ошибок прогноза: $\varepsilon^T = \varepsilon_1, \dots, \varepsilon_T$;
нулевая гипотеза: $H_0\colon r_1 = \dots = r_L = 0$;
альтернатива: $H_1\colon H_0$ неверна;
статистика: $Q\left(\varepsilon^T\right) = T\left(T+2\right)\sum\limits_{\tau=1}^{L}\frac{r_\tau^2}{T-\tau}$;
нулевое распределение: χ^2_{L-K}, K — число настраиваемых параметров модели ряда.

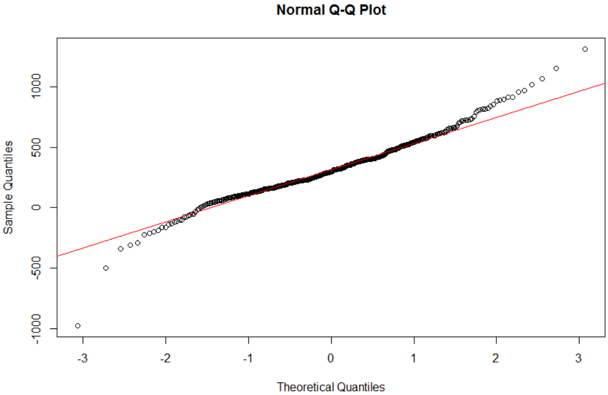


- Стационарность — отсутствие зависимости от времени:



Желательные свойства остатков прогноза

- Нормальность:



Проверка свойств остатков

- Несмещённость — критерий Стьюдента или Уилкоксона.
- Стационарность — визуальный анализ, критерий KPSS.
- Неавтокоррелированность — коррелограмма, Q-критерий Льюнга-Бокса.
- Нормальность — q-q plot, критерий Шапиро-Уилка.

Простейшие методы прогнозирования

- средним:

$$\hat{y}_{T+d} = \frac{1}{T}\sum_{t=1}^{T} y_t;$$

- средним за последние k отсчётов:

$$\hat{y}_{T+d} = \frac{1}{k}\sum_{t=T-k}^{T} y_t;$$

- наивный:

$$\hat{y}_{T+d} = y_T;$$

- наивный сезонный (s — период сезонности):

$$\hat{y}_{T+d} = y_{T+d-ks},\ k = \lfloor (d-1)/s \rfloor + 1;$$

- экстраполяции тренда:

$$\hat{y}_{T+d} = y_T + d\frac{y_T - y_1}{T-1}.$$

Простое экспоненциальное сглаживание (метод Брауна)

Наивный прогноз:

$$\hat{y}_{T+1|T} = y_T.$$

Прогноз средним значением:

$$\hat{y}_{T+1|T} = \sum_{t=1}^{T} y_t.$$

Прогноз с помощью взвешенного среднего с экспоненциально убывающими весами:

$$\hat{y}_{T+1|T} = \alpha y_T + \alpha\left(1-\alpha\right) y_{T-1} + \alpha\left(1-\alpha\right)^2 y_{T-2} + \dots$$

$\alpha \uparrow 1 \ \Rightarrow$ больший вес последним точкам,
$\alpha \downarrow 0 \ \Rightarrow$ большее сглаживание.

Наблюдение	$\alpha = 0.2$	$\alpha = 0.4$	$\alpha = 0.6$	$\alpha = 0.8$
y_T	0.2	0.4	0.6	0.8
y_{T-1}	0.16	0.24	0.24	0.16
y_{T-2}	0.128	0.144	0.096	0.032
y_{T-3}	0.1024	0.0864	0.0384	0.0064
y_{T-4}	0.08192	0.05184	0.01536	0.00128
y_{T-5}	0.065536	0.031104	0.006144	0.000256

Простое экспоненциальное сглаживание (метод Брауна)

- Метод подходит для прогнозирования рядов без тренда и сезонности:

$$\hat{y}_{t+1|t} = l_t,$$
$$l_t = \alpha y_t + (1-\alpha)\, l_{t-1} = \hat{y}_{t|t-1} + \alpha \cdot e_t.$$

$e_t = y_t - \hat{y}_{t|t-1}$ — ошибка прогноза отсчёта времени t

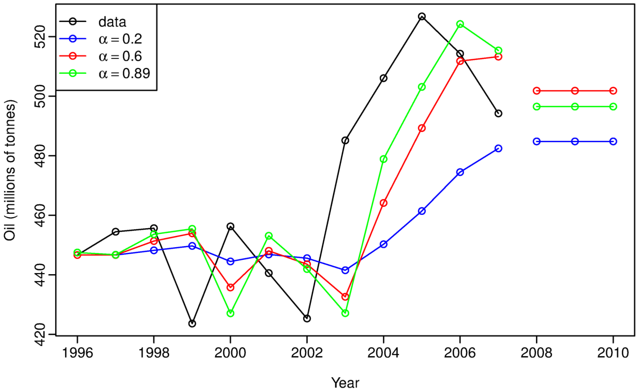
- Прогноз зависит от l_0:

$$\hat{y}_{T+1|T} = \sum_{j=1}^{T-1} \alpha\,(1-\alpha)^j\, y_{T-j} + (1-\alpha)^T\, l_0.$$

Можно взять $l_0 = y_1$ или оптимизировать его.

- Прогноз получается плоский, т. е. $\hat{y}_{t+d|t} = \hat{y}_{t+1|t}$.

Простое экспоненциальное сглаживание (метод Брауна)



Простое экспоненциальное сглаживание в применении к данным о добыче нефти в Саудовской Аравии (1996–2007).

Методы, учитывающие тренд

Аддитивный линейный тренд (метод Хольта):

$$\begin{aligned}
\hat{y}_{t+d|t} &= l_t + db_t, \\
l_t &= \alpha y_t + (1-\alpha)\left(l_{t-1} + b_{t-1}\right), \\
b_t &= \beta\left(l_t - l_{t-1}\right) + (1-\beta)\, b_{t-1}.
\end{aligned}$$

Мультипликативный линейный (экспоненциальный) тренд:

$$\begin{aligned}
\hat{y}_{t+d|t} &= l_t b_t^d, \\
l_t &= \alpha y_t + (1-\alpha)\left(l_{t-1} b_{t-1}\right), \\
b_t &= \beta \frac{l_t}{l_{t-1}} + (1-\beta)\, b_{t-1}.
\end{aligned}$$

$\alpha, \beta \in [0,1]$.

Методы, учитывающие тренд

Аддитивный затухающий тренд:

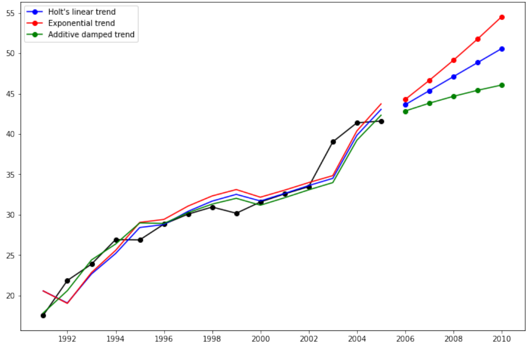
$$\begin{aligned}
\hat{y}_{t+d|t} &= l_t + \left(\phi + \phi^2 + \dots + \phi^d\right) b_t, \\
l_t &= \alpha y_t + (1-\alpha)\left(l_{t-1} + \phi b_{t-1}\right), \\
b_t &= \beta\left(l_t - l_{t-1}\right) + (1-\beta)\phi b_{t-1}.
\end{aligned}$$

Мультипликативный затухающий тренд:

$$\begin{aligned}
\hat{y}_{t+d|t} &= l_t b_t^{\left(\phi + \phi^2 + \dots + \phi^d\right)}, \\
l_t &= \alpha y_t + (1-\alpha) l_{t-1} b_{t-1}^{\phi}, \\
b_t &= \beta \frac{l_t}{l_{t-1}} + (1-\beta) b_{t-1}^{\phi}.
\end{aligned}$$

$\alpha, \beta \in [0,1], \ \phi \in (0,1)$.

Методы, учитывающие тренд



Прогнозы поголовья овец в Азии с учётом тренда.

Методы, учитывающие сезонность

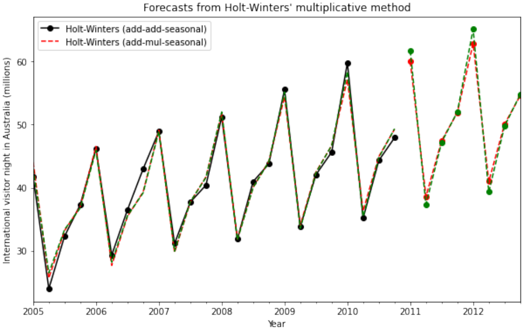
Аддитивная сезонность с периодом длины m (метод Тейла-Веджа):

$$\begin{aligned}
\hat{y}_{t+d|t} &= l_t + db_t + s_{t-m+(d \mod m)}, \\
l_t &= \alpha\left(y_t - s_{t-m}\right) + (1-\alpha)\left(l_{t-1} + b_{t-1}\right), \\
b_t &= \beta\left(l_t - l_{t-1}\right) + (1-\beta)\, b_{t-1}, \\
s_t &= \gamma\left(y_t - l_{t-1} - b_{t-1}\right) + (1-\gamma)\, s_{t-m}.
\end{aligned}$$

Мультипликативная сезонность (Хольта-Уинтерса):

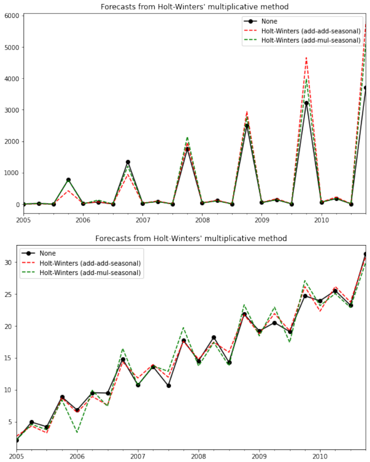
$$\begin{aligned}
\hat{y}_{t+d|t} &= \left(l_t + db_t\right) s_{t-m+(d \mod m)}, \\
l_t &= \alpha \frac{y_t}{s_{t-m}} + (1-\alpha)\left(l_{t-1} + b_{t-1}\right), \\
b_t &= \beta\left(l_t - l_{t-1}\right) + (1-\beta)\, b_{t-1}, \\
s_t &= \gamma \frac{y_t}{l_{t-1} + b_{t-1}} + (1-\gamma)\, s_{t-m}.
\end{aligned}$$

Методы, учитывающие сезонность



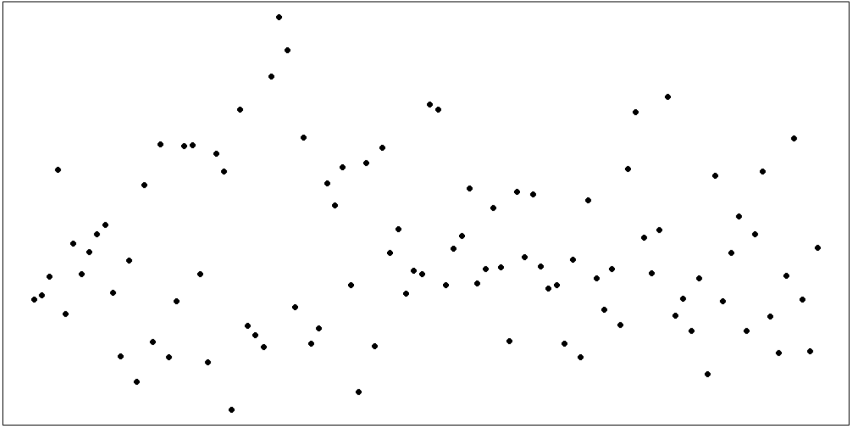
Прогнозы с учётом тренда и сезонности количества ночей, проведённых туристами в Австралии.

Методы, учитывающие сезонность



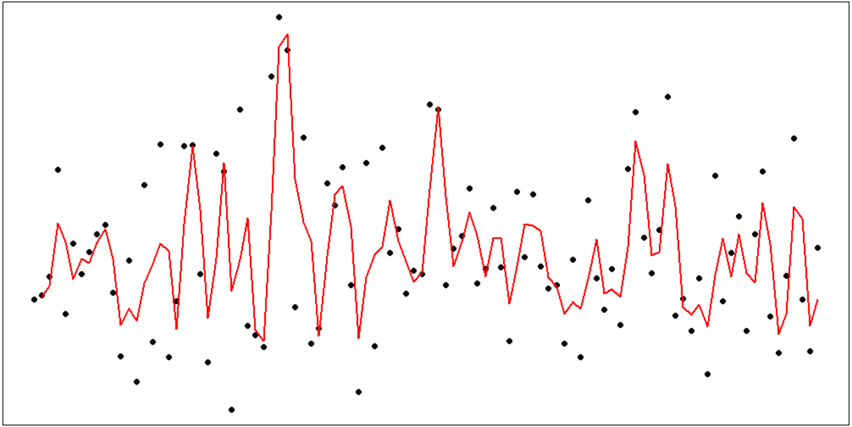
Скользящее среднее

Пусть у нас есть независимый одинаково распределённый во времени шум ε_t:



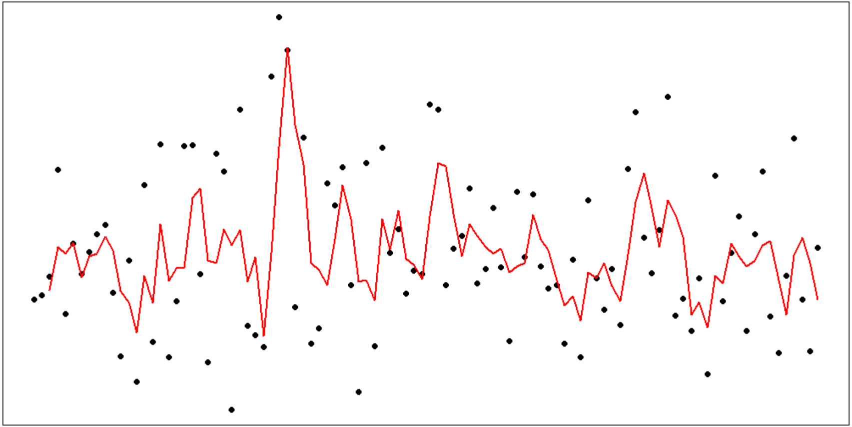
Скользящее среднее

Среднее по двум соседним точкам:



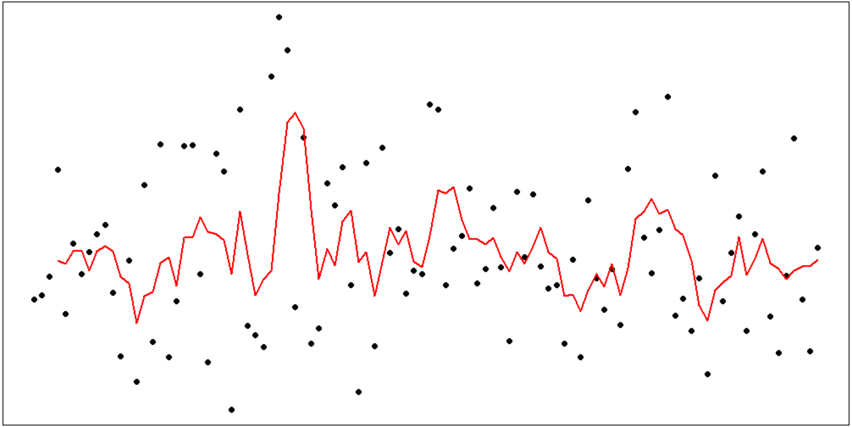
Скользящее среднее

Среднее по трём соседним точкам:



Скользящее среднее

Среднее по четырём соседним точкам:



Авторегрессия

$$AR(p): \quad y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t,$$

где y_t — стационарный ряд с нулевым средним, $\phi_1, \dots, \phi_p$ — константы ($\phi_p \neq 0$), ε_t — гауссов белый шум с нулевым средним и постоянной дисперсией σ_ε^2.

Если среднее равно μ, модель принимает вид

$$y_t = \alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t,$$

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

Другой способ записи:

$$\phi(B)\, y_t = \left(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p\right) y_t = \varepsilon_t,$$

где B — разностный оператор ($B y_t = y_{t-1}$).

Линейная комбинация p подряд идущих членов ряда даёт белый шум.

ARMA (Autogerressive moving average)

$$ARMA(p,q): \quad y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q},$$

где y_t — стационарный ряд с нулевым средним, $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$ — константы ($\phi_p \neq 0$, $\theta_q \neq 0$), ε_t — гауссов белый шум с нулевым средним и постоянной дисперсией σ_ε^2.

Если среднее равно μ, модель принимает вид

$$y_t = \alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q},$$

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

Другой способ записи:

$$\phi\left(B\right) y_t = \theta\left(B\right) \varepsilon_t.$$

Теорема Вольда: любой стационарный ряд может быть аппроксимирован моделью $ARMA(p,q)$ с любой точностью.

ARIMA (Autogerressive integrated moving average)

Ряд описывается моделью $ARIMA(p, d, q)$, если ряд его разностей

$$\nabla^d y_t = (1 - B)^d \, y_t$$

описывается моделью $ARMA(p, q)$.

$$\phi\,(B)\,\nabla^d y_t = \theta\,(B)\,\varepsilon_t.$$

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 информационные критерии
- Начальные приближения можно выбрать с помощью автокорреляций

Автокорреляционная функция (ACF)

Наблюдения временного ряда автокоррелированы.

Автокорреляция:

$$r_\tau = r_{y_t y_{t+\tau}} = \frac{\sum\limits_{t=1}^{T-\tau} (y_t - \bar{y})(y_{t+\tau} - \bar{y})}{\sum\limits_{t=1}^{T} (y_t - \bar{y})^2}, \quad \bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t.$$

$r_\tau \in [-1, 1]$, τ — лаг автокорреляции.

Проверка значимости отличия автокорреляции от нуля:

временной ряд:	$Y^T = Y_1, \dots, Y_T$;
нулевая гипотеза:	$H_0\colon r_\tau = 0$;
альтернатива:	$H_1\colon r_\tau \neq 0$;
статистика:	$T\left(Y^T\right) = \frac{r_\tau\sqrt{T-\tau-2}}{\sqrt{1-r_\tau^2}}$;
нулевое распределение:	$St(T-\tau-2)$.

Частичная автокорреляционная функция (PACF)

Частичная автокорреляция стационарного ряда y_t — автокорреляция остатков авторегрессии предыдущего порядка:

$$\phi_{hh} = \begin{cases} r\left(y_{t+1}, y_t\right), & h = 1, \\ r\left(y_{t+h} - \hat{y}_{t+h}, y_t - \hat{y}_t\right), & h \geqslant 2, \end{cases}$$

где $\hat{y}_{t+h}$ и $\hat{y}_t$ — предсказания регрессий y_{t+h} и y_t на $y_{t+1}, y_{t+2}, \dots, y_{t+h-1}$:

$$\hat{y}_t = \beta_1 y_{t+1} + \beta_2 y_{t+2} + \dots + \beta_{h-1} y_{t+h-1},$$

$$\hat{y}_{t+h} = \beta_1 y_{t+h-1} + \beta_2 y_{t+h-2} + \dots + \beta_{h-1} y_{t+1}.$$

q, Q, p, P

- В модели $ARIMA(p, d, 0)$ ACF экспоненциально затухает или имеет синусоидальный вид, а PACF значимо отличается от нуля при лаге p
- В модели $ARIMA(0, d, q)$ PACF экспоненциально затухает или имеет синусоидальный вид, а ACF значимо отличается от нуля при лаге q

$\Rightarrow$ начальные приближения для p, q, P, Q:

- q: номер последнего лага $\tau < S$, при котором автокорреляция значима
- p: номер последнего лага $\tau < S$, при котором частичная автокорреляция значима

Прогнозирование с помощью ARIMA

1. Строится график ряда, идентифицируются необычные значения.
2. При необходимости делается стабилизирующее дисперсию преобразование.
3. Если ряд нестационарен, подбирается порядок дифференцирования.
4. Анализируются ACF/PACF, чтобы понять, можно ли использовать модели AR(p)/MA(q).
5. Обучаются модели-кандидаты, сравнивается их AIC/AICc.
6. Остатки полученной модели исследуются на несмещённость, стационарность и неавтокоррелированность; если предположения не выполняются, исследуются модификации модели.
7. В финальной модели t заменяется на $T+h$, будущие наблюдения — на их прогнозы, будущие ошибки — на нули, прошлые ошибки — на остатки.

Литература

Hyndman R.J., Athanasopoulos G. *Forecasting: principles and practice*. — OTexts,
https://www.otexts.org/book/fpp