Прогнозирование временных рядов

К. В. Воронцов vokov@forecsys.ru
А.А. Романенко alexromsput@gmail.com

30 Октября 2015

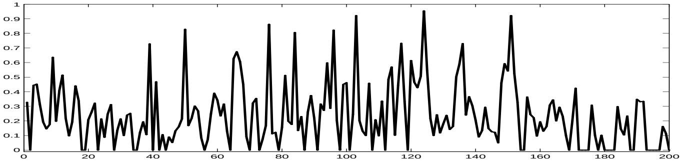
Содержание

Временной ряд

$y_0, y_1, \ldots, y_t, \ldots$ — временной ряд, $y_i \in \mathbb{R}$
$\hat{y}_{t+d}(w) = f_t(y_1, \ldots, y_t; w)$ — модель временного ряда,
где $d = 1, \ldots, D$, D — горизонт (отсрочка, delay),
w — вектор параметров модели

Особенности задачи:

- количество временных рядов $10^6 - 10^8$;
- пропуски в данных;
- нестационарность (непостоянство модели);
- несимметричная, кусочно гладкая функция потерь.



Эконометрика — основной источник задач прогнозирования

Примеры эконометрических временных рядов:

- объёмы продаж в торговых сетях
- объёмы грузовых и пассажирских перевозок
- рыночные цены
- дорожный трафик (прогнозирование пробок)
- объёмы потребления и цены электроэнергии

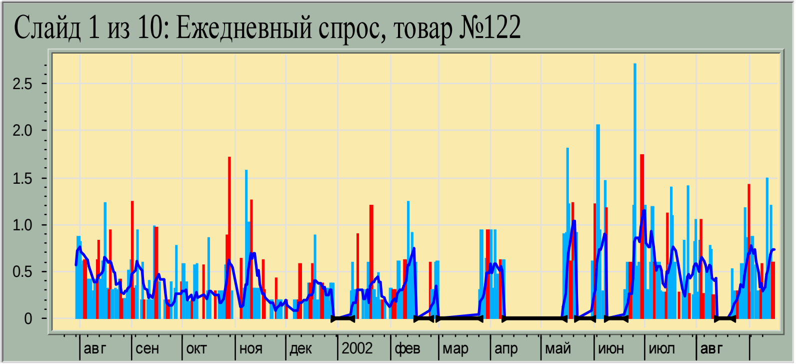
Основные явления в эконометрических временных рядах:

- тренды
- сезонности
- разладки (смены модели ряда)

Марно Вербик. Путеводитель по современной эконометрике, 2008.

Пример. Задача прогнозирования объёмов продаж

Ежедневные объёмы продаж товара



Особенности задачи: целочисленные продажи, продажи зависят от типа товара, тренды, сезонность, пропуски, праздники, промо–акции, скачки, плохо работают сложные модели

Беглый обзор методов прогнозирования

- Методы прогнозирования типа ЭСС
- Адаптивная авторегрессия
- Авторегрессионные модели
- ARMA, ARIMA, GARCH,...
- Гусеница [Голяндина, 2003]
- Адаптивная селекция моделей
- Адаптивная композиция моделей
- Нейросетевые модели
- Прогнозирование разреженных временных рядов
- Прогнозирование при несимметричном функционале
- Прогнозирование плотности распределения

Простое экспоненциальное скользящее среднее

Линейная авторегрессионная модель данных:
$y_{t+1} := \sum_{i=0}^{t} w_i y_i + \varepsilon_{t+1}$
Прогнозная модель: $\hat{y}_{t+1} := \sum_{i=0}^{t} w_i y_i$
ε_{t+1} — непрогнозируемый шум, присутствующий в данных;
$e_{t+1} = y_{t+1} - \hat{y}_{t+1}$ — ошибка прогноза $\hat{y}_{t+1}$, сделанного на шаге t
Среднее арифметическое:

$$\hat{y}_{t+1} := \frac{1}{t+1}\sum_{i=0}^{t} y_i = \hat{y}_t + \frac{1}{t+1}(y_t - \hat{y}_t) = \hat{y}_t + \frac{1}{t+1}e_t$$

Скользящее среднее (СС):

$$\hat{y}_{t+1} := \alpha_t y_t + (1-\alpha_t)\hat{y}_t = \hat{y}_t + \alpha_t(y_t - \hat{y}_t) = \hat{y}_t + \alpha_t \cdot e_t$$

При $\alpha_t = 1 - \frac{1}{t+1}$ имеем среднее арифметическое
При $\alpha_t = \mathrm{const} \in (0,1))$ имеем экспоненциальное СС

Адаптивная авторегрессионная модель

Линейная модель авторегрессии (линейный фильтр):

$$\hat{y}_{t+1}(\mathbf{w}) := \sum_{j=1}^{n} w_j y_{t-j+1}, \quad \mathbf{w} \in \mathbb{R}$$

Метод наименьших квадратов: $e_t^2 \to \min\limits_{w}$.

Один шаг градиентного спуска в каждый момент t:

$$w_j := w_j + h_t e_t y_{t-j+1}.$$

Градиентный шаг в методе скорейшего спуска:

$$h_t = \frac{\alpha}{\sum_{j=1}^{n} y_{t-j+1}^2},$$

где α — аналог параметра сглаживания.

Подбор параметра сглаживания

Чем меньше α, тем больше вес последних точек,
при $\alpha \to 0$ тривиальный прогноз $\hat{y}_{t+1} = y_t$.

Чем больше α, тем сильнее сглаживание,
при $\alpha \to 1$ тривиальный прогноз $\hat{y}_{t+1} = \bar{y}$ (или скользящее среднее).

Оптимальное α^* находим по скользящему контролю:

$$Q(\alpha) = \sum_{t=T_0}^{T_1} \big(\hat{y}_t(\alpha) - y_t\big)^2 \to \min_{\alpha}$$

Эмпирические правила:
если $\alpha^* \in (0.7, 1)$, то ряд стационарен, ЭС работает;
если $\alpha^* \in (0, 0.7)$, то ряд нестационарен, нужна модель тренда.

Следящий контрольный сигнал

$e_t = y_t - \hat{y}_t$ — ошибка прогноза $\hat{y}_t$, сделанного на шаге $t-1$
Следящий контрольный сигнал (tracking signal [Trigg, 1964])

$$K_t = \frac{\hat{e}_t}{\tilde{e}_t} \qquad \begin{aligned} \hat{e}_{t+1} &:= \gamma e_t + (1-\gamma)\hat{e}_t; \\ \tilde{e}_{t+1} &:= \gamma |e_t| + (1-\gamma)\tilde{e}_t. \end{aligned}$$

Рекомендация: $\gamma = 0.05 \ldots 0.1$

Статистический тест адекватности (при $\gamma \geqslant 0.1,\ \ t \to \infty$):
гипотеза H_0: $\mathsf{E}\varepsilon_t = 0,\ \ \mathsf{E}\varepsilon_t\varepsilon_{t+d} = 0$
принимается на уровне значимости δ, если

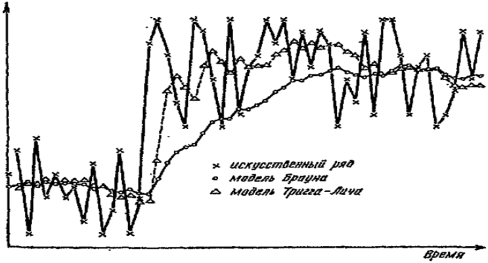
$$|K_t| \leqslant 1.2\Phi_{1-\delta/2}\sqrt{1-\gamma/(1+\gamma)},$$

$\Phi_{1-\delta/2}$ — квантиль нормального распределения,
$\Phi_{1-\delta/2} = \Phi_{0.975} = 1.96$ при $\delta = 0.05$

Модель Тригга–Лича [Trigg, Leach, 1967]

Проблема: адаптивные модели плохо приспосабливаются к резким структурным изменениям

Решение: $\alpha = |K_t|$

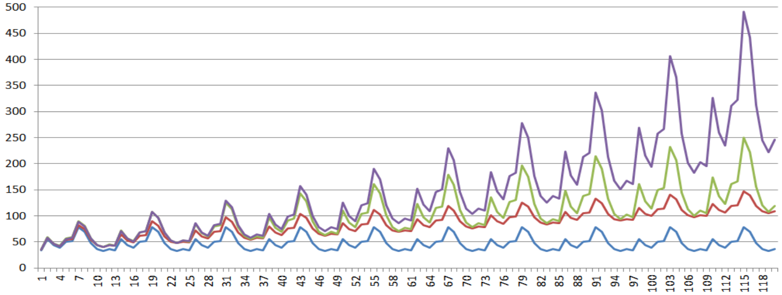


Недостатки:

1) плохо реагирует на одиночные выбросы; ($\alpha_t = |K_{t-1}|$)
2) требует подбора γ, при рекомендации $\gamma = 0.05 \ldots 0.1$.

Примеры трендов и сезонностей

Пример: сочетание тренда и сезонности (модельные данные)



Ряд 1 — сезонность без тренда

Ряд 2 — линейный тренд, аддитивная сезонность

Ряд 3 — линейный тренд, мультипликативная сезонность

Ряд 4 — экспоненциальный тренд, мультипликативная сезонность

Модель Хольта = линейный тренд

Линейный тренд без сезонных эффектов:

$$\hat{y}_{t+d} = a_t + b_t d,$$

где a_t, b_t — адаптивные коэффициенты линейного тренда

Рекуррентная формула:

$$\begin{aligned} a_t &:= \alpha_1 y_t + (1-\alpha_1)(a_{t-1} + b_{t-1}) = \hat{y}_t + \alpha_1 e_t; \\ b_t &:= \alpha_2(a_t - a_{t-1}) + (1-\alpha_2) b_{t-1} = b_{t-1} + \alpha_1 \alpha_2 e_t. \end{aligned}$$

Частный случай — модель линейного роста Брауна:

$$\alpha_1 = 1-\beta, \quad \alpha_2 = 1-\beta \;\; (\text{или } 1+\beta).$$

Модель Уинтерса = мультипликативная сезонность

Мультипликативная сезонность периода s:

$$\hat{y}_{t+d} = a_t \cdot \theta_{t-s+(d \bmod s)},$$

$\theta_0, \ldots, \theta_{s-1}$ — сезонный профиль периода s.
Рекуррентная формула:

$$a_t := \alpha_1(y_t/\theta_{t-s}) + (1-\alpha_1)a_{t-1} = a_{t-1} + \alpha_1 e_t/\theta_{t-s};$$
$$\theta_t := \alpha_2(y_t/a_t) + (1-\alpha_2)\theta_{t-s} = \theta_{t-s} + \alpha_2(1-\alpha_1)e_t/a_t.$$

Доказательство последнего равенства:

$$\theta_t := \theta_{t-s} + \alpha_2\left(y_t/a_t - \theta_{t-s}\right) = \theta_{t-s} + \alpha_2\left(y_t - \theta_{t-s}a_t\right)/a_t =$$
$$\theta_{t-s} + \alpha_2\big(y_t - \theta_{t-s}(a_{t-1} + \alpha_1 e_t/\theta_{t-s})\big)/a_t = \theta_{t-s} +$$
$$+\alpha_2\left(\underbrace{y_t - \theta_{t-s}a_{t-1}}_{e_t} - \alpha_1 e_t\right)/a_t$$

Модель Тейла–Вейджа

Линейный тренд с аддитивной сезонностью периода s:

$$\hat{y}_{t+d} = (a_t + b_t d) + \theta_{t+(d \bmod s)-s}.$$

$a_t + b_t d$ — тренд, очищенный от сезонных колебаний,
$\theta_0, \dots, \theta_{s-1}$ — сезонный профиль периода s.

Рекуррентная формула:

$$a_t := \alpha_1(y_t - \theta_{t-s}) + (1 - \alpha_1)(a_{t-1} + b_{t-1}) = a_{t-1} + b_{t-1} + \alpha_1 e_t;$$
$$b_t := \alpha_2(a_t - a_{t-1}) + (1 - \alpha_2)b_{t-1} = b_{t-1} + \alpha_1\alpha_2 e_t;$$
$$\theta_t := \alpha_3(y_t - a_t) + (1 - \alpha_3)\theta_{t-s} = \theta_{t-s} + \alpha_3(1 - \alpha_1)e_t.$$

Модель Уинтерса с линейным трендом

Мультипликативная сезонность периода s с линейным трендом:

$$\hat{y}_{t+d} = (a_t + b_t d) \cdot \theta_{t+(d \bmod s)-s},$$

$a_t + b_t d$ — тренд, очищенный от сезонных колебаний,
$\theta_0, \ldots, \theta_{s-1}$ — сезонный профиль периода s.

Рекуррентная формула:

$$a_t := \alpha_1(y_t/\theta_{t-s}) + (1-\alpha_1)(a_{t-1}+b_{t-1}) = a_{t-1} + b_{t-1} + \alpha e_t/\theta_{t-s};$$
$$b_t := \alpha_2(a_t - a_{t-1}) + (1-\alpha_2)b_{t-1} = b_{t-1} + \alpha_1\alpha_2 e_t/\theta_{t-s};$$
$$\theta_t := \alpha_3(y_t/a_t) + (1-\alpha_3)\theta_{t-s} = \theta_{t-s} + \alpha_3(1-\alpha_1)e_t/a_t.$$

Модель Уинтерса с экспоненциальным трендом

Мультипликативная сезонность с экспоненциальным трендом:

$$\hat{y}_{t+d} = a_t (r_t)^d \cdot \theta_{t+(d \bmod s)-s},$$

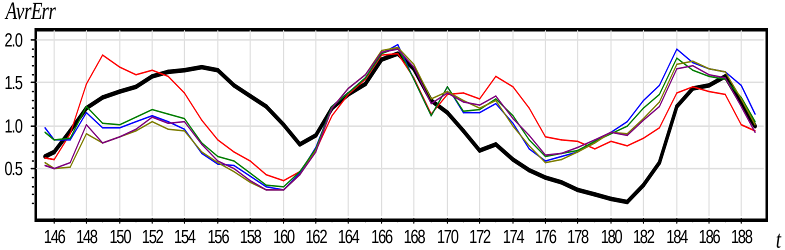
$a_t (r_t)^d$ — экспоненциальный тренд, очищенный от сезонности,
$\theta_0, \ldots, \theta_{s-1}$ — сезонный профиль периода s.

Рекуррентная формула:

$$a_t := \alpha_1 (y_t / \theta_{t-s}) + (1-\alpha_1) a_{t-1} r_{t-1} = a_{t-1} r_{t-1} + \alpha_1 e_t / \theta_{t-1};$$
$$r_t := \alpha_2 (a_t / a_{t-1}) + (1-\alpha_2) r_{t-1} = r_{t-1} + \alpha_1 \alpha_2 e_t / \theta t - 1;$$
$$\theta_t := \alpha_3 (y_t / a_t) + (1-\alpha_3) \theta_{t-s} = \theta_{t-s} + \alpha_3 (1-\alpha_1) e_t / a_t.$$

Пример

Динамика средних ошибок прогнозов для 6 моделей



AFTER [Yang Y., 2004]

ЛАВР , [Воронцов К.В., 2006]

Адаптивная селекция, [Лукашин, 2003; Timmermann, 2006]

Адаптивная композиция,

Агрегирующие алгоритмы [Вовк В., 1998].

Адаптивная селективная модель

Пусть имеется N моделей прогнозирования,
$\hat{y}_{j,t+d}$ — прогноз j-й модели на момент $t+d$,
$e_{jt} = y_t - \hat{y}_{jt}$ — ошибка прогноза в момент t,
$\tilde{e}_{jt} := \gamma|e_{jt}| + (1-\gamma)\tilde{e}_{j(t-1)}$ — сглаженная ошибка.

Неотличимо лучшие модели в момент времени t для порога качества $e^* \geqslant 0$:

$$\Omega_t^* = \left\{ j \mid e_{jt} - \arg\min_{j=1,\dots,N} \tilde{e}_{jt} < e^* \right\}.$$

Адаптивная селективная модель:

$$\hat{y}_{j,t+d} := \frac{1}{\Omega_t^*} \sum_{j \in \Omega_t^*} \hat{y}_{j_t^*,t+d}$$

Требуется подбор γ, рекомендация: $\gamma = 0.01 \dots 0.1$.

Адаптивная композиция моделей

Пусть имеется N моделей прогнозирования,
$\hat{y}_{j,t+d}$ — прогноз j-й модели на момент $t+d$,
$e_{jt} = y_t - \hat{y}_{jt}$ — ошибка прогноза в момент t,
$\tilde{e}_{jt} := \gamma|e_{jt}| + (1-\gamma)\tilde{e}_{jt}$ — экспоненциально сглаженная ошибка.

Линейная (выпуклая) комбинация моделей:

$$\hat{y}_{t+d} = \sum_{j=1}^{N} w_{jt}\hat{y}_{j,t+d}, \qquad \sum_{j=1}^{N} w_{jt} = 1, \quad \forall t.$$

Адаптивный подбор весов [Лукашин, 2003]:

$$w_{jt} = \frac{(\tilde{e}_{jt})^{-1}}{\sum_{s=1}^{N}(\tilde{e}_{st})^{-1}}.$$

Требуется подбор γ, рекомендация: $\gamma = 0.01\ldots0.1$.

Смешивание алгоритмов прогнозирования

Пусть имеется N моделей прогнозирования
$\lambda(y_t, \hat{y}_{j,t})$ — потери алгоритма j при прогнозе элемента y_t
$\mathcal{L}_j(T) = \sum_{t=1}^{T} \lambda(y_t, \hat{y}_{j,t})$ — суммарные потери алгоритма j к моменту времени T
$\mathfrak{M}$ — искомая композиция
Найти $\mathfrak{M}$ такую, что $\forall y_1, \ldots, y_T$,

$$\mathcal{L}_{\mathfrak{M}}(T) \leqslant \arg\min_{j=\overline{1,N}} f\left(\mathcal{L}_j(T)\right),$$

где $f(x)$ — мало отличается от x.
Удаётся строить композиции с теоретическими гарантиями вида:

$$\mathcal{L}_{\mathfrak{M}}(T) \leqslant c \cdot \arg\min_{j=\overline{1,N}} \mathcal{L}_j(T) + a\ln(N).$$

Агрегирующий алгоритм В. Вовка

Прогнозы композиций AA_1 и AA_2

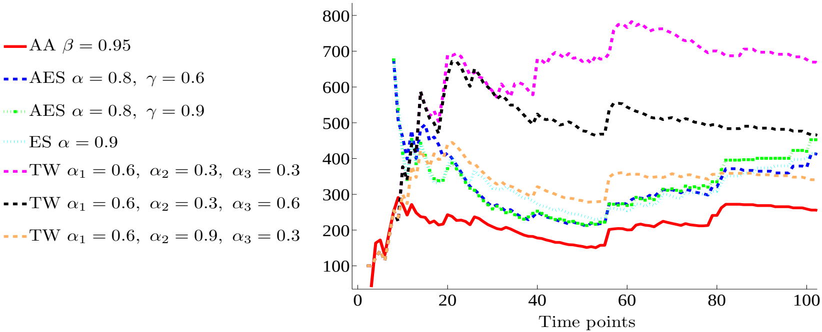
Инициализация: веса базовых алгоритмов $p_{j,0} = 1/N$

Для $t = 0, \dots, T-1$

1. получить предсказания экспертов $\hat{y}_{j,t+1}, \forall j = \overline{1, N}$;
2. построить функцию смешивания:
$$g(x) = \log_{\beta}\left(\sum_{j=1}^{N} p_{j,t} \cdot \beta^{\lambda(y, \hat{y}_{j,t+1})}\right)$$
3. $\hat{y}_{AA_1,t+1} = \frac{Y_2\sqrt{g(Y_1)}+Y_1\sqrt{g(Y_2)}}{\sqrt{g(Y_1)}+\sqrt{g(Y_2)}}$;
$\hat{y}_{AA_2,t+1} = \frac{g(Y_1)-g(Y_2)}{2(Y_2-Y_1)} + \frac{Y_1+Y_2}{2}$;
4. получить исход y_{t+1}; вычислить ошибку $\lambda(y_{t+1}, \hat{y}_{t+1})$;
5. пересчитать веса экспертов $p_{j,t+1} = \beta^{\lambda(y_{t+1}, \hat{y}_j, t+1)} \cdot p_{j,t}$.

Сравнение с базовыми алгоритмами

Эксперимент на реальных данных (1 из 1000 временных рядов)



$$\mathrm{MSE} = \frac{1}{T}\sum_{t=1}^{T} e_t^2$$

Сравнение композиций

Таблица: Сравнение различных композиций, MSE

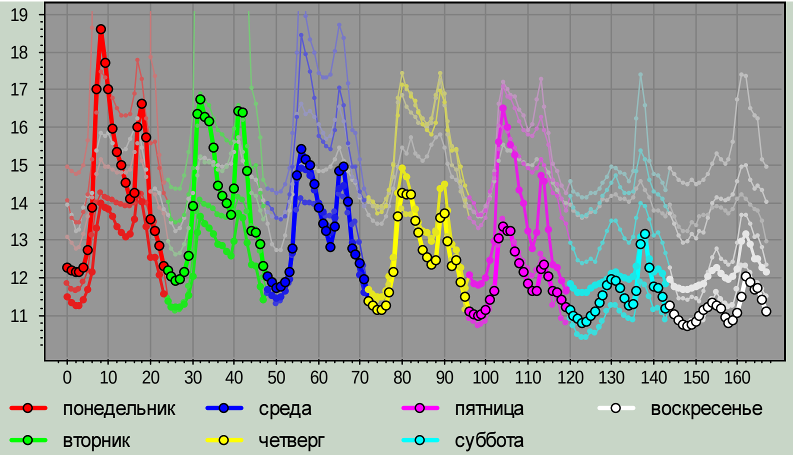
M	AFTER	AK	ЛАВР	AC	AA_1	AA_2
10	6,57	6,66	6,74	6,75	6,43	6,37
25	6,50	6,62	6,92	6,71	6,39	6,31
40	6,55	6,57	6,90	6,66	6,35	6,37
	100%	100%	105%	103%	95%	97%

Сравнение с базовыми алгоритмами (BE - лучший из базовых алгоритмов, TB - теоретическая оценка ошибки):

	1	2	5	10	15	20
AA_1	**21.69**	**32.24**	**57.33**	**94.17**	**110.4**	**139.9**
BE	22.05	32.63	58.24	95.23	111.5	140.6
TB	25.16	38.2	71.80	99.44	141.1	179.7

Авторегрессионная модель

Почасовые цены электроэнергии на бирже NordPool, 2000г.



Особенности задачи: три вложенные сезонности, скачки

Линейная модель авторегрессии

В роли признаков — n предыдущих наблюдений ряда:

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

В роли объектов $\ell = t - n + 1$ моментов в истории ряда:

$$\underset{\ell\times n}{F} = \begin{pmatrix} y_t & y_{t-1} & y_{t-2} & \cdots & y_{t-n+1} \\ y_{t-1} & y_{t-2} & y_{t-3} & \cdots & y_{t-n} \\ y_{t-2} & y_{t-3} & y_{t-4} & \cdots & y_{t-n-1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ y_n & y_{n-1} & y_{n-2} & \cdots & y_1 \end{pmatrix}, \quad \underset{\ell\times 1}{y} = \begin{pmatrix} y_{t+1} \\ y_t \\ y_{t-1} \\ \cdots \\ y_{n+1} \end{pmatrix}$$

Функционал квадрата ошибки:

$$Q_t(w, X^\ell) = \sum_{i=n+1}^{t+1} \big(\hat{y}_i(w) - y_i\big)^2 = \|Fw - y\|^2 \to \min_w$$

Модель ARMA

ARMA(p,q): $y_1, \dots, y_t$

- $y_t = c + \underbrace{\sum_{i=1}^{p} \alpha_i y_{t-i}}_{AR} + \underbrace{\sum_{j=1}^{q} \beta_j \varepsilon_{t-j}}_{MA} + \varepsilon_t;$
- $L : Ly_t = y_{t-1}$ — лаговый оператор;
  $L^i : L^i y_t = L^{i-1}(Ly_t) = y_{t-i};$
- $(1 - \sum_{i=1}^{p} \alpha_i L^i) y_t = c + (1 + \sum_{j=1}^{q} \beta_j L^j) \varepsilon_t;$

$$y_t = \mu + \frac{F(L)}{H(L)} \varepsilon_t \text{ — каноническая запись ARMA}$$

$F(L) = (1 + \sum_{j=1}^{q} \beta_j L^j)$ — оператор скользящего среднего;
$H(L) = (1 - \sum_{i=1}^{p} \alpha_i L^i)$ — оператор авторегрессии;
ε_t — случайная компонента, $E\varepsilon_t = 0, E\varepsilon_i \varepsilon_j = 0$
Ряд y_t является стационарным, если корни $H(z) = 0$ лежат вне единичного круга комплексной плоскости.

Модель ARIMA

y_t — НЕстационарный, т.е. $H(z)$ — имеет d единичных корней;

- $H(L) = (1 - \sum_{i=1}^{p} \alpha_i L^i) = (1 - \sum_{i=1}^{\tilde{p}} \alpha_i L^i)(1 - L)^d$
- $(1 - \sum_{i=1}^{\tilde{p}} \alpha_i L^i)(1 - L)^d y_t = c + (1 + \sum_{j=1}^{q} \beta_j L^j)\varepsilon_t;$

$z_t = (1-L)^d y_t = \mu + \dfrac{F(L)}{H(L)}\varepsilon_t$ — каноническая запись ARIMA(p,q,d)

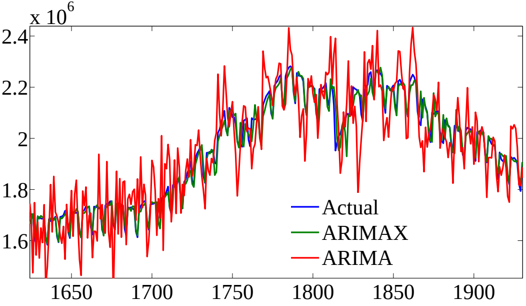
Можно также выписывать аналог для сезонных временных рядов ARIMA$(p, q, d) \times (P, Q, D)_s$:

$$(1 - L)^d (1 - L^s)^D y_t = \mu + \frac{F(L)}{H(L)} \frac{(1 + \sum_{j=1}^{Q} \gamma_j L^{s \cdot j})}{(1 - \sum_{i=1}^{P} \delta_i L^{s \cdot i})} \varepsilon_t$$

Модель ARIMAX

y_t — НЕстационарный, X_t — вектор регрессоров из $\mathbb{R}^N$, известный до начала момента прогнозирования;
ARIMAX(p,q,d):

$$z_t = \mu + \sum_{n=1}^{N} \frac{v_n(L)}{u_n(L)} X_{n,t} + \frac{F(L)}{H(L)} \varepsilon_t$$



Pro&Cons ARIMA

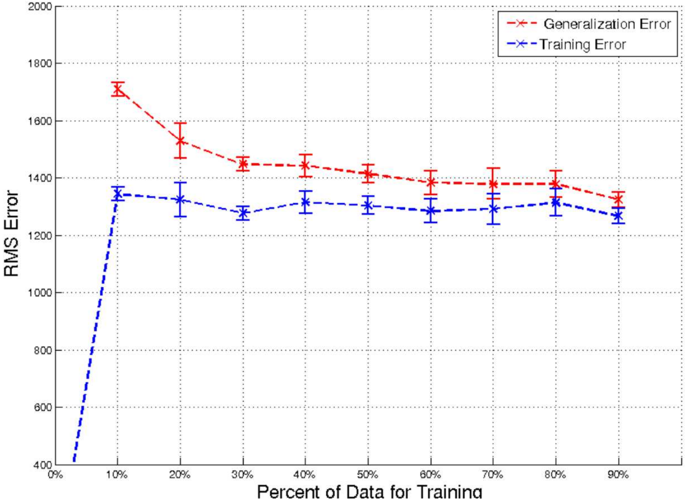
Заключение по ARIMA–моделям

- обобщают модели а-ля ЭСС (ЭСС = ARIMA(0,1,1) при $\mu = 0$)
- позволяют учитывать внешние факторы (акции, скачки цен, температуру и т.д.)
- не работает при наличии пропусков в данных;
- тяжело обучить (на практике используют перебор $p, q, d = \overline{0, 3}$;
- методы обучения опираются на нормальность ε_t
    - не для всех временных рядов удаётся найти соответствующую модель;
    - плохо работает на разреженных и коротких временных рядах;

Box, G. E. P. – Jenkins, G. M. – Reinsel, G. C.: Time Series Analysis: Forecasting and Control. John Wiley & Sons Inc., New York, 2008

Зачем нужно что-то ещё более сложное?

Прогнозирование объёмов электроэнергии



Резюме в конце лекции

- Адаптивные методы хорошо работают, когда рядов много, и прогнозировать их надо быстро
- Простота адаптивных методов компенсируется селективными и композиционными моделями
- При этом различные особенности рядов моделируются в базовых алгоритмах
- Для временных рядов со сложной структурой можно использовать более сложные алгоритмы

Лукашин Ю. П. Адаптивные методы краткосрочного прогнозирования временных рядов. Финансы и статистика, 2003.