

Прикладная статистика 11. Анализ временных рядов, часть вторая.

Рябенко Евгений
riabenko.e@gmail.com

28 апреля 2014 г.

Прогнозирование временного ряда

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

Задача прогнозирования: найти функцию f_T :

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

где $d \in \{1, 2, \dots, D\}$, D — горизонт прогнозирования.

Функция выбирается следующим образом:

$$Q_T = \sum_{t=1}^T (y_t - \hat{y}_t)^2 \rightarrow \min_{f_T}.$$

Прогнозирование временного ряда

Метод прогнозирования определяет выбор функции f_T .

Модель ряда определяет также тип отклонения ряда от прогноза; в результате задаётся случайный процесс, порождающий данные, что позволяет прогнозировать распределение y_{T+d} .

Аддитивная модель:

$$y_t = f_t + \varepsilon.$$

Мультипликативная модель:

$$y_t = f_t \cdot \varepsilon.$$

	Модель	Метод
AR(1)	$y_{t+1} = \phi_1 y_t + \varepsilon_t$	$\hat{y}_{t+1 t} = \hat{\phi}_1 y_t$
MA(1)	$y_{t+1} = \varepsilon_t + \theta_1 \varepsilon_{t-1}$	$\hat{y}_{t+1 t} = \hat{\theta}_1 (y_t - \hat{y}_t)$

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:

$$\begin{aligned}\hat{y}_{T+1|T} &= \alpha y_T + \alpha(1-\alpha)y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \dots = \\ &= \alpha y_T + (1-\alpha)\hat{y}_T = \hat{y}_T + \alpha(y_T - \hat{y}_T).\end{aligned}$$

$\alpha \uparrow 1 \Rightarrow$ больший вес последним точкам,

$\alpha \downarrow 0 \Rightarrow$ большее сглаживание.

Метод подходит для прогнозирования рядов без тренда и сезонности.

Простое экспоненциальное сглаживание (метод Брауна)

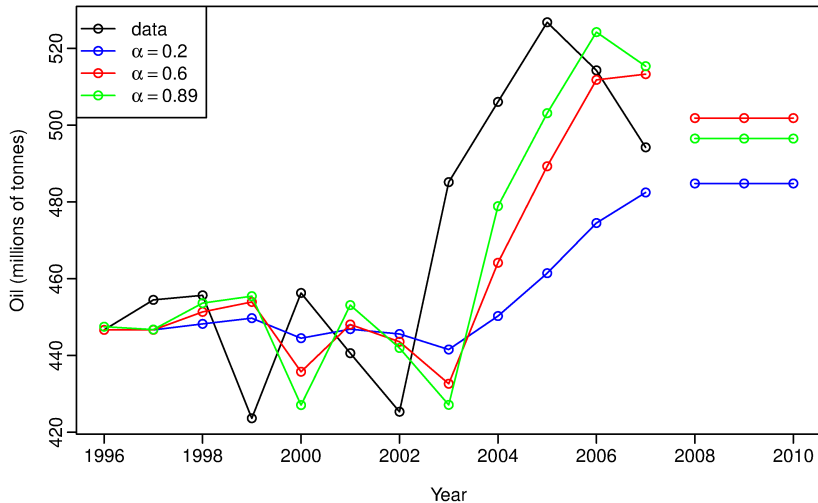
Запись в компонентной форме:

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

$$y_{t+d|t} = \hat{y}_{t+1} \quad \forall d.$$

Нужно задавать l_0 . Часто берут $l_0 = y_1$; другой вариант — оптимизировать l_0 вместе с α .

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):

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

Мультипликативный линейный (экспоненциальный) тренд:

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

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

Методы, учитывающие тренд

Аддитивный затухающий тренд:

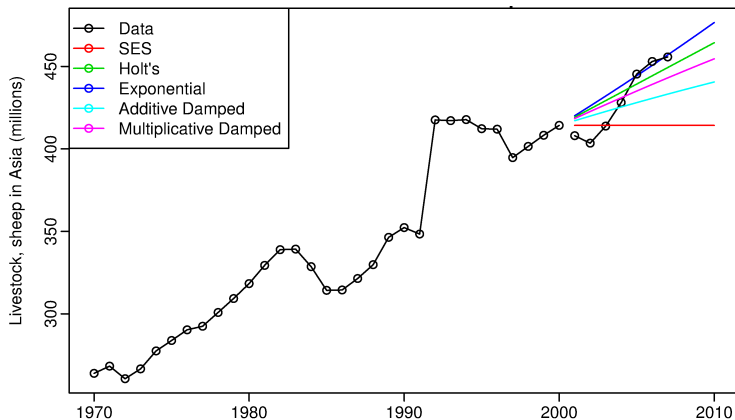
$$\begin{aligned}\hat{y}_{t+d|t} &= l_t + b_t^{(\phi + \phi^2 + \dots + \phi^d)}, \\ l_t &= \alpha y_t + (1 - \alpha) l_{t-1} b_{t-1}^\phi, \\ b_t &= \beta (l_t - l_{t-1}) + (1 - \beta) \phi b_{t-1}.\end{aligned}$$

Мультипликативный затухающий тренд:

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

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

Методы, учитывающие тренд



Прогнозы с учётом тренда поголовья овец в Азии.

	SES	Holt's	Exponential	Additive damped	Multiplicative damped
α	1	0.98	0.98	0.99	0.98
β		0	0	0	0.00
ϕ				0.98	0.98

Методы, учитывающие сезонность

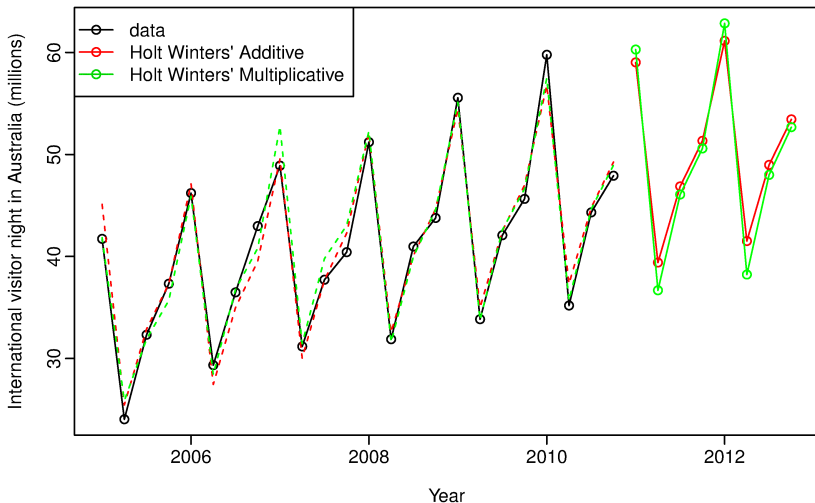
Аддитивная сезонность с периодом длины m (метод Хольта-Уинтерса):

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

Мультипликативная сезонность:

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

Методы, учитывающие сезонность



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

Модели экспоненциального сглаживания

	Сезонность		
Тренд	N (None)	A (Additive)	M (Multiplicative)
N (None)	(N,N)	(N,A)	(N,M)
A (Additive)	(A,N)	(A,A)	(A,M)
Ad (Additive damped)	(Ad,N)	(Ad,A)	(Ad,M)
M (Multiplicative)	(M,N)	(M,A)	(M,M)
Md (Multiplicative damped)	(Md,N)	(Md,A)	(Md,M)

Дополнительно можно предположить аддитивную (A) или мультипликативную (M) ошибку (тип ошибки не влияет на точечный прогноз). Мультипликативная ошибка подходит только для строго положительных рядов.

Итоговую модель можно записать в виде $ETS(\cdot, \cdot, \cdot)$.

Настройка модели экспоненциального сглаживания

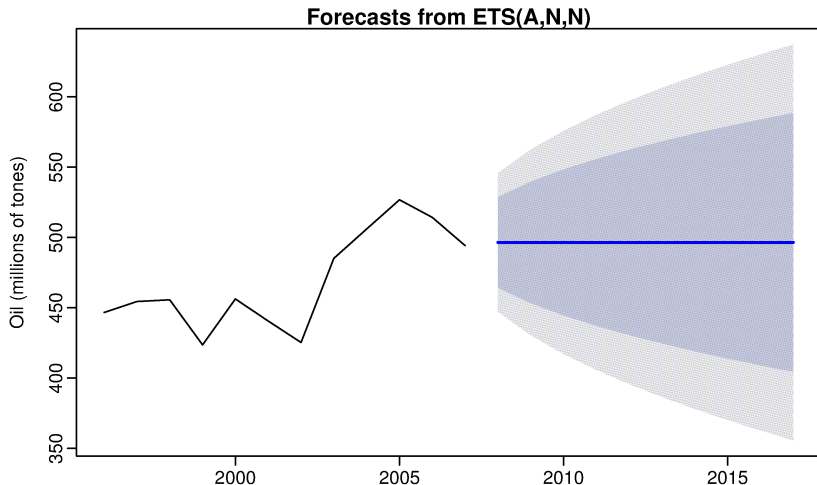
Функция `ets` пакета `forecast` в R позволяет автоматически выбрать лучшую модель экспоненциального сглаживания:

```
ets(y, model="ZZZ", damped=NULL, alpha=NULL, beta=NULL,
    gamma=NULL, phi=NULL, additive.only=FALSE, lambda=NULL,
    lower=c(rep(0.0001,3), 0.8), upper=c(rep(0.9999,3),0.98),
    opt.crit=c("lik","amse","mse","sigma","mae"), nmse=3,
    bounds=c("both","usual","admissible"),
    ic=c("aicc","aic","bic"), restrict=TRUE)
```

Построить прогноз можно с помощью функции `forecast`:

```
forecast(object, h=ifelse(object$m>1, 2*object$m, 10),
    level=c(80,95), fan=FALSE, simulate=FALSE, bootstrap=FALSE,
    npaths=5000, PI=TRUE, lambda=object$lambda, ...)
```

Настройка модели экспоненциального сглаживания



Для данных о добыче нефти в Саудовской Аравии функция ets выбирает простое экспоненциальное сглаживание.

Меры качества точечного прогноза

Mean squared error:

$$MSE = \frac{1}{D} \sum_{d=1}^D (\hat{y}_{T+d|T} - y_{T+d})^2.$$

Mean absolute error:

$$MAE = \frac{1}{D} \sum_{d=1}^D |\hat{y}_{T+d|T} - y_{T+d}|.$$

Mean absolute percentage error:

$$MAPE = \frac{100}{D} \sum_{d=1}^D \left| \frac{\hat{y}_{T+d|T} - y_{T+d}}{y_{T+d}} \right|.$$

Mean absolute scaled error:

$$MASE = \frac{1}{D} \sum_{d=1}^D |\hat{y}_{T+d|T} - y_{T+d}| \left/ \frac{1}{T-1} \sum_{t=2}^T |y_t - y_{t-1}| \right.$$

Информационные критерии

AIC — информационный критерий Акаике:

$$AIC = \log \left(\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2 \right) + \frac{T + 2k}{T},$$

где k — число параметров модели;

$AICc$ — он же с поправкой на случай небольшого размера выборки:

$$AICc = \log \left(\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2 \right) + \frac{n + k}{T - k - 2};$$

BIC (SIC) — байесовский (Шварца) информационный критерий:

$$BIC = \log \left(\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2 \right) + \frac{k \log T}{T}.$$

Относительное качество прогноза

U-коэффициент Тейла оценивает качество прогноза относительно прогноза последним значением:

$$U(d) = \sqrt{\frac{\sum_{t=R}^{T-d} (\hat{y}_{t+d|t} - y_{t+d})^2}{\sum_{t=R}^{T-d} (y_t - y_{t+d})^2}}, \quad d = 1, \dots, D.$$

Если $U(d) = 1$, то прогноз $\hat{y}_{t+d|t}$ так же хорош, как «наивный прогноз» последним значением; если $U(d) < 1$, прогноз $\hat{y}_{t+d|t}$ лучше наивного, $U(d) > 1$ — хуже.

Сравнение качества двух прогнозов

y_1, \dots, y_T — временной ряд,

$\hat{y}_{1R}, \dots, \hat{y}_{1T}$ — прогноз на период R, \dots, T первым методом,

$\hat{\epsilon}_{1R}, \dots, \hat{\epsilon}_{1T}$ — остатки первого прогноза,

$\hat{y}_{2R}, \dots, \hat{y}_{2T}$ — прогноз на период R, \dots, T вторым методом,

$\hat{\epsilon}_{2R}, \dots, \hat{\epsilon}_{2T}$ — остатки второго прогноза;

$g(y_t, \hat{y}_{it})$ — произвольная функция потерь,

$$d_t = g(y_t, \hat{y}_{1t}) - g(y_t, \hat{y}_{2t}).$$

H_0 : среднее $d_t = 0$,

H_1 : среднее $d_t < \neq > 0$.

Непараметрические критерии

$$H_0: \text{med } d_t = 0,$$

$$H_1: \text{med } d_t < \neq > 0.$$

Критерий знаков:

$$T = \sum_{t=R}^T [d_t > 0].$$

Критерий знаковых рангов Уилкоксона:

$$W = \sum_{t=R}^T r(|d_t|) \text{sign}(d_t).$$

Критерий Диболда-Мариано

Будем для простоты считать, что $R = 1$.

Пусть d_1, \dots, d_T — выборочная траектория слабо стационарного случайного процесса, тогда

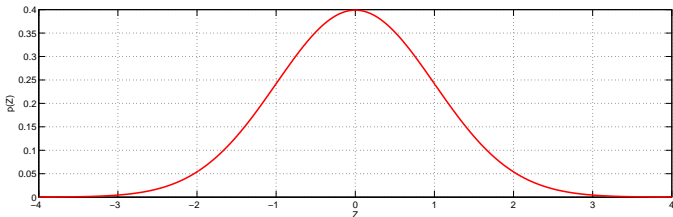
$$\sqrt{T} (\bar{d} - \mu) \xrightarrow{d} N(0, f),$$

где $\bar{d} = \frac{1}{T} \sum_{t=1}^T d_t$, μ — неизвестное среднее значение процесса, f — его дисперсия.

Критерий Диболда-Мариано

нулевая гипотеза: $H_0: \mathbb{E}d_t = 0;$ альтернатива: $H_1: \mathbb{E}d_t < \neq > 0;$

статистика: $B = \frac{\bar{d}}{\sqrt{\hat{f}/T}}, \hat{f} = \sum_{\tau=-M}^M \hat{r}_\tau, M = T^{1/3};$
 $B \sim N(0, 1)$ при $H_0;$



достигаемый уровень значимости:

$$p(b) = \begin{cases} 1 - \text{ncdf}(b, 0, 1), & H_1: \mathbb{E}d_t > 0, \\ \text{ncdf}(b, 0, 1), & H_1: \mathbb{E}d_t < 0, \\ 2(1 - \text{ncdf}(|b|, 0, 1)), & H_1: \mathbb{E}d_t \neq 0. \end{cases}$$

Критерий Диболда-Мариано

Модификация для коротких рядов (Harvey, Leybourne, Newbold):

$$B^* = \frac{B}{\sqrt{\frac{T+1-2d+\frac{d(d-1)}{T}}{T}}}$$

Сравнение качества нескольких прогнозов

Пусть имеется эталонный прогноз ряда (например, «наивным» методом) и k других прогнозов,

$$\hat{y}_{t+d} = \{\hat{y}_{j,t+d}\}_{j=0}^k.$$

Как проверить, что хотя бы один прогноз лучше эталонного?

Пусть f — мера качества прогноза относительно эталона, такая, что $f > 0$, когда качество эталона ниже, и $f < 0$, когда качество эталона выше. Пример:

$$f = L(\hat{y}_{j,t+1}) - L(\hat{y}_{0,t+1})$$

(можно добавить ещё штраф за число параметров алгоритма).

$\hat{f}_{t+d} = f(Z_{t+d}, \hat{y}_{t+d}, \hat{\beta}_t) \in \mathbb{R}^k$ — вектор оценок качества прогноза, Z_{t+d} содержит значения y_{t+d} и дополнительные предикторы x_{t+d} , $\hat{\beta}_t$ — вектор оценок параметров всех прогнозирующих алгоритмов.

Критерий reality check Уайта

нулевая гипотеза: $H_0: \max_{j=1, \dots, k} \mathbb{E}f_j^* \leq 0,$

$$\mathbb{E}f_j^* = \mathbb{E}f(Z_{t+d}, \hat{y}_{t+d}, \beta^*), \beta^* = \text{plim } \hat{\beta}_t,$$

альтернатива: $H_1: H_0$ неверна.

При выполнении ряда теоретических предположений

$$\sqrt{n}(\bar{f} - \mathbb{E}f^*) \xrightarrow{d} N(0, \Omega),$$

$$\bar{f} = \frac{1}{n} \sum_{t=R}^{T-d} \hat{f}_{t+d} \text{ — среднее относительное качество прогнозов,}$$
$$n = T - d - R + 1.$$

Для оценки Ω и вычисления достигаемого уровня значимости используется бутстреп или Монте-Карло.

Критерий reality check Уайта

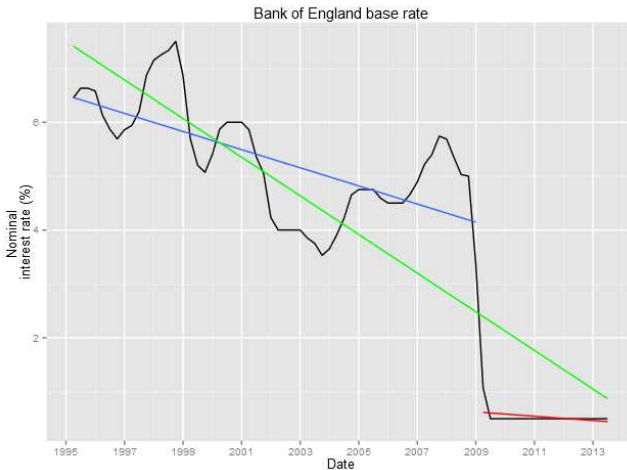
Пример (Sullivan, Timmermann, White, 1999, 2001): к ряду промышленного индекса Доу-Джонса с 1 января 1897 по 30 июня 1998 (27447 отсчётов) было применено большое количество моделей — 7846 моделей технического анализа и 9452 календарных. В качестве эталона рассматривалась стратегия долгосрочного инвестирования. Критерий качества — средний ожидаемый доход по всем инвестициям.

Критерий Уайта показал, что лучший метод технического анализа выигрывает у эталона, в то время как лучший календарный метод, по всей видимости, переобучен — он существенно лучше эталона по критерию Диболда-Мариано, но не лучше по критерию Уайта.

Модификация Романо-Вольфа

Построив на основе критерия Уайта нисходящую процедуру, можно найти все методы, дающие прогноз лучше эталона, асимптотически контролируя при этом FWER — групповую вероятность ошибки (Romano, Wolf, 2005).

Структурное изменение модели



Как проверить, не нужно ли настраивать разные модели на разных участках ряда?

Критерий Чоу

Пусть $\hat{\varepsilon}_t, k$ — остатки и число параметров общей модели, $\hat{\varepsilon}_{1t}, k_1$ — модели на первом участке, $\hat{\varepsilon}_{2t}, k_2$ — на втором.

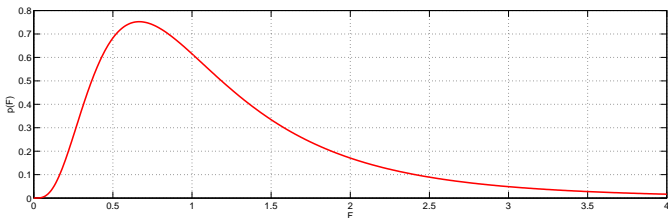
нулевая гипотеза: H_0 : структура модели стабильна;

альтернатива: H_1 : H_0 неверна;

$$\text{статистика: } RSS = \sum_{t=T_1}^{T_3} \hat{\varepsilon}_t^2, \quad RSS_1 = \sum_{t=T_1}^{T_2} \hat{\varepsilon}_{1t}^2, \quad RSS_2 = \sum_{t=T_2}^{T_3} \hat{\varepsilon}_{2t}^2,$$

$$F = \frac{(RSS - RSS_1 - RSS_2)/(k_1 + k_2 - k)}{(RSS_1 + RSS_2)/(n - k_1 - k_2)}, \quad n = T_3 - T_1 + 1;$$

$$F \sim F(k_1 + k_2 - k, n - k_1 - k_2) \text{ при } H_0;$$

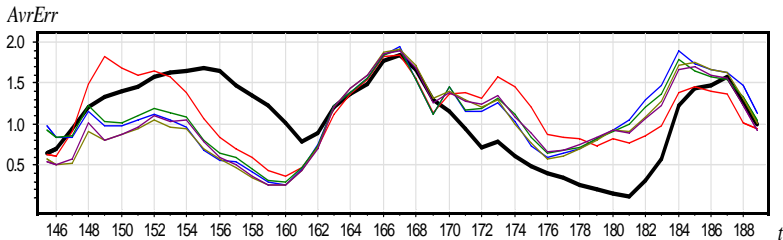


достигаемый уровень значимости:

$$p(f) = f \text{cdf}(1/f, n - k_1 - k_2, k_1 + k_2 - k).$$

Адаптивная селекция и композиция

Динамика средних ошибок прогнозов для 6 моделей (по реальным данным объёмов продаж в супермаркете):



Идея: кажется, можно успеть включить наиболее удачные модели и отключить менее удачные...

Адаптивная селективная модель

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

Лучшая модель в момент времени t :

$$j_t^* = \operatorname{argmin}_{j=1,\dots,k} \tilde{\varepsilon}_{jt}.$$

Адаптивная селективная модель:

$$\hat{y}_{j,t+d} := \hat{y}_{j_t^*,t+d}$$

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

Адаптивная композиция моделей

Пусть имеется k моделей прогнозирования,

$\hat{y}_{j,t+d}$ — прогноз j -й модели на момент $t + d$,

$\varepsilon_{jt} = y_t - \hat{y}_{jt}$ — ошибка прогноза в момент t ,

$\tilde{\varepsilon}_{jt} := \gamma|\varepsilon_{jt}| + (1 - \gamma)\tilde{\varepsilon}_{jt}$ — экспоненциально сглаженная ошибка.

Линейная (выпуклая) комбинация моделей:

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

Адаптивный подбор весов [Лукашин, 2003]:

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

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

Адаптация весов с регуляризацией

На каждом шаге t веса определяются по МНК и сглаживаются:

$$\begin{cases} \sum_{i=0}^t \beta^{t-i} \left(\sum_{j=1}^k w_j \hat{y}_{j,i} - y_i \right)^2 + \lambda \sum_{j=1}^k (w_j - w_{j,t-1})^2 \rightarrow \min_{w_j, j=1, \dots, k} \\ \sum_{j=1}^k w_j = 1. \end{cases}$$

β — коэффициент «забывания» предыстории,

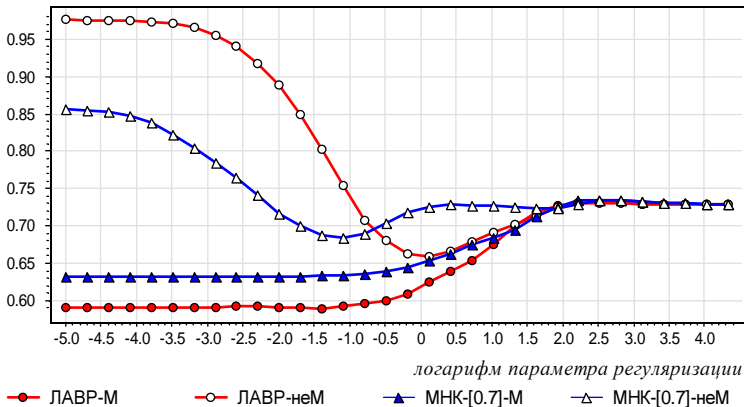
λ — коэффициент регуляризации.

Дополнительные варианты:

- $\beta \rightarrow 0$ — локальная адаптация весов с регуляризацией (оставляем в функционале только одно слагаемое, $i = t$)
- $w_j \geq 0$ — монотонный корректор

Задача прогнозирования временных рядов продаж

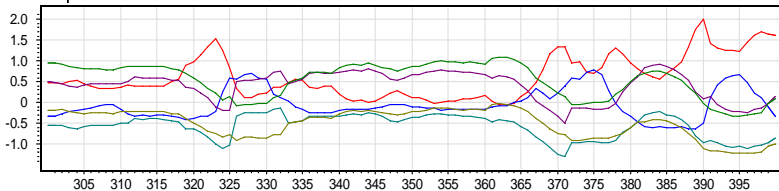
Средняя ошибка прогнозов на скользящем контроле ($T = 620$)



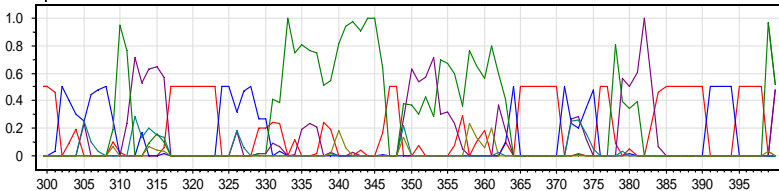
- ЛАВР-М — лучший результат, причём можно брать $\lambda = 0$
- Ограничение монотонности — сильный регуляризатор

Фрагменты динамики весов базовых моделей

Без ограничения монотонности:



С ограничением монотонности:



Сравнение моделей

Средняя ошибка прогнозов на скользящем контроле ($T = 620$)

базовый-1	0.7142
базовый-2	0.7294
базовый-3	0.7534
базовый-4	0.7624
базовый-5	0.7624
базовый-6	0.7664
базовый-7	0.7793
базовый-8	0.7793

ЛАВР+Монот	0.5899
селекция+сглаживание, γ_{opt}	0.5956
МНК+Монот, $\beta=0.7$	0.6314
ЛАВР без Монот	0.6591
МНК без Монот, $\beta=0.7$	0.6834
МНК по всем данным	0.7142
среднее	0.7294
селекция без сглаживания	0.9107

- Базовые модели, их усреднение, неадаптивный МНК по всем данным — работают плохо
- Адаптивная селекция работает хорошо, если подобрать γ
- $\gamma_{opt} = 0.2 \dots 0.3$ — усреднение по 3...5 дням

Причинность по Грейнджеру

Между рядами x_1, \dots, x_T и y_1, \dots, y_T существует **причинная связь Грейнджера** $x_t \rightarrow y_t$, если дисперсия ошибки оптимального прогноза \hat{y}_{t+1} по $y_1, \dots, y_t, x_1, \dots, x_t$ меньше, чем только по y_1, \dots, y_t .

Причинность по Грейнджеру является необходимым, но не достаточным условием причинно-следственной связи.

x_1, \dots, x_T и y_1, \dots, y_T **взаимосвязаны**, если $x_t \rightarrow y_t$ и $y_t \rightarrow x_t$.

Критерий Грейнджера

$$y_t = \alpha + \sum_{i=1}^{k_1} \phi_{1i} y_{t-i} + \sum_{i=1}^{k_2} \phi_{2i} x_{t-i} + \varepsilon_t.$$

k_1 и k_2 выбирается по информационному критерию.

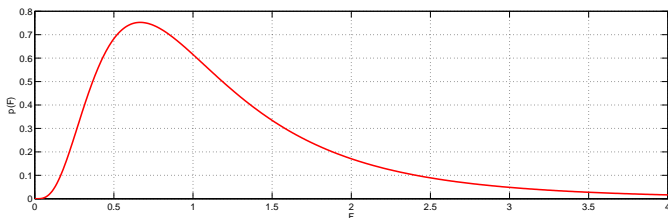
$$x_t \rightarrow y_t \Rightarrow \exists \phi_{2i} \neq 0.$$

нулевая гипотеза: $H_0: \phi_{21} = \dots = \phi_{2k_2} = 0;$

альтернатива: $H_1: H_0$ неверна;

статистика: $F = \frac{(RSS_r - RSS_{ur})/k_2}{RSS_{ur}/(T - k_1 - k_2 - 1)};$

$F \sim F(k_2, T - k_1 - k_2 - 1)$ при $H_0;$

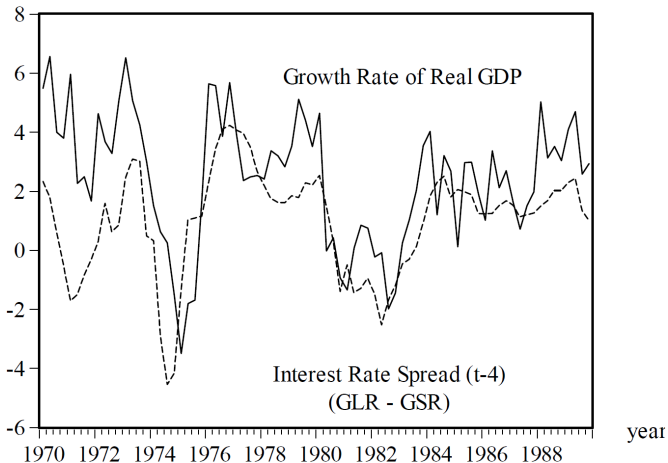


достигаемый уровень значимости:

$$p(f) = fcdf(1/f, T - k_1 - k_2 - 1, k_1).$$

Критерий Грейнджера

Маржа сверх процентной ставки и рост ВВП, ФРГ:



Критерий Грейнджера

y	x	k_1	k_2	$F(y \rightarrow x)$	$F(y \rightarrow x)$
$\Delta_4 \ln GDP_r$	$\Delta_4 \ln M1_r$	4	4	6.087***	1.918
		8	8	3.561**	1.443
$\Delta_4 \ln GDP_r$	$GLR - GSR$	4	4	3.160*	3.835**
		8	8	1.927(*)	2.077*
$\Delta_4 \ln M1_r$	$GLR - GSR$	4	4	5.615***	1.489
		8	8	2.521*	1.178

$\Delta_4 \ln GDP_r$ — годовой прирост ВВП в процентах, $\Delta_4 \ln M1_r$ — годовой прирост фактического количества денег в процентах, GLR — рост государственных облигаций, GSR — трёхмесячная ставка денежного рынка во Франкфурте.

(*), *, **, *** — значимость на уровне 0.1, 0.05, 0.01 и 0.001.

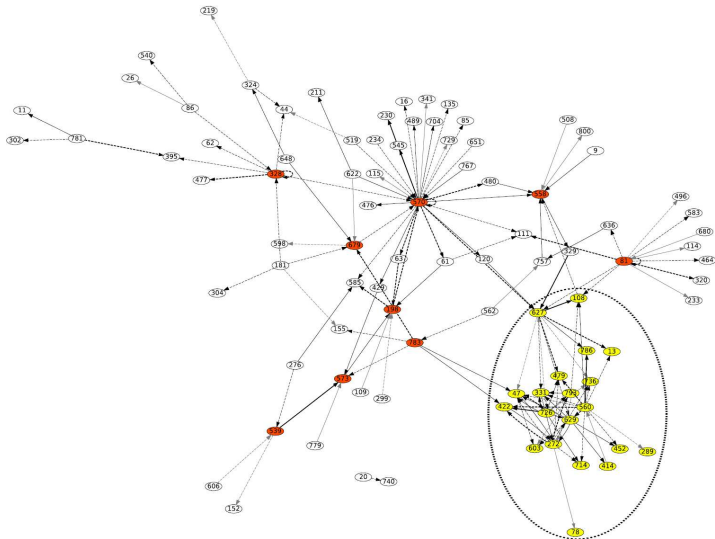
Многомерный критерий Грейнджера

Зависимость между признаками x и y может оцениваться с учётом возможной зависимости от всех остальных признаков:

$$y_t = \alpha + \sum_{i=1}^{k_1} \phi_{1i} y_{t-i} + \sum_{i=1}^{k_2} \phi_{2i} x_{t-i} + \sum_{j=1}^m \sum_{i=1}^{k_{j+2}} \phi_{(j+2)i} z_{t-i}^j + \varepsilon_t.$$

Для задач с большим количеством признаков могут использоваться регуляризаторы (лассо, ридж).

Граф причинности по Грейнджеру



Критерий Грейнджера + поправка на множественную проверку гипотез

Требования к решению задачи прогнозирования временных рядов

- визуализация данных, анализ распределения признака (оценка необходимости трансформации), оценка наличия выбросов;
- анализ автокорреляционной и частичной автокорреляционной функций;
- настройка модели ARIMA: автоматический подбор модели, проверка её соответствия особенностям ряда, при необходимости — корректировка модели, анализ остатков (нормальность, несмещённость, гомоскедастичность, неавтокоррелированность, стационарность);
- настройка модели экспоненциального сглаживания: автоматический подбор модели, проверка её соответствия особенностям ряда, корректировка, анализ остатков;
- визуальный анализ и формальная проверка наличия структурных изменений в моделях;
- сравнение и выбор лучшей модели по критерию Диболда-Мариано;
- выводы.

Литература

- модели экспоненциального сглаживания (exponential smoothing) — Hyndman, Лукашин (другие названия);
- критерий Диболда-Мариано (Diebold-Mariano) и его модификация для коротких рядов — Harvey;
- reality check Уайта (White) и нисходящая процедура на его основе — Romano;
- применение reality check к данным технического анализа — Sullivan;
- критерий Чоу (Chow test) — Chow;
- адаптивная селекция и композиция прогнозирующих алгоритмов — Лукашин, Воронцов;
- причинность по Грейнджеру (Granger causality) — Kirchgassner, глава 3;
- пример графа причинности — Opgen-Rhein.

Литература

- Воронцов К.В., Егорова Е.В. (2006). *Динамически адаптируемые композиции алгоритмов прогнозирования*. Искусственный Интеллект, Донецк, 2, 277–280.
- Лукашин Ю.П. *Адаптивные методы краткосрочного прогнозирования временных рядов*. — М.: Финансы и статистика, 2003.
- Harvey D., Leybourne S., Newbold P. (1997). *Testing the equality of prediction mean squared errors*. International Journal of Forecasting, 13, 281–291.
- Hyndman R.J., Athanasopoulos G. *Forecasting: principles and practice*. — OTexts, 2013. <https://www.otexts.org/book/fpp>
- Chow G.C. (1960). *Tests of Equality Between Sets of Coefficients in Two Linear Regressions*. Econometrica, 28(3), 591–605.
- Kirchgassner G., Wolters J., Hassler U. *Introduction to modern time series analysis*. — Heidelberg: Springer, 2013.
- Opgen-Rhein R., Strimmer K. (2007). *Learning causal networks from systems biology time course data: an effective model selection procedure for the vector autoregressive process*. BMC Bioinformatics, 8 Suppl 2, S3.
- Romano J., Wolf M. (2005). *Stepwise multiple testing as formalized data snooping*. Econometrica, 73(4), 1237–1282.
- Sullivan R., Timmermann A., White H. (2003). *Forecast evaluation with shared data sets*. International Journal of Forecasting, 19(2), 217–227.