

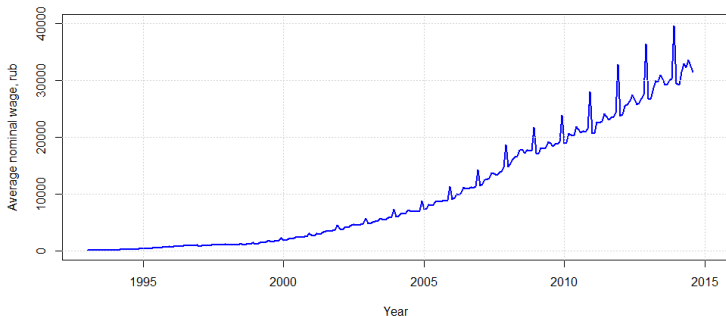
Прикладной статистический анализ данных.
10. Анализ временных рядов, часть вторая.

Рябенко Евгений
riabenko.e@gmail.com

I/2015

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 :

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

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

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):

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

- 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:

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

$\alpha \uparrow 1 \Rightarrow$ 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:

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

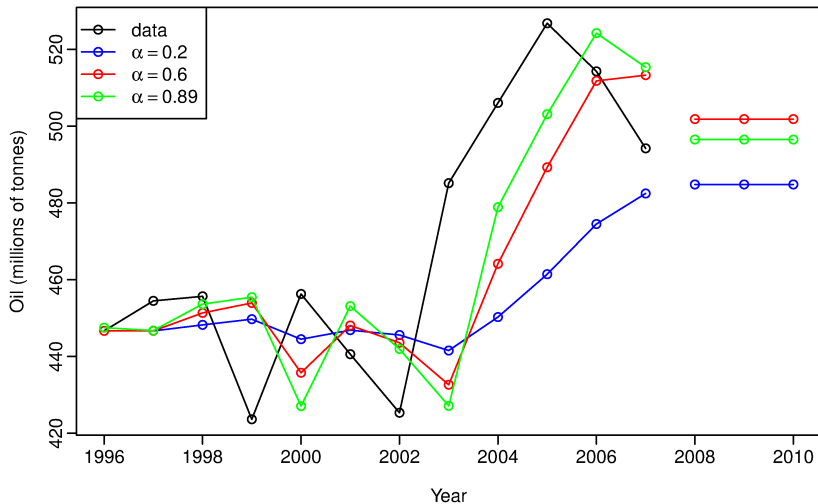
- Прогноз зависит от l_0 :

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

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):

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

Мультипликативный линейный (экспоненциальный) тренд:

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

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

Методы, учитывающие тренд

Аддитивный затухающий тренд:

$$\hat{y}_{t+d|t} = l_t + \left(\phi + \phi^2 + \dots + \phi^d \right) b_t,$$

$$l_t = \alpha y_t + (1 - \alpha) (l_{t-1} + \phi b_{t-1}),$$

$$b_t = \beta (l_t - l_{t-1}) + (1 - \beta) \phi b_{t-1}.$$

Мультипликативный затухающий тренд:

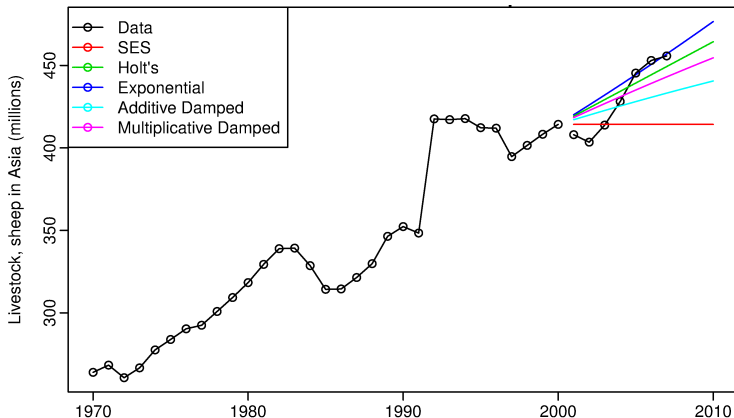
$$\hat{y}_{t+d|t} = l_t b_t^{(\phi + \phi^2 + \dots + \phi^d)},$$

$$l_t = \alpha y_t + (1 - \alpha) l_{t-1} b_{t-1}^\phi,$$

$$b_t = \beta \frac{l_t}{l_{t-1}} + (1 - \beta) b_{t-1}^\phi.$$

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

Методы, учитывающие тренд



Прогнозы поголовья овец в Азии с учётом тренда.

	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

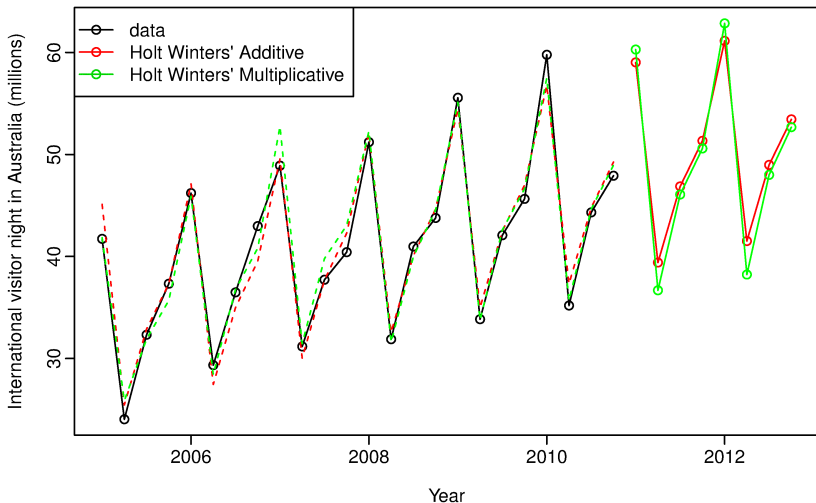
Аддитивная сезонность с периодом длины m (метод Хольта-Уинтерса):

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

Мультипликативная сезонность:

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

Методы, учитывающие сезонность



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

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

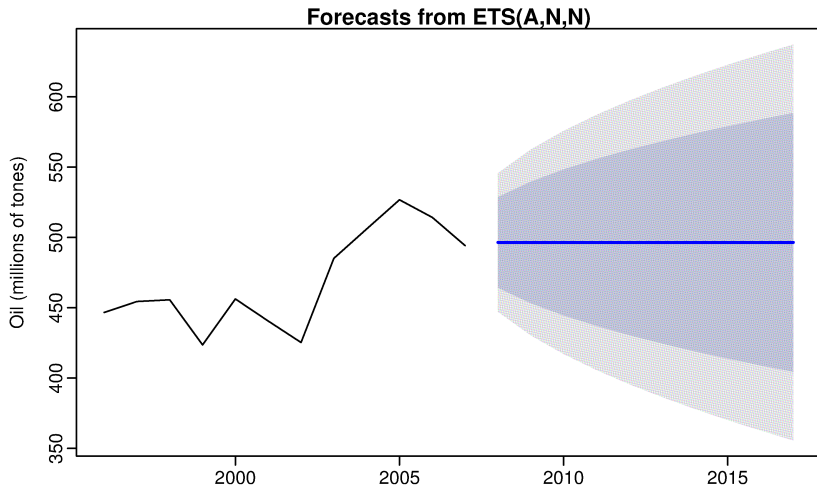
Функция `ets` пакета `forecast` в R позволяет автоматически выбрать лучшую модель экспоненциального сглаживания:

```
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, ...)
```

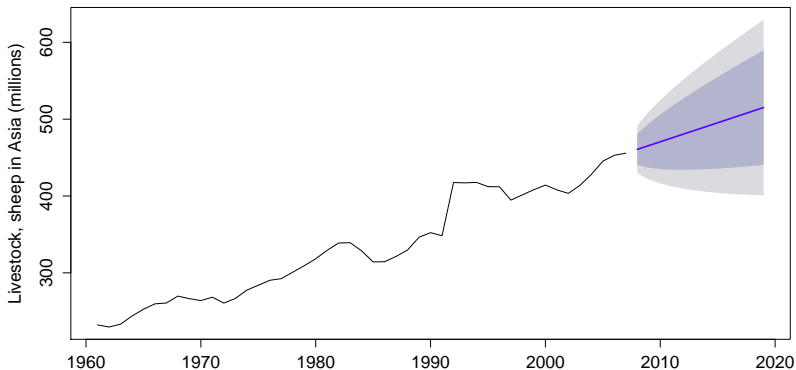
Примеры



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

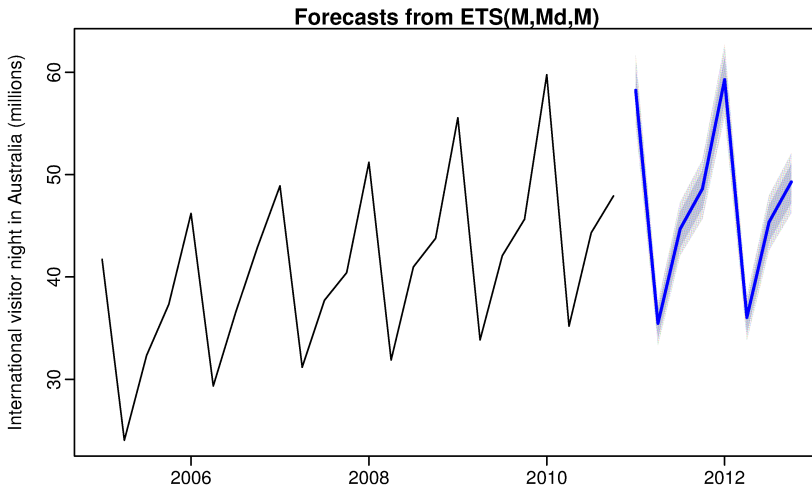
Примеры

Forecasts from ETS(M,A,N)



Для поголовья овец в Азии функция ets выбирает модель с мультипликативными ошибкой и аддитивным линейным трендом.

Примеры



Для количества ночей, проведённых туристами в Австралии, функция ets выбирает модель с мультипликативными ошибкой, сезонностью и затухающим трендом.

Меры качества точечного прогноза

Mean squared error:

$$MSE = \frac{1}{T - R + 1} \sum_{t=R}^T (\hat{y}_t - y_t)^2.$$

Mean absolute error:

$$MAE = \frac{1}{T - R + 1} \sum_{t=R}^T |\hat{y}_t - y_t|.$$

Mean absolute percentage error:

$$MAPE = \frac{100}{T - R + 1} \sum_{t=R}^T \left| \frac{\hat{y}_t - y_t}{y_t} \right|.$$

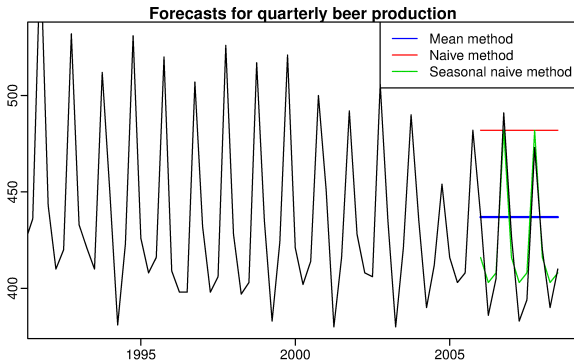
Symmetric mean absolute percentage error:

$$SMAPE = \frac{200}{T - R + 1} \sum_{t=R}^T \left| \frac{\hat{y}_t - y_t}{\hat{y}_t + y_t} \right|.$$

Mean absolute scaled error:

$$MASE = \frac{1}{T - R + 1} \sum_{t=R}^T |\hat{y}_t - y_t| \bigg/ \frac{1}{T - 1} \sum_{t=2}^T |y_t - y_{t-1}|.$$

Меры качества точечного прогноза



Метод	RMSE	MAE	MAPE	MASE
средним	38.01	33.78	8.17	2.30
наивный	70.91	63.91	15.88	4.35
наивный сезонный	12.97	11.27	2.73	0.77

Относительное качество прогноза

U-коэффициент Тейла оценивает качество прогноза относительно наивного :

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

Если $U(d) = 1$, то прогноз $\hat{y}_{t+d|t}$ так же хорош, как наивный; 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,

(например, $|\hat{\epsilon}_{it}|$ или $\hat{\epsilon}_{it}^2$),

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

H_0 : среднее $d_t = 0$,

H_1 : среднее $d_t < \neq > 0$.

Критерий знаковых рангов Уилкоксона:

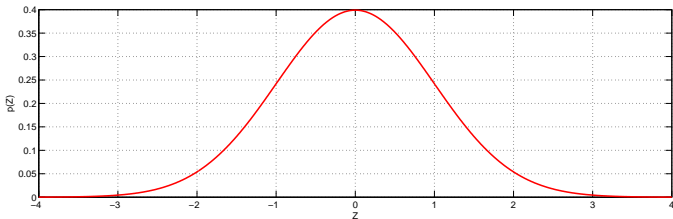
$$W = \sum_{t=R}^T \text{rank}(|d_t|) \text{sign}(d_t).$$

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 $H_0.$



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 технических 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

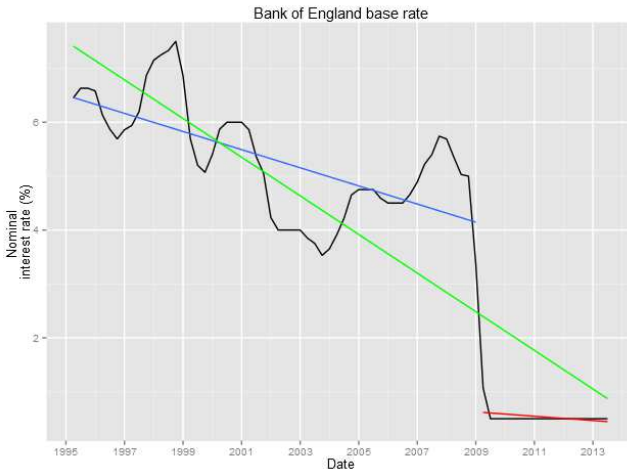
Пример (Sullivan, Timmermann, White, 1999, 2001): к ряду промышленного индекса Доу-Джонса с 1 января 1897 по 30 июня 1998 (27447 отсчётов) было применено большое количество моделей — 7846 моделей технического анализа и 9452 календарных. В качестве эталона рассматривалась стратегия долгосрочного инвестирования. 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 (Romano, Wolf, 2005).

Структурное изменение модели



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

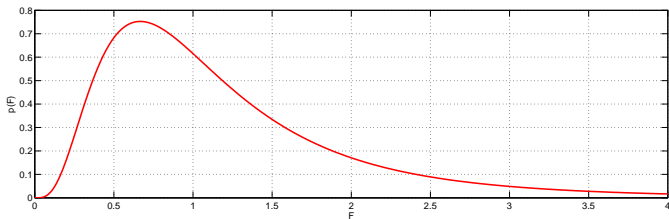
Критерий Чоу

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;

$$\text{статистика: } RSS = \sum_{t=T_1}^{T_3} \hat{\varepsilon}_t^2, \quad RSS_1 = \sum_{t=T_1}^{T_2} \hat{\varepsilon}_{1t}^2, \quad RSS_2 = \sum_{t=T_2}^{T_3} \hat{\varepsilon}_{2t}^2,$$
$$F = \frac{(RSS - RSS_1 - RSS_2)/(k_1 + k_2 - k)}{(RSS_1 + RSS_2)/(n - k_1 - k_2)}, \quad n = T_3 - T_1 + 1;$$
$$F \sim F(k_1 + k_2 - k, n - k_1 - k_2) \text{ при } H_0.$$



Критерий Чоу

Применение в R:

```
n <- length(y)
```

```
m <- ets(y)
```

```
k <- length(m$par)
```

```
rss <- sum(m$residuals^2)
```

```
m1 <- ets(y[1:24])
```

```
k1 <- length(m1$par)
```

```
rss1 <- sum(m1$residuals^2)
```

```
m2 <- ets(y[25:n])
```

```
k2 <- (m2$aic + 2*m2$loglik)/2
```

```
rss2 <- sum(m2$residuals^2)
```

```
f <- ((rss-rss1-rss2)/(k1+k2-k)) / ((rss1-rss2)/(n-k1-k2))
```

```
pf(f, k1+k2-k, n-k1-k2)
```

Примеры

Продажи австралийского вина:

<https://yadi.sk/d/Tkbp67fYVwd>

Пример для самостоятельной работы:

<https://datamarket.com/data/list/?q=provider:tsdl>

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.

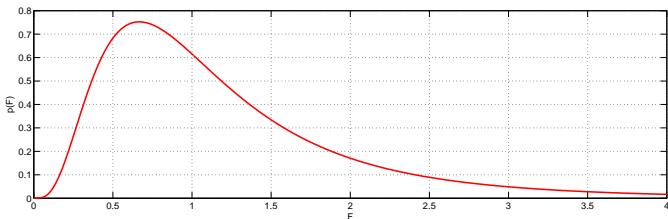
$$x_t \rightarrow y_t \Rightarrow \exists \phi_{2i} \neq 0.$$

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

альтернатива: $H_1: H_0$ неверна;

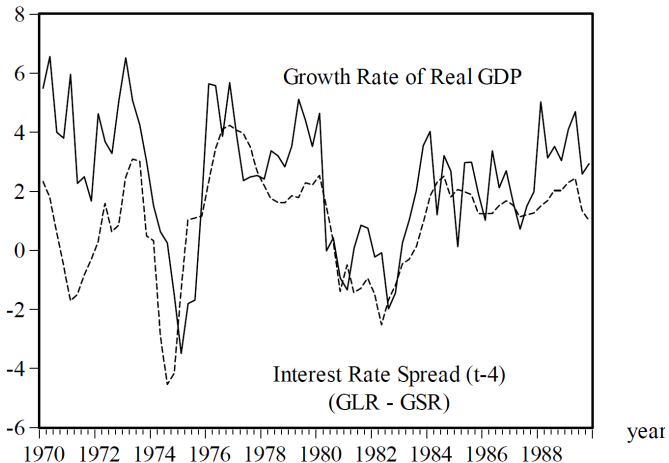
статистика: $F = \frac{(RSS_r - RSS_{ur})/k_2}{RSS_{ur}/(T - k_1 - k_2 - 1)};$

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



Критерий Грейнджера

Маржа сверх процентной ставки и рост ВВП, ФРГ:



Критерий Грейнджера

y	x	k_1	k_2	$F(y \leftarrow x)$	$F(y \rightarrow x)$
$\Delta_4 \ln GDP_r$	$\Delta_4 \ln M1_r$	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

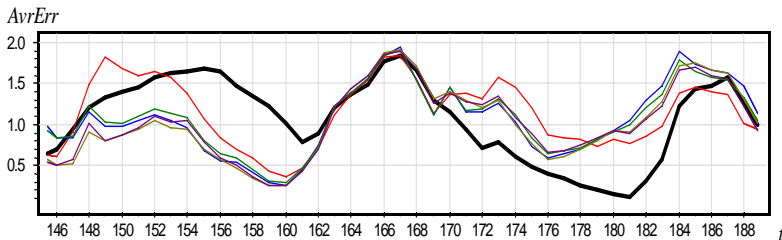
Зависимость между признаками x и y может оцениваться с учётом возможной зависимости от всех остальных признаков:

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

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

Адаптивная селекция и композиция

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



Идея: 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 ,

$\tilde{\varepsilon}_{jt} = \sum_{l=R}^{t-d} \gamma^{t-d-l} |\varepsilon_{jl}|$ — экспоненциально сглаженная ошибка.

Лучшая модель в момент времени t :

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

Адаптивный селективный прогноз:

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

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 ,

$\tilde{\varepsilon}_{jt} = \sum_{l=R}^{t-d} \gamma^{t-d-l} |\varepsilon_{jl}|$ — экспоненциально сглаженная ошибка.

Линейная 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:

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

Адаптация весов с регуляризацией

На каждом шаге t веса определяются по МНК и сглаживаются:

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

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;
- если прогнозов слишком много, индивидуальные прогнозы можно кластеризовать, усреднить по кластерам и подбирать веса средних.

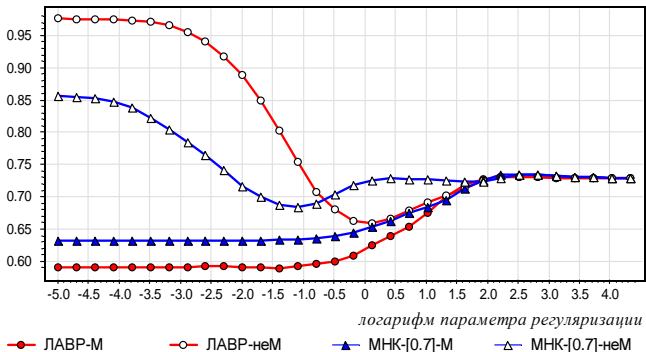
Другие способы комбинирования

- регрессия ряда на несколько первых главных компонент прогнозов в скользящем окне;
- МНК с байесовским регуляризатором на веса (априорное распределение — равномерное);
- экспоненциально сглаживаются не модули, а квадраты ошибок;
- вместо экспоненциального сглаживания используется усреднение в окне;
- ...

Сравнение способов комбинирования

(Воронцов, Егорова, 2006): лучший результаты даёт адаптация весов с монотонным корректором и без регуляризации; хорошо работает адаптивная селекция с подбором параметра сглаживания.

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



(Genre et al., 2013): относительный выигрыш сложных методов по сравнению с простым усреднением прогнозов невелик и в большинстве случаев статистически незначим (“forecast combination puzzle”).

Агрегирующий алгоритм Вовка

Пусть $y \in [0, 1]$; рассмотрим адаптивную композицию со следующими весами:

$$\begin{aligned}w_{j1} &= \frac{1}{k}, \\w_{jt}^* &= e^{-\eta \varepsilon_{jt}^2} w_{j(t-1)}, \\w_{jt} &= \frac{w_{jt}^*}{\sum_{j=1}^k w_{jt}^*},\end{aligned}$$

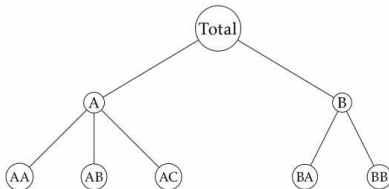
$j = 1, \dots, k$, $\eta \in (0, \infty)$ — скорость обучения.

Пусть $\varepsilon_{AGG,t}$ — остатки такого алгоритма. При $0 < \eta \leq 2$

$$\sum_{t=1}^T \varepsilon_{AGG,t}^2 \leq \sum_{t=1}^T \varepsilon_{jt}^2 + \frac{\ln k}{\eta} \quad \forall j = 1, \dots, k.$$

Иерархия рядов

Часто необходимо прогнозировать совокупности временных рядов иерархической структуры. Например, продажи могут группироваться по товарным группам, складам, поставщикам и т. д.



$$\begin{pmatrix} y_t \\ y_{A,t} \\ y_{B,t} \\ y_{AA,t} \\ y_{AB,t} \\ y_{AC,t} \\ y_{BA,t} \\ y_{BB,t} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_{AA,t} \\ y_{AB,t} \\ y_{AC,t} \\ y_{BA,t} \\ y_{BB,t} \end{pmatrix}, \quad \mathbf{y}_t = \mathbf{S}\mathbf{y}_{K,t}.$$

Подходы к прогнозированию

Снизу вверх: прогнозы рядов более высоких уровней иерархии получаются суммированием прогнозов нижнего уровня.

- информация не теряется из-за агрегирования, но
- прогнозировать ряды нижнего уровня часто сложнее.

Сверху вниз: прогноз суммарного ряда y_t распределяется согласно средним долям:

$$p_j = \frac{1}{T} \sum_{t=1}^T \frac{y_{j,t}}{y_t}$$

или долям средних:

$$p_j = \frac{\sum_{t=1}^T \frac{y_{j,t}}{T}}{\sum_{t=1}^T \frac{y_t}{T}}.$$

- прогнозировать суммированный ряд легко, но
- из-за агрегирования теряется информация (например, если компоненты имеют разную сезонность).

Подходы к прогнозированию

Оптимальная комбинация: ряд каждого уровня прогнозируется отдельно, затем прогнозы корректируются в сторону большей согласованности с помощью регрессии

$$\hat{\mathbf{y}}_h = S\beta_h + \varepsilon_h, \quad \mathbb{E}\varepsilon_h = 0, \quad \text{cov } \varepsilon = \Sigma_h;$$

$$\varepsilon_h \approx S\varepsilon_{K,h} \Rightarrow \tilde{\mathbf{y}}_h = S \left(S^T S \right)^{-1} \hat{\mathbf{y}}_h.$$

Метод реализован в пакете `hts`.

Метод с теоретическими гарантиями: (Стенина, Стрижов, 2015) если суммарные потери при прогнозировании всех рядов иерархии измеряются с помощью функции из класса дивергенций Брегмана, проецирование вектора прогнозов на множество векторов, удовлетворяющих структуре иерархии, не увеличивает суммарные потери.

Метод Хольта-Уинтерса (ETS(A,A,A)):

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

Если m велико, число сезонных коэффициентов можно сократить с помощью разложения Фурье:

$$\begin{aligned}s_t &= \sum_{j=1}^k s_{j,t}, \\ s_{j,t} &= s_{j,t-1} \cos \lambda_j + s_{j,t-1}^* \sin \lambda_j + \gamma_1 d_t, \\ s_{j,t}^* &= -s_{j,t-1} \sin \lambda_j + s_{j,t-1}^* \cos \lambda_j + \gamma_2 d_t, \\ \lambda_j &= \frac{2\pi j}{m},\end{aligned}$$

d_t — ARMA(p,q)-процесс, γ_1, γ_2 — параметры сглаживания.

- реализован с автоматическим подбором параметров в пакете `forecast`, функция `tbats`;
- позволяет моделировать ряды с нецелым периодом сезонности;
- может учитывать сезонность с несколькими разными периодами (для этого ряд должен быть объектом типа `msts`, а не `ts`);
- наследует недостаток методов экспоненциального сглаживания — не может включать регрессоры.

Пример — прогнозирование недельных данных:

```
y <- ts(x, frequency=52.18)
fit <- tbats(y)
fc <- forecast(fit)
```

regARIMA позволяет моделировать сложные сезонности за счёт включения фурье-гармоник по длинным сезонным периодам в регрессоры. Кроме того, можно добавить индикаторы плавающих праздников.

Пример — прогнозирование дневных данных с годовой сезонностью:

```
y <- ts(x, frequency=7)
z <- fourier(ts(x, frequency=365.25), K=5)
zf <- fourierf(ts(x, frequency=365.25), K=5, h=100)
fit <- auto.arima(y, xreg=cbind(z,holiday))
fc <- forecast(fit, xreg=cbind(zf,holidayf), h=100)
```

K можно выбрать, минимизируя AIC .

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, при необходимости — 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;
- причинность по Грейнджеру (Granger causality) — Kirchgassner, глава 3;
- пример построения графа причинности — Opgen-Rhein;
- адаптивная селекция и композиция прогнозирующих алгоритмов — Воронцов, Genre;
- агрегирующий алгоритм Вовка — Калнишкан;
- иерархические ряды — Hyndman, 9.4;
- TBATS — De Livera.

Литература

- Воронцов К.В., Егорова Е.В. (2006). *Динамически адаптируемые композиции алгоритмов прогнозирования*. Искусственный Интеллект, Донецк, 2, 277–280.
- Калнишкан Ю.А. *Введение в методы конкурентного предсказания*.
http://www.clrc.rhul.ac.uk/people/yura/concurrent_rus.pdf
- Лукашин Ю.П. *Адаптивные методы краткосрочного прогнозирования временных рядов*. — М.: Финансы и статистика, 2003.
- Chow G.C. (1960). *Tests of Equality Between Sets of Coefficients in Two Linear Regressions*. *Econometrica*, 28(3), 591–605.
- De Livera A.M., Hyndman R.J., Snyder R.D. (2011). *Forecasting Time Series With Complex Seasonal Patterns Using Exponential Smoothing*. *Journal of the American Statistical Association*, 106(496), 1513–1527.
- Genre V., Kenny G., Meyler A., Timmermann A. (2013). *Combining expert forecasts: Can anything beat the simple average?* *International Journal of Forecasting*, 29(1), 108–121.
- 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, 2015. <https://www.otexts.org/book/fpp>
- 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.