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

Рябенко Евгений
riabenko.e@gmail.com

14 ноября 2014 г.

Прогнозирование временного ряда

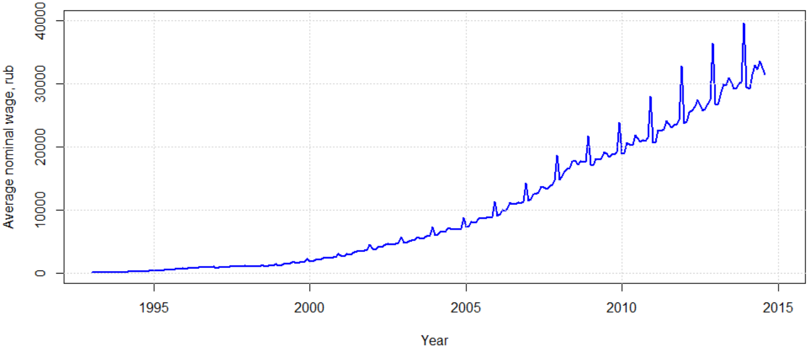
Временной ряд: $y_1, \dots, y_T, \dots, \quad y_t \in \mathbb{R}$, — значения признака, измеренные через постоянные временные интервалы.

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

Прогнозирование временного ряда



- средняя дневная цена акций IBM;
- среднемесячный уровень безработицы;
- годовой объём производства автомобилей.

Компоненты временных рядов

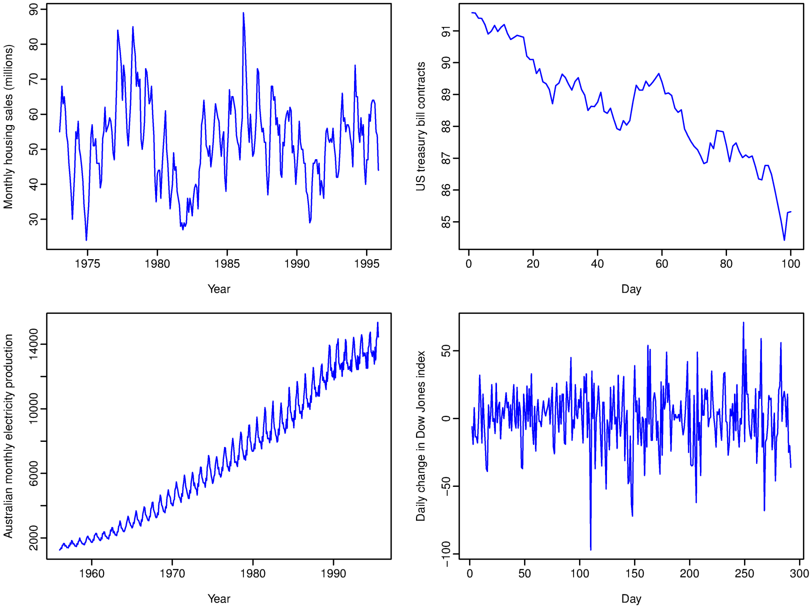
Тренд — плавное долгосрочное изменение уровня ряда.

Сезонность — циклические изменения уровня ряда с постоянным периодом.

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

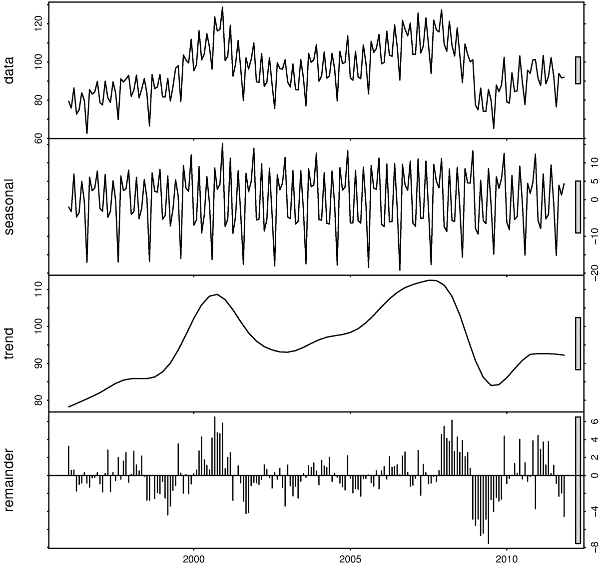
Ошибка — непрогнозируемая случайная компонента ряда.

Компоненты временных рядов



Компоненты временных рядов

STL-декомпозиция:



Автокорреляция

Автокорреляционная функция:

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

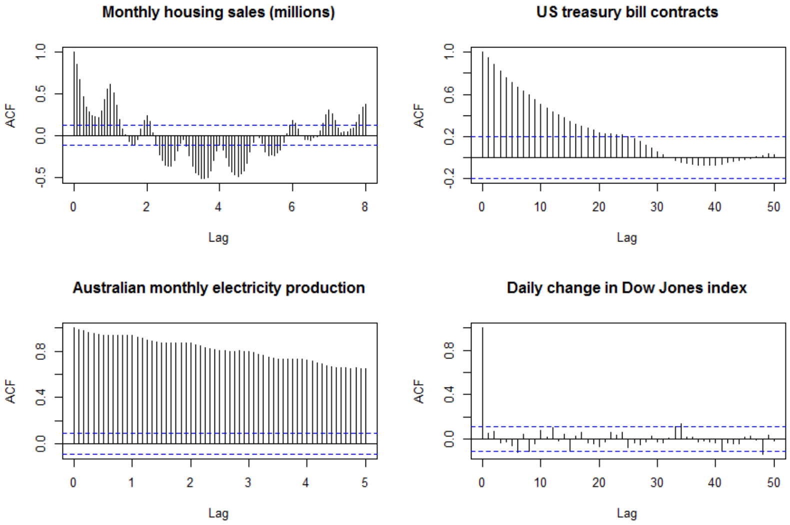
$r_\tau \in [-1, 1]$, τ — лаг автокорреляции.

Проверка значимости отличия автокорреляции от нуля:

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

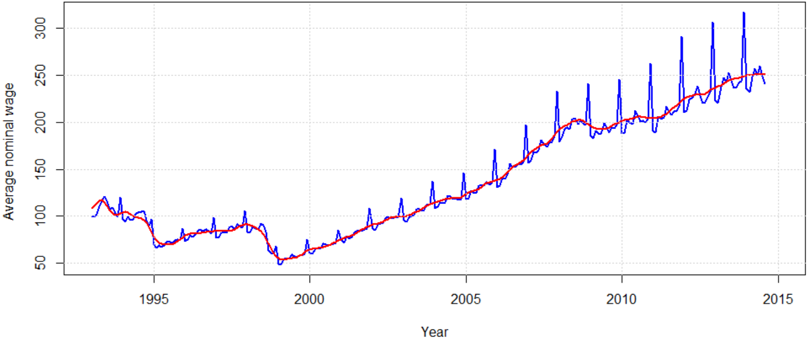
Автокорреляция

Коррелограммы:



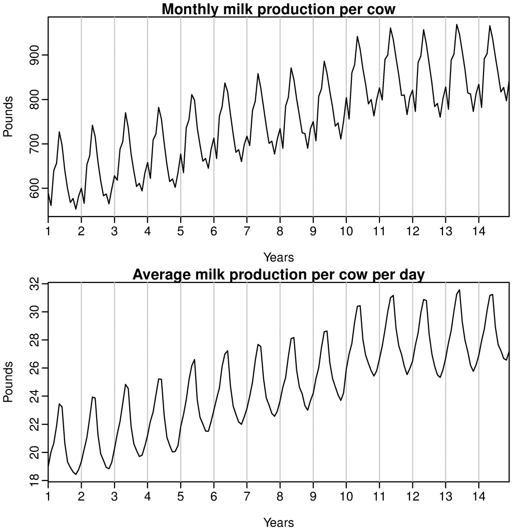
Снятие сезонности

Часто для удобства интерпретации ряда сезонная компонента вычитается:



Календарные эффекты

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



О пользе предсказательных интервалов

Пример: в апреле 1997 года в Гранд-Форкс, Северная Дакота, произошло наводнение. Город был защищён дамбой высотой в 51 фут; согласно прогнозу, высота весеннего паводка должна была составить 49 футов; истинная высота паводка оказалась равной 54 футам.

50000 жителей было эвакуировано, 75% зданий повреждено или уничтожено, ущерб составил несколько миллиардов долларов.

На исторических данных точность прогнозов метеорологической службы составляла ± 9 футов.

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:

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

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:

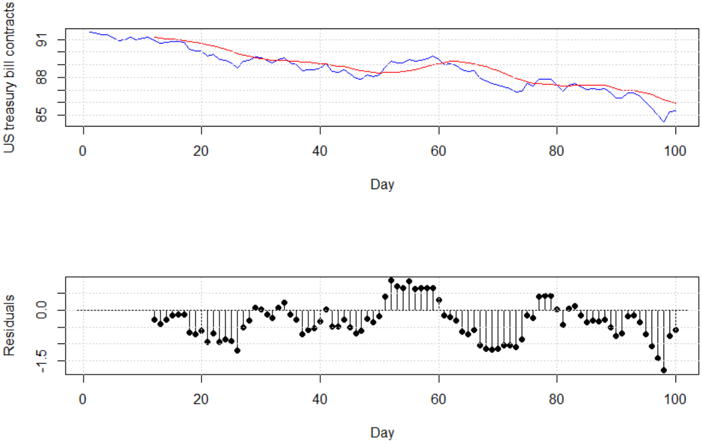
$$\hat{y}_{R+d|R}, \dots, \hat{y}_{T|T-d},$$

или с фиксированным концов истории при разных отсрочках:

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

Необходимые свойства остатков прогноза

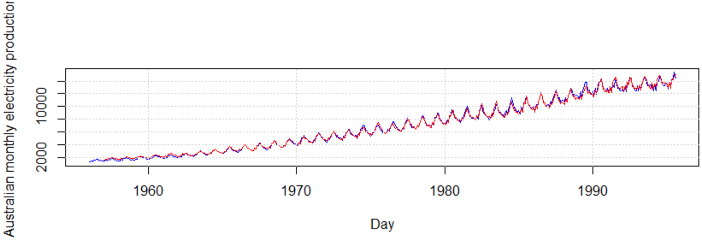
- Несмещённость — равенство среднего значения нулю:

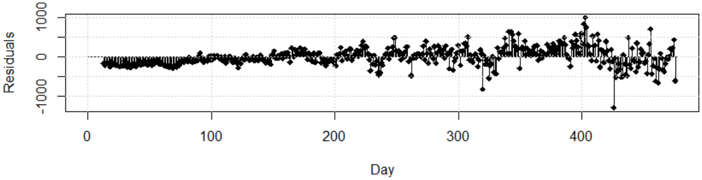


Прогноз средним за последний год отстаёт от ряда.

Необходимые свойства остатков прогноза

- Стационарность — отсутствие зависимости от времени:

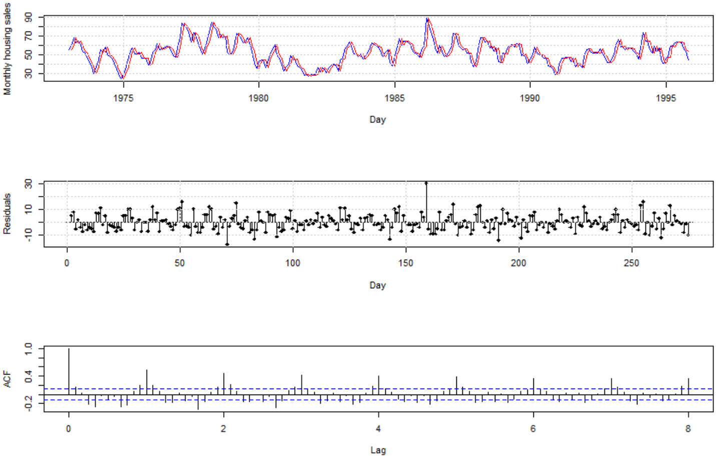




Корректировка среднего у наивного сезонного прогноза приводит к занижению прогнозных значений в начале периода и завышению в конце.

Необходимые свойства остатков прогноза

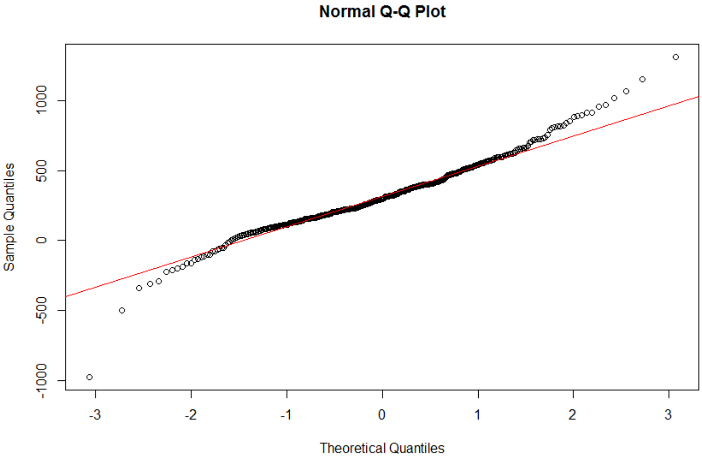
- Неавтокоррелированность — отсутствие неучтённой зависимости от предыдущих наблюдений:



Наивный прогноз не учитывает сезонность.

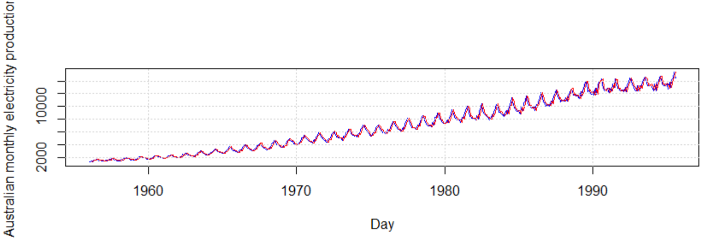
Желательные свойства остатков прогноза

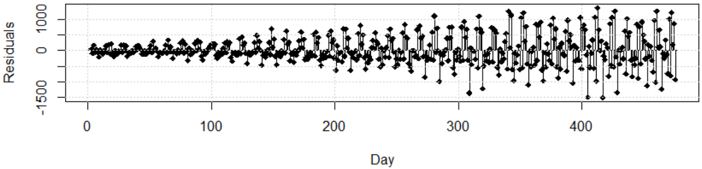
- Нормальность:



Желательные свойства остатков прогноза

- Гомоскедастичность — однородность дисперсии:





Дисперсия остатков прогноза с экстраполяцией тренда возрастает со временем.

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.
- Гомоскедастичность — визуальный анализ, критерий Бройша-Пагана.

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: $KPSS\left(\varepsilon^T\right) = \frac{1}{T^2} \sum\limits_{i=1}^{T} \left(\sum\limits_{t=1}^{i} \varepsilon_t \right)^2 \bigg/ \lambda^2$;

$KPSS\left(\varepsilon^T\right)$ при H_0 имеет табличное распределение.

Q-критерий Льюнга-Бокса

ряд ошибок прогноза: $\varepsilon^T = \varepsilon_1, \dots, \varepsilon_T;$

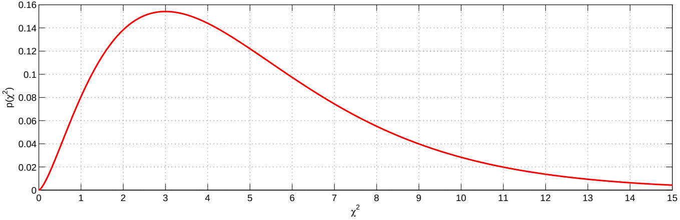
нулевая гипотеза: $H_0\colon r_1 = \dots = r_L = 0;$

альтернатива: $H_1\colon H_0$ неверна;

статистика: $Q\left(\varepsilon^T\right) = T\left(T+2\right)\sum\limits_{\tau=1}^{L}\frac{r_\tau^2}{T-\tau};$

$Q\left(\varepsilon^T\right) \sim \chi^2_{L-K}$ при H_0,

K — число настраиваемых параметров модели ряда.



Меры качества точечного прогноза

Mean squared error:

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

Mean absolute error:

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

Mean absolute percentage error:

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

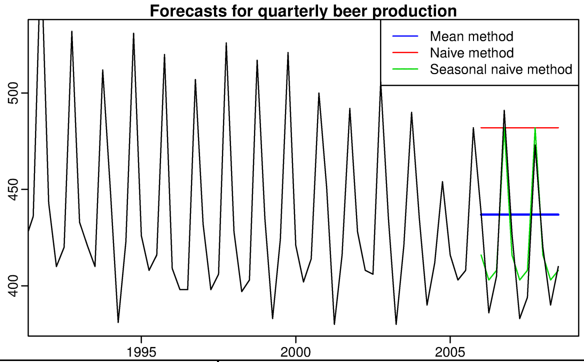
Symmetric mean absolute percentage error:

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

Mean absolute scaled error:

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

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 :

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

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

$$\begin{aligned} y_1, \dots, y_T &\text{ — временной ряд,} \\ \hat{y}_{1R}, \dots, \hat{y}_{1T} &\text{ — прогноз на период } R, \dots, T \text{ первым методом,} \\ \hat{\varepsilon}_{1R}, \dots, \hat{\varepsilon}_{1T} &\text{ — остатки первого прогноза,} \\ \hat{y}_{2R}, \dots, \hat{y}_{2T} &\text{ — прогноз на период } R, \dots, T \text{ вторым методом,} \\ \hat{\varepsilon}_{2R}, \dots, \hat{\varepsilon}_{2T} &\text{ — остатки второго прогноза;} \\ g\left(y_t, \hat{y}_{it}\right) &\text{ — произвольная функция потерь,} \\ d_t &= g\left(y_t, \hat{y}_{1t}\right) - g\left(y_t, \hat{y}_{2t}\right). \end{aligned}$$

$$\begin{aligned} H_0 &: \text{ среднее } d_t = 0, \\ H_1 &: \text{ среднее } d_t <\neq> 0. \end{aligned}$$

Критерий знаковых рангов Уилкоксона:

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

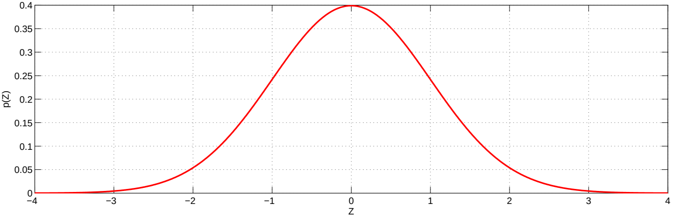
Критерий Диболда-Мариано

нулевая гипотеза: $H_0\colon \mathbb{E}d_t = 0;$

альтернатива: $H_1\colon \mathbb{E}d_t <\neq> 0;$

статистика: $B = \frac{\bar{d}}{\sqrt{\hat{f}/T}}, \quad \hat{f} = \sum\limits_{\tau=-M}^{M} \hat{r}_\tau, \quad M = T^{1/3};$

$B \sim N(0,1)$ 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,

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

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:

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

(можно добавить ещё штраф за число параметров алгоритма).
$\hat{f}_{t+d} = f\left(Z_{t+d}, \hat{y}_{t+d}, \hat{\beta}_t\right) \in \mathbb{R}^k$ — 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: $H_0: \max\limits_{j=1,\dots,k} \mathbb{E} f_j^* \leq 0,$

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

альтернатива: $H_1: H_0$ неверна.

При выполнении ряда теоретических предположений

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

$\bar{f} = \frac{1}{n} \sum\limits_{t=R}^{T-d} \hat{f}_{t+d}$ — 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

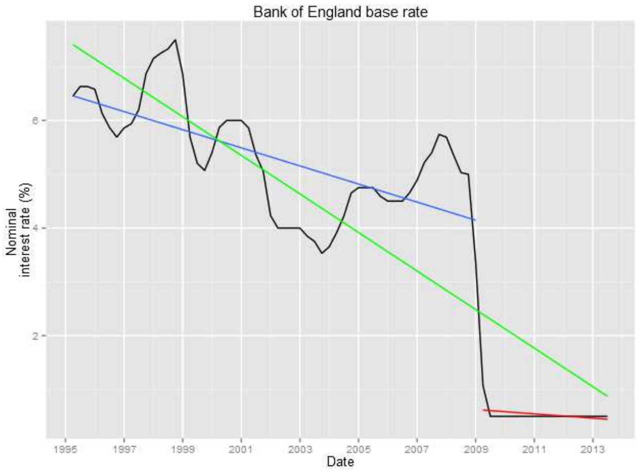
Пример (Sullivan, Timmermann, White, 1999, 2001): к ряду промышленного индекса Доу-Джонса с 1 января 1897 по 30 июня 1998 (27447 отсчётов) было применено большое количество моделей — 7846 моделей технического анализа и 9452 календарных. В качестве эталона рассматривалась стратегия долгосрочного инвестирования.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

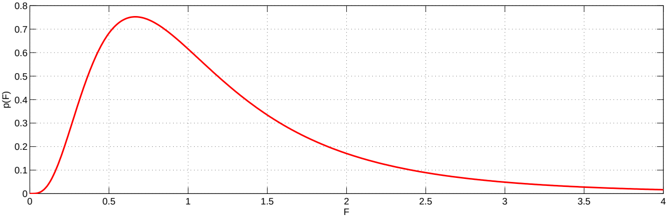
нулевая гипотеза: H_0: структура модели стабильна;

альтернатива: $H_1: H_0$ неверна;

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

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

$F \sim F(k_1 + k_2 - k, n - k_1 - k_2)$ 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)
```

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:

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

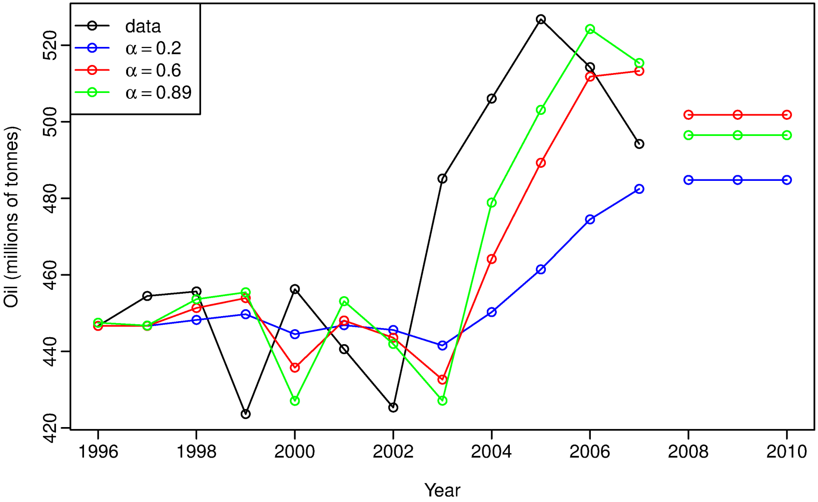
- Прогноз зависит от l_0:

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

  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:

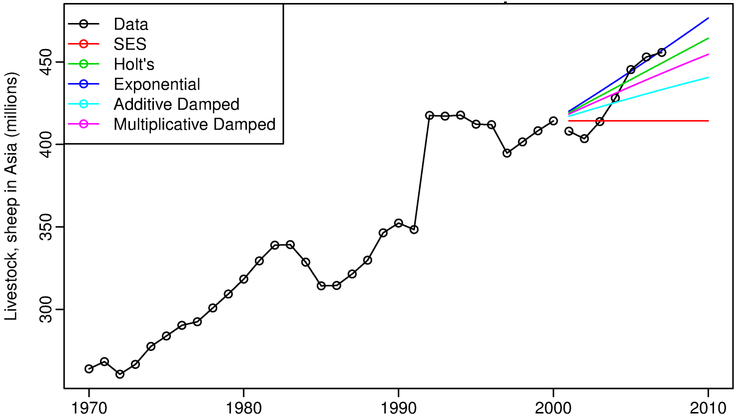
$$
\begin{aligned}
\hat{y}_{t+d|t} &= l_t + \left(\phi + \phi^2 + \dots + \phi^d\right) b_t, \\
l_t &= \alpha y_t + (1-\alpha)\left(l_{t-1} + \phi b_{t-1}\right), \\
b_t &= \beta\left(l_t - l_{t-1}\right) + (1-\beta)\,\phi b_{t-1}.
\end{aligned}
$$

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.

	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

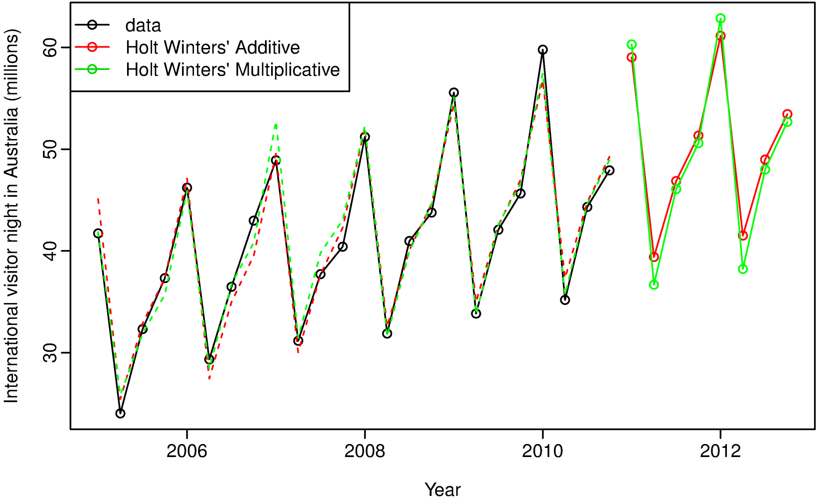
Аддитивная сезонность с периодом длины m (метод Хольта-Уинтерса):

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

Мультипликативная сезонность:

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

Методы, учитывающие сезонность



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

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 $ETS\left(\cdot,\cdot,\cdot\right)$.

Настройка модели экспоненциального сглаживания

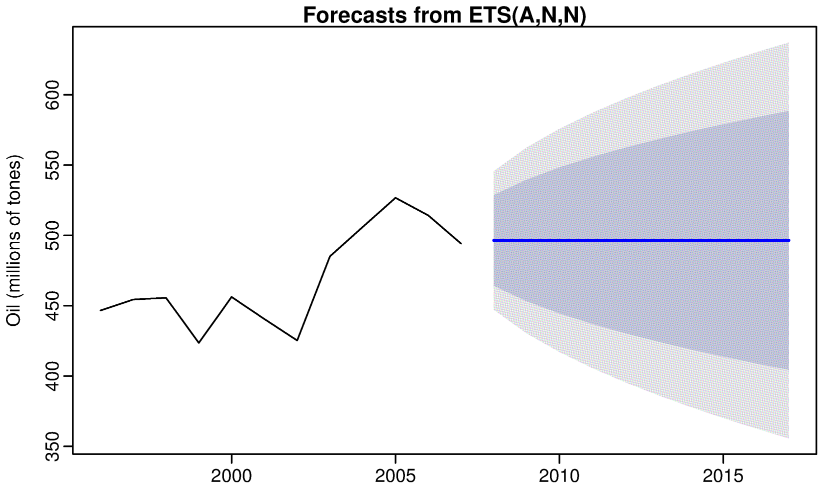
Функция ets пакета forecast в R позволяет автоматически выбрать лучшую модель экспоненциального сглаживания:

```
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:

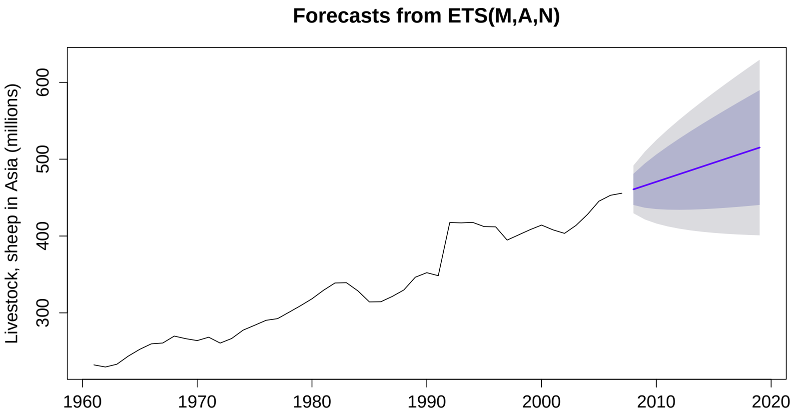
```
forecast(object, h=ifelse(object$m>1, 2*object$m, 10),
level=c(80,95), fan=FALSE, simulate=FALSE, bootstrap=FALSE,
npaths=5000, PI=TRUE, lambda=object$lambda, ...)
```

Примеры



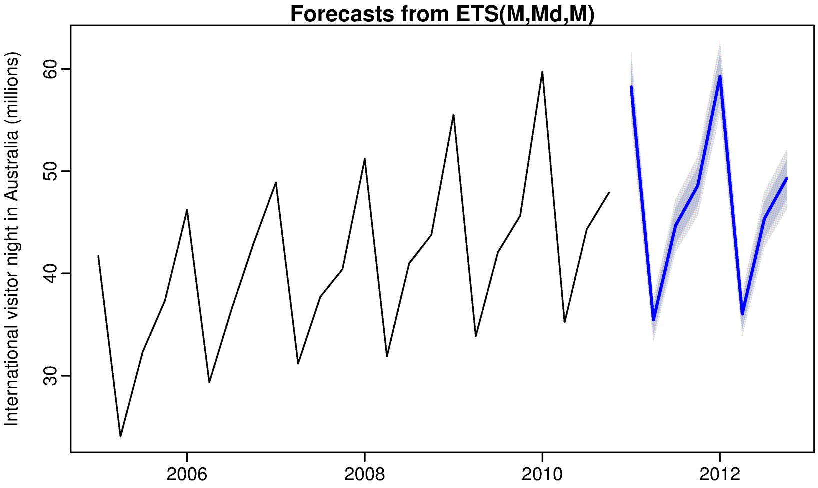
Для данных о добыче нефти в Саудовской Аравии функция ets выбирает простое экспоненциальное сглаживание.

Примеры



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.

Литература

- 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;
- модели экспоненциального сглаживания (exponential smoothing) — Hyndman, Лукашин (другие названия).

Литература

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, 2014. `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.

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.