Прикладная статистика 13. Анализ временных рядов.

6 декабря 2013 г.

Меры качества точечного прогноза

Mean squared error:

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

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:

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

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:

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

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

Сравнение качества двух прогнозов

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

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

Непараметрические критерии

$$H_0: \ \operatorname{med} d_t = 0,$$
$$H_1: \ \operatorname{med} d_t <\neq> 0.$$

Критерий знаков:

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

Критерий знаковых рангов Уилкоксона:

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

Критерий Диболда-Мариано

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

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

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

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

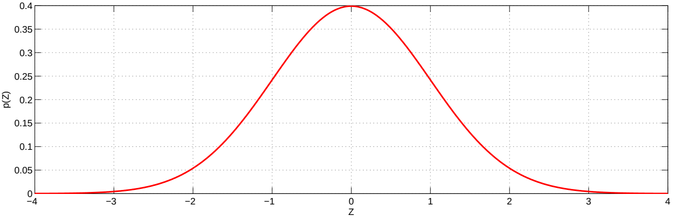
Критерий Диболда-Мариано

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 $H_0;$



достигаемый уровень значимости:

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

Критерий Диболда-Мариано

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

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

где d — отсрочка прогноза.

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:

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

(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).

Причинность по Грейнджеру

Между рядами $x_1, \dots, x_T$ и $y_1, \dots, y_T$ существует **причинная связь Грейнджера** $x_t \to y_t$, 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 $x_t \to y_t$ и $y_t \to x_t$.

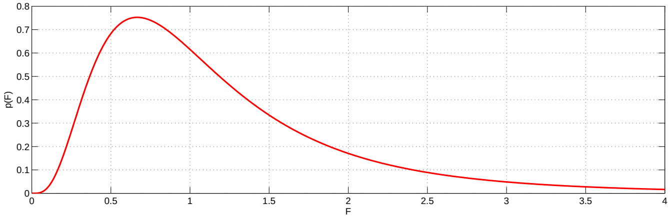
Критерий Грейнджера

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

k_1 и k_2 выбирается по информационному критерию.

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

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;

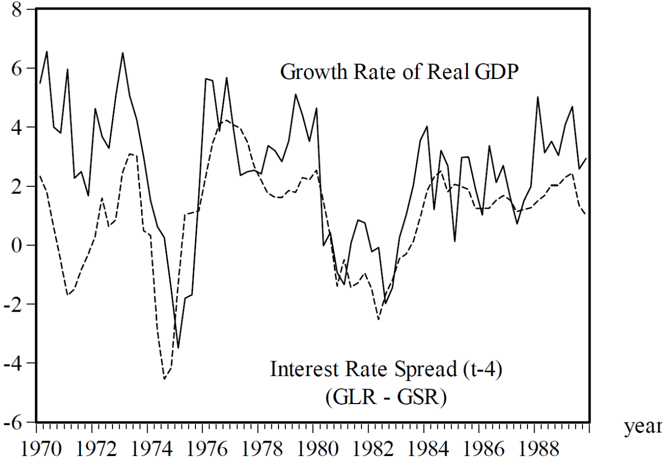


достигаемый уровень значимости:

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

Критерий Грейнджера

Маржа сверх процентной ставки и рост ВВП, ФРГ:



Критерий Грейнджера

y	x	k_1	k_2	$F(y \to x)$	$F(y \to x)$
$\triangle_4 \ln GDP_r$	$\triangle_4 \ln M1_r$	4	4	6.087^{***}	1.918
		8	8	3.561^{**}	1.443
$\triangle_4 \ln GDP_r$	$GLR - GSR$	4	4	3.160^{*}	3.835^{**}
		8	8	$1.927^{(*)}$	2.077^{*}
$\triangle_4 \ln M1_r$	$GLR - GSR$	4	4	5.615^{***}	1.489
		8	8	2.521^{*}	1.178

$\triangle_4 \ln GDP_r$ — годовой прирост ВВП в процентах, $\triangle_4 \ln M1_r$ — 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.

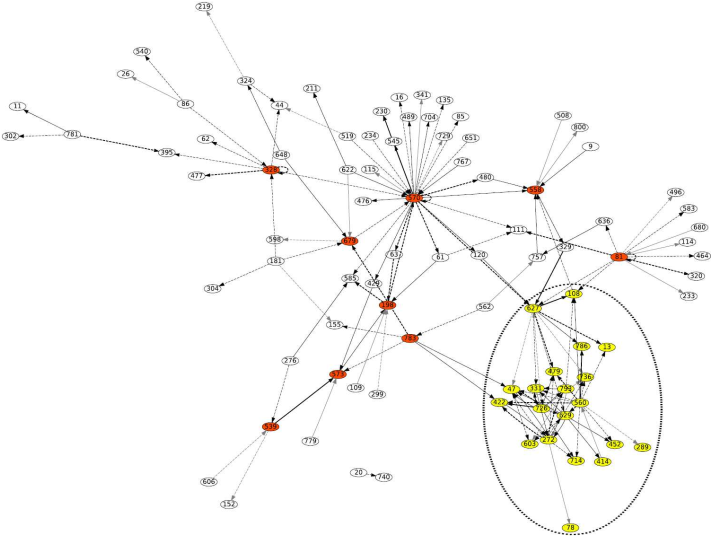
Многомерный критерий Грейнджера

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

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

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

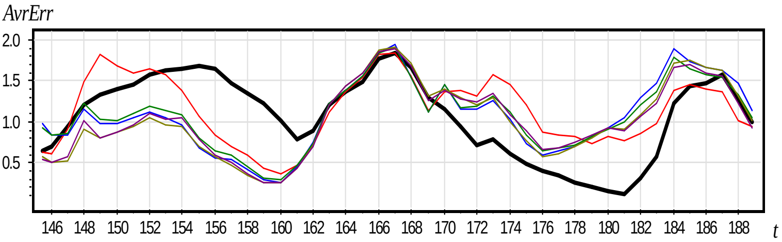
Граф причинности по Грейнджеру



Критерий Грейнджера + поправка на множественную проверку гипотез

Пример

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:

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

Адаптивный подбор весов [Лукашин, 2003]:

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

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

Адаптация весов с регуляризацией

На каждом шаге t веса определяются по МНК и сглаживаются:

$$\begin{cases} \sum\limits_{i=0}^{t} \beta^{t-i} \Bigl(\sum\limits_{j=1}^{k} w_j \hat{y}_{j,i} - y_i \Bigr)^2 + \lambda \sum\limits_{j=1}^{k} \bigl(w_j - w_{j,t-1} \bigr)^2 \to \min\limits_{w_j, j=1,\dots,k} \\ \sum_{j=1}^{k} w_j = 1. \end{cases}$$

β — коэффициент «забывания» предыстории,
λ — коэффициент регуляризации.

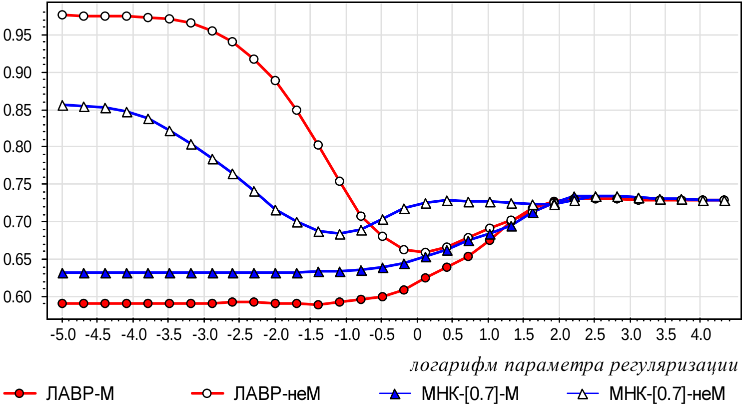
Дополнительные варианты:

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

Воронцов К. В., Егорова Е. В. Динамически адаптируемые композиции алгоритмов прогнозирования // Искусственный Интеллект, Донецк, 2006. № 2. С. 277–280.

Задача прогнозирования временных рядов продаж

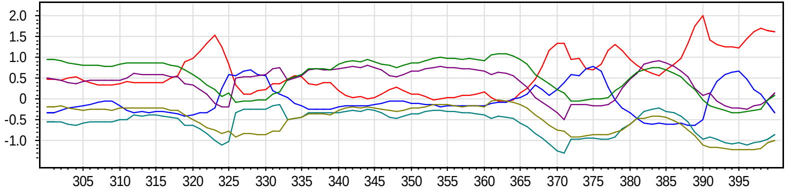
Средняя ошибка прогнозов на скользящем контроле $(T = 620)$



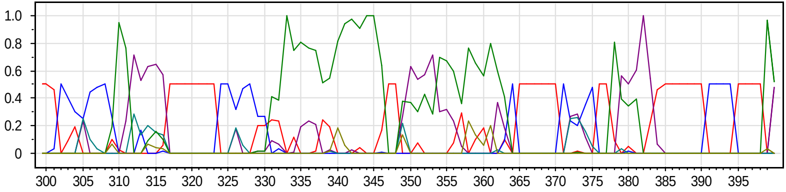
- ЛАВР-М — лучший результат, причём можно брать $\lambda = 0$
- Ограничение монотонности — сильный регуляризатор

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, β=0.7	0.6314
ЛАВР без Монот	0.6591
МНК без Монот, β=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
- $\gamma_{\text{opt}} = 0.2 \ldots 0.3$ — усреднение по 3...5 дням

Прикладная статистика
13. Анализ временных рядов.

Рябенко Евгений
riabenko.e@gmail.com