2022-06-03 TS

→ сравнение прогнозов (DM, RC)

→

→ MSTL

...

Алгоритм MSTL

STL (y_t)

период = 12

силы сглаж (сез / тренд ...)

$$y_t = trend_t + seas_t + rem_t$$

⊖ не подходит, если наличие много частот

нестат

⊕ дико устойчив, быстр

⊖ нет интервальных оценок.

$\boxed{\text{MSTL = STL много раз.}}$

Шаг 1. первичное выд-ие сезонных комп-т.

Шаг 2. корректировка.

1.1. STL(y_t, высок част) : $y_t = T_t + S_t^{(1)} + R_t$

$$y_t^{(-1)} = y_t - S_t^{(1)}$$

1.2. STL($y_t^{(-1)}$, средн. част) : $y_t^{(-1)} = T_t + S_t^{(2)} + R_t$

$$y_t^{(-1,2)} = y_t^{(-1)} - S_t^{(2)}$$

...

после шаг 1: y_t^{clean}, $S_t^{(1)}$, $S_t^{(2)}$, $S_t^{(3)}$.

2.1. Временно вернем $S_t^{(1)}$ в текущий ос. ряд.

$$y_t^{(+1)} = y_t^{clean} + S_t^{(1)}$$

уточн. сез. сост.

$$S_t^{(1)} := STL_{seas}\left(y_t^{(+1)}, \text{выс. част}\right)$$

уточн. ос. ряд.

$$y_t^{clean} := y_t^{(+1)} - S_t^{(1)}$$

2.2. Вр-нно вернем $S_t^{(2)}$ в кл. ос. ряд.

$$S_t^{(2)} := STL_{seas}\left(y_t^{(+2)}, \text{ср. част}\right)$$

$$y_t^{clean} := y_t^{(+2)} - S_t^{(2)}$$

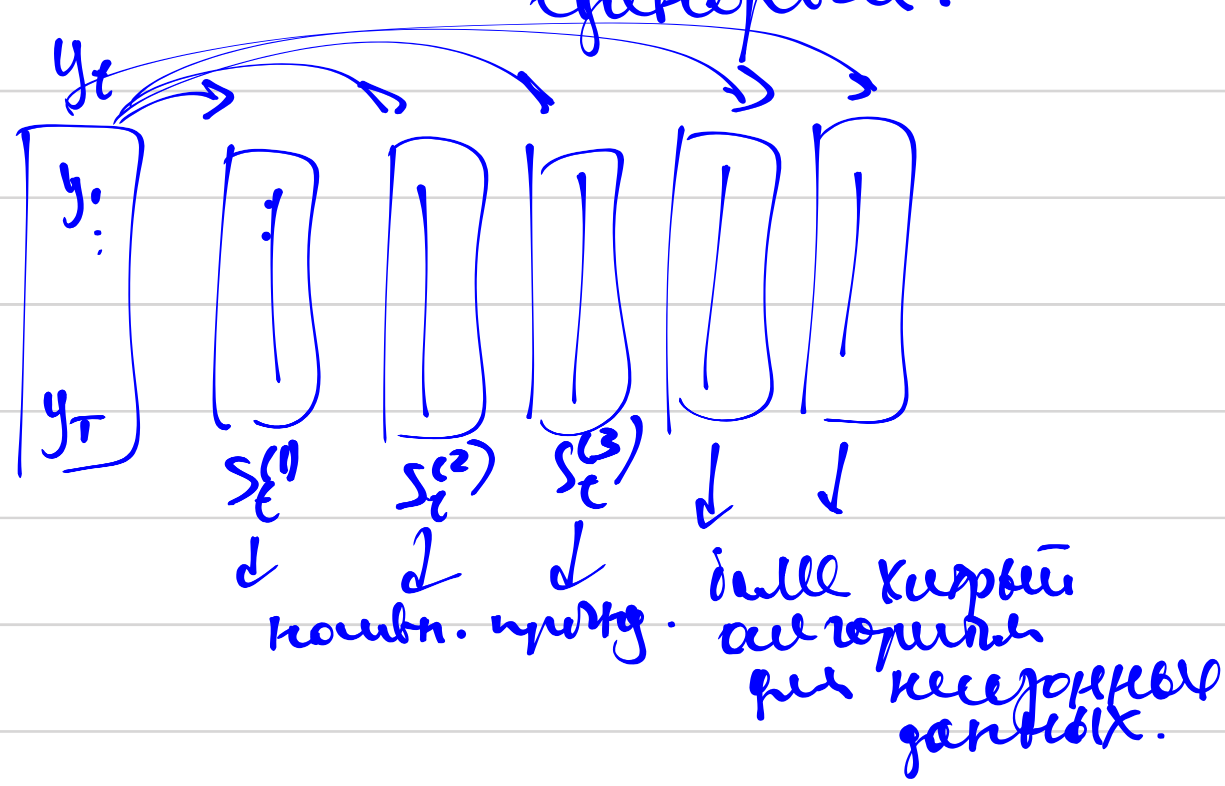
...

Шаг 2 × от полуночи до 1-го крика совы.

Финал: (уточнение)

Тренд и остаток из последнего STL-разложения.

$$\underbrace{y_t = S_t^{(1)} + S_t^{(2)} + S_t^{(3)} + T_t + R_t}_{\text{единороги!}}$$



получили:

Маша = 0.9103 Top1

Петя = 0.9101 $\neq$ Top1

⋮

А существенно ли отличие или Маше сегодня просто повезло?

тест Diebold-Mariano

→ сравнить 2 прогноза.

→ прогнозы на h шагов вперед.

[h - фикс.]

$\hat{y}^M_{T+1}$, $\hat{y}^P_{T+1}$

$\hat{y}^M_{T+2}$, $\hat{y}^P_{T+2}$

$$l_t^M = \left(\hat{y}_t^M - y_t\right)^2$$

$$l_t^P = \left(\hat{y}_t^P - y_t\right)^2$$

для примера, ф. потерь м. быть любой.

$$d_t = l_t^M - l_t^P$$

[предп.]

d_t - слабо стац. процесс

$E(d_t) = \mu_d$

$Cov(d_t, d_{t-k}) = \gamma_k$ [не зав-т от t]

в частности

$Var(d_t) = \gamma_0$

H_0: $\mu_d = 0$

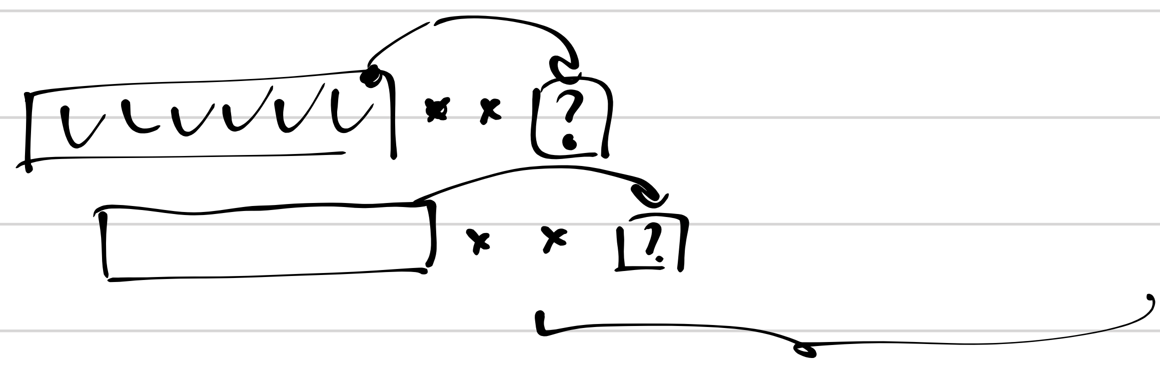
H_A: $\mu_d \neq 0$ [$\mu_d < 0$]

$$DM = \frac{\bar{d} - 0}{se(\bar{d})}$$

$$DM \xrightarrow[dist]{H_0,\ T \to \infty} N(0;1)$$

$$\bar d = \frac{d_1+d_2+\dots+d_p}{p}$$

$h = 3$



$se(\bar d)$ — сост-ая оценка для $\sqrt{Var(\bar d)}$ p раз.

$$Var(\bar d) = \frac{1}{p^2}\left(p\cdot\gamma_0 + 2(p-1)\cdot\gamma_1 + 2(p-2)\cdot\gamma_2 \dots + 2\cdot\gamma_{p-1}\right)$$

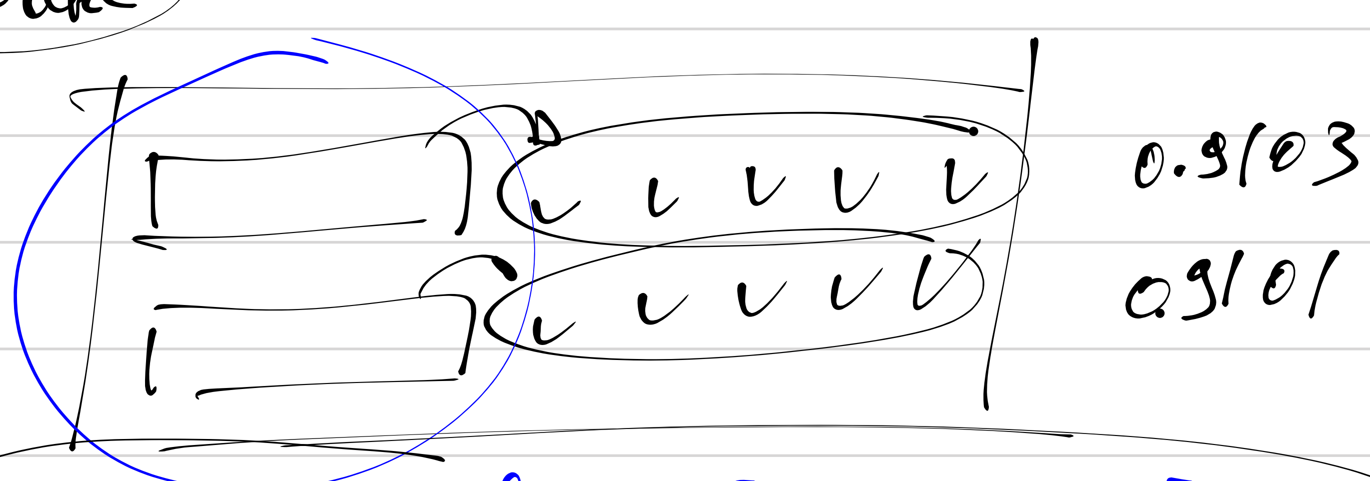
$$se(\bar d) = \sqrt{\widehat{Var}(\bar d)} = \sqrt{\frac{1}{p^2}\left(p\cdot\hat\gamma_0 + 2(p-1)\hat\gamma_1 \dots \dots 2\hat\gamma_{p-1}\right)}$$

для (d_t) используют $MA(q)$

например, $q = \sqrt{p}$

— для двух прогнозов.

нюанс



→ не учит кач-во по обуч. сет выборке.

RC reality check
SPA superior predictive ability.

цель: сравнить много [моделей]
с некоторой эталонной.

$$d_t = \begin{bmatrix} l_{A,t} - l_{bench,t} \\ l_{B,t} - l_{bench,t} \\ \vdots \end{bmatrix}$$

H_0: $\max E(d_t) \le 0$
нет „модели" лучше эталонной.
H_A: $\max E(d_t) > 0$

Бутстрэп. $RC = \max(\bar{d})$ $\bar{d} = \dfrac{\sum d_t}{P} = \begin{bmatrix} \bar{l}_A - \bar{l}_{bench} \\ \bar{l}_B - \bar{l}_{bench} \\ \vdots \end{bmatrix}$

Шаг 1. Строим бутстрэп реализацию $d_1^* \dots d_p^*$

Шаг 2. По $d_1^* \dots d_p^*$ считаем $RC^* = \max(\bar{d}^* - \bar{d})$

Шаг 1+2 × 10000 раз: $RC_1^*, \dots RC_{10000}^*$ [доб. стат]

P-значение: доля RC_b^* ок-ся выше RC.

[Куча нюансов]   [стац-ый бутстрэп.

$[1\,1\,1\,2\,3\,\dots\,p]$

выбираю случ-но t_1 из $[1 \dots p]$
далее: $t_j \to t_j + 1$ вер $(1-\alpha)$
$t_{j+1} \to$ равновер из $[1 \dots p]$ вер α

t_j $[2, 3, 4, 7, 8, 9, 1, 2, 3, 4, 8, 9]$ $d_2, d_3, d_4 \dots d_9$
$d_1^*\ d_2^* \dots d_p^*$