→ алгоритм подбора Khandakar–Hyndman [ARIMA]

→ ADF, KPSS.

ARIMA(p,d,q)

ARMA(p,q) $d=0$

$d=1,2\dots$

(y_t) ~ стац-ый

$\Delta^d y_t$ ~ стац.

$\Delta^{d-1} y_t$ ~ не стац.

$\Delta^{d-1} y_t \neq \alpha + \beta t +$ стац.

Q. Как выбрать наилучшую?

A1. если ряды длинные, много выч-х ресурсов, то кросс-валидация.

A2. алгоритм KH

Шаг 1. С помощью статист. тестов определить d. [$KPSS_c$]

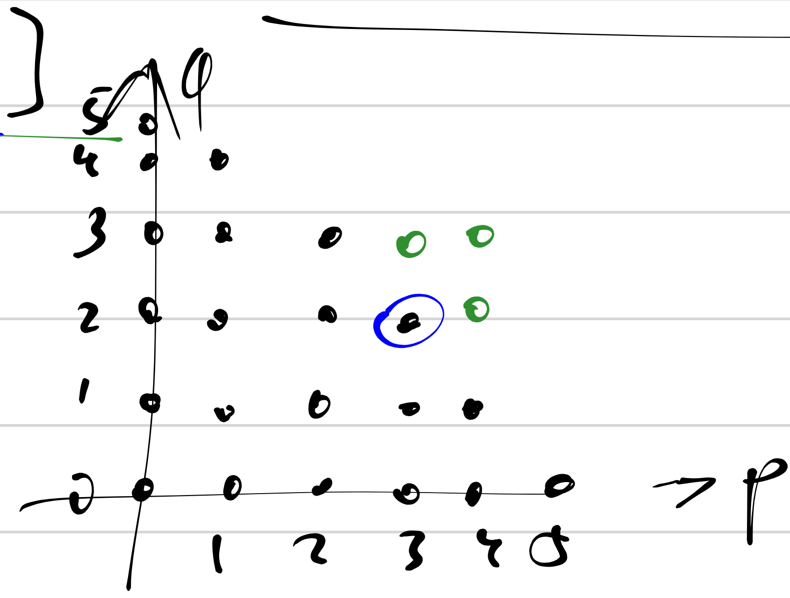
$f(y_1, \dots, y_T) \longleftarrow$ $d=0$ $y_t \sim ARMA$

$f(\Delta y_2, \dots, \Delta y_T) \longleftarrow$ $d=1$ $\Delta y_t \sim ARMA$

$f(\Delta^2 y_3 \dots \Delta^2 y_T) \longleftarrow$ $d=2$ $\Delta^2 y_t \sim ARMA$

Шаг 2. Оценивается начальн. ARMA(p,q)

например [$p+q \leq 5$]



Выбирается модель с min AICc

Пример. $M_1: \quad y_t = \alpha + u_t + \delta_1 u_{t-1} + \delta_2 u_{t-2}$ $\quad u_t \sim WN$ MA(2)

между M_1 и M_2 нельзя выб-ть по кр-ию AIC.

$M_2: \quad \Delta y_t = \alpha + u_t + \delta_1 u_{t-1} + \delta_2 u_{t-2}$ $\quad u_t \sim WN$

ARIMA(0, 1, 2)

KPSS Kwiatkowski-Phillips-Schmidt-Shin

- $KPSS_c$ „с константой"
- $KPSS_t$ „с трендом"

Опр. долгосрочная дисперсия ряда.

λ^2 – долгоср. дисп-сия y_t если

$$Var(\bar{y}) = \frac{\lambda^2}{T} + o\left(\frac{1}{T}\right)$$

$$\lim_{T\to\infty}\left(T \cdot Var(\bar{y})\right) = \lambda^2$$

Если бы $y_1 \ldots y_T$ были незав-мы и одинаково распр.

$$Var(\bar{y}) = \frac{\sigma^2}{T}$$

$$\bar{y} = \frac{y_1 + \ldots + y_T}{T}$$

Упр. $y_t = u_t + u_{t-1}$ $\quad u_t \sim WN(0; \sigma_u^2)$

а) $Var(y_t)$

б) долгосрочная дисп-ия y_t ...

$\left[\lim_{T\to\infty}\left(T \cdot Var(\bar{y})\right)\right]$

$$T \cdot Var(\bar{y}) =$$

$$= T \cdot \frac{T \cdot \overbrace{Var(y_t)}^{\gamma_0} + 2\cdot(T-1)\cdot Cov(y_{t-1}, y_t) + 2\cdot(T-2)\cdot \overbrace{\ldots}^{Cov(y_{t-2}, y_t)} + \ldots}{T^2}$$

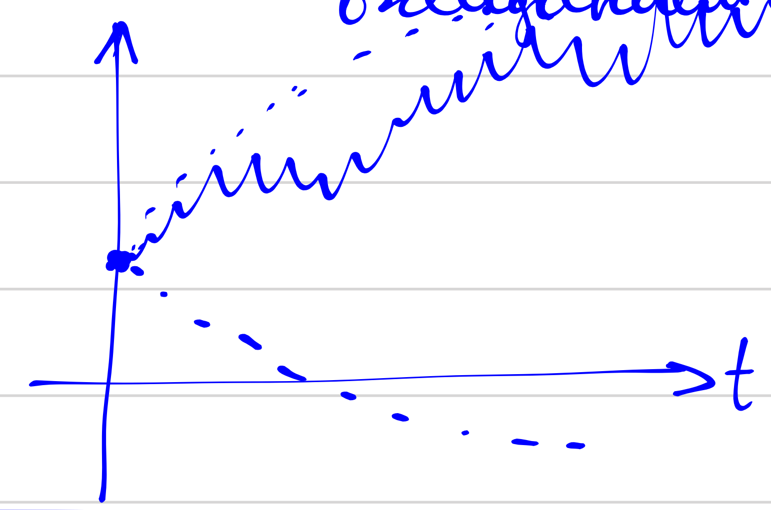
$$\lim_{T\to\infty} T\,Var(\bar{y}) = \gamma_0 + 2\gamma_1 + 2\cdot\gamma_2 + \ldots = \lambda^2$$

$\boxed{KPSS_c}$

H_0: y_t ~ стац. процесс с ненулевым ожиданием

H_A: y_t = случайное блуждание + стац-ный процесс с ненулевым ожиданием

random walk $\tau_t = \tau_{t-1} + u_t$



$$y_t = c + \tau_t + x_t$$

$c \in \mathbb{R}$

x_t ~ стац-ный с $E(x_t) = 0$

H_0: $\tau_t = 0$	H_A: $\tau_t = \tau_{t-1} + u_t$ $\tau_0 = 0$ (u_t) – б. шум незав с (x_t)

Технически:

Шаг 1. Регрессия на константу

$$\hat{c} = \bar{y} \quad \Downarrow \quad \hat{u}_t = y_t - \hat{c}$$

Шаг 2.

$$KPSS = \frac{\sum_{t=1}^{T} S_t^2}{\hat{\lambda}^2 \cdot T^2} \xrightarrow[T \to \infty]{H_0} KPSS_c$$

$$S_t = \underbrace{(y_1 - \hat{c})}_{\hat{u}_1} + \underbrace{(y_2 - \hat{c})}_{\hat{u}_2} + \ldots + \underbrace{(y_t - \hat{c})}_{\hat{u}_t}$$

$\hat{\lambda}^2$ – оценка для λ^2 $\qquad \lambda^2 = \gamma_0 + 2\gamma_1 + 2\gamma_2 + \ldots$

$$\boxed{\hat{\lambda}^2 = \hat{\gamma}_0 + w_1 \cdot 2\hat{\gamma}_1 + w_2 2\hat{\gamma}_2 \ldots}$$

Шаг 1. Алгоритм КН

тест $KPSS_c$ для y_t

H_0 ($y_t \sim$ стац) не отвергается

$\boxed{d=0}$

Шаг 2

$y_t \sim ARMA(p,q)$

H_0 отвергается в пользу H_A

H_A ($y_t =$ RW + Stat)

$y_t \to \Delta y_t$

тест $KPSS_c$ для Δy_t

H_0 не отверг. ($\Delta y_t \sim$ стац)

H_0 отверг. в пользу H_A ($\Delta y_t =$ RW + Stat)

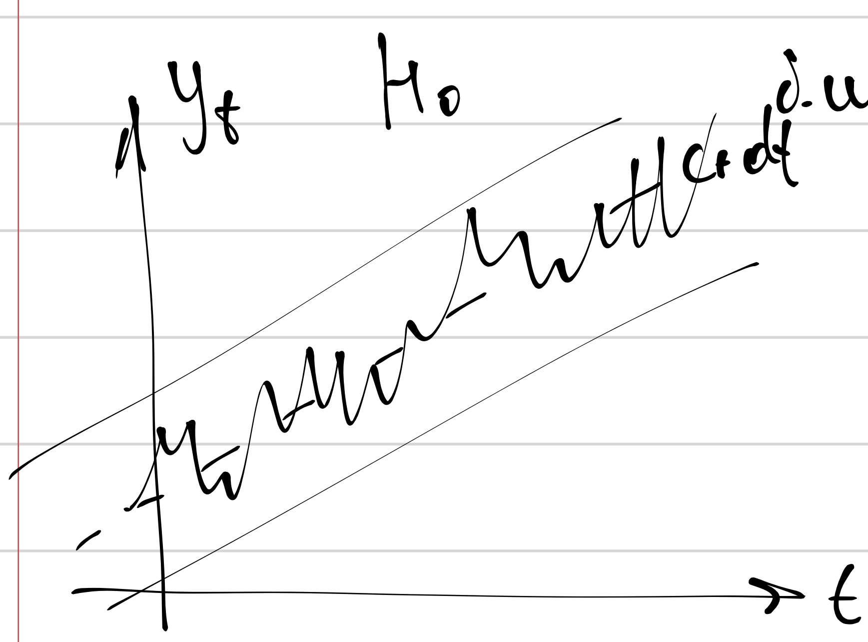
С накопленной т. зр. повторное тестирование — бяка.

$\boxed{KPSS_t}$

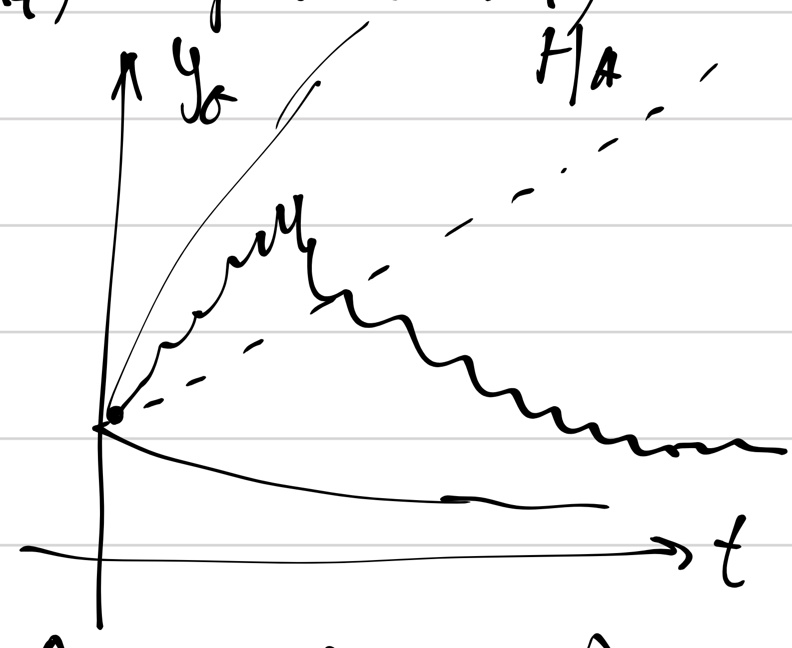
$$y_t = c + d\cdot t + \tau_t + x_t$$

$d, c \in \mathbb{R}$ $\qquad$ $x_t \sim$ стац. с $E(x_t) = 0$

H_0: $\tau_t \equiv 0$ $\qquad$ H_A: $\tau_t = \tau_{t-1} + u_t$, $\tau_0 = 0$



б. шум (u_t) некорр с (x_t)



Технически:

Шаг 1. регрессия $\hat{y}_t = \hat{c} + \hat{d}\cdot t$ $\qquad$ $\hat{u}_t = y_t - \hat{y}_t$

Шаг 2

$$KPSS = \frac{\sum_{t=1}^{T} S_t^2}{\hat{\lambda}^2 \cdot T^2} \qquad S_t = \hat{u}_1 + \hat{u}_2 + \ldots + \hat{u}_t$$

$$KPSS \xrightarrow[T\to\infty]{H_0} KPSS_t.$$

! для прогноза на 1 шаг вперёд не оч. страшно спутать стац-ый и нестац.

A. $y_t = y_{t-1} + u_t$

$\hat{y}_{T+1} = y_T$

B стац. AR(1)

$y_t = 0{,}99 \cdot y_{t-1} + u_t$

$\hat{y}_{T+1} = 0{,}99\, y_T$

! долгосрочно — страшно.

ADF = Augmented Dickey-Fuller.

↳ ADF_0 без константы
- ADF_c с константой
- ADF_T с трендом

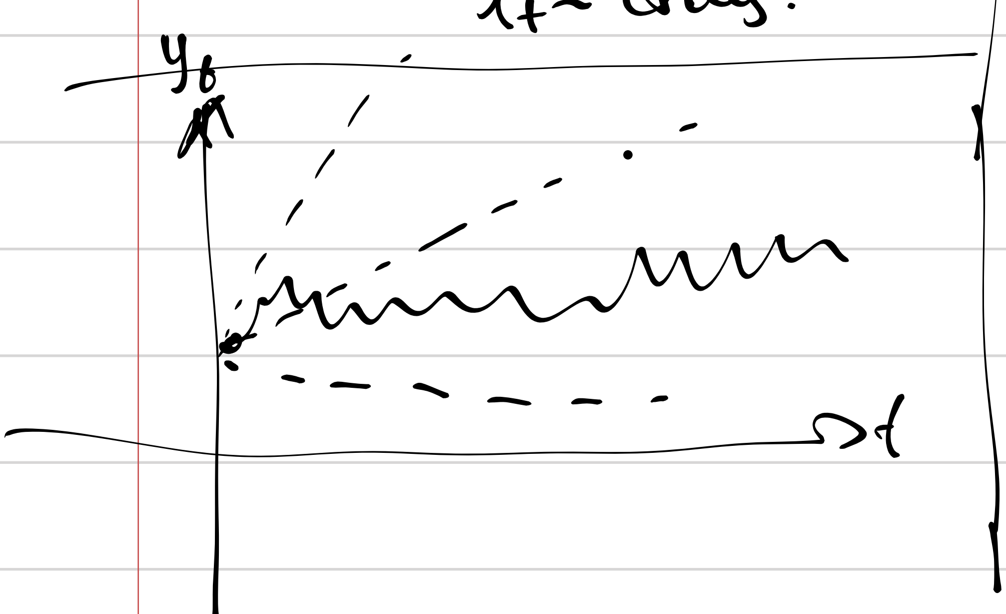
$$\Delta y_t = c + \beta y_{t-1} + \theta_1 \cdot \Delta y_{t-1} + \theta_2 \Delta y_{t-2} + \dots + \theta_p \Delta y_{t-p} + u_t$$

$H_0: \beta = 0$

Δy_t — стац-ый процесс с $E(\Delta y_t) = \mu$

$y_t = y_0 + \mu \cdot t + \sum_{k=1}^{t} x_t$

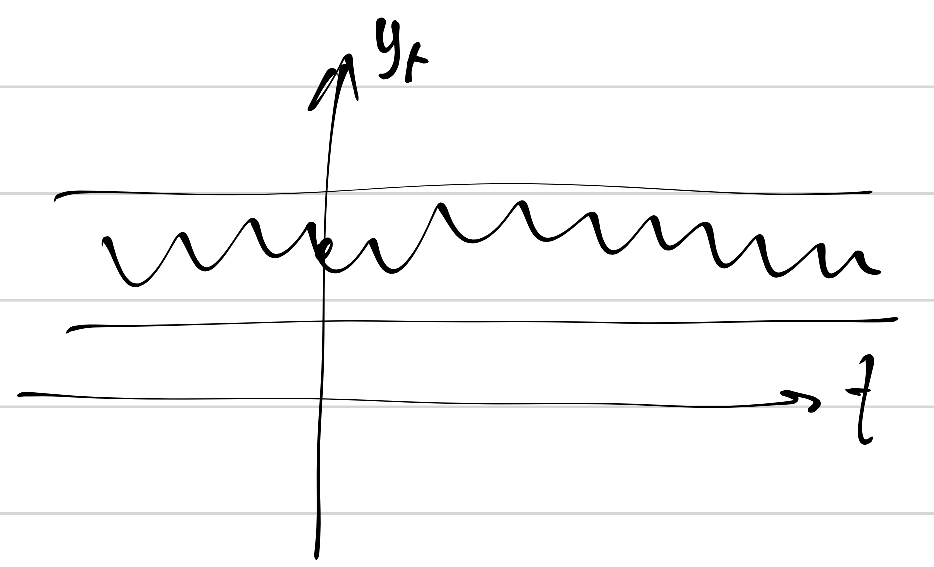
$x_t \sim$ стац.



$H_1: \beta < 0$

(y_t) — стац-ый процесс.

$y_t \sim AR(p+1)$



Технические детали

Шаг 1. Регрессия.

$$\widehat{\Delta y_t} = \hat{c} + \hat{\beta} y_{t-1} + \hat{\theta}_1 \cdot \Delta y_{t-1} + \dots + \hat{\theta}_p \Delta y_{t-p}$$

Шаг 2.

$$ADF = \frac{\hat{\beta} - 0}{se(\hat{\beta})} \xrightarrow[T \to \infty]{H_0} ADF_c$$

ADF_t

$$\Delta y_t = c + \alpha \cdot t + \beta \cdot y_{t-1} + \theta_1 \cdot \Delta y_{t-1} + \dots + \theta_p \Delta y_{t-p} + u_t$$

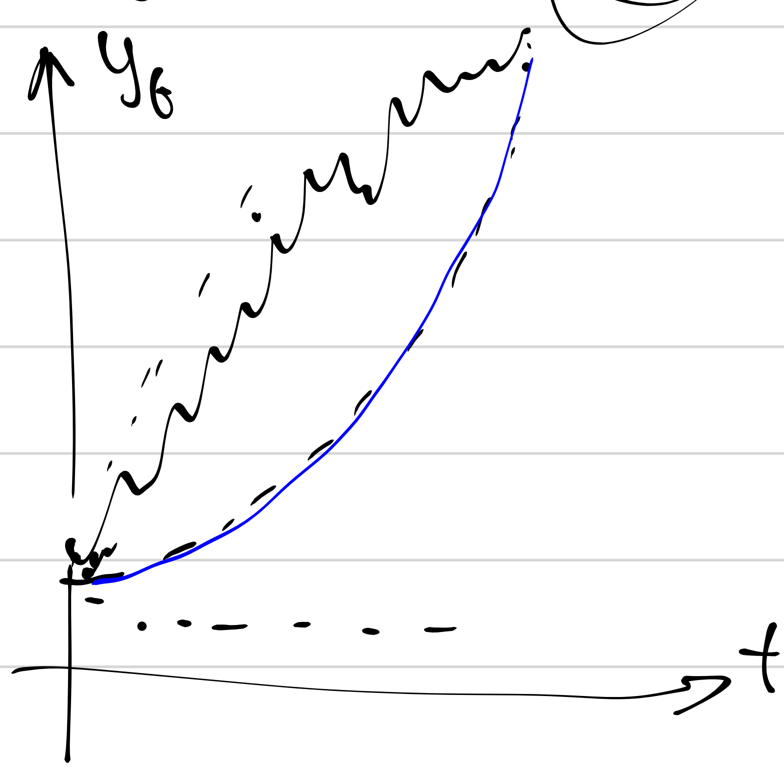
$H_0: \beta = 0$

$$\Delta y_t = \tilde{c} + \tilde{\alpha} t + x_t$$

x_t ~ стац-ый $AR(p)$

с $E(x_t) = 0$

$$y_t = y_0 + \tilde{\alpha}_1 t + \tilde{\alpha}_2 t^2 + \sum_{k=1}^{t} x_k$$

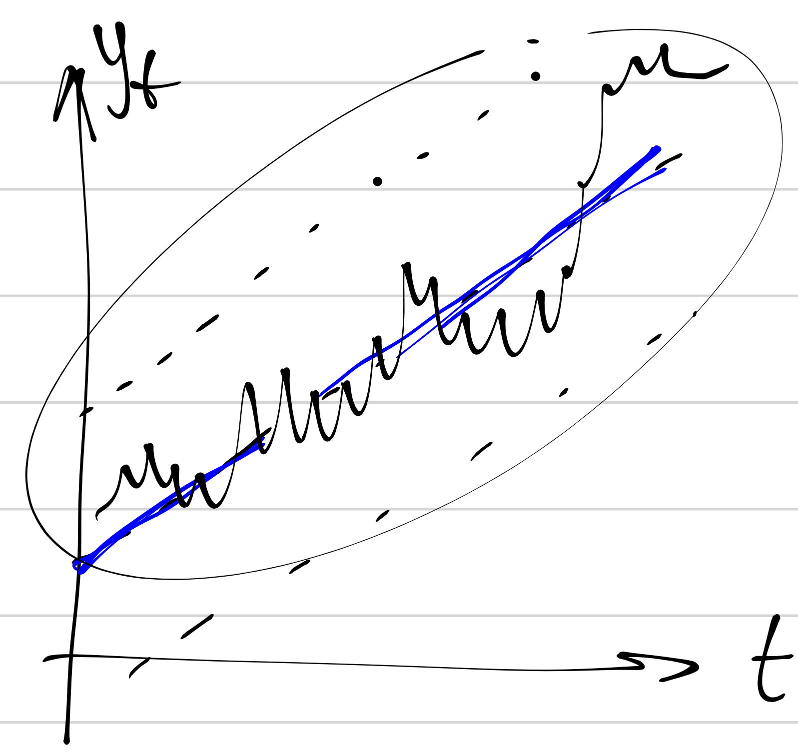


$H_1: \beta < 0$

$$y_t = \tilde{c} + \tilde{\alpha} t + x_t$$

x_t ~ стац $AR(p+1)$

с $E(x_t) = 0$



Техническое оценивание:

Шаг 1. Регрессия

$$\widehat{\Delta y_t} = \hat{c} + \hat{\alpha} \cdot t + \hat{\beta} \cdot y_{t-1} + \hat{\theta}_1 \cdot \Delta y_{t-1} + \dots + \hat{\theta}_p \cdot \Delta y_{t-p}$$

Шаг 2.

$$ADF = \frac{\hat{\beta} - 0}{se(\hat{\beta})} \xrightarrow[T \to \infty]{H_0} ADF_t.$$

Сезонная ARIMA.

Идея 1 Взять много лагов [если ряд длинный] (месячн. данные)

$(y_t) \sim AR(24)$

Идея 2

$$\underbrace{A(L)}_{A_s(L^{12})} \cdot \underbrace{\Delta^d}_{\Delta_s^D} y_t = \underbrace{M(L)}_{M_s(L^{12})} \cdot u_t \qquad ARIMA(p,d,q)$$

$d = \text{degree } A(L)$

$q = \text{degree } M(L)$

$$\boxed{A_s(L^{12}) \cdot A(L) \cdot \Delta_s^D \cdot \Delta^d y_t = M_s(L^{12}) \cdot M(L) \cdot u_t}$$

→ u_t - б. шум

$P = \text{degree } A_s$

$Q = \text{degree } M_s$

→ $\Delta_s^D \cdot \Delta^d y_t$ - стац процесс вида $MA(\infty)$ отн-но u_t

$$SARIMA\underbrace{(p,d,q)}_{\text{несезонные}} - \underbrace{(P,D,Q)}_{\text{сезонные}}$$

$$SARIMA(\underbrace{p=0}_{\deg A}, d=0, \underbrace{q=1}_{\deg M}) - (\underbrace{P=1}_{\deg A_s}, D=1, \underbrace{Q=2}_{\deg M_s})$$

$$\boxed{\underbrace{(1-\beta_1^s L^{12})}_{A_s} \cdot \Delta_s^1 y_t = \underbrace{(1+\alpha_1^s L^{12} + \alpha_2^s L^{24})}_{M_s} \cdot \underbrace{(1+\alpha L)}_{M} \cdot u_t}$$

$$\Delta_s y_t = y_t - y_{t-12}$$

$ARIMA(p,d,q)$

$SARIMA(p,d,q)-(P,D,Q)$