Rappels sur les tests

Alain Quartier-la-Tente

2022-03-12

Tests de racine unitaire

Séries temporelles et tendances

Pour modéliser une série avec tendance on peut envisager deux cas :

1. Modèle trend-stationnaire : $TS_t = DT(t) + MA_t$ avec $DT(t)$ une tendance déterministe et MA_t un processus stationnaire avec une représentation $MA(\infty)$. Par exemple :
$$TS_t = a + bt + MA_t$$
2. Modèle avec racine unité :
$$(1-B)UR_t = \underbrace{b}_{drift} + MA_t, \quad UR_0 = a$$
On a donc :
$$UR_t = a + bt + \sum_{i=1}^{t} \varepsilon_t \text{ où } \sum_{i=1}^{t} \varepsilon_t \text{ est la tendance stochastique.}$$

On a $\mathbb{V}[TS_t] = \mathbb{V}[MA_t]$ indépendante du temps mais $\mathbb{V}[UR_t] = t\sigma^2$

Tests de racine unité

Test de Dickey-Fuller

On distingue généralement 4 cas :

1. $X_t = \rho X_{t-1} + \varepsilon_t$ avec $(H_0) : \rho = 1$ (et $(H_1) : |\rho| < 1$))
2. $X_t = c + \rho X_{t-1} + \varepsilon_t$ avec $(H_0) : \rho = 1, c = 0$
3. $X_t = c + \rho X_{t-1} + \varepsilon_t$ avec $(H_0) : \rho = 1, c \neq 0$
4. $X_t = c + bt + \rho X_{t-1} + \varepsilon_t$ avec $(H_0) : \rho = 1, b = 0$

Comme le Modèle AR(1) est trop simple, on considère généralement un modèle AR(p) :
$$X_t - \mu = \sum_{i=1}^{p} \Phi_i (X_{t-i} - \mu) + \varepsilon_t$$

C'est le test de Dickey-Fuller Augmenté (ADF).

On rejette lorsque $\hat{t}_{stat} < DF_\alpha$.

Sous R, utiliser par exemple `fUnitRoots::adfTest()`, `tseries::adf.test()` ou `urca::ur.df()`[1] (qui permet une sélection automatique des retards).

[1] Voir https://new.mmf.lnu.edu.ua/wp-content/uploads/2018/03/enders_applied_econometric_time_series.pdf ou https://stats.stackexchange.com/questions/24072/interpreting-rs-ur-df-dickey-fuller-unit-root-test-results pour comprendre les sorties

Test de Phillips-Perron

Tests de (H_0) : $\rho = 1$ dans des modèles semi-paramétriques sous la forme :

$$\begin{cases} X_t = \rho X_{t-1} + u_t \\ X_t = c + \rho X_{t-1} \\ X_t = c + bt + \rho X_{t-1} \end{cases}$$

avec u_t un terme d'erreur très général.

Sous R : `tseries::pp.test` (troisième forme) ou `urca::ur.pp()`.

Test KPSS

Test d'hypothèse nulle d'un modèle trend-stationnaire

$$Y_t = \xi t + r_t + \varepsilon_t \quad r_t = r_{t-1} + u_t$$

Avec $\xi = 0$ si pas de tendance déterministe, r_0 sert de constante et u_t iid $(0, \sigma_u^2)$.

On teste (H_0) : $\sigma_u^2 = 0$ donc sous (H_0) la série est *stationnaire* (différent autres tests).

Sous R : `tseries::kpss.test` ou `urca::ur.kpss()`.

Auto.arima

Il existe beaucoup d'algorithmes automatiques pour la détection automatique de modèles ARIMA (TRAMO étant un des plus connus). Ils permettent de trouver un ARIMA$(p, d, q)(P, D, Q)_m$. Décrivons ici l'algorithme utilisé dans `forecast::auto.arima()` :

1. On choisit D par un test de Canova-Hansen puis d déterminé en utilisant des tests successifs de KPSS. Ils préfèrent utiliser des tests où (H_0) est l'absence de racine unitaire car les autres tests ont tendance à biaiser les résultats vers des modèles sur-différenciés.
2. Quatre modèles sont ensuite testés, on en sélectionne un par minimisaiton de l'AIC :

- $ARIMA(2, d, 2)(1, D, 1)$
- $ARIMA(2, d, 2)(0, D, 0)$
- $ARIMA(1, d, 0)(1, D, 0)$
- $ARIMA(0, d, 1)(0, D, 1)$

3. On considère 30 variations du modèle retenu :

- En faisant varier un seul des paramètres p, q, P ou Q de ± 1 ;
- En faisant varier p et q en même temps de ± 1 ;
- En faisant varier P et Q en même temps de ± 1 ;
- En incluant ou non la constante.
- Si un modèle minimise l'AIC on recommence.

Dans d'autres algorithmes (comme TRAMO ou pickmdl) d'autres tests sont effectués pour chaque modèle testé : tests d'autocorrélation(Ljung-Box à l'ordre 24 pour séries mensuelles), tests de sur-différenciation, de passage au log.