Struktūrinio pasikeitimo uždavinys panelinių duomenų regresijai

Vaidotas Zemlys

Vilniaus Universitetas

Université des Sciences et Technologies de Lille

2008 gegužės 13 d.

Uždavinys

- Ar modelis

$$y_{it} = \alpha_i + \mathbf{x}'_{it}\boldsymbol{\beta} + u_{it},$$

  yra adekvatus?

- Alternatyvus modelis:

$$y_{it} = \begin{cases} \alpha_i + \mathbf{x}'_{it}\boldsymbol{\beta}_0 + u_{it}, & \text{for } (i,t) \in I, \\ \alpha_i + \mathbf{x}'_{it}\boldsymbol{\beta}_1 + u_{it}, & \text{for } (i,t) \in I^c, \end{cases}$$

  čia I - indeksų aibė.

Vidurkio pasikeitimas vienmačiu atveju

- Turint imtį norime testuoti ar tam tikrame taške įvyko pasikeitimas
- Nulinė ir alternatyvi hipotezės:

$$H_0 : EX_i = \mu_0$$

prieš

$$H_A : \exists k^* \text{ toks kad } EX_i = \mu_0 + (\mu_1 - \mu_0)\mathbf{1}\,(k^* < i \leq n)$$

Jei pasikeitimas taškas žinomas

- Paprasta idėja: lyginam vidurkius.

$$\frac{1}{k^*}S_{k^*} - \frac{1}{n-k^*}\left(S_n - S_{k^*}\right) = \frac{n}{k^*(n-k^*)}\left(S_{k^*} - \frac{k^*}{n}S_n\right)$$

- Pažymėkime

$$R = (S_{k^*} - \frac{k^*}{n}S_n)$$

- Tegu $k^*/n \to c$. Tada dėl CRT turime

$$n^{-1/2}R \to N(0, \sigma^2 c(1-c))$$

- Prie alternatyvos

$$n^{-1/2}R = n^{1/2}\frac{k^*}{n}\left(1 - \frac{k^*}{n}\right)(\mu_1 - \mu_0) + O_P(1) \to \infty$$

Pasikeitimo taškas nežinomas

- Jei nežinome kur pasikeitimas, “ieškome” jo:

$$Q = \max_{1<k<n} |S_k - \frac{k}{n} S_n|$$

- Jeigu $\xi_n = S_{[nt]} + (nt - [nt])X_{[nt]+1}$:

$$Q = \max_{1<k<n} |\xi_n(k/n) - k/n\xi_n(1)|$$

- Taigi Q yra ξ_n funkcionalas

$$Q = f(\xi_n), \text{ čia } f(x) = \sup_{0<t<1} |x(t) - x(1)|$$

- Funkcionalas $f : C[0,1] \to R$ yra tolydus erdvėje $C([0,1])$.
- Pritaikius invariantiškumo principą kartu su tolydaus atvaizdžio teorema

$$n^{-1/2} Q \xrightarrow{D} \sup_{0<t<1} |W(t) - tW(1)|$$

Epideminės alternatyvos

- Norim testuoti epidemiją: vidurkis pasikeitė ir po to grįžta.
- Alternatyvi hipotezė: $EX_i = \mu_0 + (\mu_1 - \mu_0)\mathbf{1}\left(k^* < i \leq m^*\right)$
- Ta pati argumentacija:

$$UI(n, \alpha) = \max_{1 \leq i < j \leq n} \frac{|S_j - S_i - (j-i)/nS_n|}{[(j-i)/n]^\alpha}$$

- Funkcionalas netolydus $C[0,1]$, bet tolydus $\mathrm{H}^o_\alpha([0,1])$

Kodėl dalinti?

- Nustatytos epidemijos ilgis priklauso nuo α.
- Prie alternatyvos

$$n^{-1/2} UI(n,\alpha) \geq \frac{l^{*(1-\alpha)}}{n^{1/2-\alpha}} \left(1 - \frac{l^*}{n}\right) |\mu_1 - \mu_0| + O_P(1)$$

- Jei $l^* = n^\gamma$ tai

$$n^{-1/2} UI(n,\alpha) \to \infty \text{ kai } \gamma > \frac{1/2 - \alpha}{1 - \alpha}$$

Dvimatis atvejis

- Kažkuriame laiko intervale pasikeitimas įvyksta visiems individams
- Stebėjimo pradžioje pasikeitimas įvyksta tik kai kuriems individams
- Stebėjimo pabaigoje pasikeitimas įvyksta tik kai kuriems individams

Stačiakampės alternatyvos

Panelinių duomenų imtis $\{X_{ij}, i = 1, \ldots, n; j = 1, \ldots, m\}$

- (H_0) : Visų X_{ij} vidurkis yra μ_0.
- (H_A) : Egzistuoja sveiki skaičiai $1 < a^* \leq b^* < n$, $1 < c^* \leq d^* < m$ ir konstanta $\mu_1 \neq \mu_0$ tokia, kad

$$\mathbf{E}\, X_{ij} = \mu_0 + \mu_1 \mathbf{1}\left((i,j) \in [a^*, b^*] \times [c^*, d^*] \cap \mathbb{N}^2\right)$$

Pasikeitimo stačiakampis žinomas

- Lyginam vidurkius:

$$R = \sum_{i=1}^{n} \sum_{j=1}^{m} X_{ij} \mathbf{1} \left(\left(\frac{i}{n}, \frac{j}{m} \right) \in D^* \right) - \frac{k^* l^*}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} X_{ij},$$

  čia $k^* = b^* - a^*$, $l = d^* - c^*$.

- Prie nulinės hipotezės

$$(nm)^{-1/2} R \to N(0, \sigma^2 |D^*|(1 - |D^*|))$$

  kai $n \wedge m \to \infty$.

- Prie alternatyvos

$$(nm)^{-1/2} R = (nm)^{1/2} \frac{k^* l^*}{nm} \left(1 - \frac{k^* l^*}{nm} \right) (\mu_1 - \mu_0) + O_P(1).$$

- Statistika diverguos, kai $k^* = O(n^\gamma)$ ir $l^* = O(m^\delta)$ su $\gamma, \delta > 1/2$.

Kaip normuoti?

- Vienmačiu atveju normuojame skirtumu, funkcionalas tolydus Hiolderio erdvėje.
- Dvimačiu atveju normuojame diametru, funkcionalas tolydus Hiolderio erdvėje.

Normavimas diametru

- Testinė statistika

$$DUI(nm,\alpha) = \max_{\substack{1\le a<b\le n \\ 1\le c<d\le m}} \frac{|\Delta^1_{b-a}\Delta^2_{d-c}S_{bd} - (s_b - s_a)(t_d - t_c)S_{nm}|}{\max\{s_b - s_a, t_d - t_c\}^{\alpha}}$$

čia $s_i = i/n$, $t_j = j/m$, $i = 1, \ldots, n$, $j = 1, \ldots, m$ ir

$$\Delta^1_{b-a}\Delta^2_{d-c}S_{bd} = S_{bd} - S_{ad} - S_{bc} + S_{ac}$$

Ribinis pasiskirstymas

- Kiekvienam n ir m $DUI(nm,\alpha)$ yra tolydus funkcionalas $\mathrm{H}^o_\alpha([0,1]^2)$, kurio riba yra funkcionalas T_α

$$T_\alpha(x) = \sup_{\mathbf{0}\le \boldsymbol{s}<\boldsymbol{t}\le \mathbf{1}} \frac{|x(\boldsymbol{t})-x(s_1,t_2)-x(t_1,s_2)+x(\boldsymbol{s})-(t_1-s_1)(t_2-s_2)x(1,1)|}{|\boldsymbol{t}-\boldsymbol{s}|^\alpha}$$

- Invariantiškumo principas duoda

$$(nm)^{-1/2} DUI(nm,\alpha) \to T_\alpha(W),$$

čia W - Vynerio paklodė.

Elgesys prie alternatyvos

Jei X_{ij} yra nepriklausomi ir $\sigma_0^2 = \sup_{\boldsymbol{n}} var(X_{\boldsymbol{n}}) < \infty$ bei

$$\lim_{\boldsymbol{n}\to\infty}(nm)^{1/2}\frac{h_{nm}}{d_{nm}^{\alpha}}|\mu_1 - \mu_0| \to \infty,$$

čia

$$h_{nm} = \frac{k^* l^*}{nm}\left(1 - \frac{k^* l^*}{nm}\right) \text{ ir } d_{nm} = \max\left\{\frac{k^*}{n}, \frac{l^*}{m}\right\},$$

tai

$$(nm)^{-1/2} DUI(nm, \alpha) \to \infty$$

Epidemijų ilgumas

- Vienmačiu atveju galima epidemijos ilgis $n^{\frac{1-2\alpha}{2-2\alpha}}$.
- Dvimačiu atveju tarkime, kad $k^* = n^\gamma$, $l^* = m^\delta$ ir kad $\mu_1 - \mu_0$ nepriklauso nuo (n, m). Tada

$$\frac{n^{\gamma-1/2} m^{\delta-1/2}}{[n^{\gamma-1} \vee m^{\delta-1}]^\alpha} \to \infty$$

- Jei $n^{\gamma-1} > m^{\delta-1}$ tai

$$n^{\gamma(1-\alpha)+\alpha-1/2} m^{\delta-1/2} \to \infty$$

Regresija vienmatis atvejis

Nulinė hipotezė

$$y_t = \mathbf{x}_t'\boldsymbol{\beta} + u_t \tag{1}$$

Alternatyva

$$y_t = \begin{cases} \mathbf{x}_t'\boldsymbol{\beta}_0 + u_t, t = 1, \dots, t_0, \\ \mathbf{x}_t'\boldsymbol{\beta}_1 + u_t, t = t_0 + 1, \dots, T. \end{cases} \tag{2}$$

- CUSUM statistikos, Brown, Durbin, Watson (1975) normalių paklaidų atveju, rekursinėm liekanoms.
- Sen (1982), i. i. d. paklaidų atveju.
- Kramer, Ploberger (1992), įprastoms liekanoms.

Panelinių duomenų regresija

- Įprasta regresija

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + u_{it}, \quad i = 1, \ldots, n; t = 1, \ldots, T$$

- Fiksuotų efektų regresija

$$y_{it} = \mu_i + \mathbf{x}_{it}\boldsymbol{\beta} + u_{it}, \quad i = 1, \ldots, n; t = 1, \ldots, T$$

Invariantiškumo principas

Jeigu u_{it} tenkina invariantiškumo principą:

$$(nT)^{-1/2}\xi_{nT} \xrightarrow{D} W$$

tai

- Įprastai regresijai

$$(nT)^{-1/2}\xi_{nT}^{(OLS)}(u,v) \to W(u,v) - uvW(1,1)$$

- Fiksuotų efektų regresijai

$$(nT)^{-1/2}\xi_{nT}^{(FE)}(u,v) \to W(u,v) - vW(u,1)$$

Lokalios alternatyvos

Suppose

$$\boldsymbol{\beta}_{ij} = \boldsymbol{\beta} + \frac{1}{\sqrt{nm}} \boldsymbol{g}\left(\frac{i}{n}, \frac{j}{m}\right)$$

- Įprasta regresija

$$(nT)^{-1/2} \xi_{nT}^{(OLS)}(t,s) \xrightarrow{D} W(t,s) - tsW(1,1)$$
$$+ \int_0^t \int_0^s \boldsymbol{c}' \boldsymbol{g}(u,v) du dv - ts\boldsymbol{c}' \int_0^1 \int_0^1 \boldsymbol{g}(u,v) du dv$$

- Fiksuotų efektų regresija

$$(nT)^{-1/2} \xi_{nT}^{(FE)}(t,s) \xrightarrow{D} W(t,s) - sW(t,1)$$
$$+ \int_0^t \int_0^s \boldsymbol{c}' \boldsymbol{g}(u,v) du dv - s \int_0^t \int_0^1 \boldsymbol{c}' \boldsymbol{g}(u,v) du dv,$$