O DETEKCI ZMĚN V PANELOVÝCH DATECH

J. ANTOCH (a další)

ROBUST 2016
16. ZÁŘÍ 2016

Model

Uvažujeme lineární regresní model

$$y_{i,t} = \boldsymbol{x}_{i,t}^{\top}\big(\boldsymbol{\beta}_i + \boldsymbol{\delta}_i I\{t \geq t_0\}\big) + e_{i,t}, \qquad var\, e_{i,t} = \sigma_i^2 \tag{1}$$

$$1 \leq i \leq N,\ 1 \leq t \leq T,\ 1 \leq t_0 \leq T \text{ fixed}$$

Model

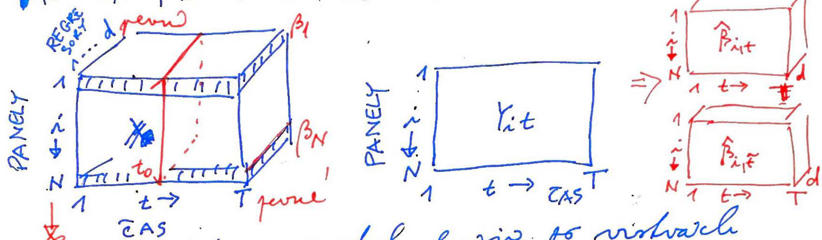
Uvažujeme lineární regresní model

$$y_{i,t} = \boldsymbol{x}_{i,t}^\top\big(\boldsymbol{\beta}_i + \boldsymbol{\delta}_i I\{t \geq t_0\}\big) + e_{i,t}, \qquad var\ e_{i,t} = \sigma_i^2 \tag{1}$$

$$1 \leq i \leq N,\ 1 \leq t \leq T,\ 1 \leq t_0 \leq T \text{ fixed}$$

$\boxed{\text{MODEL MANINA}}$

(1.1) $\quad y_{i,t} = x_{i,t}\left(\beta_i + \delta_i I[t \geq t_0]\right) + e_{i,t} \qquad 1 \leq i \leq N, \ 1 \leq t \leq T$



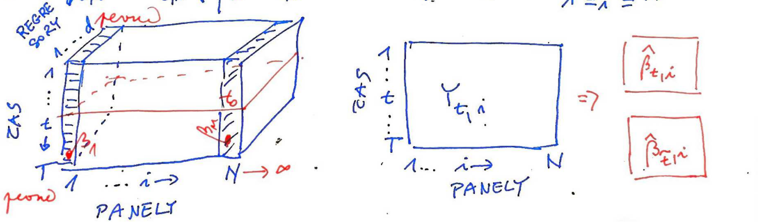
X je 3D table (three-way), kterou model krájí po vrstvách

Model - poznámka

(three-way)

MODEL JÁRA (by se mu líbil mnohem více / je lip se dá uživatelům mnohem snaze „uchopit")

$$(\tilde{1.1}) \quad Y_{t,i} = x_{t,i}^T \left(\beta_i + \delta_i \, I[t \geq t_0]\right) + \varepsilon_{t,i} \qquad 1 \leq t \leq T, \quad 1 \leq i \leq N$$



Problémy k řešení

- jak odhadnout neznámé parametry modelu
- rozhodnout, zda naměřená data jsou „homogenní“, tj. lze je popsat modelem (1), nebo zda došlo v průbehu času ke změně
- v případě zamítnutí hypotézy homogenity odhadnout, kdy ke změně došlo
- rozhodnout, zda neošlo případně k více změnám

Problémy k řešení

- jak odhadnout neznámé parametry modelu
- rozhodnout, zda naměřená data jsou „homogenní", tj. lze je popsat modelem (1), nebo zda došlo v průbehu času ke změně
- v případě zamítnutí hypotézy homogenity odhadnout, kdy ke změně došlo
- rozhodnout, zda neošlo případně k více změnám

- jaký vliv by mělo, kdyby došlo ke změně jenom v některých panelech
- jaký vliv by mělo, kdyby došlo ke změně nikoliv ve stejné době t_0
- jaký je vliv vybraných skupin panelů, např. odvětví průmyslu
- jak moc mohou být data závislá
- atd.

Testová statistika

$$\sum_{i=1}^{N} \left(\widehat{\boldsymbol{\beta}}_{i,t} - \widehat{\boldsymbol{\beta}}_{i,T}\right)^{\top} \boldsymbol{C}_{i,t} \left(\widehat{\boldsymbol{\beta}}_{i,t} - \widehat{\boldsymbol{\beta}}_{i,T}\right) \tag{2}$$

kde

$$\boldsymbol{C}_{i,t} = \boldsymbol{Z}_{i,t} \boldsymbol{Z}_{i,T}^{-1} \boldsymbol{Z}_{i,t}, \quad \boldsymbol{Z}_{i,t} = \sum_{i=1}^{t} \boldsymbol{x}_{i,t} \boldsymbol{x}_{i,t}^{\top}.$$

- $\widehat{\boldsymbol{\beta}}_{i,t}$ je odhad MNČ založený na prvních t-pozorováních pro i-tý panel
- $\widehat{\boldsymbol{\beta}}_{i,T}$ je odhad MNČ založený na všech pozorováních pro i-tý panel

$$\widehat{\sigma}_i^2 = \frac{1}{T-d} \sum_{t=1}^{T} \left(y_{i,t} - \boldsymbol{x}_{i,t}^{\top} \widehat{\boldsymbol{\beta}}_{i,T}\right)^2. \tag{3}$$

Poznámky: • Volba $\boldsymbol{C}_{i,t}$ velmi ovlivňuje chování testové statistiky, byť asymptotika je mnohem „velkorysejší“.

• Nešlo by σ_i^2 odhadovat lépe? Především za případné alternativy.

Co umíme

- najít rozdělení našeho odhadu
  (podmínky regularity jsou na dvě hustě popsané strany A4)
- máme „napočítanou“ představu o tom, které volby matice $\boldsymbol{C}_{i,t}$ fungují
- máme napočítány síly vybraných testů pro různé modely a různé typy změn
- máme představu o výpočetní náročnosti
- máme představu o tom, jaký bootstrap použít a co přináší

Vybrané předpoklady

Assumption 1.7.

$$\frac{1}{N^{1/2}}\sum_{i=1}^{N}\|\boldsymbol{\gamma}_i\|^2 = o(1).$$

Define

$$\begin{aligned}
a^2_{i,t,1} &= \sigma_i^2\left(\sum_{s=1}^{t}\mathbf{x}_{i,s}^\top(\mathbf{Z}_{i,t}^{-1}-\mathbf{Z}_{i,T}^{-1})\mathbf{C}_{i,t}(\mathbf{Z}_{i,t}^{-1}-\mathbf{Z}_{i,T}^{-1})\mathbf{x}_{i,s}+\sum_{s=t+1}^{T}\mathbf{x}_{i,s}^\top\mathbf{Z}_{i,T}^{-1}\mathbf{C}_{i,t}\mathbf{Z}_{i,T}^{-1}\mathbf{x}_{i,s}\right)\\
&= \sigma_i^2\left\{\mathrm{tr}((\mathbf{Z}_{i,t}^{-1}-\mathbf{Z}_{i,T}^{-1})\mathbf{C}_{i,t}(\mathbf{Z}_{i,t}^{-1}-\mathbf{Z}_{i,T}^{-1})\mathbf{Z}_{i,t})+\mathrm{tr}(\mathbf{Z}_{i,T}^{-1}\mathbf{C}_{i,t}\mathbf{Z}_{i,T}^{-1}\widetilde{\mathbf{Z}}_{i,T-t})\right\}\\
&= \sigma_i^2\mathrm{tr}(\mathbf{C}_{i,t}(\mathbf{Z}_{i,t}^{-1}-\mathbf{Z}_{i,T}^{-1})),\\
A_N^{(1)}(t) &= \sum_{1\le i\le N} a^2_{i,t,1},\\
\mathbf{S}_{i,t}(j) &= \sum_{s=1}^{t}\mathbf{x}_{i,s}\varepsilon_{i,s}.
\end{aligned}$$

For the existence of the asymptotic variance of $U_N(t)$ we require

Assumption 1.8. *the function*

$$\begin{aligned}
\Gamma^{(1)}(t,t') = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\Big[E\Big\{&(\mathbf{Z}_{i,t}^{-1}\mathbf{S}_{i,t}-\mathbf{Z}_{i,T}^{-1}\mathbf{S}_{i,T})^\top\mathbf{C}_{i,t}(\mathbf{Z}_{i,t}^{-1}\mathbf{S}_{i,t}-\mathbf{Z}_{i,T}^{-1}\mathbf{S}_{i,T})\\
&\times(\mathbf{Z}_{i,t'}^{-1}\mathbf{S}_{i,t'}-\mathbf{Z}_{i,T}^{-1}\mathbf{S}_{i,T})^\top\mathbf{C}_{i,t'}(\mathbf{Z}_{i,t'}^{-1}\mathbf{S}_{i,t'}-\mathbf{Z}_{i,T}^{-1}\mathbf{S}_{i,T})\Big\}-a^2_{i,t,1}a^2_{i,t',1}\Big]
\end{aligned}$$

exists for all $\underline{t}_0\le t,t'\le T-\bar{t}_0$, *where* $\underline{t}_0=\max(d,t_1,t_2)$ *and* $\bar{t}_0=t_3$.

Somewhere should be discussion when this is fulfilled. ??? Here probably one could take into account assumptions in Baltagi et al (2016) First we consider the case when the effect of the common factors is negligible i.e., A 1.5 is assumed.

Vybraná tvrzení

Theorem 1.1. *If H_0 and Assumptions* as-1 ass-gamma *1.1–1.8 hold, then we have*

$$\left\{N^{-1/2}\left(U_N(t) - A_N^{(1)}(t)\right),\ \underline{t}_0 \le t \le T - \bar{t}_0\right\} \quad \xrightarrow{\mathcal{D}} \quad \left\{\xi_t^{(1)}, \underline{t}_0 \le t \le T - \bar{t}_0\right\}, \tag{1.7}$$

where $\xi_t^{(1)}, \underline{t}_0 \le t \le T - \bar{t}_0$ is jointly normal with zero mean and covariance $\Gamma^{(1)}(t,t')$ and $\underline{t}_0, \bar{t}_0$ are defined in Assumption ass-ga-1 *??.*

Next we consider the partial sums process $V_N(t)$ of the squares of the residuals, i.e. $V_N(t)$ defined in (V_N 1.6). Let

$$A_N^{(2)}(t) = \sum_{i=1}^{N} a_{i,t,2}^2 \quad \text{with} \quad a_{i,t,2}^2 = \sigma_i^2 \sum_{s=1}^{t} \left(1 - \mathbf{x}_{i,s}^\top \mathbf{Z}_{i,T}^{-1} \mathbf{x}_{i,s}\right).$$

It is easy to see that

$$a_{i,t,2}^2 = \sigma_i^2 \left(t - \mathrm{tr}(\mathbf{Z}_{i,t} \mathbf{Z}_{i,T}^{-1})\right).$$

To get the limit distribution of $V_N(t)$ we need to replace Assumption ass-ga-1 ?? with

Assumption 1.9. *the function*

$$\Gamma^{(2)}(t,t') = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \left[\sum_{s=1}^{t} \sum_{s'=1}^{t'} E\left\{(\varepsilon_{i,s} - \mathbf{x}_{i,s}^\top \mathbf{Z}_{i,T}^{-1} \mathbf{S}_{i,T})(\varepsilon_{i,s'} - \mathbf{x}_{i,s'}^\top \mathbf{Z}_{i,T}^{-1} \mathbf{S}_{i,T})\right\}^2 - a_{i,t,2}^2 a_{i,t',2}^2\right]$$

exists for all $1 \le t, t' \le T$.

Theorem 1.2. *If H_0 and Assumptions* as-1 *1.1,* as-0 as-6 *1.3–1.2,* as-8 *1.7 and* ass-ga-2 *1.9 hold, then we have that*

$$\left\{N^{-1/2}\left(V_N(t) - A_N^{(2)}(t)\right),\ 1 \le t \le T\right\} \quad \xrightarrow{\mathcal{D}} \quad \left\{\xi_t^{(2)}, 1 \le t \le T\right\}, \tag{1.8}$$

where $\xi_t^{(2)}, 1 \le t \le T$ is jointly normal with zero mean and covariance $\Gamma^{(2)}(t,t')$.

Here discussion about the related tests it means take max over t but it contains unknown

	change in intercept with $\boldsymbol{\delta} = (0.25, 0)^\top$					change in slope with $\boldsymbol{\delta} = (0, 0.25)^\top$				
	number of panels					number of panels				
	200	400	600	800	1000	200	400	600	800	1000
$T = 50$	0.074	0.154	0.214	0.292	0.370	0.320	0.678	0.868	0.946	0.988
$T = 100$	0.268	0.498	0.760	0.848	0.962	0.888	0.996	1	1	1
$T = 150$	0.504	0.892	0.972	0.996	1	0.996	1	1	1	1
$T = 200$	0.738	0.984	1	1	1	1	1	1	1	1

Tab. 7. Nominal rejection proportions; $t_0 = T/2$; $\vartheta = 0.5$; rejection rate at 5%.

	change in intercept with $\boldsymbol{\delta} = (0.25, 0)^\top$					change in slope with $\boldsymbol{\delta} = (0, 0.25)^\top$				
	number of panels					number of panels				
	200	400	600	800	1000	200	400	600	800	1000
$T = 50$	0.042	0.060	0.116	0.092	0.096	0.082	0.200	0.320	0.422	0.518
$T = 100$	0.090	0.164	0.238	0.342	0.366	0.360	0.652	0.894	0.964	0.984
$T = 150$	0.132	0.286	0.500	0.644	0.742	0.680	0.962	0.998	1	1
$T = 200$	0.282	0.494	0.736	0.860	0.930	0.890	0.994	1	1	1

Tab. 8. Nominal rejection proportions; $t_0 = T/2$; $\vartheta = 0.25$; rejection rate at 5%.

Špatné zprávy

Funguje nám to pěkně na simulovaných datech, ale jakmile kolega Hanousek přinese reálná ekonomická data, tak jsme . . .