Základy ekonometrie

VI. Uvolnění klasických předpokladů – autokorelace náhodných složek

Obsah tématu

Motivace

- Předpoklad: chyby regrese pro různá pozorování vzájemně nekorelovaná.
- Rozumné pro průřezová data.
- Časové řady: korelace sousedních pozorování.
- Příklad: vývoj makroekonomických agregátů.
- Korelace mezi pozorováními $\rightarrow$ korelaci mezi různými náhodnými členy.

Východiska

- Značení: $t = 1, \ldots, T$.
- Pro časové řady i lepší metody a modely než LRM.
- Pro naše potřeby:

$$Y_t = \alpha + \beta_1 X_{1t} + \ldots + \beta_k X_{kt} + \epsilon_t.$$

- Splněny všechny klasické předpoklady, jen $cov(\epsilon_t, \epsilon_{t-s}) \neq 0$ pro některá $s \neq 0$.

Důsledky

- Při splnění všech klasických předpokladů $\rightarrow$ Gaussův-Markovův teorém říká, že OLS je BLUE.
- Při nesplnění předpokladu o nekorelovanosti: OLS estimátor není BLUE.
- OLS nestranný, ale ne nejlepší.
- Využití GLS estimátoru $\rightarrow$ vhodná transformace (a OLS).

Obsah tématu

$AR(1)$ proces

- Náhodné chyby: *autoregresní proces řádu 1 (AR(1))*:

$$\epsilon_t = \rho\epsilon_{t-1} + u_t.$$

- u_t: $E(u_t) = 0$, $var(u_t) = \sigma_u^2$ a $cov(u_t, u_{t-s}) = 0$ (pro $s \neq 0$).
- Předpokládáme $-1 < \rho < 1 \rightarrow$ *stacionarita*.
- Obecný $AR(p)$ procesu náhodných chyb:

$$\epsilon_t = \rho_1\epsilon_{t-1} + \rho_2\epsilon_{t-2} + \ldots + \rho_p\epsilon_{t-p} + u_t.$$

Vlastnosti ϵ_t – střední hodnota

- Klasické předpoklady pro ϵ_t.
- $AR(1)$ model pro všechny časové okamžiky (od $-\infty$ do ∞).
- Pozorujeme $t = 1, \ldots, T$ a platí $\epsilon_{t-s} = \rho\epsilon_{t-s-1} + u_{t-s}$ pro jakékoli t a s.
- Postupné substituce $\rightarrow$ jen $u_t, u_{t-1}, \ldots$.

$$\epsilon_t = \sum_{i=0}^{\infty} \rho^i u_{t-i}.$$

- Chyby regrese mají nulovou střední hodnotu:

$$E(\epsilon_t) = E\left(\sum_{i=0}^{\infty} \rho^i u_{t-i}\right) = 0.$$

Vlastnosti ϵ_t – rozptyl

- Využijeme pro c menší než 1 v absolutní hodnotě:

$$\sum_{i=0}^{\infty} c^i = \frac{1}{1-c}.$$

- Vlastnosti operátoru rozptylu a u_t splňuje klasické předpoklady:

$$\begin{aligned} var\left(\epsilon_t\right) &= var\left(\sum_{i=0}^{\infty} \rho^i u_{t-i}\right) \\ &= \sum_{i=0}^{\infty} \rho^{2i} var(u_{t-i}) = \sigma_u^2 \sum_{i=0}^{\infty} \rho^{2i} \\ &= \frac{\sigma_u^2}{1-\rho^2}, \end{aligned}$$

- Autokorelované náhodné chyby jsou homoskedastické.

Vlastnosti ϵ_t – kovariance

- S využitím definice operátoru kovariance a $E(\epsilon_t) = 0$:

$$cov(\epsilon_t, \epsilon_{t-s}) = E(\epsilon_t \epsilon_{t-s}) - E(\epsilon_t)E(\epsilon_{t-s}) = E(\epsilon_t \epsilon_{t-s})$$
$$= E\left[\left(\sum_{i=0}^{\infty} \rho^i u_{t-i}\right)\left(\sum_{i=0}^{\infty} \rho^i u_{t-s-i}\right)\right].$$

- $cov(u_t, u_{t-s}) = E(u_t u_{t-s}) = 0$ (pro $s \neq 0$) $\rightarrow$ eliminace většiny členů:

$$cov(\epsilon_t, \epsilon_{t-s}) = E(\rho^s u_{t-s}^2 + \rho^{s+2} u_{t-s-1}^2 + \dots)$$
$$= E\left(\sum_{i=0}^{\infty} \rho^{s+2i} u_{t-s-i}^2\right) = \sum_{i=0}^{\infty} \rho^{s+2i} E(u_{t-s-i}^2)$$
$$= \sigma_u^2 \sum_{i=0}^{\infty} \rho^{s+2i} = \sigma_u^2 \rho^s \sum_{i=0}^{\infty} \rho^{2i}$$
$$= \frac{\sigma_u^2 \rho^s}{1 - \rho^2} = \rho^s var(\epsilon_t).$$

Příklad

- Gujarati a Porter (2009) – příklad 12.5.
- Datový soubor `Table_12.4` na záložce `Gujarati`.
- Level-level i log-log specifikace.
- Vztah mezi mzdami a produktivitou v sektoru obchodu v USA (1960–2005), datový soubor obsahuje data období 1959-1998 .
- Y – reálná kompenzace za hodinu (index).
- X – výstup za hodinu (index).

Obsah tématu

Transformace modelu

- „Kvazi diferencování“.

$$Y_t = \alpha + \beta_1 X_{1t} + \ldots + \beta_k X_{kt} + \epsilon_t.$$

- Platí v každém časovém okamžiku $\rightarrow$ čas $t-1$ a násobení parametrem ρ:

$$\rho Y_{t-1} = \rho\alpha + \rho\beta_1 X_{1t-1} + \ldots + \rho\beta_k X_{kt-1} + \rho\epsilon_{t-1}.$$

- Odečtení od původní regrese:

$$Y_t - \rho Y_{t-1} = \alpha - \rho\alpha + \beta_1(X_{1t} - \rho X_{1t-1}) + \ldots + \beta_k(X_{kt} - \rho X_{kt-1}) + \epsilon_t - \rho\epsilon_{t-1}$$

- Nové značení:

$$Y_t^* = \alpha^* + \beta_1 X_{1t}^* + \ldots + \beta_k X_{kt}^* + u_t.$$

- u_t splňuje klasické předpoklady $\rightarrow$ metoda OLS je BLUE (GLS estimátor pro model s $AR(1)$ náhodnými složkami).

Problém

- GLS estimátor zahrnuje regresi Y^* na $X_1^*, \ldots, X_k^*$, kdy proměnné „kvazi-diferencované“:

$$
\begin{aligned}
Y_t^* &= Y_t - \rho Y_{t-1}, \\
X_{1t}^* &= X_{1t} - \rho X_{1t-1}, \\
&\text{atd.}
\end{aligned}
$$

- Problém: data od $t = 1, \ldots, T \Rightarrow Y_1^* = Y_1 - \rho Y_0$ zahrnuje Y_0 (totéž pro vysvětlující proměnné).
- Nejsme schopni pozorovat *počáteční podmínku* jako je Y_0.
- Řešení: data od $t = 2, \ldots, T$ (data pro $t = 1$ jako počáteční podmínky) $\rightarrow$ ztráta pozorování.
- Mnohem sofistikovanější metody jako je odhad metodou maximální věrohodnosti.

Praisova-Winstenova transformace

- Model: $Y_t = \beta_1 + \beta_2 X_t + \epsilon_t$.
- Náhodné $AR(1)$ složky: $\epsilon_t = \rho\epsilon_{t-1} + u_t$.
- Kvazi-diferencování:

$$(Y_t - \rho Y_{t-1}) = \beta_1(1-\rho) + \beta_2(X_1 - \rho X_{t-1}) + \epsilon_t - \rho\epsilon_{t-1}.$$

- Ztráta jednoho pozorování.
- Vyhnutí se ztráty = transformace prvních pozorování:

$$Y_1\sqrt{1-\rho^2} \quad X_1\sqrt{1-\rho^2}.$$

- Pokud necháme první pozorování netransformované → problém s homoskedasticitou!

Metoda prvních diferencí

- Pokud je ρ blízké jedné: silná pozitivní autokorelace (> 0.8).
- Model již bez autokorelace: $\Delta Y_t = \beta_2 \Delta X_t + u_t$.
- Co když do regrese dáme úrovňovou konstantu?

$$\Delta Y_t = \beta_1 + \beta_2 \Delta X_t + u_t.$$

- Možnost testování trendu v původní časové řadě.
- První diference i pro stacionarizaci časových řad.

$$\epsilon_t = \epsilon_{t-1} + u_t \rightarrow \epsilon_t - \epsilon_{t-1} = \Delta\epsilon_t = u_t).$$

- Berenbluttův-Webbův test: $H_0 : \rho = 1 \rightarrow$ g statistika

$$g = \frac{\sum_{t=2}^{N} \widehat{u}_t^2}{\sum_{t=1}^{N} \widehat{\epsilon}_t^2.}$$

- Tabulky DW testu $\times$ nulová hypotéze $\rho = 1$ (ne $\rho = 0$).

Obsah tématu

Motivace

- Pokud známe $\rho \rightarrow$ transformace.
- Jinak odhad, $\widehat{\rho}$.
- Využití OLS metody (nestranné odhady) + získání reziduí k odhadu parametru ρ.
$$\epsilon_t = \rho\epsilon_{t-1} + u_t.$$
- Další zobecnění pro $AR(p)$ procesy: odhady koeficientů $\rho_1, \ldots, \rho_p$ + vynechání p pozorování (počáteční podmínky).
- Použití $\widehat{\rho} \rightarrow$ feasible GLS (FGLS) resp. estimated GLS (EGLS).
- Dobré u velkých vzorků, u malých vzorků potřeba opatrnosti při interpretaci (plus ponechání prvních pozorování – á la Praisova-Winstonova transformace = tzv. full EGLS resp FEGLS).

Postup

1. Regrese Y_t na úrovňovou konstantu, $X_1, \ldots, X_k$ s využitím OLS a získání reziduí, $\widehat{\epsilon}_t$.
2. Regrese $\widehat{\epsilon}_t$ na $\widehat{\epsilon}_{t-1}$ využitím OLS $\rightarrow$ získání $\widehat{\rho}$.
3. Kvazi-diferencování všech proměnných:

$$
\begin{aligned}
Y_t^* &= Y_t - \widehat{\rho} Y_{t-1}, \\
X_{1t}^* &= X_{1t} - \widehat{\rho} X_{1t-1}, \\
&\text{atd.}
\end{aligned}
$$

4. Regrese Y_t^* na úrovňovou konstantu, $X_1^*, \ldots, X_k^*$ metodou OLS $\rightarrow$ GLS odhady koeficientů.

Iterační varianta

1. Regrese Y_t na úrovňovou konstantu, $X_1, \ldots, X_k$ pomocí OLS a získání reziduí, $\widehat{\epsilon}_t$.
2. Regrese $\widehat{\epsilon}_t$ na $\widehat{\epsilon}_{t-1}$ využitím OLS a získání $\widehat{\rho}$.
3. Kvazi-diferencování všech proměnných k získání:

$$
\begin{aligned}
Y_t^* &= Y_t - \widehat{\rho} Y_{t-1}, \\
X_{1t}^* &= X_{1t} - \widehat{\rho} X_{1t-1}, \\
&\text{atd.}
\end{aligned}
$$

4. Regrese Y_t^* na úrovňovou konstantu, $X_1^*, \ldots, X_k^*$ metodou OLS a získání GLS odhadů koeficientů $\widehat{\alpha}, \widehat{\beta}_1, \ldots, \widehat{\beta}_k$.
5. Vytvoření nových reziduí použitím GLS odhadu z kroku 4, $\widehat{\epsilon}_t = Y_t - \widehat{\alpha} - \widehat{\beta}_1 X_{1t} - \ldots - \widehat{\beta}_k X_{kt}$.
6. Návrat ke kroku 2 opakování postupu stále dokola dokud se odhad, $\widehat{\rho}$ nepřestane měnit (s danou tolerancí).

Obsah tématu

Motivace

- Podobně jako u heteroskedasticity (HCE).
- Korektní vztah pro $var(\widehat{\beta})$ = korektní použití OLS metody.
- Druhé nejlepší řešení.
- Obvykle tzv. *heteroskedasticitě a autokorelaci konzistentními estimátory (HAC)*, např. *Neweyho-Westův estimátor*.
- Problém malých vzorků!

Obsah tématu

Motivace

- Pokud je $\rho = 0 \rightarrow$ korektní OLS metoda.
- Test nulové hypotézy: $H_0 : \rho = 0$ oproti alternativní hypotéze $H_1 \neq 0$.
- Řada testů (i obecnější než jen $AR(1)$).
- Proč tolik testů? $\rightarrow$ jako u heteroskedasticity není prokázáno, který je nejlepší.
- Jednoduchý test – vykreslení (standardizovaných) reziduí.

Testy věrohodnostního poměru

- $H_0 : \rho = 0$ oproti alternativě $H_1 : \rho \neq 0$.
- Potřeba maximálně věrohodného odhadu jak omezeného, tak i neomezeného modelu.
- Software udělá práci za nás – nelineární optimalizace.
- Z důvodu potřeby odhadu obou modelů není preferován.

Breuschův-Godfreyho test

- Test Lagrangeových multiplikátorů (odhad pouze omezeného modelu).
- Omezený model = model vícenásobné regrese splňující klasické předpoklady.
- Pro obecný případ $AR(p)$ procesu:

$$\epsilon_t = \rho_1 \epsilon_{t-1} + \rho_2 \epsilon_{t-2} + \ldots + \rho_p \epsilon_{t-p} + u_t.$$

- Test sdružené hypotézy $H_0 : \rho_1 = 0, \rho_2 = 0, \ldots, \rho_p = 0$:
  1. Regrese Y_t na úrovňovou konstantu a $X_1, \ldots, X_k$ použitím OLS a zísjkání reziduí, $\widehat{\epsilon}_t$.
  2. Regrese $\widehat{\epsilon}_t$ na úrovňovou konstantu a $X_1, \ldots, X_k, \widehat{\epsilon}_{t-1}, \ldots, \widehat{\epsilon}_{t-p}$ metodou OLS a získání koeficientu determinace, R^2.
  3. Spočítání testové statistiky $LM = TR^2$
- Při platnosti nulové hypotézy, H_0, má LM statistika (aproximativně) $\chi^2(p)$ rozdělení.

Boxův-Pierceův test

- Myšlenka: pokud nejsou náhodné složky autokorelován, potom by korelace mezi různými náhodnými složkami měla být nulová.
- Odhady v podobě reziduí, $\widehat{\epsilon}_t$, pro $t = 1, \ldots, T$, na základě OLS regrese Y na úrovňovou konstantu a $X_1, \ldots, X_k$.
- Odhad korelace mezi ϵ_t a ϵ_{t-s}:

$$r_s = \frac{\sum_{t=s+1}^{T} \widehat{\epsilon}_t \widehat{\epsilon}_{t-s}}{\sum_{t=s+1}^{T} \widehat{\epsilon}_t^2}.$$

- *Boxova-Pierceova testová statistika* (*Q-statistika*):

$$Q = T \sum_{j=1}^{p} r_j^2,$$

- Výběr p označuje test existence $AR(p)$.

Ljungův test

- Stejný postup jako Boxův-Pierceův test.
- *Ljungova testová statistika* (*Ljungovo Q*):

$$Q^* = T(T+2)\sum_{j=1}^{p} \frac{r_j^2}{T-j}.$$

- Obě statistiky mají (aproximativně) $\chi^2(p)$ (při platnosti $H_0 : \rho_1 = 0, \rho_2 = 0, \ldots, \rho_p = 0$).
- Úskalí: mezi regresory zpožděná vysvětlovaná proměnná $\Rightarrow$ oba testy nejsou vhodné.

Durbinův-Watsonův test

- Pro $AR(1)$ proces.
- OLS regrese vysvětlované veličin, Y, na úrovňovou konstantu a vysvětlující proměnné, $X_1, \ldots, X_k \rightarrow$ OLS rezidua, $\widehat{\epsilon}_t$.
- Testová statistika:

$$DW = \frac{\sum_{t=2}^{T}(\widehat{\epsilon}_t - \widehat{\epsilon}_{t-1})^2}{\sum_{t=1}^{T}\widehat{\epsilon}_t^2}.$$

- Platí: $0 \leq DW \leq 4$.
- DW aproximativně rovna $2(1-\widehat{\rho}) \Longleftrightarrow \widehat{\rho} = 1 - DW/2$.
- Nestandardní rozdělení (hodnoty tabelovány); z programů i p-hodnoty (H_0: nekorelovanost).

Vyhodnocení DW testu

- Z tabulek hodnoty dolní meze (d_L) a horní meze (d_U).
- Závisí na počtu pozorování, T, a počtu vysvětlujících proměnných, k.

Hodnota testu	Vyhodnocení
$4 - d_L < DW < 4$	Zamítáme H_0; závěr $\rho < 0$
$4 - d_U < DW < 4 - d_L$	Výsledek neurčitý
$2 < DW < 4 - d_U$	Nezamítáme H_0; závěr $\rho = 0$
$d_U < DW < 2$	Nezamítáme H_0; závěr $\rho = 0$
$d_L < DW < d_U$	Výsledek neurčitý
$0 < DW < d_L$	Zamítáme H_0; závěr $\rho > 0$

- Není zaručeno jednoznačné zamítnutí $H_0 : \rho = 0$.
- Nevhodný v případě zpožděné závisle proměnné jako vysvětlující proměnné a při náhodných regresorech (i pro velké vzorky).

Durbinova *h*-statistika

- Pro případ, kdy jedna z vysvětlujících proměnných je zpožděná vysvětlovaná proměnná:

$$Y_t = \alpha + \beta Y_{t-1} + \gamma X_t + \epsilon_t.$$

- Postup:
  1. Provedení OLS regrese Y na úrovňovou konstantu, zpožděnou závisle proměnnou a X (případně další vysvětlující proměnné v regresi).
  2. Získání DW statistiky a rozptylu odhadu parametru β (tj. $var(\widehat{\beta})$).
  3. Výpočet testové statistiky (*Durbinovo h*):

$$h = \left(1 - \frac{DW}{2}\right)\sqrt{\frac{T}{1 - Tvar(\widehat{\beta})}}.$$

  4. Platí: $h \sim N(0,1)$.
- Test nebude fungovat pro $Tvar(\widehat{\beta}) > 1$.

Obsah tématu

Praktická doporučení

- Jestliže testy indikují přítomnost autokorelace: zkusit model transformovat či jinak specifikovat.
- Hrubé pravidlo: práce s prvními diferencemi pokud $DW < R^2$.
- Možné použití Cochranovy-Orcuttovy procedury (případ $AR(1)$ náhodných složek) $\times$ část ekonometrů proti tomuto postupu (zavádějící výsledky pokud jiný než $AR(1)$ proces).
- Lze zobecnit i pro $AR(p)$ procesy.
- Po každé transformaci vhodné testovat vyřešení problému.
- Pokud nedokážeme nalézt transformaci $\rightarrow$ raději HAC estimátor.

Problém malých vzorků

- FGLS a HAC oproti OLS efektivní $\times$ vlastnosti u malých vzorků málo zdokumentované.
- Malé vzorky: studie (Monte Carlo simulace) $\rightarrow \rho < 0.3 \Rightarrow$ raději OLS.
- Co je velké? $\rightarrow$ relativní $\rightarrow$ 15 pozorování asi ne $\times$ 50+ už rozumně velké.

Další otázky a metody

- Durbinova dvou kroková procedura.
  1. $Y_t = \beta_1(1-\rho) + \beta_2 X_t - \beta_2 \rho X_{t-1} + \rho Y_{t-1} + u_t$.
  2. Odhad $\widehat{\rho}$ (u Y_{t-1}) $\rightarrow$ standardní transformace původního modelu a odhad parametrů.
- Theilův-Nagarův odhad ρ založený na DW statistice (pro malé vzorky):

$$\widehat{\rho} = \frac{N^2(1 - DW/2) + k^2}{N^2 - k^2}.$$

- Hildrethova-Luova procedura skenování resp. hledání (odhad ρ pro $AR(1)$ metodou hledání pro $\rho \in (-1, 1)$).
- Rozlišení čisté autokorelace a chybné specifikace!

Shrnutí přednášky na závěr