

MULTIKOLINEARITA, HETEROSKEDASTICITA A AUTOKORELOVANOST REZIDUÍ

29.10.2012 A 5.11.2012

NÁPLŇ CVIČENÍ

- Multikolinearita.
- Ověřování předpokladu homoskedasticity a možná řešení v případě jeho porušení.
- Ověřování předpokladu nezávislosti a možná řešení v případě jeho porušení.

DATA: Na dnešní hodinu budeme potřebovat data `sazba.txt`, `avgexp.txt` a `AAA.xls`, která si proto stáhněte z internetu k sobě do adresáře.

MULTIKOLINEARITA

1. Načtěte do EViews data `sazba.txt`
[File](#) → [Open](#) → [Foreign Data as Workfile](#).
2. Odhadněte lineární model pro úrokovou sazbu v závislosti na spotřebě a investicích.
[Např. otevřete si proměnné jako skupinu, pak Proc](#) → [Make Equation](#)
3. Nechte si vypsát korelační matici pro naše data, která nám naznačí, v čem je problém.
[V otevřené skupině View](#) → [Covariance analysis](#) vybrat [Correlation](#).
Jaký je v datech vztah mezi spotřebou a investicemi?
4. Změňte poslední hodnotu proměnné spotřeba z 236 na 240.
[View](#) → [Spreadsheet](#) a zapnout [Edit+/-](#)
Odhadněte model s těmito změněnými daty. Všimněte si zejména
 - významnosti regresních koeficientů,
 - co se stane, pokud z modelu vyloučíme investice.Souhlasí to s tím, co jste si o důsledcích multikolinearity říkali na přednášce?
5. Z přednášky nejspíš znáte kritérium VIF (variance inflation factor). Jak vychází pro tento model?
[View](#) → [Coefficients Diagnostics](#) → [Variance Inflation Factor](#).

HETEROSKEDASTICITA

6. Načtěte si data `avgexp.txt`. Data obsahují měsíční výdaje 100 náhodně vybraných osob na kreditní kartě. K dispozici máme věk v letech, roční příjem (v 10 000 USD), průměrný měsíční výdaj na kreditní kartě a informace o tom, zda má daná osoba vlastní dům (1-ano, 0-je v podnájmu) a zda je OSVČ (1-ano, 0-ne). Kladný (nenulový) výdaj na kartě mají pouze osoby, jejichž žádost o kreditní kartu byla kladně vyřízena. V následné analýze proto osoby, kterým kreditní karta vydána nebyla a mají tudíž na ni nulový příjem, uvažovat nebudeme.
[Nastavíme Sample na if AVGEXP>0](#).

7. Budeme modelovat výdaje (AVGEXP) v závislosti na příjmu (INCOME). Podíváme se proto nejprve na vhodný obrázek znázorňující tuto závislost.

Ve Workfile si otevřte AVGEXP a INCOME jako skupinu, View → Graph → Scatter. Aby bylo AVGEXP na y-ové ose, musí být ve View → Group members na druhém místě.

8. Odhadněte lineární model

$$E(AVGEXP) = \beta_0 + \beta_1 INCOME. \quad (M1)$$

Posuďte

- koeficient determinace,
- histogram reziduí a jejich normalitu,
- graf reziduí v závislosti na INCOME

Ve Workfile si proměnné RESID a INCOME otevřete jako skupinu a View → Graph → Scatter

Co pozorujeme na posledním obrázku? Jsme spokojeni s modelem?

9. Jaké jsou důsledky porušení předpokladu homoskedasticity (viz přednáška)?
10. Otestujte předpoklad homoskedasticity pomocí Whiteova testu:
V rovnici View → Residual diagnostics → Heteroscedasticity test zde vybrat White
Jaký je závěr testu?
11. V některých situacích se spokojíme s klasickým OLS modelem a pouze modifikujeme odhady rozptylů $\hat{\beta}_i$ tak, aby příslušné p -hodnoty t -testů byly platné. To je možné provést pomocí Whiteova odhadu:
V rovnici Estimate → záložka Options zde vybrat Coefficient covariance matrix: White
Co se změní použitím tohoto odhadu?
12. Odhadneme nyní model (M1) pomocí vážené metody nejmenších čtverců.
V rovnici Estimate → záložka Options zde musíme zadat správné váhy (weights). V našem případě uvažujeme, že rozptyl chyb je proporcionální INCOME. Ekvivalentně, $1/(INCOME)^{1/2}$ je proporcionální Inv. Std., což je v EViews preferovaná volba.
Jak se změnil odhady parametrů?
13. Ověřte, že vážený odhad nejmenších čtverců je totéž jako klasický odhad v modelu

$$E\left(\frac{AVGEXP}{\sqrt{INCOME}}\right) = \beta_0 \frac{1}{\sqrt{INCOME}} + \beta_1 \frac{INCOME}{\sqrt{INCOME}}.$$

14. Vyřešil se váženým odhadem problém s heteroskedasticitou? Posuďte
- na základě testů,
 - na základě grafu reziduí
15. Vyzkoušejte, zda se problém s heteroskedasticitou nevyřeší použitím logaritmu. Odhadněte model

$$E(\log(AVGEXP)) = \beta_0 + \beta_1 \log(INCOME)$$

a podívejte se na diagnostické testy a obrázek.

AUTOKORELOVANOST REZIDUÍ

16. Načtěte si data AAA.xls. Data obsahují čtvrtletní výnosnost korporátních AAA dluhopisů v letech 1990 až 1994.
17. Budeme chtít modelovat AAA v závislosti na TBILL. Odhadneme proto příslušný lineární model.
18. Ve výstupu si budeme všimnout zejména hodnoty Durbin-Watsonovy statistiky. Jaká její hodnota odpovídá nezávislosti? Jakou situaci naznačuje hodnota, která vychází pro naše data?
19. Naše podezření si ověříme pomocí Breusch-Godfrey testu:
[View](#)→[Residual diagnostics](#) →[Serial correlation LM test](#) zde zvolíme $p = 1$
Jaký je výsledek testu?
20. Ke stejným závěrům bychom měli dojít z ACF a PACF reziduí:
[View](#)→[Residual diagnostics](#) →[Correlogram - Q-statistics](#)
21. Jaké jsou důsledky autokorelovanosti reziduí?
22. Jednou z možností nápravy je opět pouze modifikace odhadu varianční matice $\hat{\beta}_i$. To lze provést pomocí Neweyho-Westova odhadu:
V rovnici [Estimate](#)→ záložka [Options](#) zde vybrat [Coefficient covariance matrix: Newey-West](#)
23. Nyní odhadneme model s chybami, které se řídí vhodným $AR(p)$ modelem. Začneme nejjednodušším $AR(1)$ modelem.
Do odhadu rovnice napíšeme navíc člen $AR(1)$, zadání rovnice tedy bude [aaa c tbill ar\(1\)](#) .
24. Zkontrolujte, zda se již problém s autokorelovaností vyřešil, nebo zda je potřeba zahrnout i AR členy vyšších řádů.

POZNÁMKY K ZAMYŠLENÍ

- Mělo by smysl ptát se na autokorelovanost reziduí (ve smyslu AR modelů) v případě dat [av-gexp.txt](#)?