

BAYESIÁNSKÁ ANALÝZA – CVIČENÍ 7

Toto cvičení je založeno na znalosti sedmé kapitoly z učebnice Koop (2003): *Bayesian econometrics*, případně na odpovídající kapitole podkladového učebního textu *Bayesiánská analýza*.

Co bude náplní cvičení?

- ✎ Odhad a posteriorní analýza modelů s panelovými daty.

Zadání příkladů

1. **Nákladová funkce aerolinií** K dispozici máme data z panelové studie skupiny amerických aeroline. Pokoušíme se odhadnout jednoduchý model celkových nákladů produkce těchto aerolinií:

$$\log(cost_{it}) = \beta_1 + \beta_2 \log(output_{it}) + \beta_3 \log(fuel_price_{it}) + \beta_4 load_factor_{it} + \epsilon_{it}$$

Výstup (output) je měřen v příjmech z osobomil. Nákladní faktor (load factor) je míra využití kapacit, jedná se o průměrnou míru s jakou jsou sedadla v letadlech naplněna. Komplikovanější modely nákladů by mohly zahrnovat ceny jednotlivých faktorů (materiál, kapitál) případně kvadratický člen v logaritmu výstupu, který by umožňoval proměnlivé úspory z rozsahu. Příklad i data pocházejí z Greeneovy „Econometric Analysis“ z roku 2002, kapitola 13, příklad 13.1. Data jsou obsažena přímo ve zdrojovém souboru `panel_greene.m`. Jedná se o část většího souboru dat poskytnutých panu Greeneovi profesorem Moshe Kimem, původně byla data sestavena asociací Christensen Associates of Madison z Wisconsinu. Výsledky odhadu metodou nejmenších čtverců jsou následující (v závorce jsou odhadnuté směrodatné odchylky):

$$\begin{aligned} \log(cost_{it}) &= 9.5169(0.22924) + 0.88274(0.013255) \log(output_{it}) \\ &+ 0.45398(0.020304) \log(fuel_price_{it}) \\ &- 1.62751(0.34540) load_factor_{it} \\ R^2 &= 0.9882898, s^2 = 0.015528, e'e = 1.335442193 \end{aligned}$$

V souboru `panel_greene.m` je tento model odhadován bayesovsky v rámci pooled modelu, individual effects modelu (s nehierarchickou a hierarchickou apriorní hustotou) a v rámci modelu náhodných koeficientů. Prohlédněte si vytvořený algoritmus, porovnejte ho s postupem z Koop (2003) a analyzujte a diskutujte dosažené výsledky. Pro úsporu času není řešena problematika porovnávání modelů (Chibovou metodou) a konvergenční diagnostiky, pokuste se tak v rámci některého z vybraných modelů implementovat Chibovu metodu výpočtu marginální věrohodnosti a ověřit konvergenci s využitím konvergenčních diagnostik.

Apriorní hustoty jsou spíše neinformativního charakteru a snaží se být pro všechny modely stejné pro účely případné komparace výsledků.

2. **Nákladová funkce nemocnic** Soubor `data_hospital.mat` obsahuje panelová data relevantní pro odhad nákladové funkce nemocni. Obsahuje data $N = 382$ nemocnic za $T = 5$ let. Všechny proměnné (kromě umělých proměnných) jsou logaritmovány. Proměnné jsou:

- $costs$ = logaritmus provozních nákladů nemocnice (v tisících dolarech);
- $beds$ = logaritmus počtu lůžek nemocnice;
- $inpatient$ = logaritmus počtu hospitalizovaných návštěvníků;
- $outpatient$ = logaritmus počtu ambulantně ošetřených návštěvníků;
- $case_mix$ = logaritmus indexu vážností případů (vyšší hodnoty odpovídají náročnějším případům);
- K = logaritmus kapitálové zásoby (v tisících dolarech);

- *nonprofit* = 1 pro neziskové nemocnice (= 0 jinak);
 - *forprofit* = 1 pro „ziskové“ nemocnice (= 0 jinak).
- (a) Odhadněte souhrnný regresní model s využitím těchto dat.
 - (b) Odhadněte model individuálních vlivů (s nehierarchickou a hierarchickou apriorní hustotou) s využitím těchto dat.
 - (c) Zkuste odhadnout model náhodných koeficientů s využitím těchto dat.
 - (d) Interpretujte dosažené výsledky.
 - (e) Pokuste se v rámci některého z vybraných modelů implementovat Chibovu metodu výpočtu marginální věrohodnosti a ověřte konvergenci s využitím konvergenčních diagnostik.