BAYESIÁNSKÁ ANALÝZA – CVIČENÍ 5

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

Co bude náplní cvičení?

✎ Odhad a posteriorní analýza normálního nelineárního regresního modelu s obecnou kovarianční maticí, a to s využitím reálných dat.

✎ Příklad na heteroskedasticitu využívající efektivnější kódy než původní empirická ilustrace.

✎ Příklad na autokorelaci a SUR model odpovídající empirické ilustraci z Koop (2003).

Zadání příkladů

1. Předpokládejme, že spotřeba energie v čase roste s tím, jak narůstá populace a její důchod, a naopak klesá s růstem cen energií. Využijte roční data pro Spojené státy od roku 1960 do roku 1985 a odhadněte regresní model závislosti distribuované energie na čase, ceně a důchodu domácností. Datový soubor je `EX64.dat`, výchozí m-files postupně `energya.m`, `energyb.m` a `energyc.m`. Model je definována jako

   $$Q = \beta_1 + \beta_2 \cdot YEAR + \beta_3 \cdot P + \beta_4 \cdot INC$$

   kde Q je logaritmus množství dodané energie ($10^{15} BTU$), $YEAR$ nabývá hodnoty 1 v roce 1960 a 26 v roce 1960, P je logaritmus ceny energií (za jednotku, v dolarech roku 1975) a INC je logaritmus důchodu na jednu domácnost (v dolarech roku 1975, !v původním datovém souboru není logaritmován!).

   Vyjděte z nezávislé normální-gama apriorní hustoty pro odhadované parametry. Výsledky metodou nejmenších čtverců prezentované Pindyckem a Rubinfeldem[1] (Příklad 6.4, strana 158) jsou (t-statistiky jsou uvedeny v závorkách):

   $$\begin{array}{ccccccccc} \hat{Q} & = & -36.48 & + & 0.18YEAR & - & .67P & + & .55INC \\ & & (-1.98) & & (1.76) & & (-8.99) & & (2.03) \\ & & R^2 = .970 & & & & s = .032 & & \end{array}$$

   Řešte tento příklad bez předpokladu heteroskedasticity a potom i s předpokladem známé kovarianční matice a neznámé kovarianční matice (náhodné chyby budou z t-rozdělení). V případě známé kovarianční matice testujte (porovnáním modelů) případ, neexistence heteroskedasticity. Inspiraci pro postup můžete hledat v Koop (2003), využijte rovněž i hotové skripty z předchozích cvičení (nebo se jimi alespoň inspirujte). V ramci regrese pro prvky obecné kovarianční matice Ω můžete využít i neinformativní prior $\alpha \propto 1$, v tom případě nelze počítat Bayesův faktor (nedává správné výsledky) pro testování hypotéz zahrnující α.

2. Proveďte bayesiánskou analýzu v rámci příkladu (empirické ilustrace) z Koop (2003), příklad (6.5.3). Tento regresní model se snaží vysvětlit procento výher baseballového týmu New York Yankees v každém roce (v období 1903 až 1999) v závislosti na statistikách charakterizujících ofenzivní a defenzivní výkonnost tohoto týmu. Datový soubor je `yankees.txt`, výchozí m-files pak `chapter06c_noauto.m`, `chapter06c.m` a zcela obecný `chapter06c_expand.m`. Vysvětlovanou proměnnou je winning percentage (PCT) v roce t, tedy $PCT = wins/(wins+losses)$. Vysvětlujícími veličinami jsou on-base percentage (OBP), team slugging average (SLG) a team earned run average (ERA) vše za rok t. Bližší vysvětlení těchto veličin je obsaženo v souboru `baseball statistics.pdf`.

   Předpokládáme v modelu autokorelaci náhodných složek. Vzhledem k tomu, že subjektivní určení apriorních hustot je obtížné, použijte neinformativní apriorní hustotu pro β a zvolte tedy $\underline{V}^{-1} = 0_{k \times k}$. Stejně tak užijte neinformativní apriorní hustotu pro h nastavením $\underline{\nu} = 0$. Totuo volbou se stává hodnota $\underline{\beta}$ a $\underline{s}^{-2}$ irelevantní.

[1] Pindyck, Rubinfeld (1997) Econometric models and economic forecast. 4. vydání

Spočítejte Savage-Dickey density ratio porovnávající modely s $\rho_j = 0$ pro $j = 1, \ldots, p$ vzhledeme k neomezenému modelu. To vyžaduje formulaci informativní apriorní hustoty pro ρ, vhodnou volbou je zde $\underline{\rho} = 0$ a $V_\rho = c * I_p$. V rámci analýzy citlivosti apriorních hustot pak volíme různá c.

Koop ukazuje, že Bayesův faktor a HPDI nepodopořili hypotézu o autokorelaci vyššího řádu než je jedna. Můžete tudíž rovnou zvolit $p = 1$, což bude znamenat, že stacionární oblast je pro $|\rho_1| < 0$. Zkuste však odhadnout model i s předpokladem autokorelace vyššího řádu a samozřejmě neopomeňte odhad bez předpokladu autokorelace (pro srovnání výsledků).

3. Proveďte bayesiánskou analýzu v rámci příkladu (empirické ilustrace) z Koop (2003), příklad (6.6.4). Tento regresní model se snaží vysvětlit procento výher baseballového týmu New York Yankees a Red Sox v každém roce (v období 1903 až 1999) v závislosti na statistikách charakterizujících ofenzivní a defenzivní výkonnost tohoto týmu. Datový soubor je `yankees.txt` a `yankees.txt`, výchozí m-file pak `chapter06d.m`. Vysvětlovanou proměnnou je winning percentage (PCT) v roce t, tedy $PCT = wins/(wins + losses)$. Vysvětlujícími veličinami jsou on-base percentage (OBP), team slugging average (SLG) a team earned run average (ERA) vše za rok t. Bližší vysvětlení těchto veličin je obsaženo v souboru `baseball statistics.pdf`. Model bereme jako model zdánlivě nesouvisejících regresí.

   Vzhledem k tomu, že subjektivní určení apriorních hustot je obtížné, použijme neinformativní apriorní hustotu pro β a zvolme tedy $\underline{\beta} = 0_k$, $\underline{V}^{-1} = 4I_k$ a $\underline{H}^{-1} = 0_{2\times 2}$. Na základě výsledků korelace náhodných složek mezi oběma rovnicemi pak můžeme rozhodnout, jestli bylo vůbec nutné pracovat se SUR modelem.