

# Bayesiánská analýza

## VII. Lineární regresní model s panelovými daty

# Obsah tématu

- 1 Souhrnný model
- 2 Model individuálních vlivů
- 3 Model náhodných koeficientů
- 4 Chibova metoda porovnání modelů
- 5 Empirická ilustrace
- 6 Model stochastických hranic

- Panelová data — časová i prostorová dimenze.
- Souhrnný model (pooled model), model individuálních vlivů (individual effects model), model náhodných koeficientů (random coefficients model).
- Chibova metoda marginální věrohodnosti.
- Model stochastických mezí (stochastic frontier model).

# Značení

- $y_{it}$  a  $\epsilon_{it}$ :  $t$ -té pozorování (pro  $t = 1, \dots, T$ ) pro  $i$ -tého jednotlivce ( $i = 1, \dots, N$ ).
- $X_i = [\iota_T \tilde{X}_i]$ .
- $TN$ -rozměrné vektory:

$$y = \begin{bmatrix} y_1 \\ \cdot \\ \cdot \\ y_N \end{bmatrix} \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \cdot \\ \cdot \\ \epsilon_N \end{bmatrix}$$

- Matice  $TN \times K$ :

$$X = \begin{bmatrix} X_1 \\ \cdot \\ \cdot \\ X_N \end{bmatrix}$$

# Obsah tématu

- 1 Souhrnný model
- 2 Model individuálních vlivů
- 3 Model náhodných koeficientů
- 4 Chibova metoda porovnání modelů
- 5 Empirická ilustrace
- 6 Model stochastických hranic

# Princip

- Stejný regresní vztah pro všechny jednotlivce:

$$y_i = X_i\beta + \epsilon_i.$$

- Předpoklady:

- 1  $\epsilon_i \sim N(0_T, h^{-1}I_T)$ .
- 2  $\epsilon_i$  a  $\epsilon_j$  jsou nezávislé pro  $i \neq j$ .
- 3 Všechny prvky  $X_i$  jsou pevná čísla (tj. nenáhodné veličiny) nebo v případě, že jsou náhodnými veličinami, jsou nezávislé na všech prvcích  $\epsilon_j$  a mají hustotu pravděpodobnosti  $p(X_i|\lambda)$  kde  $\lambda$  je vektor parametrů, který neobsahuje  $\beta$  ani  $h$ .

- $\epsilon_{it}$  a  $\epsilon_{is}$  jsou vzájemně nezávislé pro  $t \neq s \rightarrow$  zobecnění  $\epsilon_i$  má kovarianční matici  $\Omega$  (řešení jako SUR model).

# Bayesovská analýza

- Věrohodnostní funkce:

$$p(y|\beta, h) = \prod_{i=1}^N \frac{h^{\frac{T}{2}}}{(2\pi)^{\frac{T}{2}}} \left\{ \exp \left[ -\frac{h}{2} (y_i - X_i\beta)' (y_i - X_i\beta) \right] \right\}.$$

- Přepsání do podoby:

$$p(y|\beta, h) = \frac{h^{\frac{NT}{2}}}{(2\pi)^{\frac{NT}{2}}} \left\{ \exp \left[ -\frac{h}{2} (y - X\beta)' (y - X\beta) \right] \right\}.$$

- Např. nezávislá apriorní normální-gama hustota  $\beta \sim N(\underline{\beta}, \underline{V})$  a  $h \sim G(\underline{s}^{-2}, \underline{\nu}) \rightarrow$  Gibbsův vzorkovač.

# Obsah tématu

- 1 Souhrnný model
- 2 Model individuálních vlivů**
- 3 Model náhodných koeficientů
- 4 Chibova metoda porovnání modelů
- 5 Empirická ilustrace
- 6 Model stochastických hranic



# Princip

- Příklad z oblasti marketingu:  $y_{it}$  je prodej nápoje značky  $i$  v čase  $t$ .
- Prodeje závisí např. na ceně + existují i nezachytitelné kvantify (věrnost značce):

$$y_{it} = \alpha_i + \beta x_{it} + \epsilon_{it}$$

- $\alpha_i$  = individuální vliv (*individual effect*).

# Věrohodnostní funkce

- Regresní model:

$$y_i = \alpha_i \iota_T + \tilde{X}_i \tilde{\beta} + \epsilon_i$$

- Z předpokladů:

$$p(y|\alpha, \tilde{\beta}, h) = \prod_{i=1}^N \frac{h^{\frac{T}{2}}}{(2\pi)^{\frac{T}{2}}} \left\{ \exp \left[ -\frac{h}{2} (y_i - \alpha_i \iota_T - \tilde{X}_i \tilde{\beta})' (y_i - \alpha_i \iota_T - \tilde{X}_i \tilde{\beta}) \right] \right\}.$$

- $\alpha = (\alpha_1, \dots, \alpha_N)'$ .

# Nehierarchická apriorní hustota

- Model:

$$y = X^* \beta^* + \epsilon.$$

- $X^*$ : matice rozměru  $TN \times (N + k - 1)$ .

$$X^* = \begin{bmatrix} \iota_T & 0_T & \cdot & \cdot & 0_T & \tilde{X}_1 \\ 0_T & \iota_T & \cdot & \cdot & \cdot & \tilde{X}_2 \\ \cdot & 0_T & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 0_T & \cdot \\ 0_T & \cdot & \cdot & \cdot & \iota_T & \tilde{X}_N \end{bmatrix} \quad \beta^* = \begin{bmatrix} \alpha_1 \\ \cdot \\ \cdot \\ \alpha_N \\ \tilde{\beta} \end{bmatrix}.$$

- Klasická ekonometrie = fixed effects model ( $X^*$  s umělými proměnnými).
- Např. nezávislá normální-gama apriorní hustota:

$$\beta^* \sim N(\underline{\beta}, \underline{V}),$$

$$h \sim G(\underline{s}^{-2}, \underline{\nu}).$$

# Hierarchická apriorní hustota

- Velká dimenze vektoru parametrů  $\rightarrow$  hierarchická apriorní hustota.
- $N + k$  parametrů  $\rightarrow$  problém pokud  $T$  relativně malé vzhledem k  $N$ .
- Obvyklý předpoklad pro  $i = 1, \dots, N$ :

$$\alpha_i \sim N(\mu_\alpha, V_\alpha)$$

- $\alpha_i$  a  $\alpha_j$  vzájemně nezávislé pro  $i \neq j$ .
- Hierarchická struktura: pokud  $\mu_\alpha$  a  $V_\alpha$  neznámé parametry.

# Konkretizace apriorních hustot

- Předpokládáme nezávislost  $\mu_\alpha$  a  $V_\alpha$ :

$$\begin{aligned}\tilde{\mu}_\alpha &\sim N(\underline{\mu}_\alpha, \underline{\sigma}_\alpha^2), \\ V_\alpha^{-1} &\sim G(\underline{V}_\alpha^{-1}, \underline{\nu}_\alpha).\end{aligned}$$

- Zbylé parametry s nehierarchickou apriorní hustotou (nezávislé normální-gama rozdělení):

$$\begin{aligned}\tilde{\beta} &\sim N(\underline{\beta}, \underline{V}_\beta), \\ h &\sim G(\underline{s}^{-2}, \underline{\nu}).\end{aligned}$$

- Klasická ekonometrie: tzv. random effects model.

# Posterioční analýza při nehierarchickém prioru

- LRM s nezávislou normální-gama apriorní hustotou (Gibbs):

$$\beta^* | y, h \sim N(\bar{\beta}^*, \bar{V}),$$

$$h | y, \beta^* \sim G(\bar{s}^{-2}, \bar{\nu}),$$

$$\bar{V} = (\underline{V}^{-1} + hX^{*'}X^*)^{-1},$$

$$\bar{\beta}^* = \bar{V}(\underline{V}^{-1}\underline{\beta}^* + hX^{*'}y),$$

$$\bar{\nu} = TN + \underline{\nu},$$

$$\bar{s}^2 = \frac{\sum_{i=1}^N (y_i - \alpha_{i,T} - \tilde{X}_i \tilde{\beta})'(y_i - \alpha_{i,T} - \tilde{X}_i \tilde{\beta}) + \underline{\nu} \underline{s}^2}{\bar{\nu}}.$$

- Standardní analýza konvergence, predikční analýza a porovnání modelů.
- Numerický problém, pokud  $N$  příliš velké ( $\bar{V}$  matice rozměru  $(N + k - 1) \times (N + k - 1)$  + inverze)  $\rightarrow$  teorém o inverzi dělené matici (snížení dimenze invertovaných matic).

# Posterioční analýza při hierarchickém prioru

- Odvození = násobení věrohodnostní funkce a apriorních hustot a analýzu výsledného výrazu pro  $\tilde{\beta}$ ,  $h$ ,  $\alpha$ ,  $\mu_\alpha$  a  $V_\alpha \rightarrow$  jádrové (podmíněné) hustoty  $\rightarrow$  Gibbsův vzorkovač.
- Posterioční rozdělení pro  $\tilde{\beta}$  a  $h$  podmíněné veličinou  $\alpha$  analogické jako LRM s nezávislou normální-gama apriorní hustotou.
- $p(\tilde{\beta}|y, h, \alpha, \mu_\alpha, V_\alpha)$  a  $p(h|y, \tilde{\beta}, \alpha, \mu_\alpha, V_\alpha)$  nezávisí na  $\mu_\alpha$  a  $V_\alpha \rightarrow$  ekvivalence vzhledem k  $p(\tilde{\beta}|y, h, \alpha)$  a  $p(h|y, \tilde{\beta}, \alpha)$ .

# Posteriorní analýza pro $\beta$ a $h$

$$\tilde{\beta} | y, h, \alpha, \mu_\alpha, V_\alpha \sim N(\bar{\beta}, \bar{V}_\beta),$$

$$h | y, \tilde{\beta}, \alpha, \mu_\alpha, V_\alpha \sim G(\bar{s}^{-2}, \bar{\nu}),$$

$$\bar{V}_\beta = \left( \underline{V}_\beta^{-1} + h \sum_{i=1}^N \tilde{X}_i' \tilde{X}_i \right)^{-1},$$

$$\bar{\beta} = \bar{V}_\beta \left( \underline{V}_\beta^{-1} \underline{\beta} + h \sum_{i=1}^N \tilde{X}_i' [y_i - \alpha_i t_T] \right),$$

$$\bar{\nu} = TN + \underline{\nu},$$

$$\bar{s}^2 = \frac{\sum_{i=1}^N (y_i - \alpha_i t_T - \tilde{X}_i \tilde{\beta})' (y_i - \alpha_i t_T - \tilde{X}_i \tilde{\beta}) + \underline{\nu} s^2}{\bar{\nu}}.$$



# Posteriorní analýza pro $\alpha$

- Podmíněná posteriorní hustota pro  $\alpha_i$  je nezávislá na  $\alpha_j$  pro  $i \neq j$ :

$$\alpha_i | y, \tilde{\beta}, h, \mu_\alpha, V_\alpha \sim N(\bar{\alpha}_i, \bar{V}_i),$$

$$\bar{V}_i = \frac{V_\alpha h^{-1}}{TV_\alpha + h^{-1}},$$

$$\bar{\alpha}_i = \frac{V_\alpha (y_i - \tilde{X}_i \tilde{\beta})' \iota_T + h^{-1} \mu_\alpha}{TV_\alpha + h^{-1}}.$$

# Posterioční analýza pro hierarchické parametry

- Podmíněné hustoty pro hierarchické parametry  $\mu_\alpha$  a  $V_\alpha$ :

$$\mu_\alpha | y, \tilde{\beta}, h, \alpha, V_\alpha \sim N(\bar{\mu}_\alpha, \bar{\sigma}_\alpha^2),$$

$$V_\alpha^{-1} | y, \tilde{\beta}, h, \alpha, \mu_\alpha, V_\alpha \sim G(\bar{V}_\alpha^{-1}, \bar{\nu}_\alpha),$$

$$\bar{\sigma}_\alpha^2 = \frac{V_\alpha \underline{\sigma}_\alpha^2}{V_\alpha + N \underline{\sigma}_\alpha^2}$$

$$\bar{\mu}_\alpha = \frac{V_\alpha \underline{\mu}_\alpha + \underline{\sigma}_\alpha^2 \sum_{i=1}^N \alpha_i}{V_\alpha + N \underline{\sigma}_\alpha^2},$$

$$\bar{\nu}_\alpha = \underline{\nu}_\alpha + N,$$

$$\bar{V}_\alpha = \frac{\sum_{i=1}^N (\alpha_i - \mu_\alpha)^2 + V_\alpha \underline{\nu}_\alpha}{\bar{\nu}_\alpha}.$$

# Obsah tématu

- 1 Souhrnný model
- 2 Model individuálních vlivů
- 3 Model náhodných koeficientů**
- 4 Chibova metoda porovnání modelů
- 5 Empirická ilustrace
- 6 Model stochastických hranic

# Princip

- Uvolněný předpoklad o společném sklonu regresní (nad)roviny:

$$y_i = X_i\beta_i + \epsilon_i.$$

- Problém odhadu pro malé  $T$  (vzhledem k  $N$ )  $\rightarrow$  hierarchická konstrukce apriorní hustoty.
- Motivace: příklad z marketingu  $\rightarrow$  navíc odlišný marginální efekt změny ceny na prodeje (věrnost znače skrze mezní vlivy ceny).

# Věrohodnostní funkce

- Z předpokladů o chybovém členu a tvaru regresního modelu:

$$p(y|\beta, h) = \prod_{i=1}^N \frac{h^{\frac{T}{2}}}{(2\pi)^{\frac{T}{2}}} \left\{ \exp \left[ -\frac{h}{2} (y_i - X_i \beta_i)' (y_i - X_i \beta_i) \right] \right\}.$$

- $\beta = (\beta_1', \dots, \beta_N)'$  označuje všechny regresní koeficienty.

# Hierarchická apriorní hustota

- $\beta_i$  pro  $i = 1, \dots, N$  jsou nezávislé výběry z normálního rozdělení:

$$\beta_i \sim N(\mu_\beta, V_\beta).$$

- Druhá fáze hierarchické apriorní hustoty:

$$\mu_\beta \sim N(\underline{\mu}_\beta, \underline{\Sigma}_\beta),$$

$$V_\beta^{-1} \sim W(\underline{\nu}_\beta, \underline{V}_\beta^{-1}).$$

- Wishartovo rozdělení:  $E(V_\beta^{-1}) = \underline{\nu}_\beta \underline{V}_\beta^{-1} +$  neinformativní varianta pro  $\underline{\nu}_\beta = 0$ .
- Přesnost chyby:

$$h \sim G(\underline{s}^{-2}, \underline{\nu}).$$

Posterioční analýza při nehierarchickém prioru –  $\beta$ 

- Gibbsův vzorkovač (standardní odvození):
- Vzájemně nezávislé podmíněné posterioční hustoty parametrů  $\beta_i$ , pro  $i = 1, \dots, N$ :

$$\beta_i | y, h, \mu_\beta, V_\beta \sim N(\bar{\beta}_i, \bar{V}_i),$$

$$\bar{V}_i = (hX_i'X_i + V_\beta^{-1})^{-1},$$

$$\bar{\beta}_i = \bar{V}_i(hX_i'y_i + V_\beta^{-1}\mu_\beta).$$

# Posteriorní analýza při nehierarchickém prioru – $\mu_\beta$ a $V_\beta$

- Relevantní hustoty pro  $\mu_\beta$  a  $V_\beta$ :

$$\mu_\beta | y, \beta, h, V_\beta \sim N(\bar{\mu}_\beta, \bar{\Sigma}_\beta)$$

$$V_\beta^{-1} | y, \beta, h, \mu_\beta \sim W(\bar{\nu}_\beta, [\bar{\nu}_\beta \bar{V}_\beta]^{-1}),$$

$$\bar{\Sigma}_\beta = \left( N V_\beta^{-1} + \underline{\Sigma}_\beta^{-1} \right)^{-1},$$

$$\bar{\mu}_\beta = \bar{\Sigma}_\beta \left( V_\beta^{-1} \sum_{i=1}^N \beta_i + \underline{\Sigma}_\beta^{-1} \underline{\mu}_\beta \right),$$

$$\bar{\nu}_\beta = N + \underline{\nu}_\beta,$$

$$\bar{V}_\beta = \sum_{i=1}^N (\beta_i - \mu_\beta)(\beta_i - \mu_\beta)' + \underline{V}_\beta.$$

- Výraz  $\sum_{i=1}^N \beta_i$ :  $k$ -rozměrný vektor obsahující součty prvků  $\beta_i$ .



# Posterioční analýza při nehierarchickém prioru – $h$

- Podmíněná posterioční hustota pro přesnost chyby:

$$h|y, \beta, \mu_\beta, V_\beta \sim G(\bar{s}^{-2}, \bar{\nu}),$$

$$\bar{\nu} = TN + \underline{\nu},$$

$$\bar{s}^2 = \frac{\sum_{i=1}^N (y_i - X_i \beta_i)' (y_i - X_i \beta_i) + \underline{\nu} s^2}{\bar{\nu}}.$$

- Gibbsův vzorkovač: výběry z normálního, gama a Wishartova rozdělení.
- Predikční analýza a konvergenční testy proveditelné standardním způsobem.

# Obsah tématu

- 1 Souhrnný model
- 2 Model individuálních vlivů
- 3 Model náhodných koeficientů
- 4 Chibova metoda porovnání modelů**
- 5 Empirická ilustrace
- 6 Model stochastických hranic

# Úvod

- Problém Savage-Dickeyho poměru hustot např. pro test  $V_\alpha = 0$  (volba  $V_\alpha^{-1} = \infty \rightarrow$  volba velkého čísla a s tím problém hrubé aproximace).
- Problém vysoké dimenze parametrů (metoda Gelfanda a Deye nepřesná).
- Řešení v podobě Chibovy metody.

## Značení a princip

- Obecné značení:  $\theta$ ,  $p(y|\theta)$ ,  $p(\theta)$  a  $p(\theta|y)$ .
- Jednoduché zjištění:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}.$$

- $p(y)$  nezávisí na parametrech  $\theta \Rightarrow$  vyhodnocení pravé strany rovnice v jakémkoliv bodě  $\theta^*$ .
- Výsledkem bude marginální věrohodnost  $\rightarrow$  pro jakýkoliv bod  $\theta^*$ :

$$p(y) = \frac{p(y|\theta^*)p(\theta^*)}{p(\theta^*|y)}.$$

- *Základní identita marginální věrohodnosti (basic marginal likelihood identity);  $p(\theta^*)$  je zkrácený zápis pro  $p(\theta = \theta^*)$ .*

# Problém

- Neznáme všechny plné hustoty.
- Jak vyhodnotit posteriorní hustotu v bodě (tj. spočítat  $p(\theta^* | y)$ )?
- Zde pro případ více dimenzionálního vektoru parametrů a Gibbsova vzorkovače.

# Problém dimenzionality

- Modelová struktura s nízko dimenzionálním vektorem  $\theta$  a vysoce dimenzionálním vektorem  $z$  (často *latentní data*).
- *Gibbsův vzorkovač s rozšířenými daty*: sekvenční výběry z  $p(\theta|y, z)$  a  $p(z|y, \theta)$ ,
- Zde  $z$  jako vektor individuálních vlivů (tedy  $z = \alpha$ ) nebo vektor náhodných koeficientů ( $z = \beta$ ).
- Vyintegrování vysoce dimenzionálního vektoru  $z$  a práce s nízko dimenzionálním vektorem  $\theta$ .

# Obecný postup

- Zákony pravděpodobnosti implikují:

$$p(\theta^*|y) = \int p(\theta^*|y, z)p(z|y)dz.$$

- Výpočet  $p(\theta^*|y, z^{(s)})$  pro každý výběr  $s = 1, \dots, S$  a výsledek se zprůměruje.
- Pokud  $z^{(s)}$  pro  $s = 1, \dots, S$  jsou výběry z Gibbsova vzorkovače, potom

$$\widehat{p(\theta^*|y)} = \frac{1}{S} \sum_{s=1}^S p(\theta^*|y, z^{(s)})$$

konverguje k  $p(\theta^*|y)$  pro  $S$  jdoucí k nekonečnu.

- Výpočet marginální věrohodnosti pokud jsme schopni spočítat  $p(y|\theta^*)$ ,  $p(\theta^*)$  a  $p(\theta^*|y, z)$  (plné hustoty).
- Použití jakékoliv hodnoty  $\theta^*$  (nejlépe např.  $\theta^*$  jako posteriorní střední hodnota).

# Problém modelu panelových dat

- Neznáme  $p(\theta^*|y, z)$ .
- Snadné řešení  $\rightarrow$  předpokládejme rozdělení  $\theta$  do dvou bloků,  $\theta_1$  a  $\theta_2$  a máme Gibbsův vzorkovač se sekvenčními výběry z  $p(\theta_1|y, z, \theta_2)$ ,  $p(\theta_2|y, z, \theta_1)$  a  $p(z|y, \theta_1, \theta_2)$ .
- Jsme schopni získat  $\theta_1^{(s)}$ ,  $\theta_2^{(s)}$  a  $z^{(s)}$  pro  $s = 1, \dots, S$ .
- K využití Chibovy metody musíme spočítat  $p(\theta_1^*, \theta_2^*|y)$ , kde  $\theta_1^*$  a  $\theta_2^*$  jsou jakékoliv prvky.



# Řešení problému I

- Pravidla pravděpodobnosti říkají:

$$p(\theta_1^*, \theta_2^* | y) = p(\theta_1^* | y) p(\theta_2^* | y, \theta_1^*),$$

$$p(\theta_1^* | y) = \iint p(\theta_1^* | y, \theta_2, z) p(\theta_2, z | y) d\theta_2 dz,$$

$$p(\theta_2^* | y, \theta_1^*) = \int p(\theta_2^* | y, \theta_1^*, z) p(z | y, \theta_1^*) dz.$$

- Odhady pro  $p(\theta^* | y)$  a  $p(\theta_1^* | y)$ .

$$\widehat{p(\theta_1^* | y)} = \frac{1}{S} \sum_{s=1}^S p(\theta_1^* | y, z^{(s)}, \theta_2^{(s)})$$

konverguje k  $p(\theta_1^* | y)$  pro  $S$  jdoucí k nekonečnu.

## Řešení problému II

- Jak spočítat  $p(\theta_2^*|y, \theta_1^*)$ ?
- Samostatně s využitím Gibbsova vzorkovače.
- Druhý Gibbsův vzorkovač pro sekvenční výběry z  $p(\theta_2|y, z, \theta_1^*)$  a  $p(z|y, \theta_1^*, \theta_2^*)$ .
- Výstup k získání odhadu  $p(\theta_2^*|y, \theta_1^*)$ .
- Vzorky  $\theta_2^{(s^*)}$  a  $z^{(s^*)}$  pro  $s^* = 1, \dots, S^* \rightarrow$

$$p(\widehat{\theta_2^*|y, \theta_1^*}) = \frac{1}{S^*} \sum_{s^*=1}^{S^*} p(\theta_2^*|y, z^{(s^*)}, \theta_1^{(s^*)})$$

konverguje k  $p(\theta_2^*|y, \theta_1^*)$  pro  $S^*$  jdoucí k nekonečnu.

- Odhady  $\widehat{p(\theta_1^*|y)}$  a  $\widehat{p(\theta_2^*|y, \theta_1^*)} \rightarrow p(\theta_1^*, \theta_2^*|y)$ .

# Zobecnění

- Rozdělení vektoru  $\theta$  do  $B$  bloků (tj.  $\theta = (\theta'_1, \dots, \theta'_B)'$ ).
- Platí

$$p(\theta_1^*, \theta_2^*, \dots, \theta_B^* | y) = p(\theta_1^* | y) p(\theta_2^* | y, \theta_1^*) \dots p(\theta_B^* | y, \theta_1^*, \dots, \theta_{B-1}^*)$$

- Původní Gibbsův vzorkovač k výpočtu  $p(\theta_1^* | y)$ , druhý Gibbsův vzorkovač k výpočtu  $p(\theta_2^* | y, \theta_1^*)$  a analogicky až  $B$ -tý Gibbsův vzorkovač pro výpočet  $p(\theta_B^* | y, \theta_1^*, \dots, \theta_{B-1}^*)$ .
- Chibova metoda výpočetně náročnější  $\times$  všech  $B$  Gibbsových vzorkovačů s totožnou strukturou.

# Ukázka pro modely panelových dat

- Model individuálních vlivů s hierarchickou apriorní hustotou:  $B = 4$  (tj.  $\theta_1 = \tilde{\beta}$ ,  $\theta_2 = h$ ,  $\theta_3 = \mu_\alpha$  a  $\theta_4 = V_\alpha^{-1}$ ) a  $z = \alpha$ .
- Model náhodných koeficientů:  $B = 3$  (tj.  $\theta_1 = h$ ,  $\theta_2 = \mu_\beta$  a  $\theta_3 = V_\beta^{-1}$ ) a  $z = \beta$ .
- Známá podoba podmíněných hustot  $\rightarrow$  výpočet  $p(\theta^*|y)$ .
- Potřeba vyhodnocení  $p(\theta^*)$  a  $p(y|\theta^*)$ .
- Apriorní hustoty přímo spočitatelné  $\times$  vyhodnocení věrohodnostní funkce o něco složitější (máme  $p(y|\theta, z)$  a tedy ne  $p(y|\theta)$ ).

# Úprava věrohodnostní funkce

- Analytická integrace pro přechod z  $p(y|\theta, z)$  na  $p(\theta|y)$  na základě vlastností vícerozměrného normálního rozdělení.
- Model individuálních vlivů s hierarchickou apriorní hustotou:

$$p(y|\tilde{\beta}, h, \mu_\alpha, V_\alpha^{-1}) = \prod_{i=1}^N p(y_i|\tilde{\beta}, h, \mu_\alpha, V_\alpha^{-1}), \text{ kde}$$

$$p(y_i|\tilde{\beta}, h, \mu_\alpha, V_\alpha^{-1}) \sim N(\mu_i, V_i),$$

$$\mu_i = \mu_\alpha \iota_T + \tilde{X}_i \tilde{\beta}$$

$$V_i = V_\alpha \iota \iota'_T + h^{-1} I_T.$$

- Model náhodných koeficientů:

$$p(y|h, \mu_\beta, V_\beta^{-1}) = \prod_{i=1}^N p(y_i|h, \mu_\beta, V_\beta^{-1}), \text{ kde}$$

$$p(y_i|h, \mu_\beta, V_\beta^{-1}) \sim N(\mu_i, V_i),$$

$$\mu_i = X_i \mu_\beta$$

$$V_i = X_i V_\beta X_i' + h^{-1} I_T.$$

# Obsah tématu

- 1 Souhrnný model
- 2 Model individuálních vlivů
- 3 Model náhodných koeficientů
- 4 Chibova metoda porovnání modelů
- 5 Empirická ilustrace**
- 6 Model stochastických hranic

# Úvod

- Bude časem doplněno. . .

# Obsah tématu

- 1 Souhrnný model
- 2 Model individuálních vlivů
- 3 Model náhodných koeficientů
- 4 Chibova metoda porovnání modelů
- 5 Empirická ilustrace
- 6 Model stochastických hranic**



# Princip

- Ekonomická teorie: stochastic frontier model.
- Model individuálních vlivů s odlišnou hierarchickou apriorní hustotou.
- Analýza efektivity produkce firem či jiných agentů.
- Ekonomická teorie → ekonometrický model.

# Úvod do modelu

- Ekonomický model produkce: výstup firmy  $i$  v čase  $t$ ,  $Y_{it}$ , je vyráběn s využitím vektoru vstupů,  $X_{it}^*$ , kde  $i = 1, \dots, N$  a  $t = 1, \dots, T$ .
- Firmy využívají běžnou, nejlepší možnou dostupnou technologii závislou na neznámých parametrech,  $\beta$ :

$$Y_{it} = f(X_{it}^*; \beta).$$

- *Hranice výrobních možností (production frontier).*
- Odchylka skutečného výstupu od maximálně dosažitelného = měřítko neefektivity:

$$Y_{it} = f(X_{it}^*; \beta) \tau_i.$$

- $0 < \tau_i \leq 1$ : míra efektivity specifická pro jednotlivé firmy.
- $\tau_i = 1$ : firma  $i$  plně efektivní.
- Předpoklad: každá firma úroveň efektivity neměnnou v čase (lze uvolnit).

# Úvod do modelu (pokračování)

- Chybový člen  $\zeta_{it}$ :

$$Y_{it} = f(X_{it}^*; \beta) \tau_i \zeta_{it}.$$

- Zahrnutí chybového členu (chyb měření) = stochastická hranice.
- Pokud hranice výrobních možností,  $f()$ , v log-lineární podobě (např. Cobb-Douglasova produkční funkce nebo produkční funkce TRANSLOG)  $\rightarrow$  logaritmování:

$$y_{it} = X_{it}\beta + \epsilon_{it} - z_i,$$

- $\beta = (\beta_1, \dots, \beta_k)'$ ,  $y_{it} = \ln(Y_{it})$ ,  $\epsilon_{it} = \ln(\zeta_{it})$ ,  $z_i = -\ln(\tau_i)$  a  $X_{it}$  je protějškem  $X_{it}^*$  (vstupy transformovány logaritmy).
- Veličina  $z_i$  = neefektivita a díky  $0 < \tau_i \leq 1$  nezáporná náhodná veličina.
- $X_{it}$  s úrovnovou konstantou a koeficientem  $\beta_1$ .

# Úvod do modelu (dokončení)

- Podoba modelu individuálních vlivů: výraz  $\beta_1 - z_i$  odpovídá  $\alpha_i$ .
- Ekonomická teorie dává vodítko k výběru hierarchické apriorní hustoty.
- Pro nelog-lineární produkční funkci (např. CES produkční funkce) potřeba kombinace s technikami M-H algoritmu.
- Setřídění proměnných a matic:

$$y_i = X_i\beta + \epsilon_i - z_i\ell_T.$$

# Věrohodnostní funkce

- Předpoklad nezávislosti  $z_i$  a  $\epsilon_j$  pro všechna  $i$  a  $j$ :

$$p(y|\beta, h, z) = \prod_{i=1}^N \frac{h^{\frac{T}{2}}}{(2\pi)^{\frac{T}{2}}} \left\{ \exp \left[ -\frac{h}{2} (y_i - X_i\beta + z_i\iota_T)' (y_i - X_i\beta + z_i\iota_T) \right] \right\}.$$

- $z = (z_1, \dots, z_N)'$ .
- $z$ : vektor neznámých parametrů.
- „klasická“ ekonometrie: věrohodnostní funkce definována jako  $p(y|\beta, h, \theta) = \int p(y|\beta, h, z)p(z|\theta)dz$ , kde  $p(z|\theta)$  odpovídá předpokladu o rozdělení neefektivity (závisí na vektoru neznámých parametrů  $\theta$ ).
- Matematicky ekvivalentní postup bayesovskému přístupu využívajícímu  $p(z|\theta)$  jako hierarchickou apriorní hustotu  $\Rightarrow$  volba označení „věrohodnostní funkce“ a „hierarchická apriorní hustota“ je čistě sémantickou záležitostí.

# Hierarchická apriorní hustota

- Koeficienty hranice výrobních možností a přesnost chyby:

$$\beta \sim N(\underline{\beta}, \underline{V}),$$

$$h \sim G(\underline{s}^{-2}, \underline{\nu}).$$

- Míra neefektivity: hierarchická apriorní hustota.
- $z_i > 0 \rightarrow$  ne hierarchická hustota odpovídající normální hustotě pravděpodobnosti.
- Obvykle omezené normálnímu rozdělení nebo rozdělení z rodiny gama rozdělení, zde exponenciální rozdělení ( $z_i$  a  $z_j$  a priori nezávislé pro  $i \neq j$ ):

$$z_i \sim G(\mu_z, 2).$$

- $z_i > 0 \Rightarrow \mu_z > 0$  (snadnější práce s  $\mu_z^{-1}$  než přímo s  $\mu_z$ ):

$$\mu_z^{-1} \sim G(\underline{\mu}_z^{-1}, \underline{\nu}_z).$$

## Apriorní hustota pro hyperparametry

- Apriorní hyperparametry pro  $\underline{\mu}_z^{-1}$  a  $\underline{\nu}_z$  na základě předpokladů o rozdělení efektivity.
- Např. nechtě  $\tau^*$  označuje apriorní medián tohoto rozdělení.
- Pokud očekáváme spíše efektivní firmy v našem vzorku: hodnota  $\tau^*$  vysoká (např. 0.95), jinak nižší.
- Literatura:  $\underline{\nu}_z = 2$  implikuje relativně neinformativní prior.
- Z nastavení  $\underline{\mu}_z = -\ln(\tau^*)$ : medián apriorního rozdělení efektivity  $\tau^*$ .
- Strategie stanovení priorů prostřednictvím snadno interpretovatelných hyperparametrů v kontextu výchozí ekonomické teorie (např.  $\tau^*$ ) + následná zpětná transformace pro nalezení hyperparametrů použitých v modelu (např.  $\underline{\mu}_z$  a  $\underline{\nu}_z$ ).
- Omezení z ekonomické teorie: např. restrikce, že hranice výrobních možností je monotónně rostoucí ve vstupech nebo nákladová funkce je konkávní nebo možnost technologického úpadku  $\Rightarrow$  omezení parametrů ve tvaru nerovností.

# Bayesovský výpočet

- Gibbsův vzorkovač: podmíněné hustoty jako v modelu individuálních vlivů s hierarchickou apriorní hustotou (s výjimkou  $z$  a  $\mu_z$ ).
- Parametry hranice výrobních možností:

$$\beta|y, h, z, \mu_z \sim N(\bar{\beta}, \bar{V}),$$

$$\bar{V} = \left( \underline{V}^{-1} h \sum_{i=1}^N X_i' X_i \right)^{-1},$$

$$\bar{\beta} = \bar{V} \left( \underline{V}^{-1} \underline{\beta} + h \sum_{i=1}^N X_i' [y_i + z_i t_T] \right).$$



## Bayesovský výpočet (pokračování)

- Standardní výsledky pro přesnost chyby:

$$h|y, \beta, z, \mu_z \sim G(\bar{s}^{-2}, \bar{\nu}),$$

$$\bar{\nu} = TN + \underline{\nu},$$

$$\bar{s}^{-2} = \frac{\sum_{i=1}^N (y_i + z_i \iota_T - X_i \beta)' (y_i + z_i \iota_T - X_i \beta) + \underline{\nu} s^2}{\bar{\nu}}.$$

- Nezávislé podmíněné posteriorní hustoty pro neefektivitu odpovídají normálnímu rozdělení omezenému na kladné hodnoty:

$$p(z_i | y_i, X_i, \beta, h, \mu_z) \propto f_N(z_i | \bar{X}_i \beta - \bar{y}_i - (Th\mu_z)^{-1}, (Th)^{-1}) \mathbf{1}(z_i \geq 0).$$

- $\bar{y}_i = \frac{\sum_{t=1}^T y_{it}}{T}$  a  $\bar{X}_i$  je matice rozměru  $(1 \times k)$  obsahující průměrné hodnoty každé vysvětlující proměnné pro každého jednotlivce  $i$ .
- $\mathbf{1}(z_i \geq 0)$  je indikační funkce rovna jedničce pokud  $z_i \geq 0$  a nule v ostatních případech.

## Bayesovský výpočet (dokončení)

- Podmíněná posteriorní hustota pro  $\mu_z^{-1}$ :

$$\mu_z^{-1} | y, \beta, z \sim G(\bar{\mu}_z, \bar{\nu}_z),$$

$$\bar{\nu}_z = 2N + \underline{\nu}_z,$$

$$\bar{\mu}_z = \frac{N + \frac{\underline{\nu}_z}{2}}{\sum_{i=1}^N z_i + \underline{\mu}_z}.$$

- Výběry z omezeného normálního rozdělení (neomezené + vyhození  $z_i < 0$  nebo specifické algoritmy).
- Tradiční způsob predikční analýzy a provedení MCMC diagnostik; porovnání modelů např. pomocí Chibovy metody.
- Metody i pro čistě průřezovou verzi tohoto modelu (tzn.  $T = 1 \times$  nepřijatelné použití určitých nepravých priorů  $\rightarrow$  nepravé posteriory).
- Intuitivně:  $T = 1 \Rightarrow$  parametry  $z, \mu_z, \beta, h = N + K + 2$  parametrů  $\times$  jen  $N$  pozorování.

# Empirická ilustrace

- Bude časem doplněno. . .