PREDYKCJA
W BAYESOWSKIM MODELU NORMALNEJ REGRESJI LINIOWEJ

Łukasz Kwiatkowski

Katedra Ekonometrii i Badań Operacyjnych

Plan wykładu

1) Podstawy predykcji bayesowskiej
2) Predykcja w BMNRL
3) Przykład

Podstawy predykcji bayesowskiej 1

- Ogólne podstawy teoretyczne predykcji na gruncie bayesowskim – na razie bez kontekstu regresji
- **Wektor obserwacji niedostępnych** – podlegających prognozie, „niezaobserwowanych obserwacji”; (a także ewentualnych *braków danych* „w próbie”):

$$y^f = \begin{bmatrix} y_1^f \\ y_2^f \\ \vdots \\ y_h^f \end{bmatrix} \in Y^f \subseteq \mathbb{R}^h$$

→ W przypadku **danych przekrojowych**:

- y_i^f $(i \in \{1, 2, \dots, h\})$ – wartość zmiennej przewidywana dla jakiegoś i-tego „obiektu” (np. prognozowana cena mieszkania) – w regresji: przy **zadanych** wartościach zmiennych objaśniających, charakteryzujących ten obiekt (np. wielkość powierzchni, liczba pokoi, nr piętra)
- h – ilość takich jednocześnie rozważanych prognoz (rozważamy różne obiekty jednocześnie, np. mieszkania o różnych charakterystykach); jeżeli $h = 1$, wówczas y^f jest skalarem, więc indeks i w oznaczeniu y_i^f staje się zbyteczny

→ W przypadku **szeregów czasowych**:

- y_i^f $(i \in \{1, 2, \dots, h\})$ – wartość zmiennej przewidywana na (zwykle) i okresów wprzód poza okres próby (and. i-*step-ahead forecast*), tak więc $y_i^f \equiv y_{T+i}$ (z reguły używa się tego ostatniego symbolu)
- h – maksymalny horyzont prognozy

Podstawy predykcji bayesowskiej 2

➢ Predykcja / Prognozowanie = wnioskowanie o y^f:

- **na gruncie klasycznym** – poprzez charakterystyki tzw. **(teorio-)próbkowego rozkładu predyktywnego**: $p(y^f|y;\theta)$, a dokładniej $p(y^f|y;\theta=\hat{\theta})$ ← „zakotwiczony” w pojedynczym punkcie przestrzeni parametrów (ocenie θ); zatem nie uwzględnia niepewności estymacji θ:
  - ✓ Prognoza punktowa: $\hat{y}_i^f \equiv \hat{E}(y_i^f|y;\theta=\hat{\theta})$ → w regresji: $\hat{y}_i^f = x_i^f\hat{\beta}$
  - ✓ Średni błąd predykcji *ex ante*:

$$D(e_i^f) = \sqrt{s^2\left[1 + x_i^f(X'X)^{-1}x_i^{f\prime}\right]},$$

    gdzie $e_i^f = y_i^f - \hat{y}_i^f$ to błąd prognozy
  - ✓ „Klasyczne” przedziały prognozy – w KMNRL: $\left(\hat{y}_i^f - t_{\frac{\alpha}{2}} \cdot D(e_i^f), \hat{y}_i^f + t_{\frac{\alpha}{2}} \cdot D(e_i^f)\right)$

- **na gruncie bayesowskim** – y^f ma tu taki sam status jak „zwykłe” parametry, θ, tj. stanowi nieobserwowalną (przynajmniej w momencie dokonywania prognozy) zmienną losową:
  - θ – zawsze nieobserwowalne
  - y^f – nieobserwowalne przynajmniej w momencie dokonywania prognozy (w przypadku prognozowania szeregów czasowych wystarczy „zaczekać”; w przypadku danych przekrojowych możemy się nigdy nie doczekać… ;)

→ Bayesowskie wnioskowanie o y^f (= predykcja) – poprzez tzw. **(bayesowski) rozkład predyktywny:**

$$p(y^f|y)$$

Podstawy predykcji bayesowskiej 3

- Bayesowskie wnioskowanie o y^f (predykcja) – poprzez tzw. (bayesowski) **rozkład predyktywny:**

$$p(y^f|y)$$

→ Rozkład ten:

- jest rozkładem łącznym wszystkich wielkości prognozowanych („siedzących" w y^f)
- odzwierciedla całą naszą wiedzę (niepewność!) o y^f po wglądzie w dane y (i – jak się za chwilę okaże – także z uwzględnieniem niepewności związanej z parametrami, θ)
- co do konstrukcji jest rozkładem *a posteriori* (warunkowanie względem zaobserwowanych danych), ale termin „rozkład *a posteriori*" rezerwujemy dla $p(\theta|y)$

→ Prognoza bayesowska prowadzi do wyznaczenia całego **rozkładu prawdopodobieństwa** możliwych do zaobserwowania wartości zjawiska (pod warunkiem informacji zawartej w próbie), co idealnie odpowiada paradygmatowi tzw. **prognozowania probabilistycznego** (← *currently, most trendy* ;)

→ ALE: Jak wyznaczyć rozkład predyktywny?

Podstawy predykcji bayesowskiej 4

- **Model bayesowski bez predykcji [przypomnienie]:**

$$p(y,\theta) = p(y|\theta)p(\theta)$$

- **Model bayesowski z uwzględnieniem predykcji:**

$$p(y, y^f, \theta) = \underbrace{p(y^f|y,\theta)}_{\substack{(teorio-)\\ próbkowy\\ rozkład\\ predyktywny}} \underbrace{p(y,\theta)}_{\substack{model\\ bayes.\\ bez\\ predykcji}}$$

- **(Bayesowski) rozkład predyktywny:**

$$p(y^f|y) = \int_\Theta p(y^f,\theta|y)d\theta = \int_\Theta p(y^f|\theta,y) \underbrace{p(\theta|y)}_{\substack{rozkład\\ a\ post.}} d\theta$$

→ Rozkład predyktywny jest *mieszanką* tych tzw. (teorio-)próbkowych rozkładów predyktywnych, z rozkładem *a posteriori* jako rozkładem mieszającym (różne gęstości $p(y^f|\theta,y)$ – z uwagi na możliwe różne wartości θ – są „ważone” gęstością *a posteriori* reprezentującą niepewność co do tych różnych wartości θ). Zatem – w odróżnieniu od (teorio-)próbkowego (in. „klasycznego”) – bayesowski rozkład predyktywny uwzględnia w sobie niepewność związaną z parametrami.

⇒ W konsekwencji można się spodziewać, że bayesowskie rozkłady predyktywne z reguły będą (trochę) bardziej rozproszone od tych wyznaczonych na gruncie „klasycznym” → a to dobrze czy niedobrze?!

Podstawy predykcji bayesowskiej 5

➢ **Wnioskowanie o y^f – praktyka:**

→ Jeżeli $h > 1$, tj. y^f jest wektorem, a nie skalarem, wówczas przechodzimy na brzegowe rozkłady predyktywne poszczególnych współrzędnych y^f:

$$p(y_i^f|y) = \int_{\mathbb{R}^{h-1}} p(y^f|y)\, dy_{\backslash i}^f$$

gdzie $y_{\backslash i}^f$ oznacza wektor y^f po usunięciu i-tej współrzędnej.

→ Rozkład brzegowy $p\left(y_i^f|y\right)$ uwzględnia w sobie niepewność związaną z pozostałymi współrzędnymi

→ Charakterystyka $p\left(y_i^f|y\right)$ poprzez:

- Predyktywne wartości oczekiwane: $E(y_i^f|y)$
- Mediany predyktywne: $Me(y_i^f|y)$
- Modalne predyktywne: $Mo(y_i^f|y)$

Bayesowskie prognozy punktowe

- Predyktywne odchylenia standardowe: $D(y_i^f|y)$
- Kwantyle predyktywne: $Q_\alpha(y_i^f|y)$
- Predyktywne przedziały kwantylowe: $\left(Q_{\frac{\alpha}{2}}(y_i^f|y), Q_{1-\frac{\alpha}{2}}(y_i^f|y)\right)$
- Przedziały najwyższej gęstości predyktywnej: $HPredD\left(y_i^f|y\right)$

Przy wielookresowej prognozie szeregów czasowych, zwykle prezentowane w formie wykresu wachlarzowego

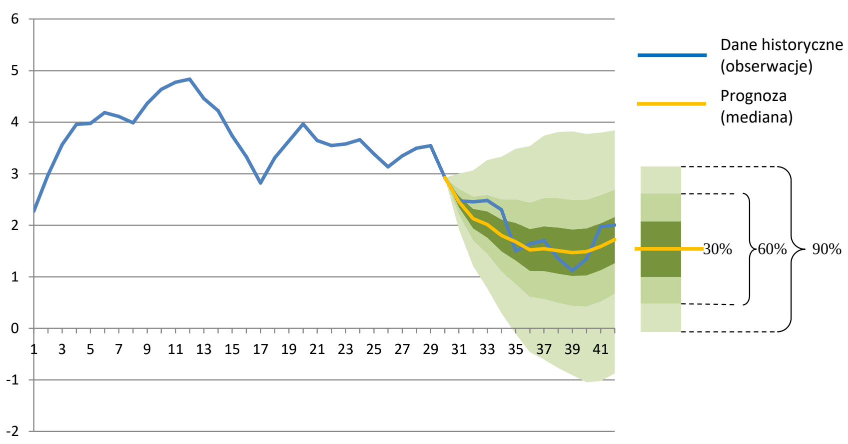
Podstawy predykcji bayesowskiej 6

➢ **Wykres wachlarzowy (ang. *fan chart, fan plot*):**

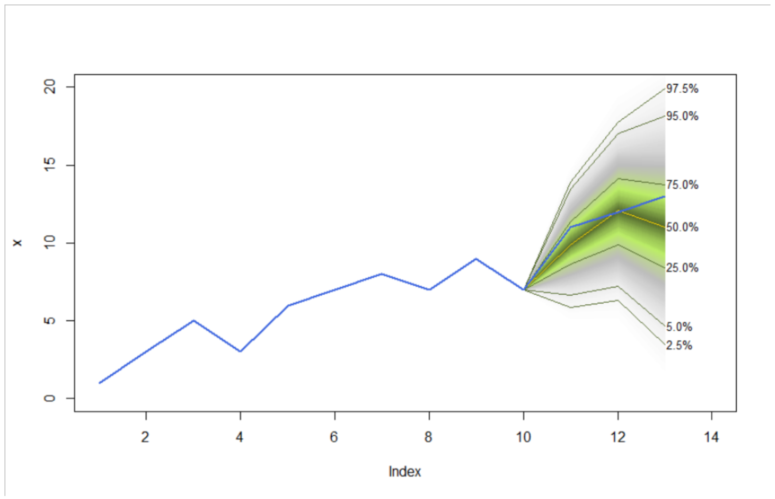
→ Termin sformułowany i spopularyzowany przez Bank Anglii w raportach prognostycznych inflacji, podawanych do informacji publicznej

→ Świetnie nadaje się do graficznej prezentacji wielookresowej, probabilistycznej prognozy szeregu czasowego (czy to bayesowskiej, czy „klasycznej")

Fan chart w wersji <u>przedziałowej</u> (ang. *interval*)



Fan chart w wersji <u>percentylowej</u> (ang. *percentile*)



→ R: funkcja **`fan`** w bibliotece **`fanplot`** – wymaga wcześniejszego uzyskania <u>losowań z rozkładu predyktywnego</u>

→ **Paleta barw – str. 2** w: http://www.biecek.pl/R/PrzewodnikPoPakiecieRWydanieIVinternet.pdf

Predykcja w BMNRL 1

- Przechodzimy do modelu regresji – równanie:

1) **W próbie**:

- Skalarnie ($t = 1, 2, \ldots, T$):

$$y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \cdots + \beta_k x_{tk} + \varepsilon_t = \underbrace{[x_{t1} \;\; x_{t2} \ldots x_{tk}]}_{\substack{x_t \\ (1\times k)}} \underbrace{\begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}}_{\substack{\beta \\ (k\times 1)}} + \varepsilon_t = x_t \beta + \varepsilon_t$$

- Macierzowo: $y = X\beta + \varepsilon$

2) **Poza próbą – predykcja**:

- Pojedyncza współrzędna y_i^f ($i = 1, 2, \ldots, h$): $y_i^f = x_i^f \beta + \varepsilon_i^f$, x_i^f – ustalony:

$$x_i^f = [x_{i1}^f \;\; x_{i2}^f \;\; \ldots \;\; x_{ik}^f]$$

- Wektor y^f ($h \geq 1$ prognoz): $y^f = X^f \beta + \varepsilon^f$, $\underbrace{X^f}_{(h\times k)} = \begin{bmatrix} x_1^f \\ \vdots \\ \underbrace{x_h^f}_{wiersze} \end{bmatrix}$, $\underbrace{\varepsilon^f}_{(h\times 1)} = \begin{bmatrix} \varepsilon_1^f \\ \vdots \\ \varepsilon_h^f \end{bmatrix}$

Predykcja w BMNRL 2

➢ **Model bayesowski – z predykcją:**

$$p(y, y^f, \theta) = \underbrace{p(y^f|y, \theta)}_{\substack{(teorio-) \\ próbkowy \\ rozkład \\ predyktywny}} \underbrace{p(y, \theta)}_{\substack{model \\ bayes. \\ bez \\ predykcji}}$$

a tak naprawdę

$$p(y, y^f, \theta | X, X^f) = p(y^f | y, \theta, X^f) p(y, \theta | X)$$

→ dla zwięzłości zapisu pomijamy dalej warunkowanie względem X oraz X^f

→ $p(y, \theta)$ – omówiony wcześniej (4 przypadki w zależności od rozkładu *a priori*)

→ Pytanie: jak wygląda $p(y^f|y, \theta)$?

Założenia KMNRL – dotyczące rozkładu próbkowego (czyli rozkładu danych pod warunkiem parametrów) „obowiązują" dla wszystkich obserwacji – zarówno tych w próbie, jak i tych spoza niej (w przeciwnym razie model nie byłby spójny!)

Zatem skoro [przypomnienie]: $p(y|\theta) = f_N^T(y|X\beta, \tau^{-1}I_T)$,

więc $p(y^f|y, \theta) = p(y^f|\theta) = f_N^h(y^f|X^f\beta, \tau^{-1}I_h)$

Predykcja w BMNRL 3

➢ **Rozkład predyktywny [przypomnienie]:**

$$p\left(y^f \middle| y\right) = \int_\Theta p\left(y^f, \theta \middle| y\right) d\theta = \int_\Theta p\left(y^f \middle| \theta, y\right) \underbrace{p(\theta|y)}_{\substack{rozkład \\ a\ post.}} d\theta$$

→ Szczegóły odn. $p\left(y^f \middle| y\right)$ zależą od przyjętego rozkładu *a priori*:

1) Reguła Jeffreysa z zależnością [RJz] – nie rozważamy…
2) Reguła Jeffreysa z niezależnością [RJn] – da się wyprowadzić analityczna postać $p\left(y^f \middle| y\right)$
3) Gamma-normalny z zależnością [GNz] – da się wyprowadzić analityczna postać $p\left(y^f \middle| y\right)$
4) Gamma-normalny z niezależnością [GNn] – $p\left(y^f \middle| y\right)$ nie jest żadnym znanym rozkładem prawdopodobieństwa → konieczność zastosowania podejścia symulacyjnego

Ad [RJn] i [GNz]: $\boxed{p\left(y^f \middle| y\right) = f_{St}^h(y^f | \bar{a}^f, \bar{P}^f, \bar{n}),}$

gdzie $\bar{a}^f = X^f \bar{a}$, $\bar{P}^f = \frac{\bar{n}}{\bar{s}} \bar{G}$, $\bar{G} = \left[I_h + X^f \bar{C}^{-1} X^{f\prime}\right]^{-1}$,

natomiast $\bar{a}, \bar{C}, \bar{n}, \bar{s}$ – patrz wcześniejsze wykłady (RJn i GNz):

RJn	**GNz**
$\bar{a} = \hat{\beta} = (X'X)^{-1}X'y$	$\bar{a} = \bar{C}^{-1}(Ca + X'X\hat{\beta})$
$\bar{C} = X'X$	$\bar{C} = C + X'X$
$\bar{n} = T - k$	$\bar{n} = n_0 + T$
$\bar{s} = S(\hat{\beta}) = \left(y - X\hat{\beta}\right)'\left(y - X\hat{\beta}\right)$	$\bar{s} = s_0 + y'y + a'Ca - \bar{a}'\bar{C}\bar{a}$

Predykcja w BMNRL 4

- **Charakterystyki łącznego rozkładu predyktywnego (w przypadkach [RJn] i [GNz]):**

$$\boxed{p\left(y^f \middle| y\right) = f_{St}^{h}(y^f|\bar{a}^f, \bar{P}^f, \bar{n})}$$

$$\bar{a}^f = X^f \bar{a}, \qquad \bar{P}^f = \frac{\bar{n}}{\bar{s}} \bar{G}, \qquad \bar{G} = \left[I_h + X^f \bar{C}^{-1} X^{f\prime}\right]^{-1},$$

- $E\left(y^f \middle| y\right) = \bar{a}^f,$ dla $\bar{n} > 1$
- $Mo\left(y^f \middle| y\right) = \bar{a}^f$ ($\bar{n}$ – dowolne)
- mediana w przypadku rozkładów wielowymiarowych – za trudne…
- $V\left(y^f \middle| y\right) = \frac{\bar{n}}{\bar{n}-2} {\bar{P}^f}^{-1} = \frac{\bar{s}}{\bar{n}-2} \bar{G}^{-1}$, dla $\bar{n} > 2$ $\rightarrow$ $Var\left(y_i^f | y\right) = \left[V\left(y^f \middle| y\right)\right]_{ii}$

Predykcja w BMNRL 5

➢ **Brzegowe rozkłady predyktywne (tj. dla pojedynczych $\boldsymbol{y_i^f}$, $\boldsymbol{i = 1, 2, \dots, h}$):**

$$p(y_i^f|y) = \int_{\mathbb{R}^{h-1}} p(y^f|y) dy_{\setminus i}^f = f_{St}^1(y_i^f|\bar{a}_i^f, \bar{P}_i^f, \bar{n})$$

gdzie $y_{\setminus i}^f$ oznacza wektor powstały z y^f poprzez pominięcie *i*-tej jego współrzędnej, $\bar{a}_i^f$ jest i-tą współrzędną wektora $\bar{a}^f$, natomiast

$$\bar{P}_i^f = Prec(y_i^f|y) = \frac{\bar{n}}{\bar{n}-2}\left(Var(y_i^f|y)\right)^{-1} = \frac{\bar{n}}{\bar{s}}([\bar{G}^{-1}]_{ii})^{-1}$$

Uwaga: $\bar{P}_i^f \neq [\bar{P}^f]_{ii}$

- Predyktywna wartość oczekiwana: $E(y_i^f|y) = \bar{a}_i^f$ dla $\bar{n} > 1$
- Modalna i mediana predyktywna: $Mo(y_i^f|y) = Me(y_i^f|y) = \bar{a}_i^f$ ($\bar{n}$ – dowolne)
- Wariancja predyktywna: $Var(y_i^f|y) = [V(y^f|y)]_{ii}$ dla $\bar{n} > 2$
- Predyktywne odchylenie standardowe: $D(y_i^f|y) = \sqrt{Var(y_i^f|y)}$

Predykcja w BMNRL 6

➢ **Brzegowe rozkłady predyktywne (tj. dla pojedynczych $\boldsymbol{y_i^f}$, $\boldsymbol{i = 1, 2, \dots, h}$):**

- $HPredD_{1-\alpha}(y_i^f) = \left(\bar{a}_i^f - t_{\frac{\alpha}{2}}\sqrt{\bar{P}_i^{f^{-1}}},\ \bar{a}_i^f + t_{\frac{\alpha}{2}}\sqrt{\bar{P}_i^{f^{-1}}}\right),$

  gdzie $t_{\frac{\alpha}{2}}$ – dwustronna wartość kryt. w $St(0, 1, \bar{n})$:

  - $\Pr\left(|x| \geq t_{\frac{\alpha}{2}}\right) = 1 - \alpha$, gdzie $x \sim St(0, 1, \bar{n})$;
  - $t_{\frac{\alpha}{2}}$ w Excelu: → =ROZKŁAD.T.ODW(*prawdopodobieństwo* = α; *stopnie_swobody* = $\bar{n}$) (starsza wersja)

    → =ROZKŁ.T.ODWR.DS(*prawdopodobieństwo* = α; *stopnie_swobody* = $\bar{n}$) (nowsza wersja)