

Kalibrace scoringových modelů

Finanční matematika v praxi II
Hotel Podlesí
11. – 13. září 2012

Pavel Plát
pavel.plat@rb.cz
Basel II & Scoring department

- **Scoringový model**
- **Kalibrace scoringového modelu**
- **„Forward looking“ kalibrace**
- **Příklady**

- **Scoringový model**
- Kalibrace scoringového modelu
- „Forward looking“ kalibrace
- Příklady

Business pohádka...

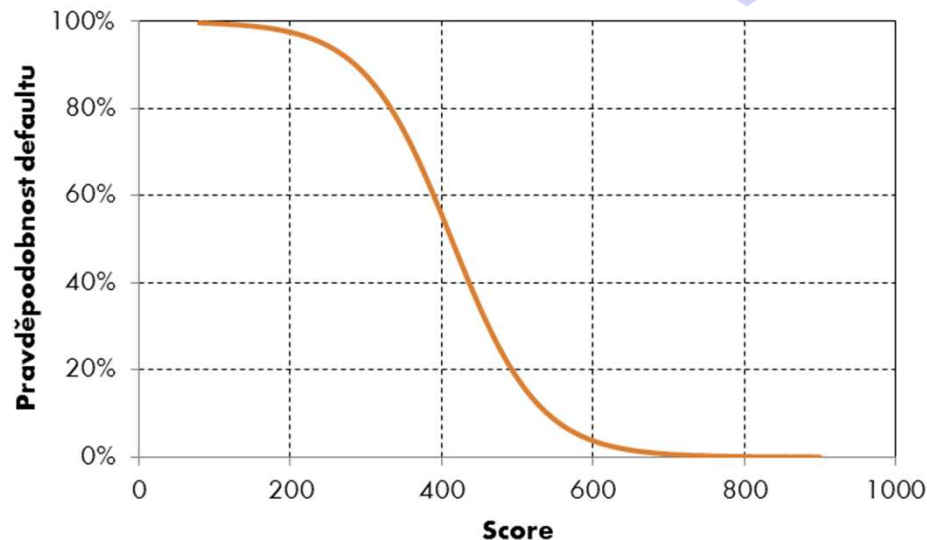
- Banka poskytuje klientům úvěry a spravuje úvěrové portfolio
- Z hlediska plnění závazků vůči bance je každý klient/kontrakt hodnocen buď jako dobrý nebo jako špatný (default)
- Někteří klienti mají vyšší šanci, že se dostanou do problémů než jiní.
- O každém klientu/kontraktu má banka nějaké informace
- Na základě aktuální situace klienta/kontraktu chceme predikovat schopnost plnit závazky vůči bance v následujícím období
- Hodnocení klienta je jednoznačně definované a opakovatelné

...a trocha formalismu

- Označme \mathbf{K} množinu všech klientů/ kontraktů
- $$D_k = \begin{cases} 0 & \text{jestliže } k \in \mathbf{K} \text{ je dobrý} \\ 1 & \text{jestliže } k \in \mathbf{K} \text{ je špatný} \end{cases}$$
- Používáme pravděpodobnostní model, tj. chceme predikovat pravděpodobnost defaultu
$$PD = P(D_k = 1)$$
- Situace klienta/kontraktu $k \in \mathbf{K}$ je popsána bodem ve vhodně zvoleném měřitelném prostoru - obvykle \mathbf{R}^n
$$x_k = (x_{k,1}, \dots, x_{k,n})$$
- $$P(D_k = 1) = f(x_k)$$
- $$\text{SCORE: } \mathbf{R}^n \rightarrow \mathbf{R}$$

Pravděpodobnost defaultu

- definice špatného klienta/kontraktu
 - události selhání (dny po splatnosti, insolvence, restrukturalizace, podvod,...)
 - časový horizont selhání



$$P(D_k = 1) = f(\text{SCORE}_k)$$

- $f: R \rightarrow [0,1]$
- f je monotonní

Score

- Ukazatele:
 - Socio-demografické (vzdělání, věk, stav...)
 - Behaviorální (historická platební morálka, čerpání,...)

- Scoringový model
- **Kalibrace scoringového modelu**
- „Forward looking“ kalibrace
- Příklady

Master scale

Score	Poměr šancí	ln(poměr šancí)	PD
460	2.25	0.811	30.77%
500	4.5	1.504	18.18%
540	9	2.197	10.00%
580	18	2.890	5.26%
620	36	3.584	2.70%
660	72	4.277	1.37%
700	144	4.970	0.69%
740	288	5.663	0.35%
780	576	6.356	0.17%
820	1152	7.049	0.09%

- score 660 = poměr šancí 72
- dvojnásobný poměr šancí každých 40 bodů

$$P(D_k = 1) = f(\text{SCORE}_k)$$

- $$f(x) = \frac{1}{1 + \exp\left(\ln(72) - \frac{660}{40} * \ln(2) + \frac{\ln(2)}{40} * x\right)}$$
- Lineární závislost mezi score a logaritmem poměru šancí

Kalibrace scoringového modelu

Krok 0/ Scoringový model vyvinutý pomocí logistické regrese → předpokládáme model ve tvaru:

$$PD = \frac{1}{1 + \exp(\text{intercept} + \text{slope} * \text{SCORE}^*)}$$

Krok 1/ Odhad parametrů logistického modelu:

- $\widehat{\text{intercept}}$
- $\widehat{\text{slope}}$

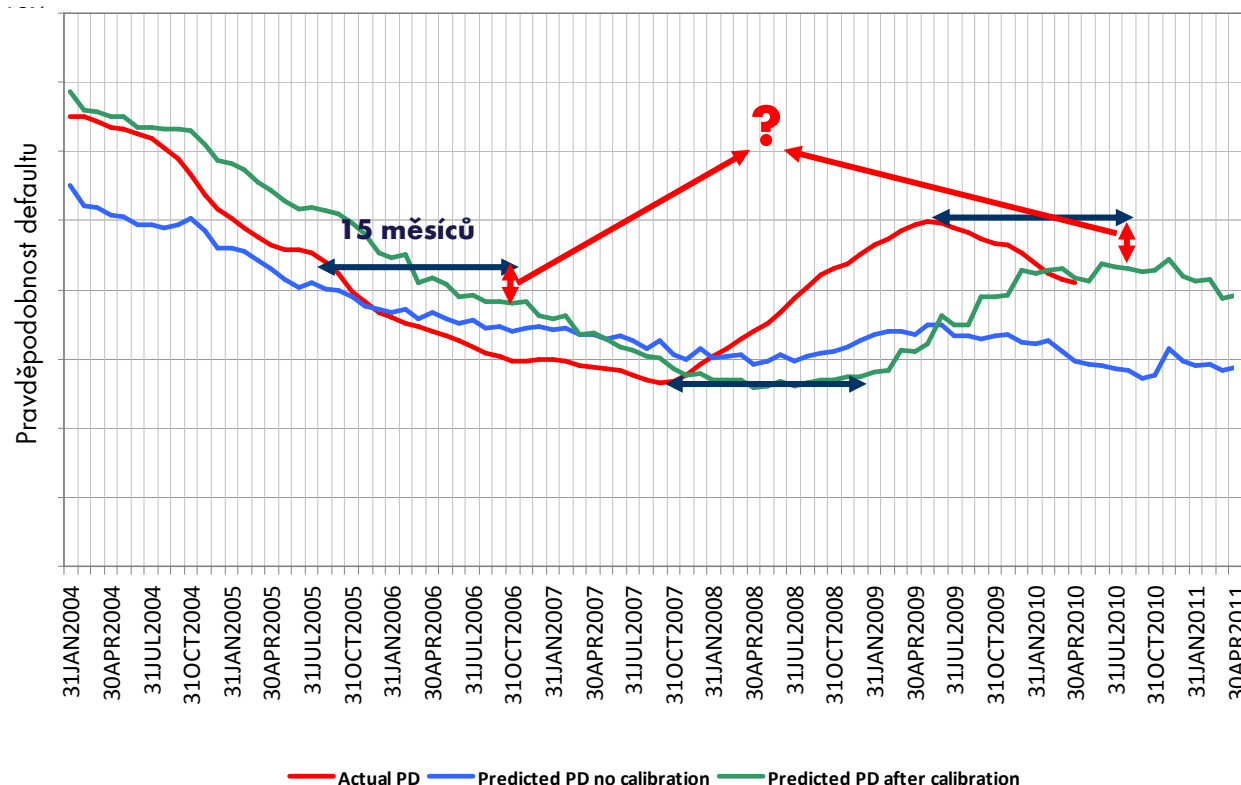
Krok 2/ $\text{SCORE} = \text{intercept}_{\text{cal}} + \text{slope}_{\text{cal}} * \text{SCORE}^*$

$$\text{intercept}_{\text{cal}} = \frac{\widehat{\text{intercept}} - \ln(72) + \frac{660}{40} * \ln(2)}{\frac{\ln(2)}{40}}$$

$$\text{slope}_{\text{cal}} = \frac{\widehat{\text{slope}}}{\frac{\ln(2)}{40}}$$

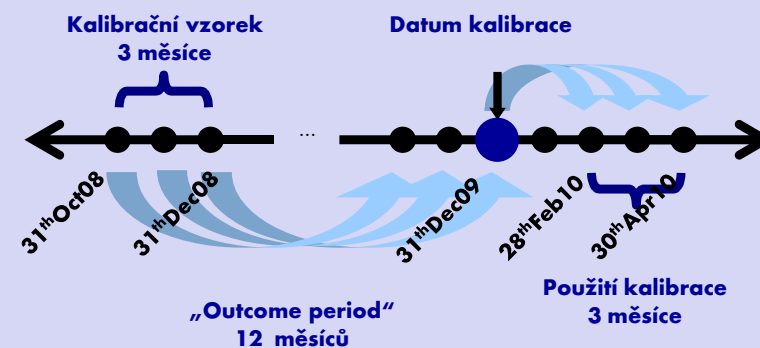
- Scoringový model
- Kalibrace scoringového modelu
- **„Forward looking“ kalibrace**
- Příklady

Kalibrace založená na historii



Chyba kalibrace založené na historii

Hlavní důvod je zpoždění mezi kalibračním vzorkem a obdobím použití kalibrace.



Je možné kalibrovat lépe? Snad ano!

Základní myšlenka:

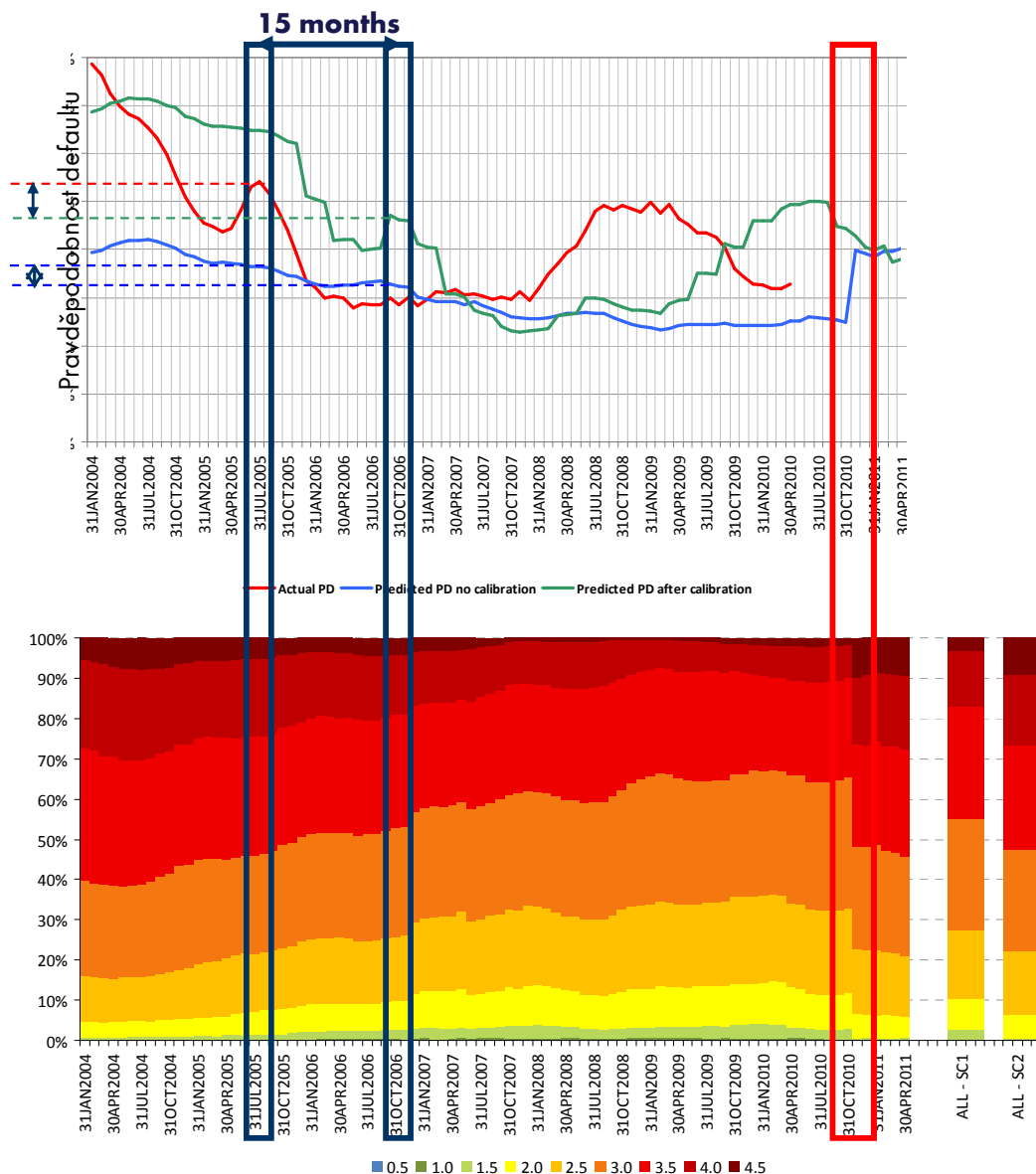
- 1/ Vyvinout prediktivní model pro PD (červená křivka) s použitím makroekonomických veličin
- 2/ Použít makroekonomickou predikci PD k úpravě kalibrační transformace ($\text{intercept}_{\text{cal}}$)

Není to tak jednoduché...

- 1/ Chyba kalibrace není jenom časový posun
- 2/ Pro vývoj makroekonomického modelu PD máme krátké řady

Vážené PD

Vysvětlovaná proměnná pro makroekonomický model – část 1



Další problémy:

1/ Změna ve struktuře portfolia

- “Predicted PD no calibration” (modrá křivka) je vypočtena s použitím logistické funkce na nekalibrované score, tzn.

$$\frac{1}{1 + \exp\left(\ln(72) - \frac{660}{40} \ln(2) + \frac{\ln(2)}{40} \text{SCORE}^*\right)}$$

- Změny v “Predicted PD no calibration” ukazují na vývoj ve struktuře portfolia z pohledu ukazatelů vstupujících do scoringového modelu
- Predikce změn ve struktuře portfolia není cíl makroekonomického modelu

2/ Implementace nové scorekarty

- Struktura portfolia založená na různých scoringových modelech nemusí být jednoduše porovnatelná

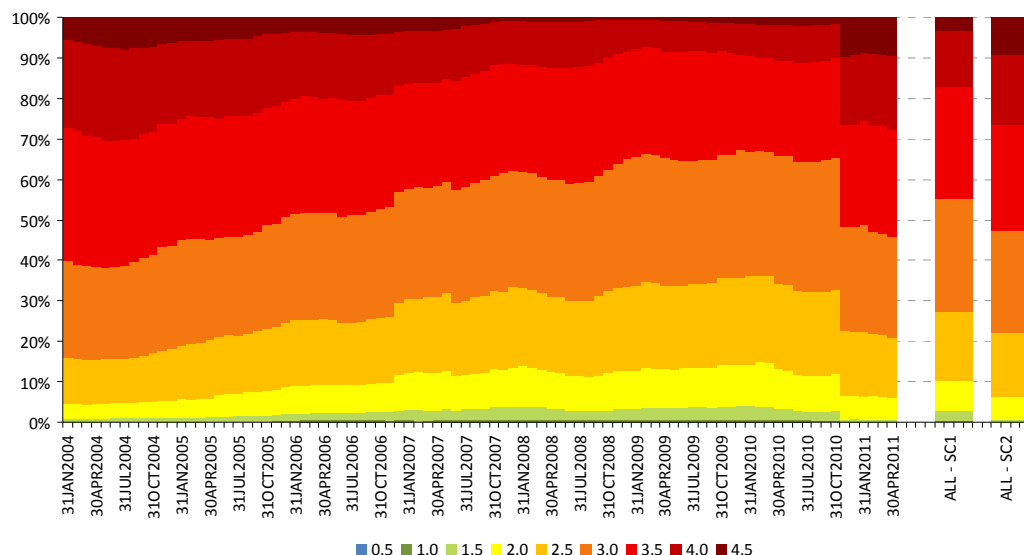
Návrh řešení:

Ad 1/ Jako vysvětlovanou proměnou pro makroekonomický model použít vážené PD

Ad 2/ Váhy definovat pro každou scorekarty samostatně

Vážené PD

Vysvětlovaná proměnná pro makroekonomický model – část 2



Cíl

Pro všechny pozorování definovat váhy $w_{m,p}^{(sc)}$ závislé na skorkartě (sc), měsíci (m) příslušnosti do rizikové skupiny definované na základě nekalibrovaného score (p) tak, že vážená struktura portfolia bude invariantní v čase a rovná „průměrné“ struktuře celého neváženého portfolia. Tyto váhy použít pro výpočet váženého PD.

Definice vah:

Označme:

$n_{m,p}^{(sc)}$ - Počet pozorování ve skupině p v měsíci m, scorovaných modelem sc

$$n_{m,\bullet}^{(sc)} = \sum_p n_{m,p}^{(sc)}, \quad n_{\bullet,p}^{(sc)} = \sum_m n_{m,p}^{(sc)}, \quad n_{\bullet,\bullet}^{(sc)} = \sum_{m,p} n_{m,p}^{(sc)}$$

a definujeme:

$$w_{m,p}^{(sc)} = \frac{n_{\bullet,p}^{(sc)} \cdot n_{m,\bullet}^{(sc)}}{n_{\bullet,\bullet}^{(sc)} \cdot n_{m,p}^{(sc)}}$$

Weighted actual PD:

Označme:

$d_{m,p}^{(sc)}$ - Počet defaultních pozorování ve skupině p v měsíci m, scorovaných modelem sc

a definujeme:

$$wPD_m^{(sc)} = \frac{\sum_p d_{m,p}^{(sc)} \cdot w_{m,p}^{(sc)}}{\sum_p n_{m,p}^{(sc)} \cdot w_{m,p}^{(sc)}}$$

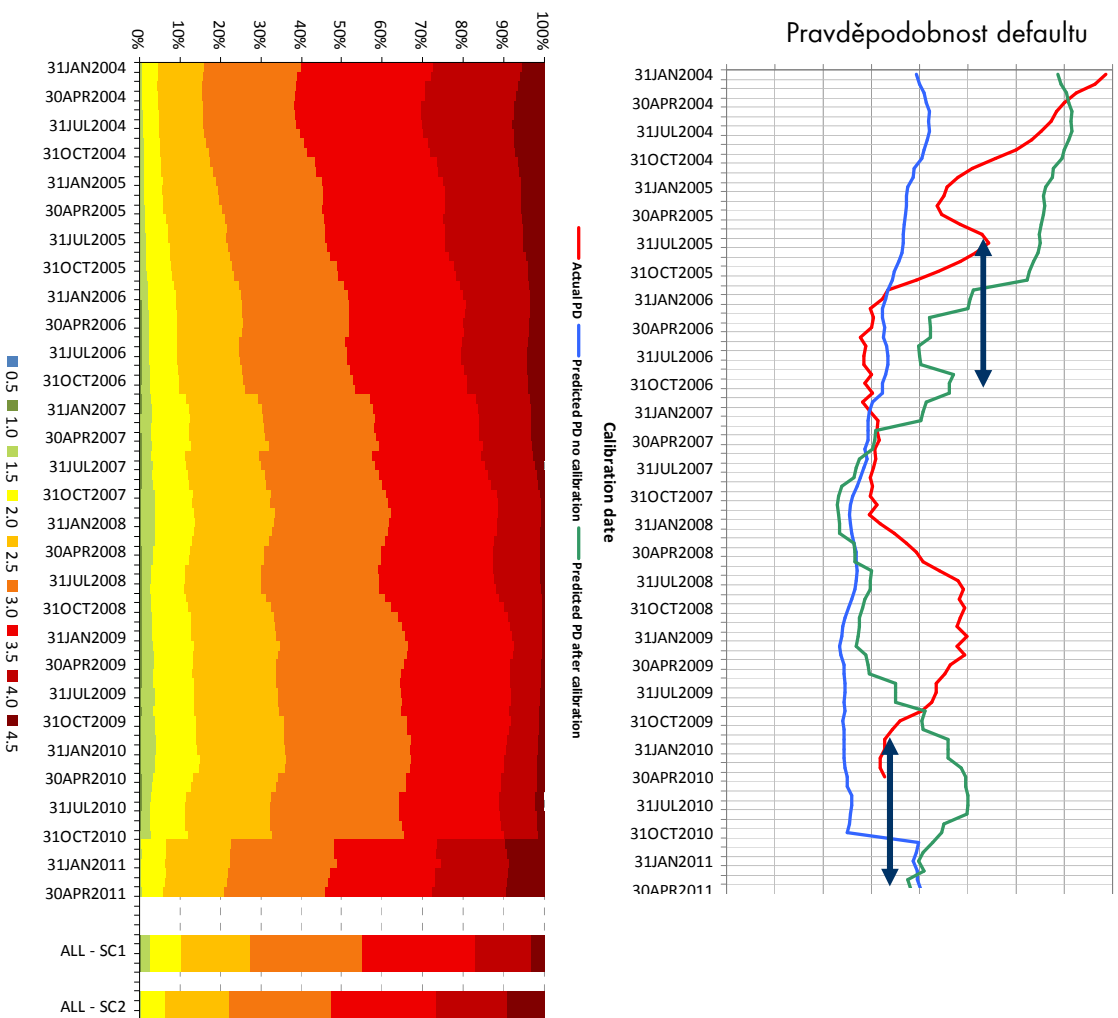
Vlastnosti:

- Vážená struktura portfolia je časově invariantní
- Vliv změn ve struktuře portfolia je z váženého wPD odstraněn
- V každé skupině (p) jsou si PD a vážené wPD rovny

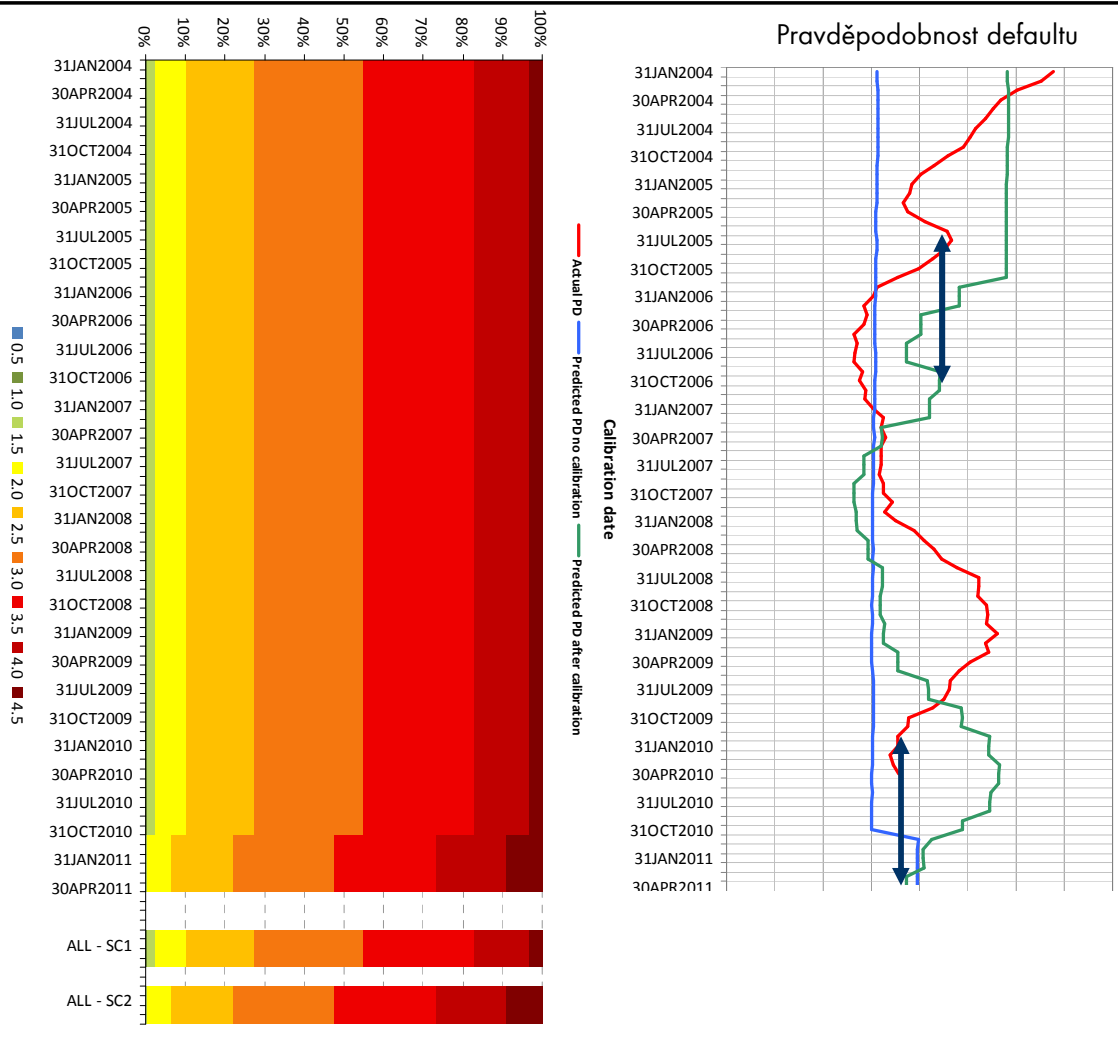
$wPD_m^{(sc)}$ je vhodná
vysvětlovaná proměnná pro
makroekonomický model

Původní vs. vážené PD

Původní



Vážené



- Vážená struktura je časově invariantní time invariant
- Vliv změny struktury portfolia je z váženého PD odstraněn
- "Predicted PD no calibration" je časově invariantní
- "Predicted PD after calibration" je pouze zpožděné za "Actual PD"

Vysvětlujíví proměnné

zdroj: ARAD (ČNB), ČSU, predikce makroekonomů Raiffeisen bank



Banka inspirovaná klienty

Price indices (corresponding period of the previous year = 100)

- Consumer price index
- Industry producer prices index (PPI)

Unemployment

- Unemployment rate
- 12 months difference

Indices of confidence indicators

- Business confidence indicator index
- Consumers confidence indicator index
- Composite indicator

Interbank deposit market interest rate fixing – PRIBOR (Prague Interbank Offered Rate)

- PRIBOR – 1M, 3M, 6M, 1Y

Gross Domestic Product

- Gross domestic product
- Final consumption expenditure of households

Bond yield

- Czech bonds – 5Y, 10Y
- Eurozone bonds – 3Y, 5Y, 10Y

Retail trade (corresponding period of the last year = 100, constant prices)

- Retail trade excl. retail sale of automotive fuel
- Retail trade of food, beverages and tobacco, constant prices
- Retail trade, other goods, constant prices

Industrial production index (corresponding month previous year = 100)

- Industry, total - working day adjusted data

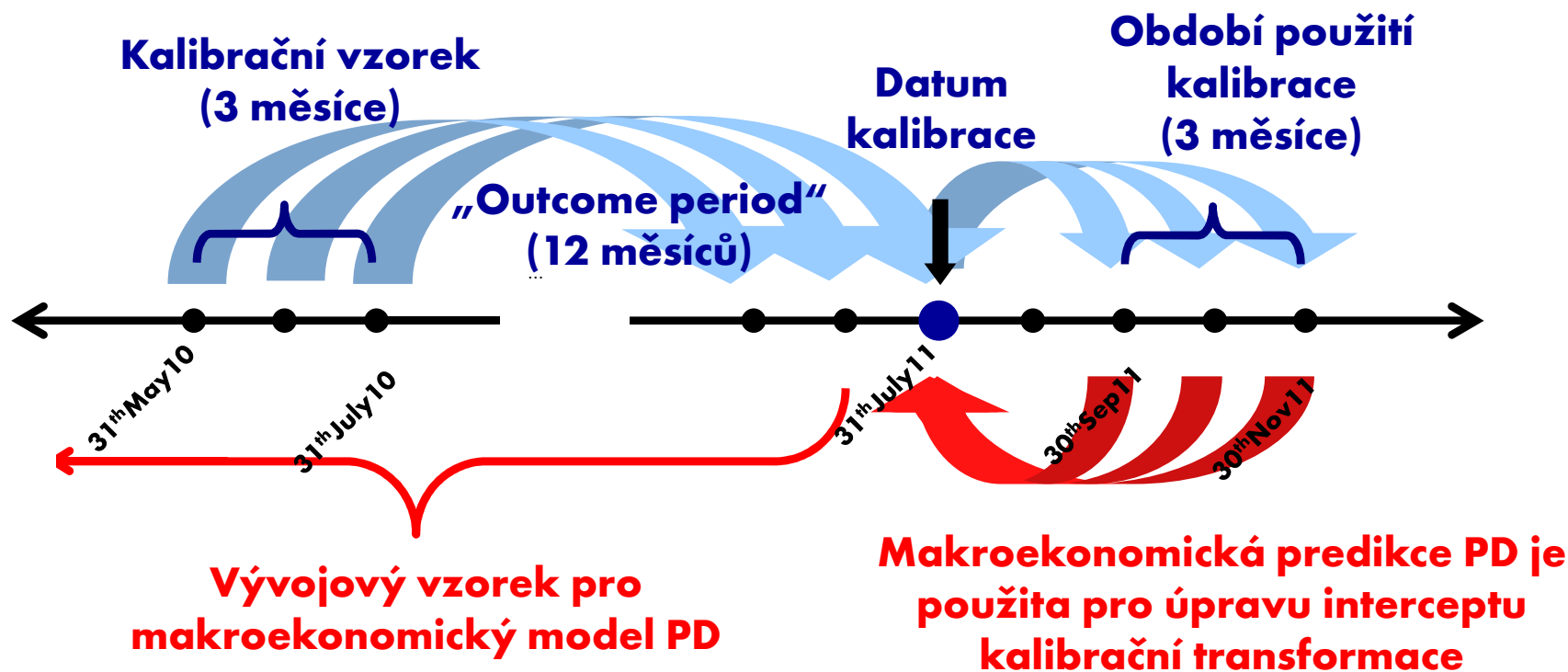
Exchange rates (monthly average)

- CZK / EUR, CZK / USD, CZK / GBP

Očekávaný trend PD v případě rostoucího trendu makroekonomického ukazatele



„Forward looking“ kalibrace



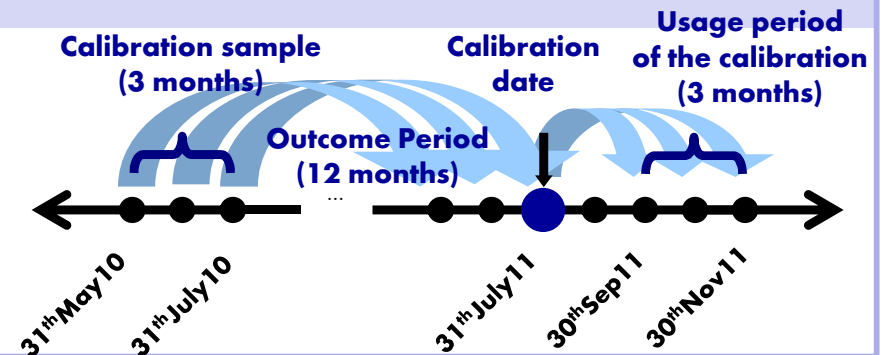
„Forward looking“ kalibrace

Krok 1 – Historická kalibrace

- Kalibrace založená na historickém kalibračním vzorku.
- Logistická regrese s jedním prediktorem SCORE* (nekalibrované score).

Výstup:

- $\text{intercept}_{\text{cal}}^{(\text{history})}, \text{slope}_{\text{cal}}$
- $\text{SCORE}^{(\text{history})} = \text{intercept}_{\text{cal}}^{(\text{history})} + \text{slope}_{\text{cal}} * \text{SCORE}^*$

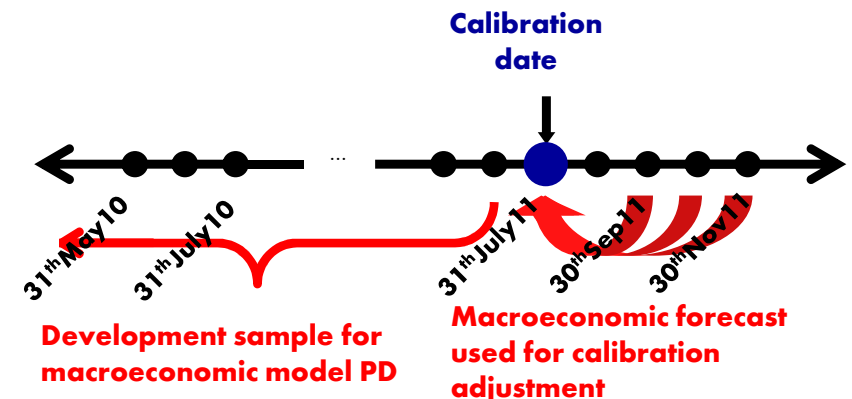


Krok 2 – Makroekonomická predikce PD

- Predikce váženého PD založená na makroekonomickém modelu.

Výstup:

- $\text{wPD}_m^{(\text{macro})}$ - makroekonomická predikce váženého PD v měsíci m
- $\text{wPD}^{(\text{target})} = \frac{1}{3} \sum_{m \in \text{období použití kalibrace}} \text{wPD}_m^{(\text{macro})}$



„Forward looking“ kalibrace

Krok 3 – Úprava historické kalibrace dle makroekonomické predikce PD

- Korekce interceptu kalibrační transformace získané z historické kalibrace tak aby vážené wPD predikované pomocí výsledného kalibrovaného score bylo na kalibračním vzorku rovno požadované hodnotě dle makroekonomické predikce. Tzn je třeba nalézt takové Δ aby platilo

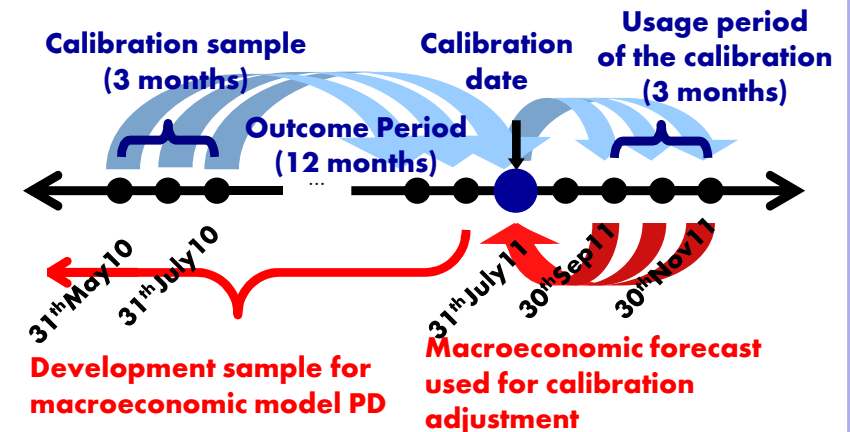
$$wPD^{(target)} = \frac{1}{\sum_{\text{contract } i \in \text{calibration sample}} w_i} \cdot \sum_{\text{contract } i \in \text{calibration sample}} \frac{w_i}{1 + \exp\left(\ln(72) - \frac{660}{40} \ln(2) + \frac{\ln(2)}{40} \left(\Delta + \text{intercept}_{cal}^{(history)} + \text{slope}_{cal} \cdot \text{SCORE}_i^*\right)\right)}$$

Iterativní algoritmus pro výpočet Δ (ukončit když $\Delta^{(j+1)} - \Delta^{(j)}$ je dostatečně malé):

$$\Delta^{(0)} = 0, wPD^{(j)} = \frac{1}{\sum w_i} \sum \frac{w_i}{1 + \exp\left(\ln(72) - \frac{660}{40} \ln(2) + \frac{\ln(2)}{40} \left(\Delta^{(j)} + \text{intercept}_{cal}^{(history)} + \text{slope}_{cal} \cdot \text{SCORE}_i^*\right)\right)}, \Delta^{(j+1)} = \Delta^{(j)} + \frac{40}{\ln 2} \left(\ln \frac{1 - wPD^{(target)}}{wPD^{(target)}} - \ln \frac{1 - wPD^{(j)}}{wPD^{(j)}} \right)$$

Výstup:

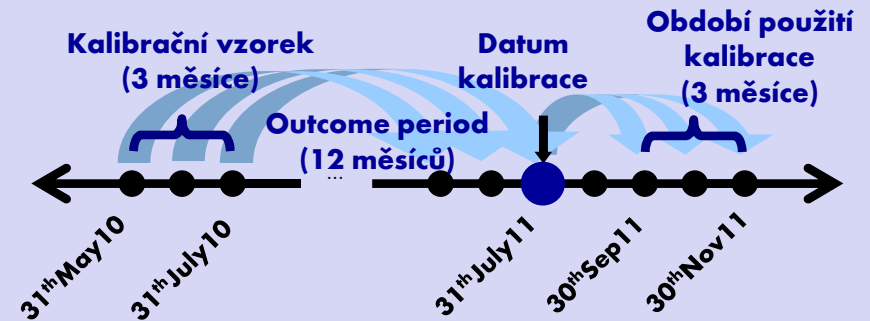
- $\text{intercept}_{cal} = \Delta + \text{intercept}_{cal}^{(history)}$, slope_{cal}
- SCORE = $\text{intercept}_{cal} + \text{slope}_{cal} \cdot \text{SCORE}^*$**
(final calibrated score)



- Scoringový model
- Kalibrace scoringového modelu
- „Forward looking“ kalibrace
- **Příklady**

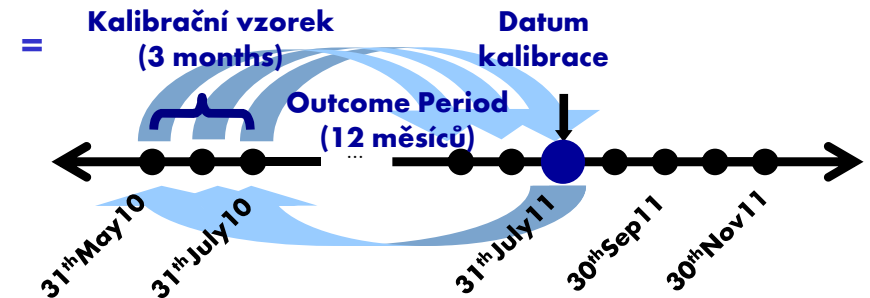
Porovnávání přístupů

“history based”

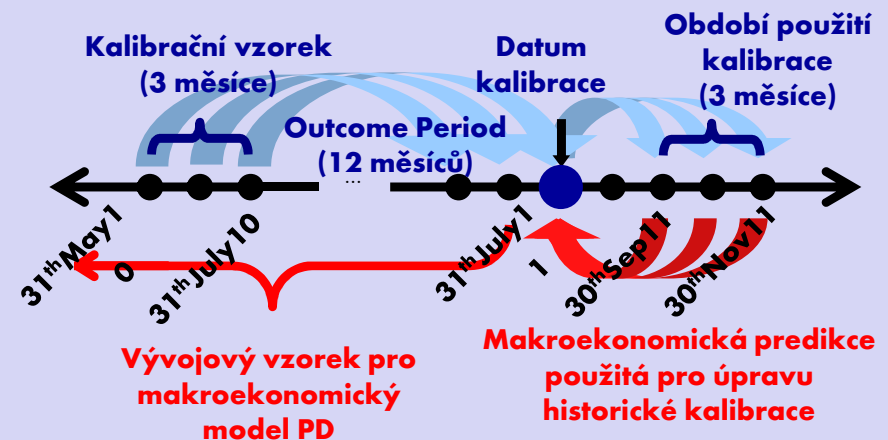


“ideal”

Období použití kalibrace (3 měsíce)



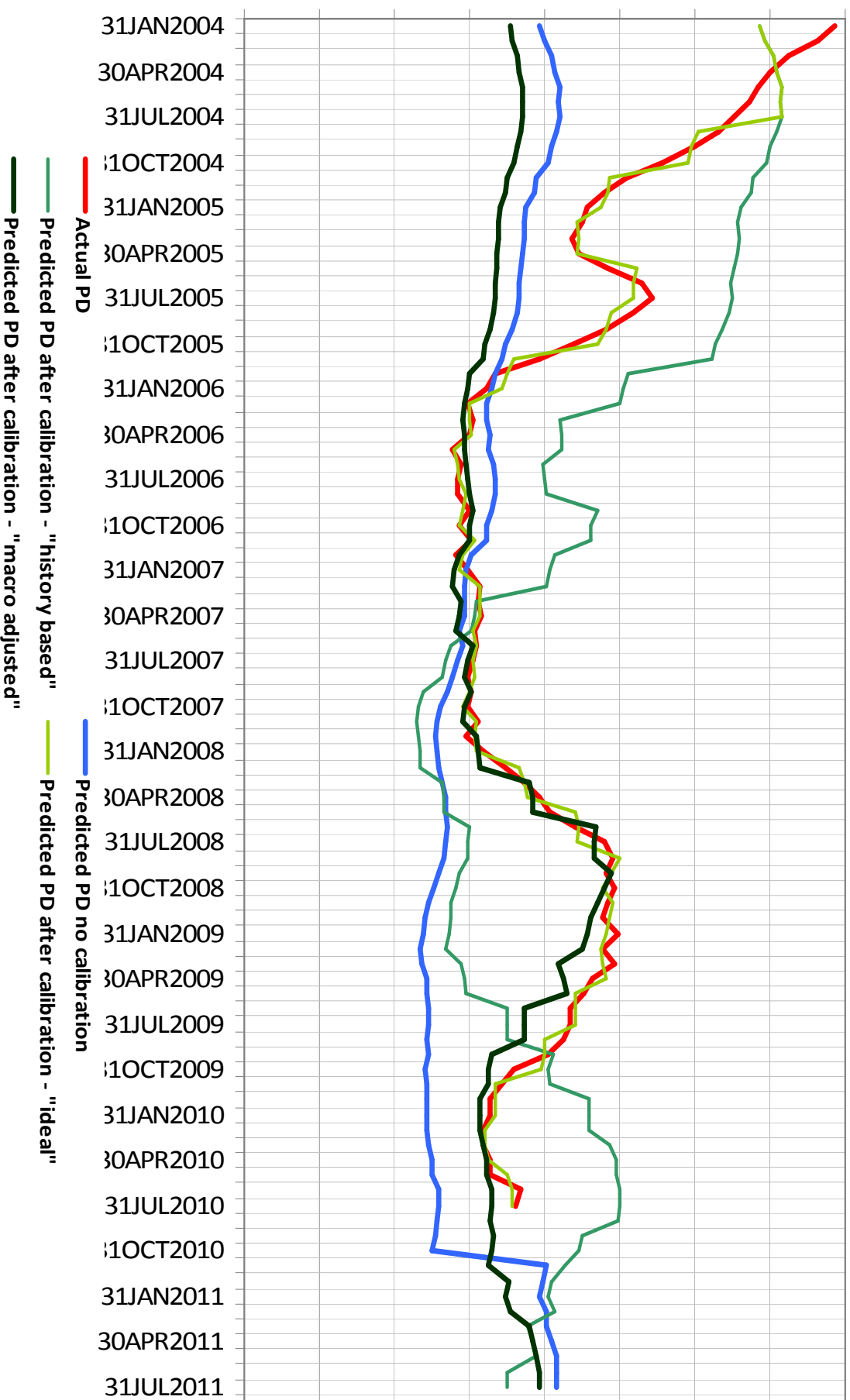
“forward looking”



Příklad 1

PD = f(HDP, Nezaměstnanost)

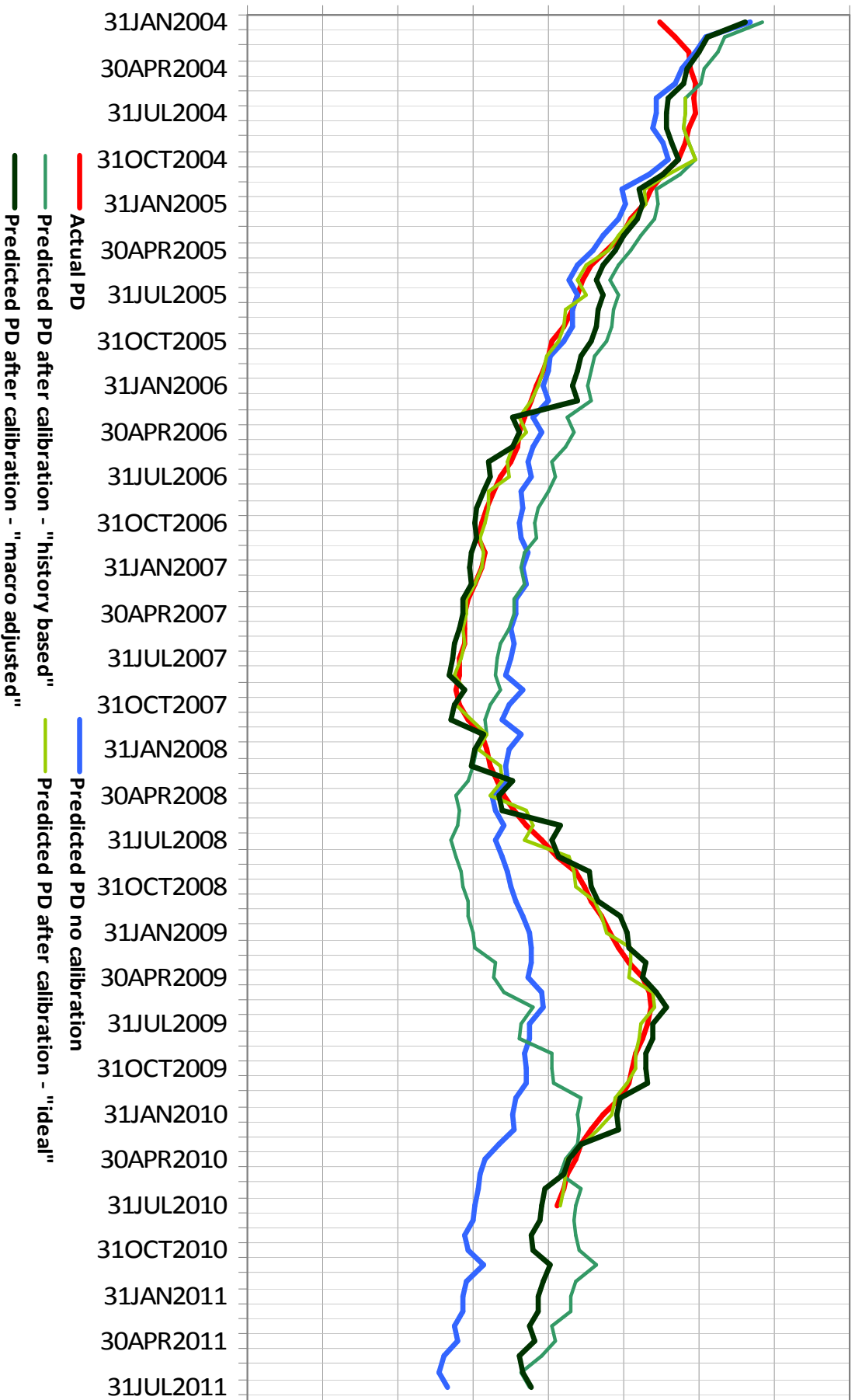
Pravděpodobnost defaultu



Příklad 2

$$PD = f(HDP, IPI)$$

Pravděpodobnost defaultu



Nestabilita makroekonomické predikce PD

- Krátké časové řady
- Vývoj modelu prováděn pro napozorované hodnoty makroekonomických ukazatelů, predikce PD používá makroekonomické predikce

Změny v procesech banky

- Významné změny v procesech se nedají vyhladit pouze vážením

Predikce PD na základě extrapolace již pozorovaných pravděpodobností selhání v krátkém horizontu

- Aktuálně nepoužíváme