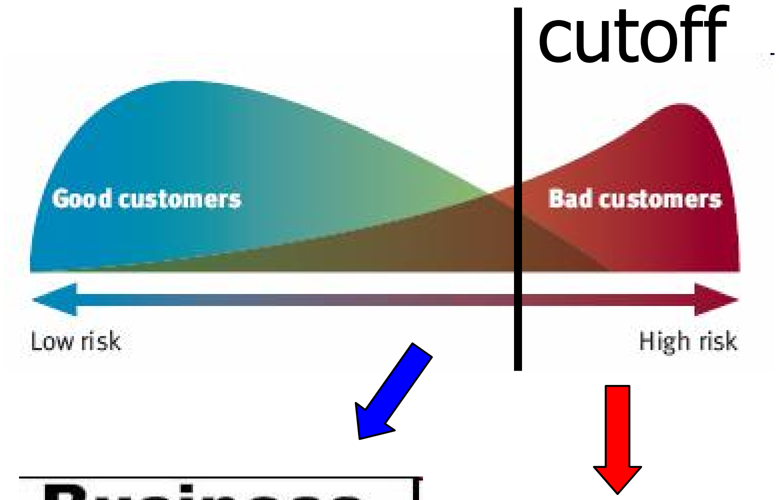
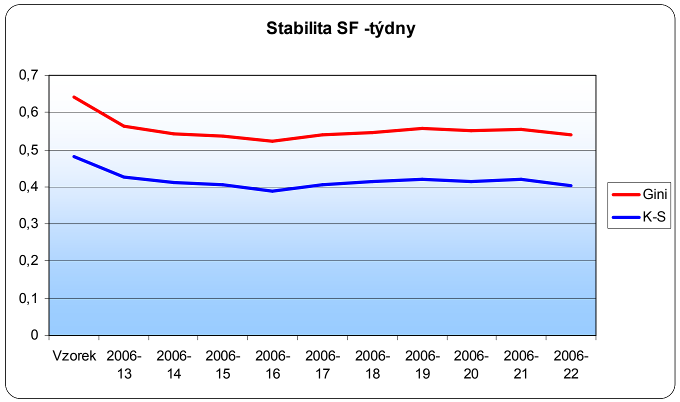
Stanovení cutoff, monitoring



	výv. vzorek [1]	týden1 [2]	[3]=[2] -[1]	[4]=[2]/[1]	[5]=ln[4]	[6]=[3]*[5]
skóre_1	10,00%	5,63%	-0,044	0,563	-0,574	0,025
skóre_2	10,00%	11,21%	0,012	1,121	0,114	0,001
skóre_3	10,00%	11,00%	0,010	1,100	0,095	0,001
skóre_4	10,00%	10,97%	0,010	1,097	0,092	0,001
skóre_5	10,00%	10,31%	0,003	1,031	0,031	0,000
skóre_6	10,00%	10,12%	0,001	1,012	0,012	0,000
skóre_7	10,01%	9,62%	-0,004	0,961	-0,039	0,000
skóre_8	10,00%	9,89%	-0,001	0,989	-0,011	0,000
skóre_9	10,00%	10,31%	0,003	1,031	0,030	0,000
skóre_10	10,00%	10,94%	0,009	1,095	0,091	0,001
					PSI	0,030



Možné zamítací škály – cutoff

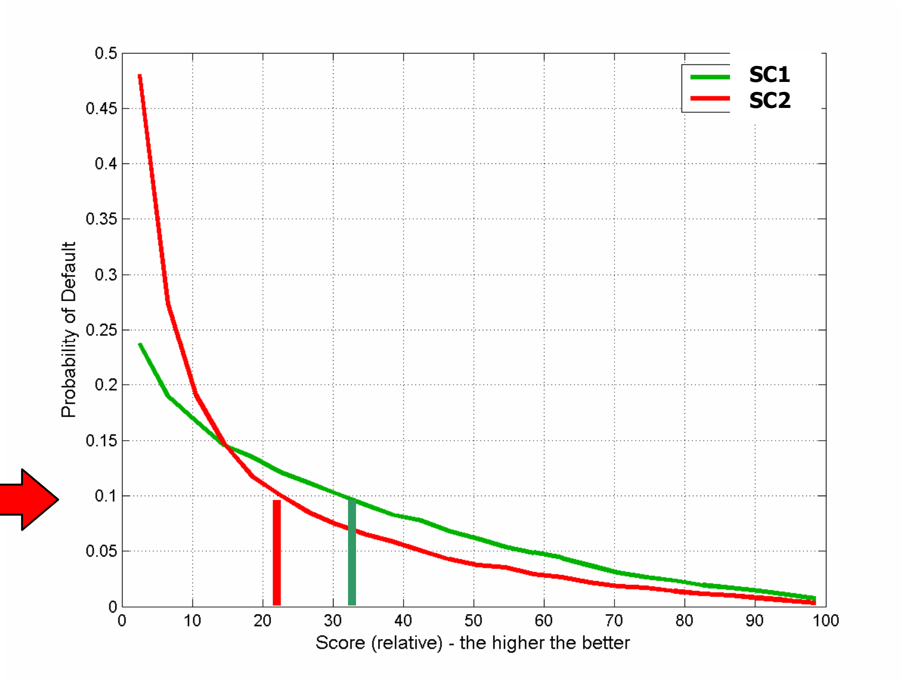
- cutoff hodnota určuje mez, při které je žádost o úvěr schválena/zamítnuta

- Je možné použít tyto zamítací škály:
  - **PD**
  - **KRN**
  - **Margin**
  - **RAROA**
  - **...**

cutoff na škále PD

cutoff = 0.1 (tj. zamítám všechny s pravděpodobností defaultu větší než 10 %)

Cut off



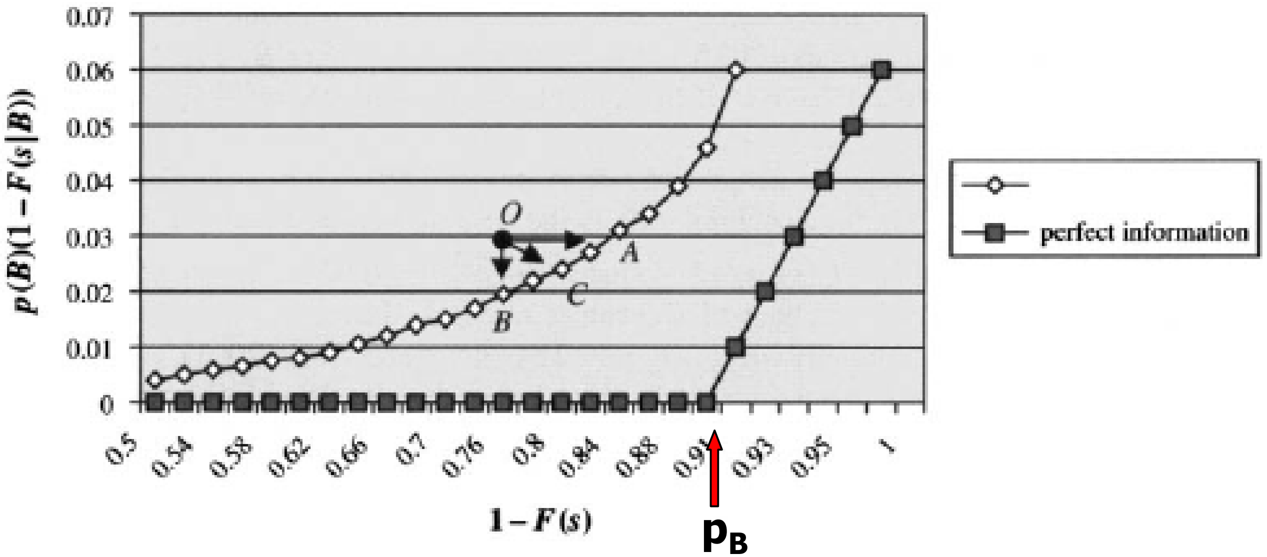
- Pro SC1 je reject rate 22 %.
- Pro SC2 je reject rate 33 %.

Strategy curve

$$\text{Bad acceptance rate} = p_B\left(1 - F\left(s \middle| B\right)\right)$$

$$\text{Acceptance rate} = 1 - F(s)$$

$$\text{Actual bad rate} = \frac{p_B\left(1 - F\left(s \middle| B\right)\right)}{1 - F(s)}$$



What often happens on introducing a new scorecard is that the existing operating policy gives a point *O* that is above the new strategy curve. The question then is where on the strategy curve one wants to go by choosing the appropriate cutoff. If one moves to *A,* then one keeps the bad acceptance rate the same but accepts more people, while moving to *B* would accept the same numbers but lower the bad acceptance rate and hence the bad rate. Moving to *C* would keep the bad rate the same and again increase the numbers accepted.

Nastavení cutoff maximalizující profit

Profit:

$$R = \begin{cases} 0 & \text{if the account is rejected,} \\ L & \text{if the account is accepted and becomes good } (L \text{ is lost profit from ruling out a good),} \\ -D & \text{if the account is accepted and becomes bad (so } D \text{ is the default amount).} \end{cases}$$

The expected profit per consumer if one accepts those with score s is

$$E\{R|s\} = Lq(G|s) - D(1 - q(G|s)) = (L + D)q(G|s) - D.$$

Thus to maximize profit, one should accept those with scores s if $q(G|s) \geq \frac{D}{D+L}$. Let A be the set of scores where the inequality holds; then the expected profit per consumer from the whole population is

$$E^*\{R\} = \sum_{s \in A} ((L + D)q(G|s) - D)p(s).$$

Nastavení cutoff maximalizující profit

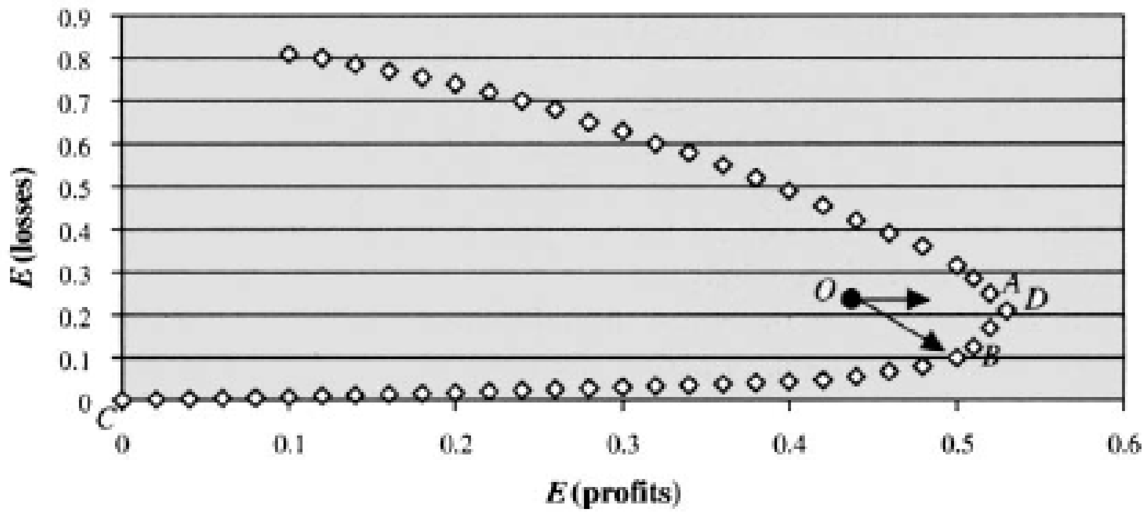
Notice that if the profits and default losses $L(s)$ and $D(s)$ were score dependent, the decision rule would be to accept if $q(G|s) \geq \frac{D(s)}{D(s)+L(s)}$. In that case, A may consist of several regions of scores, i.e., accept if the score is between 300 and 400, accept if between 500 and 550, and accept if over 750; reject otherwise. However, we assume that the profits and losses are independent of the score and that $q(G|s)$ is monotonically increasing in s. This means that $A = \{s|s \geq c\}$, where $q(G|s) \geq \frac{D}{D+L}$, so c is the cutoff score. In this case, define $F(s|G)$, $F(s|B)$ to be the probabilities a good or a bad has a score less than s:

$$\begin{aligned} E^*\{R\} &= \sum_{s \geq c} ((L+D)q(G|s) - D)p(s) = \sum_{s \geq c} (Lp_G p(s|G) - Dp_B p(s|B)) \\ &= Lp_G(1 - F(c|G)) - Dp_B(1 - F(c|B)) \\ &= Lp_G - Dp_B + (Dp_B F(c|B) - Lp_G F(c|G)). \end{aligned} \tag{14.4}$$

The first term on the right-hand side, $Lp_G - Dp_B$, is the profit if we accept everyone, so the second term is the profit that the scorecard brings. One can rewrite (14.4) in a different way to say how far away $E^*\{R\}$ is from the expected profit if there were perfect information $E\{PI\}$. With perfect information, one would only accept goods, so $E\{PI\} = Lp_G$. Hence

$$\begin{aligned} E^*\{R\} &= Lp_G - (Dp_B(1 - F(c|B)) + Lp_G F(c|G)) \\ &= E\{PI\} - (Dp_B(1 - F(c|B)) + Lp_G F(c|G)). \end{aligned}$$

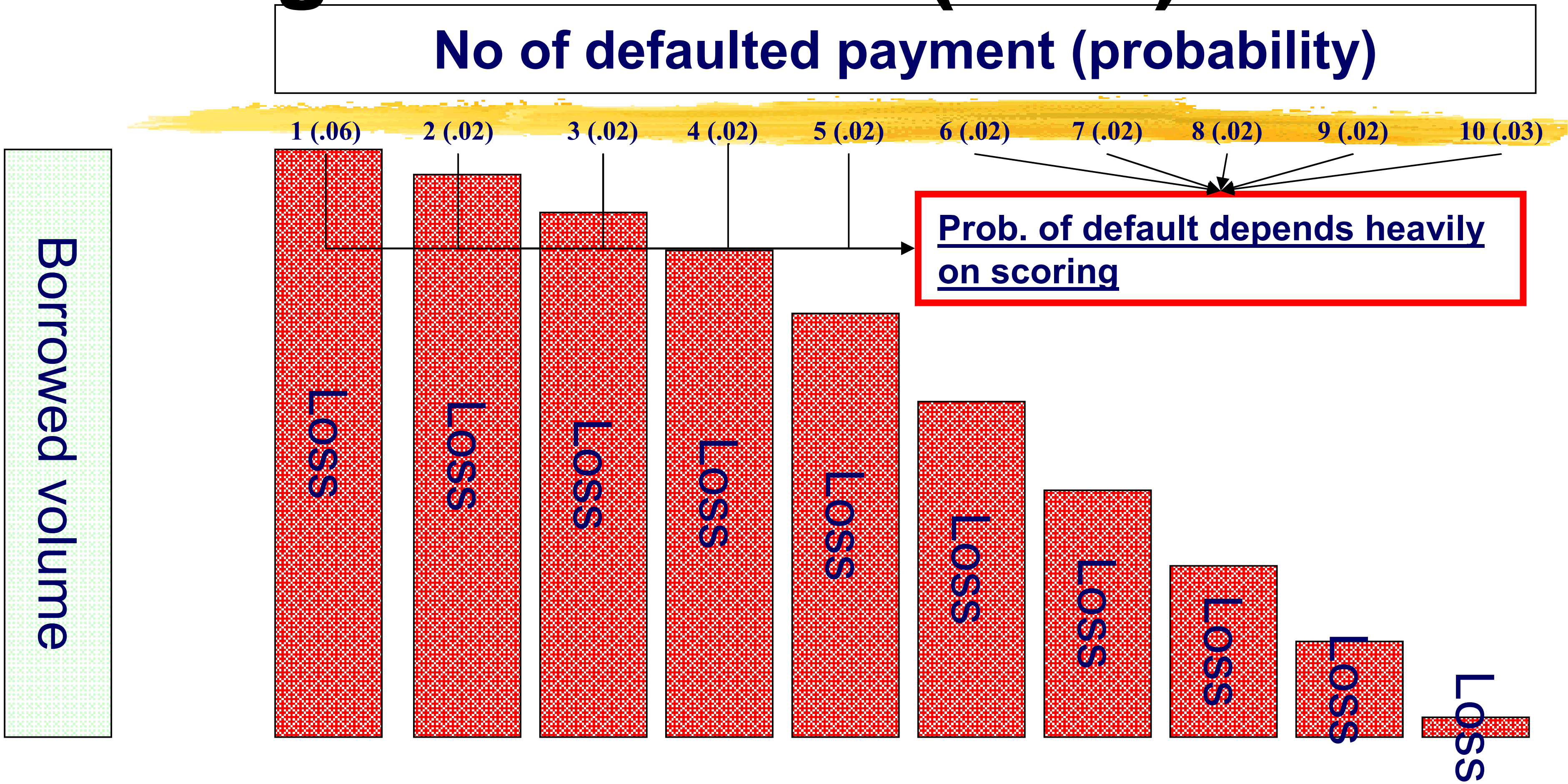
Nastavení cutoff maximalizující profit



The points on the lower part of the curve have higher cutoff scores, so less bads are accepted; those on the higher part of the curve correspond to lower cutoff scores with higher numbers of bads accepted. The efficient frontier of this curve is the lower part from C to D. These give points with an expected profit and an expected loss, in which the former cannot be raised without the latter also being raised.

If a lender is at present operating at point *O,* then again he can move onto the new scorecard curve either by keeping the bad acceptance rate the same, i.e., *A,* or by keeping the acceptance rate the same (which would move to point *B* on the curve). In this case, one would suggest that the move to *A* is less sensible because it is not an efficient point and one could have the same expected profit with lower expected losses.

Background of CRE (KRN)



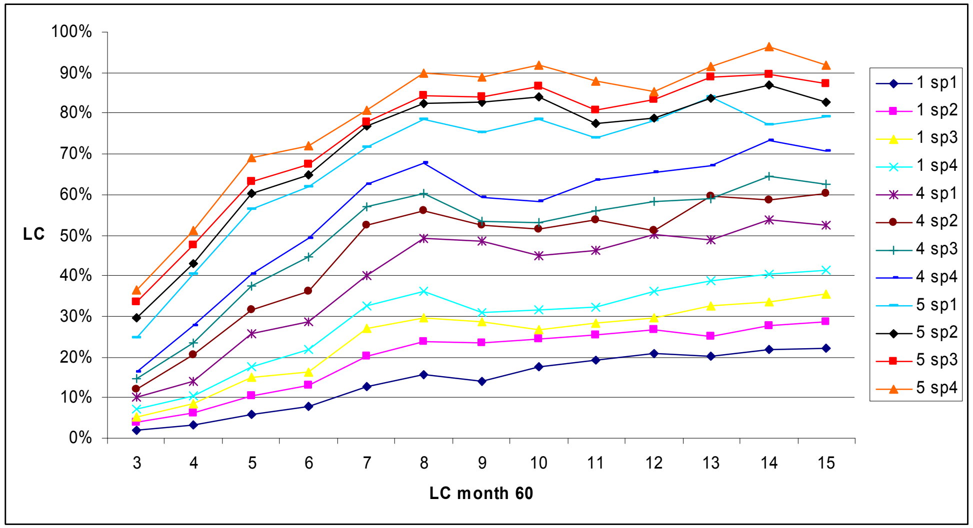
CRE = ((1-Recovery) * SUM(PD * Loss))/(Expected Average Volume)

Profit = (Interest rate – CRE)*Expected Average Volume

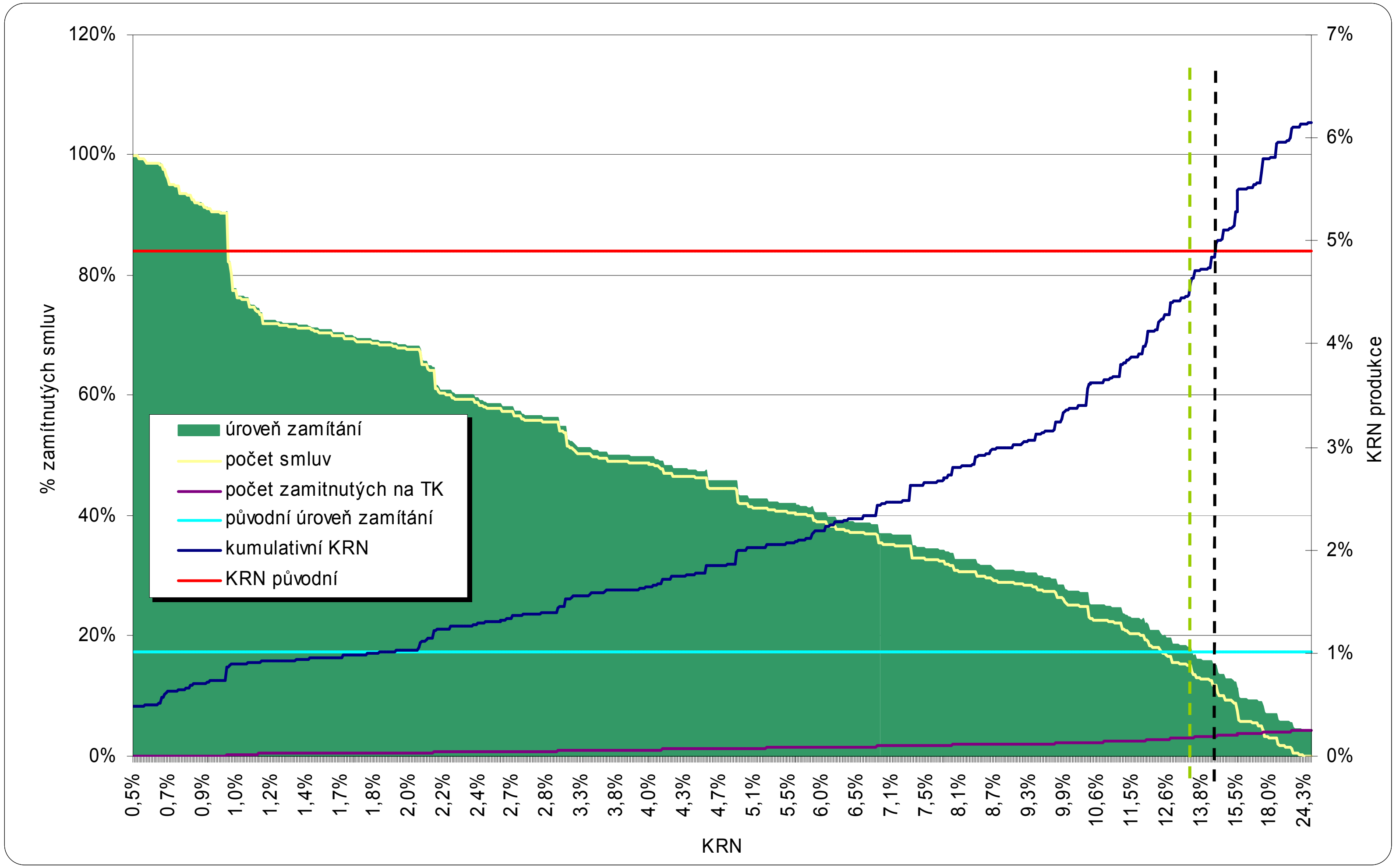
Recovery (=Late collection(LC))

Default occured at payment	Fraud score			
	band1	band2	band3	band4
1.	**20%**	**25%**	**30%**	**35%**
2.-4.	**50%**	**55%**	**60%**	**65%**
5. +	**75%**	**80%**	**85%**	**90%**

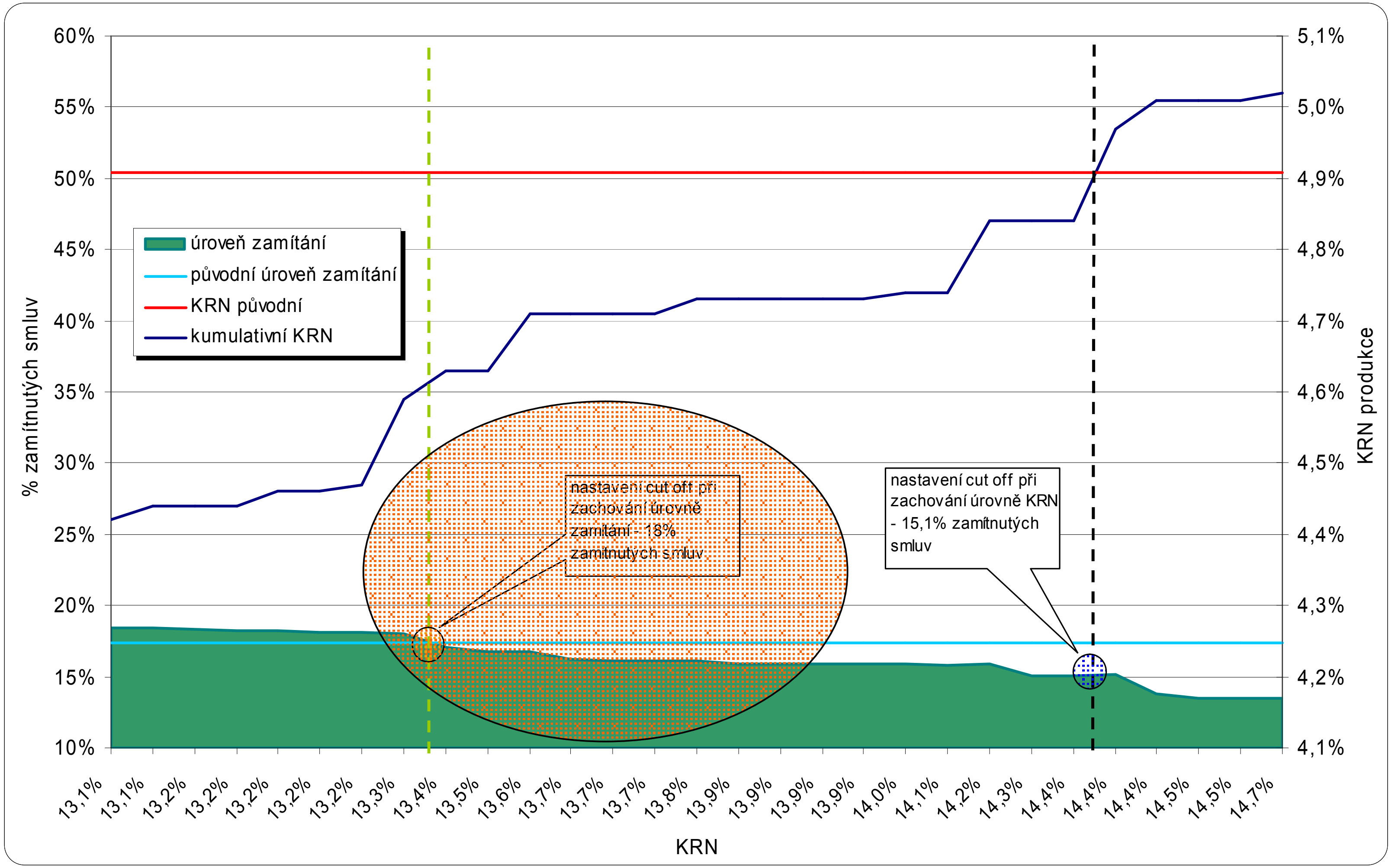
estimation



cutoff na škále KRN



cutoff na škále KRN

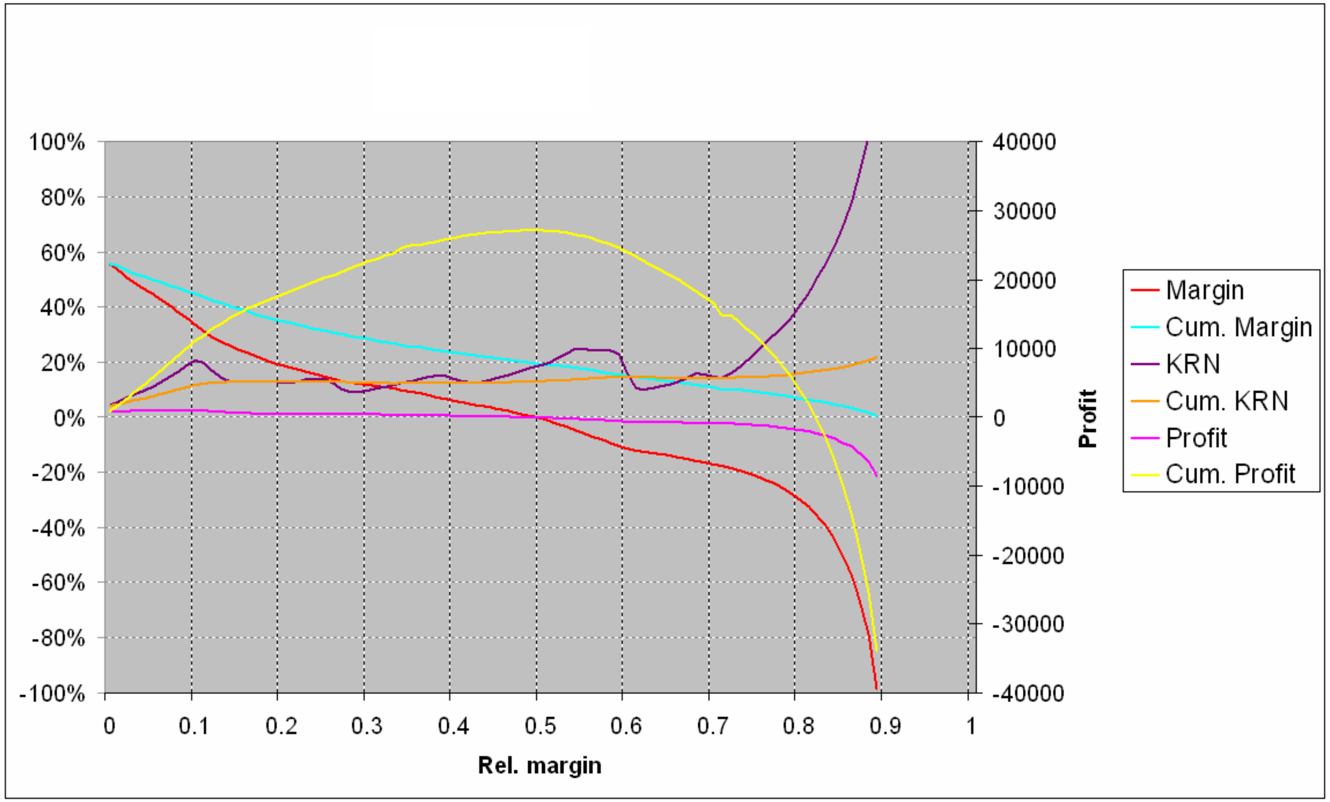


(Expected) Margin

(Expected) Margin = Interest Rate (incl. commiss.) – KRN – OPEX

- ***Interest rate***
  - *effective rate of the ideal flow (-amount-commiss.; annuity; annuity; ... ; annuity)*
- ***KRN***
  - *see previsous slides*
- ***OPEX***
  - *Cost of money*
  - *Overhead costs, var. costs, partner support*
  - *Administrators*

Margin



- Optimální cutoff: *marže=0*

RAROA (Risk Adjusted Return On Assets)

RAROA = (expeted income –expected loss) /borrowed volume

RAROA

- $t \ldots$ pořadí splátky úvěru, 0 je okamžik poskytnutí úvěru
- $T \ldots$ počet splátek
- $x(t) \ldots$ nesplacená část úvěru podle splátkového plánu v čase t, $\ldots t = 0, \ldots, T$. $x(0)$ je výše úvěru, $x(T) = 0$.
- $u(t) \ldots$ úroková část anuity t, $t = 1, \ldots, T$.
- $j(t) \ldots$ část anuity odpovídající splátce jistiny t, $t = 1, \ldots, T$.
- $k(t) \ldots$ komise od klienta v čase t, $t = 0, \ldots, T$.
- $A \ldots$ výše anuity (absolutně). $A = u(t) + j(t), t = 1, \ldots, T$.
- $p(t) \ldots$ pravděpodobnost 90 denního defaultu úvěru na splátce t, $t = 1, \ldots, T$
- $\mathrm{E}\, Z \ldots$ očekávaná ztráta z úvěru
- $\mathrm{E}\, P \ldots$ očekávaný úrokový příjem z úvěru
- $RC \ldots$ absolutní výše z dlužné částky klienta 90 dní po splatnosti, která je klientem splacena v budoucnu, přepočtena přes NPV k okamžiku devadesátidenního defaultu klienta
- $r(t, f) \ldots$ procento výtěžnosti z dlužné částky klienta, který je poprvé 90 dní po splatnosti na splátce t a klient má hodnotu podvodnického skóre (nesplacení první splátky) f. Procento zohledňuje NPV všech budoucích splátek klienta po okamžiku defaultu.

RAROA

- $GM\ldots$ hrubý očekávaný zisk z klienta
- s je sazba úvěru p.a.
- $i\ldots$ cena zdrojů vyjádřená v procentu p.a.
- $c\ldots$ komise z obchodu poskytnutá obchodnímu partnerovi vyjádřená jako procento z jistiny
- $NM_I\ldots$ čistý očekávaný zisk typu I z klienta po odečtení ceny zdrojů
- $NM_{II}\ldots$ čistý očekávaný zisk z klienta typu II po odečtení ceny zdrojů a komisí z obchodu.""
- $ROA\ldots$ ukazatel Return on Asset počítaného z hrubého zisku
- $ROA_I\ldots$ ukazatel Return on Asset typu I počítaný z čistého zisku typu I
- $ROA_{II}\ldots$ ukazatel Return on Asset typu II počítaný z čistého zisku typu II
- KRN je úroková míra p.a. vyjadřující rizikovost úvěru.

RAROA

$$\mathrm{E}\,Z = \sum_{t=1}^{T} p(t)\cdot x(t-1).$$

$$\mathrm{E}\,P = k(0) + \sum_{t=1}^{T}\left(1-\sum_{s=1}^{t} p(s)\right)(u(t)+k(t)).$$

$$GM = \mathrm{E}\,P - \mathrm{E}\,Z + RC.$$

$$RC = \sum_{t=1}^{T} p(t)\cdot r(t,f)\cdot x(t-1)$$

$$NM_I = GM - \sum_{t=1}^{T}\left(1-\sum_{s=1}^{t} p(s)\right)\frac{i}{12}\cdot x(t-1).$$

$$NM_{II} = NM_I - c\cdot x(0).$$

$$ROA = \frac{GM}{x(0)}.$$

$$ROA_I = \frac{NM_I}{x(0)}.$$

$$ROA_{II} = \frac{NM_{II}}{x(0)}.$$

$$\frac{\mathrm{KRN}}{12}\cdot\sum_{t=1}^{T}\left(1-\sum_{s=1}^{t} p(s)\right)x(t-1) = EZ - RC.$$

$$\sum_{t=1}^{T}\left(1-\sum_{s=1}^{t} p(s)\right)x(t-1) = \frac{\sum_{t=1}^{T}\left(1-\sum_{s=1}^{t} p(s)\right)u(t)}{s/12},$$

$$\mathrm{KRN} = \frac{EZ-RC}{EP}\cdot s.$$

Výhody RAROA

	Case A		Case B	
	Ideal flow	Expected flow	Ideal flow	Expected flow
	-1000	-1000	-1000	-1000
1	400	200	150	110
2	400	180	150	100
3	400	170	150	90
4	400	160	150	80
5			150	70
6			150	60
7			150	50
8			150	40
9			150	30
10			150	16
11			150	10
12			150	0

- Case A – short term loan with high fraud risk
- Case B – long term loan with high default risk

Interest rate (A) = 22%
Interest rate (B) = 10%

CRE(A) = 44%
CRE(B) = 20%
cutoff on CRE prefers B

Gross margin (A) = -22%
Gross margin (B) = -10%
cutoff on Margin prefers B

RAROA (A) = -0.29
RAROA (B) = -0.36
cutoff on RAROA prefers A

Case A is better since its return is larger (710>656) and sooner

cutoff Segmentace

- ***Možná segmentace podle:***
  - ***Prodejní síť (skupina obchodních míst)***
  - ***Profitabilita produktu***
  - ***Kvalita prodejního místa***
  - ***Typ zboží (pro spotřebitelské úvěry)***
  - ***Výše úvěru***
  - ***...***

cutoff scénáře

all

scenario	All credits		Approved credits			
	Reject rate		Avg. margin		Avg. KRN	Avg. profit
	Credits	Volume	Credits	Volume		
0reject	0.0%	0.00%	-7.83%	-21.34%	30.70%	-96302.30%
tk	22.4%	24.74%	-3.97%	-15.72%	26.33%	-69473.37%
current	36.8%	46.79%	2.32%	-4.58%	19.11%	-19232.49%
all30	29.5%	37.07%	3.18%	-3.36%	19.78%	-17071.60%
total30	30.3%	38.55%	4.11%	-1.29%	19.26%	-9208.76%
total32	31.5%	39.99%	4.63%	-0.77%	18.96%	-7032.25%
total35	35.8%	46.01%	7.32%	3.22%	16.07%	9703.64%
all40	38.9%	48.97%	8.17%	3.41%	15.19%	10427.31%
tk_ekonom	59.3%	70.03%	19.39%	17.14%	13.47%	54317.19%
ekonom	50.9%	63.64%	19.24%	17.03%	14.23%	54614.34%

new

scenario	All credits		Approved credits			
	Reject rate		Avg. margin		Avg. KRN	Avg. profit
	Credits	Volume	Credits	Volume		
0reject	0.0%	0.00%	-4.19%	-16.40%	26.19%	-72361.94%
tk	9.0%	10.25%	-2.39%	-13.71%	24.50%	-60136.89%
current	24.5%	34.89%	3.26%	-3.42%	17.99%	-14785.60%
all30	16.5%	23.87%	4.07%	-2.34%	18.60%	-12845.12%
total30	17.3%	25.42%	4.85%	-0.59%	18.19%	-6273.57%
total32	18.6%	27.16%	5.36%	-0.03%	17.87%	-3975.72%
total35	23.2%	33.80%	7.74%	3.52%	15.34%	10723.25%
all40	26.7%	37.35%	8.61%	3.88%	14.45%	12045.96%
tk_ekonom	50.7%	62.90%	19.53%	17.26%	13.05%	53999.46%
ekonom	47.5%	60.46%	19.48%	17.23%	13.26%	54059.69%

Monitoring scoringových modelů

- Není překvapivé, že prediktivní modely se ve statistickém slova smyslu chovají nejlépe na vývojovém vzorku dat. Výstupy těchto modelů, např. skóre nebo rating klienta, jsou počítány pomocí jistých vzorců, jejichž koeficienty příslušející nezávislým proměnným (prediktorům) jsou odvozeny na datech vývojového vzorku. Posun distribuce výstupu daného modelu je pak zapříčiněn právě změnou vstupních hodnot modelu, tj. prediktorů, v průběhu času. V podstatě ihned (alespoň většinou) po nasazení prediktivního modelu do praxe dochází k jistému poklesu jeho prediktivní síly, který je způsoben určitou změnou vstupních hodnot modelu. Zásadní je v praxi nastavení takových procesů, které odhalí, že se tak děje, proč se tak děje a jak vážný problém to ve svých důsledcích znamená.

Monitoring scoringových modelů

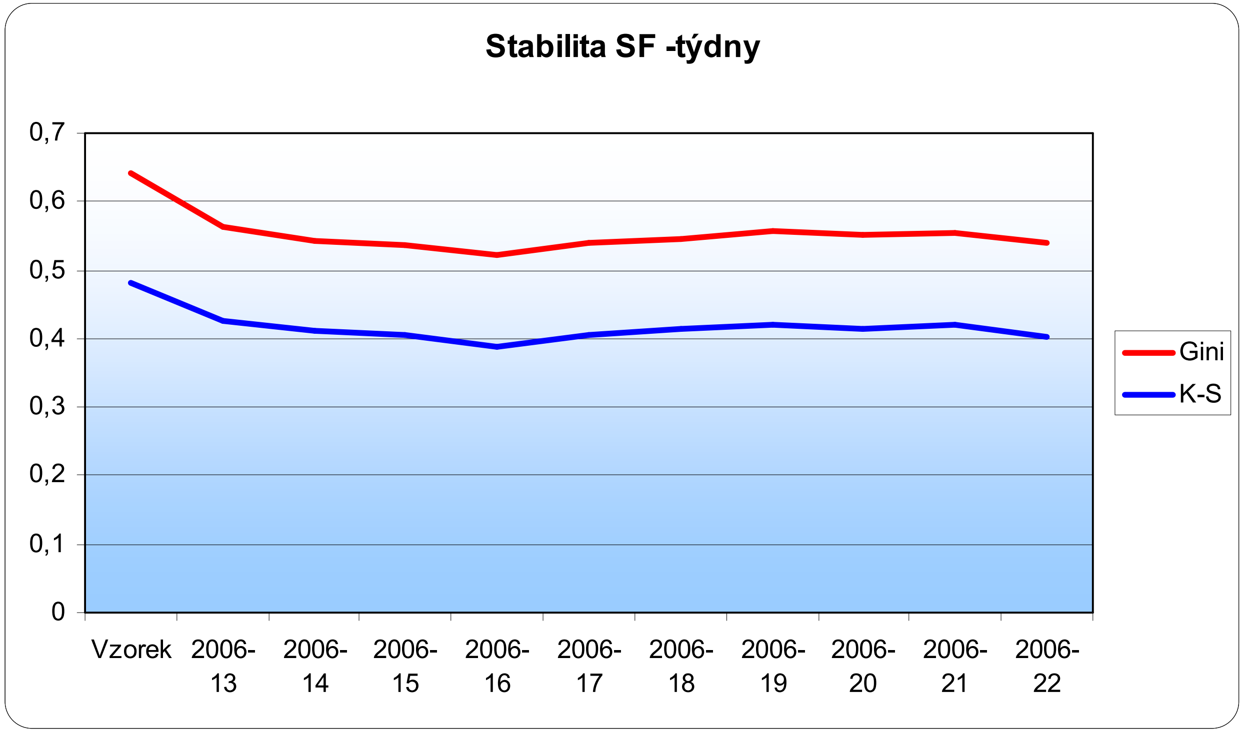
- Faktorů způsobujících posun v distribuci prediktorů, a následně posun v distribuci výstupu prediktivního modelu, je několik:
  - Přirozený posun v datech/změna demografické struktury dat
  - Databázové chyby
  - Změna datového zdroje
  - Změna definice/formátu vstupních dat
  - Změna datového univerza
  - Ostatní

Monitoring scoringových modelů

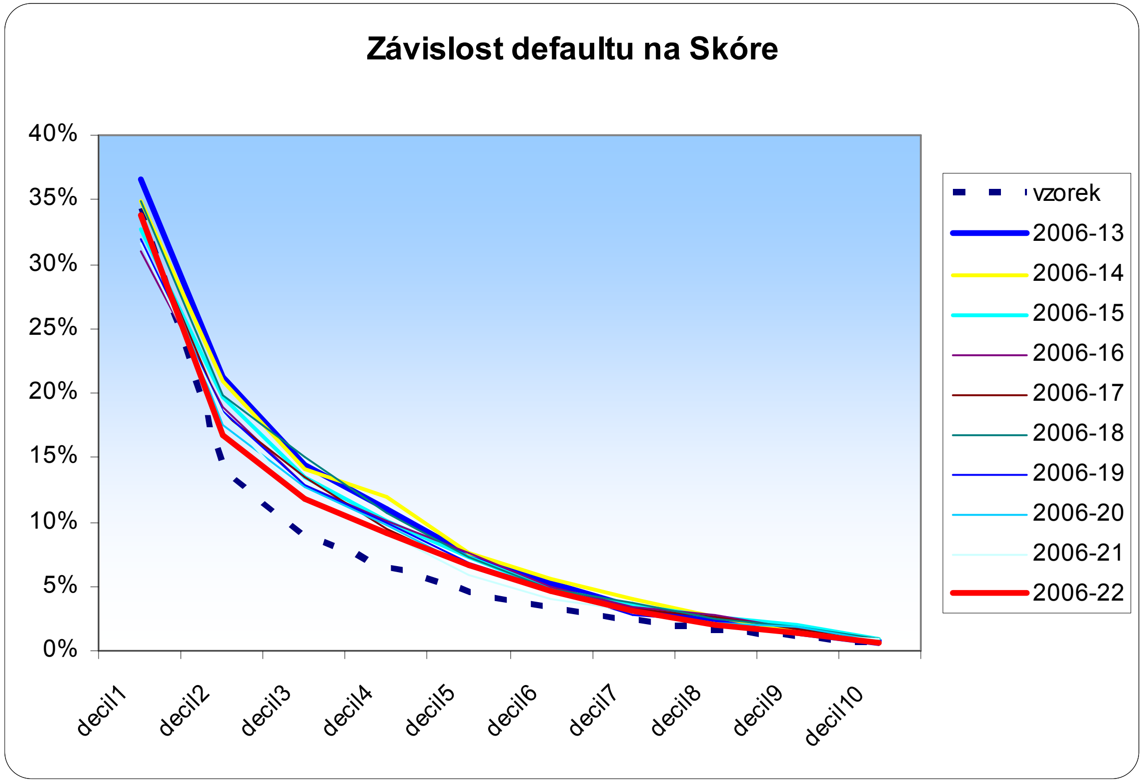
- Typickým příkladem prvního uvedeného důvodu je příjem klienta (všeobecným trendem je růst příjmu populace). Změnou definice/formátu vstupních dat je myšlena například situace, kdy je rozšířen číselník hodnot, kterých může vstupní proměnná nabývat. Změnou datového univerza je myšlen případ kdy je vyvinutý prediktivní model použit např.pro odlišný/nový segment portfolia nebo odlišný/nový produkt.

Monitoring scoringových modelů

- K-S, Gini:



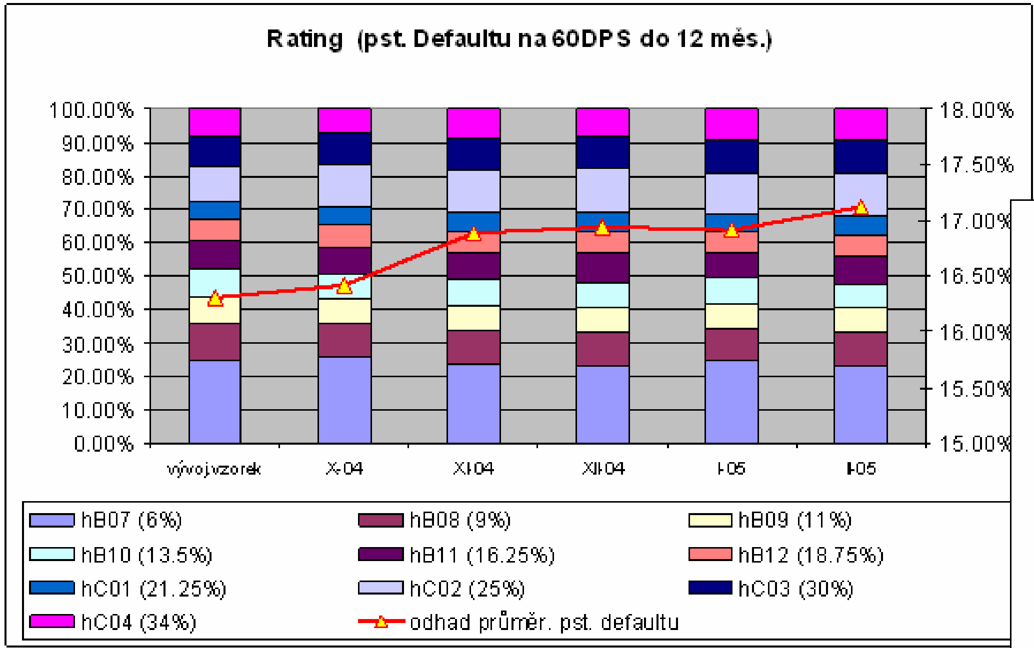
Monitoring scoringových modelů

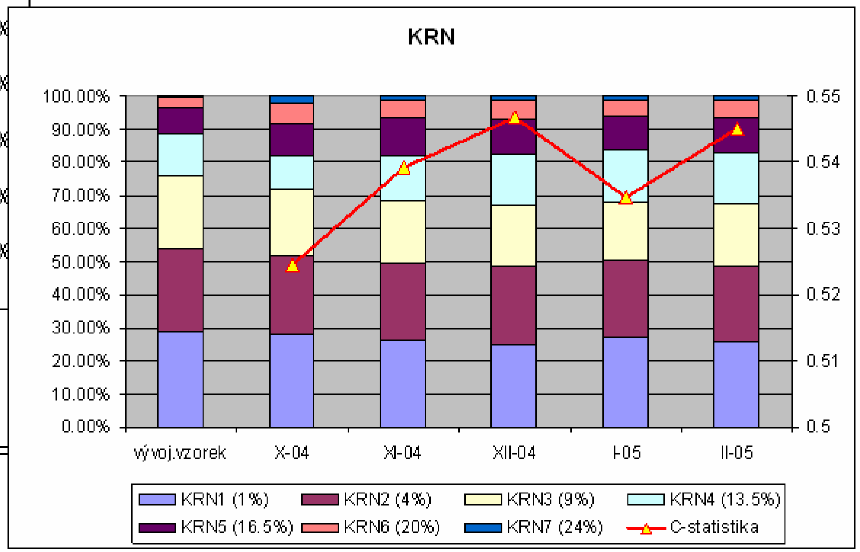


- Čím strmější křivka tím lépe.
- V průběhu času se splošťuje – jde o to, jak moc.

Monitoring scoringových modelů

- c-statistika:





Monitoring scoringových modelů

- Chceme posoudit zda se distribuce skóre na vývojovém vzorku liší od distribuce skóre v daném časovém intervalu:

$$\chi^2 = \sum_{i=1}^{r} \frac{(O_i - E_i)^2}{E_i}$$

$$PSI = \sum_{i=1}^{r} (O_i - E_i) \ln\left(\frac{O_i}{E_i}\right)$$

Monitoring scoringových modelů

	výv. vzorek [1]	týden1 [2]	[3]=[2] -[1]	[4]=[2]/[1]	[5]=ln[4]	[6]=[3]*[5]
skóre_1	10,00%	5,63%	-0,044	0,563	-0,574	0,025
skóre_2	10,00%	11,21%	0,012	1,121	0,114	0,001
skóre_3	10,00%	11,00%	0,010	1,100	0,095	0,001
skóre_4	10,00%	10,97%	0,010	1,097	0,092	0,001
skóre_5	10,00%	10,31%	0,003	1,031	0,031	0,000
skóre_6	10,00%	10,12%	0,001	1,012	0,012	0,000
skóre_7	10,01%	9,62%	-0,004	0,961	-0,039	0,000
skóre_8	10,00%	9,89%	-0,001	0,989	-0,011	0,000
skóre_9	10,00%	10,31%	0,003	1,031	0,030	0,000
skóre_10	10,00%	10,94%	0,009	1,095	0,091	0,001
					PSI	0,030

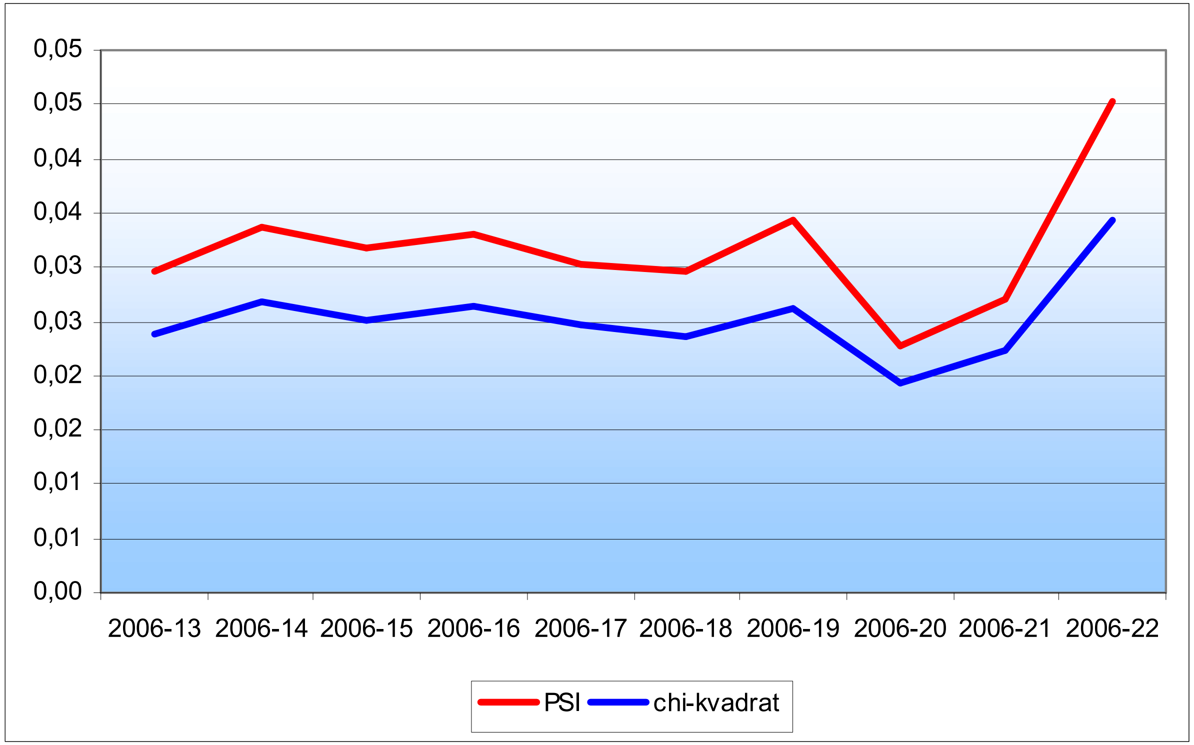
Monitoring scoringových modelů

$PSI \leq 0{,}1$ značí žádný nebo jen velmi malý rozdíl daných distribucí skóre.

$0{,}1 < PSI \leq 0{,}25$ znamená, že došlo k nějakému posunu distribuce, nicméně nikterak významnému.

$PSI > 0{,}25$ signalizuje významný posun v distribuci skóre, tj. zamítáme hypotézu o shodě daných distribucí.

Monitoring scoringových modelů

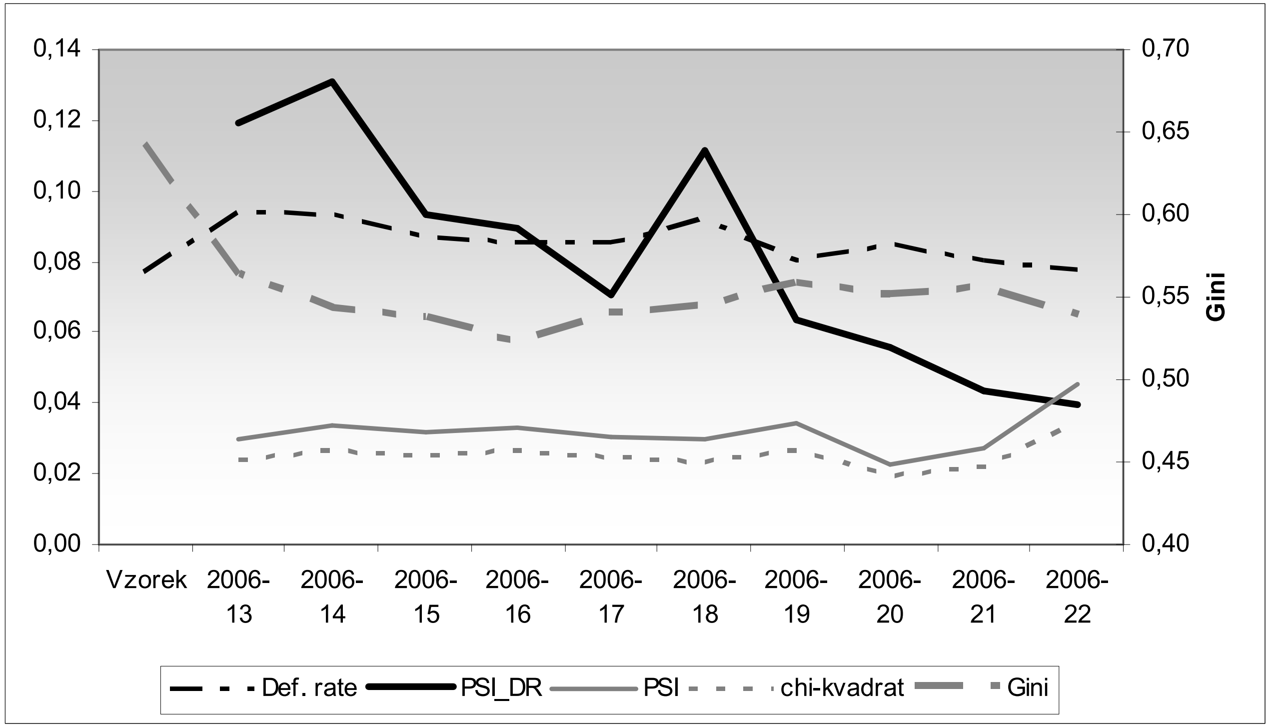


Monitoring scoringových modelů

$$PSI_{DR} = \sum_{i=1}^{r} (DR2_i - DR1_i) \ln\left(\frac{DR2_i}{DR1_i}\right)$$

	def_rate	Gini	PSI_DR	PSI	chi-kvardat
vzorek	7,69%	0,643			
200613	9,38%	0,564	0,120	0,030	0,024
200614	9,35%	0,542	0,131	0,034	0,027
200615	8,70%	0,537	0,093	0,032	0,025
200616	8,57%	0,523	0,089	0,033	0,026
200617	8,59%	0,540	0,071	0,030	0,025
200618	9,19%	0,544	0,111	0,030	0,024
200619	8,03%	0,558	0,063	0,034	0,026
200620	8,52%	0,552	0,055	0,023	0,019
200621	8,05%	0,555	0,043	0,027	0,022
200622	7,76%	0,539	0,039	0,045	0,034

Monitoring scoringových modelů



Champion-challenger

- The use of champion and challenger strategies became much more widespread through the 1990s and is now more accepted through credit and other areas. The principle is quite simple. There is an accepted way of doing something. This is known as the champion. However, there are one or more alternative ways to achieve the same (or a very similar) objective. These are known as the challengers. However, there is no evidence of the effect of the challengers. Therefore, on a random sample of cases, we try the challengers. This trial not only will test effectiveness in comparison with the champion but also will allow us the opportunity to identify the existence and extent of side effects. Eventually, we decide that one of the challengers is better overall than the champion, and this challenger becomes the new champion.