



cutting through complexity

Dynamické chování pojistníků

První jarní setkání České společnosti
aktuárů

31. květen 2013



Co znamená dynamické chování pojistníků?

Typy opcí a garancí

Motivace pro zahrnutí opcí a garancí do ocenění

Faktory ovlivňující chování pojistníků

Způsoby zahrnutí opcí a garancí v ocenění

- Explicitní vzorec
- Statistické metody
- Další důsledky dynamického chování pojistníků



Typy opcí a motivace

Příklady opcí a garancí

Storno;

Garantovaná anuitizační opce

Možnost dodatečného pojistného případně prodloužení pojistné doby

Zproštění od placení

Variable annuities

- **garantované plnění při úmrtí, odkup...**

Ze kterých požadavků vyplývá nutnost zohlednění dynamického chování pojistníků?

Market consistent embedded value, princip č. 7

„Allowance must be made in the MCEV for the potential impact on future shareholder cash flows of all financial options and guarantees within the in-force covered business. The allowance for the time value of financial options and guarantees must be based on stochastic techniques using methods and assumptions consistent with the underlying embedded value. All projected cash flows should be valued using economic assumptions that are consistent with the assumptions used to value the underlying business. They are valued in line with the price of similar cash flows in the market.”

Assumptions for future annual lapses are based on the historical experience of the company in Mexico.

MCEV model takes into account the dynamic nature of the business, including the impact of lapses on the value of the business.

Lapse rates from policyholders have been dynamically modeled. For traditional business, lapse rates depend on the difference between the credited rate to the policyholders and the anticipated policyholders' expectations.

Swiss Life

Směrnice Solvency II, pův. článek 78

„When calculating technical provisions, insurance and reinsurance undertakings shall take account of the value of financial guarantees and any contractual options included in insurance and reinsurance policies. Any assumptions made by insurance and reinsurance undertakings with respect to the likelihood that policyholders will exercise contractual options, including lapses and surrenders, shall be realistic and based on current and credible information. The assumptions shall take account, either explicitly or implicitly, of the impact that future changes in financial and non-financial conditions may have on the exercise of those options

IFRS 4, odst. 15

„...The test considers current estimates of all contractual cash flows, and of related cash flows such as claims handling costs, as well as cash flows resulting from embedded options and guarantees...“

- jaké přírážky zohledňují dobře cenu opcí?
- Je dobré používat konstantní přírážku i při změně nejlepšího odhadu předpokladu?

Ocenění produktů

nastavení srážek při odbytém, resp. jiné technické změně

Příklady faktorů ovlivňujících chování pojistníků

Standardní - produkt, délka trvání smlouvy, zprostředkovatel, lze věk, frekvence placení...

Typ zákazníka

- u movitějších, případně korporátních lze předpokládat racionálnější chování

Role zprostředkovatele

- lze lépe ovlivňovat interní síť, ale nelze přeceňovat
- nelze nutit nepříznivou variantu pro klienta.

Daně

- chovají se daňově uznatelné pojistky jinak?

Regulace

- zvýšení povinných informačních povinností
- dopad nového občanského zákoníku (nejen sjednání, ale i odkup)
- penzijní reforma

Příklady faktorů ovlivňujících chování pojistníků

Předchozí masové storno

- ekvivalent equity dampeneru
- obecně lze říci, že mnoho pojistek je odkupováno v souvislosti s potřebou peněžních prostředků
- US, 1951 – 45%,
 - jen 37% z důvodu, že není pojištění potřeba

Půjčky v pojistkách, sekundární trh s pojistkami

Marketingové akce (retence, nové smlouvy),

Vnímání pojistitele veřejností

- kdo by si nechal peníze v MSD, kdyby si je mohl převést?

Možné ekonomické faktory ovlivňující rozhodování

Úrokové sazby

Vývoj akciových indexů

Míra nezaměstnanosti

Růst HDP

Inflace

Otázky

- Existují spolehlivé projekce pro tyto proměnné?
- Jak rychle reaguje pojistník?
- Jsou proměnné zachyceny pomocí ESGs?

Krátký příklad

Parametry

- Pojištění na smrt a dožití s jednorázovým pojistným, pojistná doba 10 let
- Muž, 30 let, pojistná částka 10 000; nákladové přírážky

- Technická úroková míra nastavena variantně na 2% / 4% / 6%
- Storna pro ilustraci možného dopadu nastavena v prvních letech na 20%, 15%, 10% a potom 5% do konce.

- 80% investičního výnosu nad garanci je připsáno pojistníkům k rezervě

- Scénáře kalibrované na současné ekonomické podmínky.
- Aktiva kryjí rezervy amortizovanou hodnotou
- Na počátku investice do desetiletého zero-coupon dluhopisu, poté již jen prodeje z důvodu storen (převažují nad podíly na zisku)

Výsledky

- PVFP se snižuje se zvyšující se částkou
- Časová cena opcí a garancí roste se zvyšující se TIR
 - Ztráty jsou už zohledněny v PVFP, pouze v některých případech má

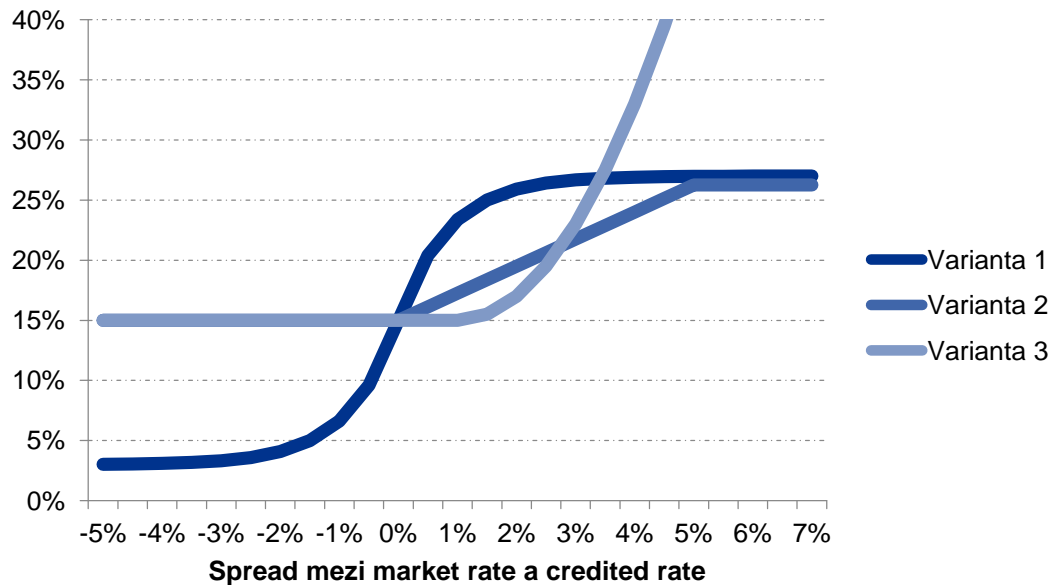
TÚM	PVFP	AVG (stoch. PVFP)	TVOG
2%	824	733	90
4%	693	618	75
6%	575	512	62

- S dynamickými storny – pojistníci mají tendence držet vyšší garanci (zahrnutí nákladovosti do rozhodování není v tomto zjednodušeném příkladu uvažováno)

TÚM	PVFP	AVG (stoch. PVFP)	TVOG
2%	623	551	72
4%	367	331	36
6%	279	256	23

Explicitně nastavené parametry, jak ekonomické proměnné ovlivňují storna lze považovat za nejjednodušší přístup

Příklady typů stornovosti jako funkce spreadu



- **Varianta 1 ...15%* (20%+80% * IF (spread <=0; exp(120%*100*spread); 2-exp(120%*100*spread))**
- **Varianta 2 ... IF (spread<=0; 15% ; 15% *(1+min(spread*15 ; 75%))**
- **Varianta 3 IF (spread<=1%; 15% ; 15% + min (200 * spread^2 – spread*4+2%) ; 90%)**



Metody založené na statistické analýze

Lineární regrese

- standardní metoda matematické statistiky

$$y_i = \beta x_i + \varepsilon_i, \text{ kde } \varepsilon_i \sim N(0, \sigma^2)$$

- Závislost stornovosti na dříve uvedených ekonomických faktorech
- Modelují nadstorno resp. podstorno.
- Pokud neuvažují vektor závislých proměnných (tzn. zanedbávám možnou změnu kmene a jeho neekonomických faktorů), situace se ještě „zjednoduší“ a lze fakticky standardně testovat nulovou hypotézu, zda stornovost závisí na vybrané tržní proměnné
- **Jak stanovit nejlepší odhad stornovosti bez zohlednění ekonomických faktorů?**
- Obecně zřejmě narazím se splněním předpokladů normality i lineární závislosti na parametru.

Tobit model

- Pokud modelují přímo stornovost, lze využít tzv. tobit model
- Model předpokládá, že existuje latentní (nepozorovatelná) proměnná, která je lineárně závislá na determinujících proměnných.
- Chyba je normálně rozložená
- Pozorovatelná proměnná je rovna latentní v případě, že je nezáporná, jinak je pozorovatelná proměnná nulová

$$y_i = \begin{cases} y_i^* & y_i^* > 0 \\ 0 & y_i^* \leq 0 \end{cases}$$

- $y_i^* = \beta x_i + \varepsilon_i$, kde $\varepsilon_i \sim N(0, \sigma^2)$
- Věrohodnostní funkce pro pozorování, které obsahuje právě k nul.
- $L = F_i^*(0)^k \prod_{y_i > 0} f_i(y_i) = [1 - \Phi\left(\frac{\beta' x_i}{\sigma}\right)]^k \prod_{y_i > 0} \Phi\left(\frac{y_i - \beta' x_i}{\sigma}\right)_i$
- V literatuře se zkoumá závislost stornovosti jednorázově placených odložených rent. Navržený model zahrnuje surrender charge, trvání smlouvy, rozdíl mezi 5ti letým státním dluhopisem a credited rate a růst HDP.



Zobecněný lineární model

Formální shoda řady regresních modelů:

- lineární modely,
- ANOVA,
- logistická regrese,
- loglineární modely,
- multinomické modely...

Snaha co nejúplněji postihnout vzájemnou souvislost různých jevů:

- škodní frekvence v závislosti na segmentaci,
- průměrná výše škody v závislosti na segmentaci,
- stornovost v závislosti na čemkoliv,
- marketing...

Metoda schopná správných předpovědí, zohledňující korelace i interakce.

Prakticky použitelná, tj. v běžné praxi nepřiliš složitá.

GLM – struktura modelu 1a

- Pozorujeme náhodnou veličinu Y , jejíž každou realizaci y_i (výsledek měření) považujeme za kombinaci systematické složky $E[Y]_i$ a náhodné složky ε_i .

$$y_i = E[Y]_i + \varepsilon_i = \mu_i + \varepsilon_i$$

- Systematickou složku se snažíme vyjádřit pomocí vysvětlujících veličin X , náhodná složka je generována podkladovým náhodným dějem, který je zodpovědný za rozdělení $\rho(y_i)$ veličiny Y .
- GLM umožňuje na základě historie (n měření) předpovídat systematickou složku pomocí zvolených vysvětlujících veličin a zároveň respektovat náhodnost podkladového děje.
- Bohužel ani závislost $\mu_i(x_1, \dots, x_p)$ ani rozdělení $\rho(y_i)$ nemohou být libovolné.

GLM – struktura modelu 2

- Předpokládáme, že systematická složka μ_i je prostřednictvím prosté a diferencovatelné funkce g , tzv. *spojovací (link) funkce*, spojena s tzv. *lineárním prediktorem* η_i , tj. lineární funkcí parametrů modelu.

$$g(\mu_i) = \eta_i \quad \Rightarrow \quad \mu_i = g^{-1}(\eta_i)$$

- V rámci GLM je tedy systematická složka μ_i funkcí lineárního prediktoru η_i .
- Dále předpokládáme, že rozdělení ρ veličiny Y je z tzv. *exponenciální rodiny* rozdělení. Pro tato rozdělení platí, že jsou plně určena střední hodnotou a rozptylem (mají až 2 volné parametry) a rozptyl je funkcí střední hodnoty.
- V modelu zvolíme spojovací funkci g , vysvětlující veličiny X , a na základě předpokladu o rozdělení ρ náhodné veličiny Y hledáme takové koeficienty lineárního prediktoru, aby model co nejlépe vystihoval výsledky měření.

GLM – vysvětlující proměnné 1

- Lineární prediktor je následující funkce

$$\eta_i = \sum_{j=1}^p x_{ij} \beta_j + \xi_i; i = 1, \dots, n$$

$$\vec{\eta} = \mathbf{X}\vec{\beta} + \vec{\xi}$$

- \mathbf{X} je tzv. konstrukční matice (*design matrix*) neboli matice $n \times p$, jejíž řádky odpovídají jednotlivým měřením a sloupce tvoří jednotlivé vysvětlující proměnné. Aby byl model jednoznačně definován, musí mít matice \mathbf{X} plnou sloupcovou hodnost.
- jsou koeficienty, které vyjadřují vliv jednotlivých vysvětlujících proměnných na modelovanou veličinu a jejichž hodnoty hledáme.
- ξ je tzv. *offset* neboli člen shrnující vlivy, jejichž efekt na modelovanou veličinu známe a nepotřebujeme jej tedy odhadovat.
- Vysvětlující veličiny, resp. proměnné, mohou být jak kvantitativní (spojité), například hmotnost, tak kvalitativní (kategoriální), například barva. Toto rozlišení je však často dáno spíše kontextem a volbou.

GLM – vysvětlující proměnné 2

- Kategoriálními proměnnými jsou hladiny (*levels*) jednotlivých kategoriálních veličin, faktorů (*factors*). Například veličina barva může mít několik hladin, které pak tvoří jednotlivé proměnné.
- Kategoriální proměnné jsou takové, pomocí nichž sledujeme, zda měření patří nebo nepatří do nějaké kategorie. Nabývají tedy typicky hodnot 1 – patří, 0 – nepatří (Dummy variables).
- Hladiny lze zakódovat různě (1,0;-1,1;...) \Rightarrow matice kontrastů (*contrast matrix*).
- U kategoriálních proměnných může snadno dojít k lineární závislosti. Například pro proměnné muž a žena, by platilo muž=1-žena. Tyto závislosti ohrožují hodnost design matrix, a tedy určitost modelu \Rightarrow je třeba správně zvolit kontrasty.
- Absolutní člen (*intercept*) β_0 , který v sobě obsáhne všechny základní hladiny faktorů reprezentovaných kategoriálními proměnnými takové obtíže řeší. Všechna měření pak obsahují tento absolutní člen (základní hladinu) a proměnné popisují pouze odlišnost od této reference. Máme pak jen nezávislé proměnné a absolutní člen.

$$\eta_i = \sum_{j=1}^{p'} x_{ij} \beta_j + \beta_0 + \xi_i; i = 1, \dots, n$$

GLM – exponenciální rodina rozdělení

- Hustota pravděpodobnosti exponenciální rodiny rozdělení má obecně tvar

$$\rho(y_i; \theta_i, \phi) = \text{Exp} \left(\frac{y_i \theta_i - b(\theta_i)}{a_i(\phi)} + c(y_i, \phi) \right)$$

- θ je *kanonický parametr* související se střední hodnotou, ϕ je *rozptylový parametr* související s rozptylem, $a_i(\phi)$ je spojitá a kladná funkce, $b(\theta)$ (kumulantová funkce) je dvakrát diferencovatelná konvexní funkce a $c(y, \phi)$ je funkce normující ρ , nezávislá na θ .

$$E[y_i] = \mu_i = \left. \frac{db}{d\theta} \right|_{\theta_i} = b'(\theta_i)$$

$$\text{var}(y_i) = a_i(\phi) \left. \frac{d^2 b}{d\theta^2} \right|_{\theta_i} = a_i(\phi) b''(\theta_i) = a_i(\phi) V(\mu_i)$$

- V je varianční funkce, obvykle $a_i(\phi) = \phi / w_i$, kde w_i je apriorní váha i -tého měření

GLM – exponenciální rodina rozdělení 1

	Označení	Definiční obor	ϕ	$b(\theta)$	$c(y, \phi)$	$\mu(\theta)$	$\theta(\mu)$ (kanonický link)	$V(\mu)$
Normální	$N(\mu, \sigma^2)$	$(-\infty, \infty)$	σ^2	$\frac{\theta^2}{2}$	$-\frac{1}{2} \left(\frac{y^2}{\phi} + \ln(2\pi\phi) \right)$	θ	μ	1
Poissonovo	$P(\mu)$	0,1,2,...∞	1	e^θ	$-\ln(y!)$	e^θ	$\ln(\mu)$	μ
Binomické	$Bi(m, \pi)/m$	$\frac{0,1,2,\dots,m}{m}$	$\frac{1}{m}$	$\ln(1 + e^\theta)$	$\ln \binom{m}{my}$	$\frac{e^\theta}{1 + e^\theta}$	$\ln \left(\frac{\mu}{1 - \mu} \right)$	$\mu(1 - \mu)$
Gamma	$G(\mu, \nu)$	$(0, \infty)$	$\frac{1}{\nu}$	$-\ln(-\theta)$	$\nu \ln(\nu y) - \ln(y) - \ln(\Gamma(\nu))$	$-\frac{1}{\theta}$	$\frac{1}{\mu}$	μ^2
Inverzní Gaussovo	$IG(\mu, \sigma^2)$	$(0, \infty)$	σ^2	$-\sqrt{-2\theta}$	$-\frac{1}{2} \left(\ln(2\pi\phi y^3) + \frac{1}{\phi y} \right)$	$\frac{1}{\sqrt{-2\theta}}$	$\frac{1}{\mu^2}$	μ^3

„Blízcí příbuzní:“ negativně binomické, Weibulovo, ... (Lognormální NE)

Sestavení a vyhodnocení modelu

Rozdělení

- Analýza rozdělení sledované veličiny, porovnání výsledků modelu se skutečností

Spojovací funkce

- Praktičnost
- Realističnost

Vysvětlující proměnné, design matrix

- Volba veličin
- Volba hladin kategoriálních veličin
- Zahrnutí interakcí
- Analýza vlivu jednotlivých proměnných na výsledky modelu

Sestavení modelu – rozdělení

- Volba rozdělení vychází z předchozí znalosti, zkušeností a podstaty podkladového náhodného děje.
- Správnost volby lze (ne nezávisle na zbytku modelu) ověřit pomocí různých měr rozdílů, reziduí, mezi měřenými a modelem předpovídanými hodnotami.
- Vhodnou volbou jsou tzv. devianční rezidua, která jsou při správné volbě modelu velmi dobře normálně rozdělena.

$$D = \sum_{i=1}^N r_{Di}^2, \quad r_{Di} = \text{sign}(y_i - \mu_i) \sqrt{d_i} = \text{sign}(y_i - \mu_i) \sqrt{2 \int_{\mu_i}^{y_i} \frac{y_i - t}{V(t)} dt}$$

- Standardizovaná devianční rezidua mají navíc jednotkový rozptyl.

$$r_{DSi} = \frac{r_{Di}}{\sqrt{\phi(1-h_i)}} = \frac{\text{sign}(y_i - \mu_i)}{\sqrt{\phi(1-h_i)}} \sqrt{2 \int_{\mu_i}^{y_i} \frac{y_i - t}{V(t)} dt}$$

- h_i jsou diagonální prvky vlivové matice (*hat-matrix*) tzv. páky (*leverage*), které popisují vliv i -tého měření na model, 1 – velký vliv, 0 – malý vliv

Sestavení modelu – spojovací funkce

- Kanonický link zjednodušuje tvar věrohodnostní funkce, a jeho použití má i jiné příznivé důsledky, které však dnes, díky počítačům, nejsou rozhodující.
- Rozhodují data a praktičnost – v pojišťovnictví je zpravidla příjemný multiplikativní model s logaritmem jako spojovací funkcí. Pro binomické modely je třeba link, který zobrazuje hodnoty z intervalu $\langle 0, 1 \rangle$ na $\langle -\infty, \infty \rangle$ - např. kvantilové funkce.
- Testovat lze maximum věrohodnostní funkce, kterého je možné dosáhnout s různými spojovacími funkcemi.

$$g(x; \lambda) = \begin{cases} \frac{x^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \ln(x), & \lambda = 0 \end{cases}$$

- $g(x; \lambda)$ přechází od inverzní, pro $\lambda = -1$, přes logaritmickou, pro $\lambda = 0$, do identické, pro $\lambda = 1$, spojovací funkce, a nabízí tak možnost určit vhodnou spojovací funkci nalezením maxima věrohodnostní funkce v závislosti na λ , a vybrat tak spojovací funkci maximalizující věrohodnost.

Sestavení modelu – proměnné – testování

- Přidávány by měly být pouze proměnné, které model signifikantně vylepší.
- Standardní mírou dobré shody modelu je deviance D , repektive škálovaná deviance D^*

$$D = 2 \sum_{i=1}^n \int_{\mu_i}^{y_i} \frac{y_i - t}{V(t)} dt$$

$$D^* = 2 \sum_{i=1}^n \frac{1}{\phi} \int_{\mu_i}^{y_i} \frac{y_i - t}{V(t)} dt$$

- Dva vnořené modely lze tedy porovnávat srovnáním jejich škálovaných deviancí, pokud je parametr ϕ známý (např. u Poissonova rozdělení) (model ω je podmodelem modelu Ω).

$$D_{\omega}^* - D_{\Omega}^* = 2(\ell_{\Omega} - \ell_{\omega}) \sim \chi_{df_{\omega} - df_{\Omega}}^2, \quad df_{\omega} > df_{\Omega}$$

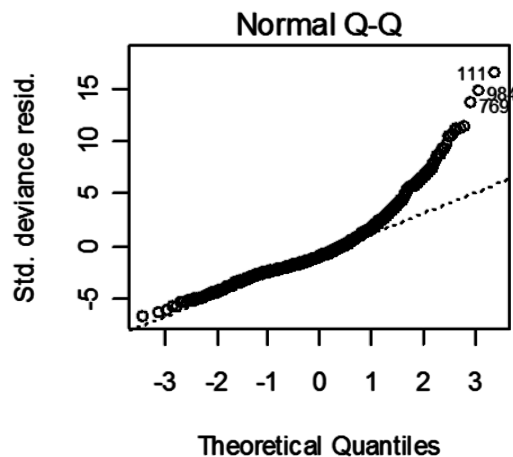
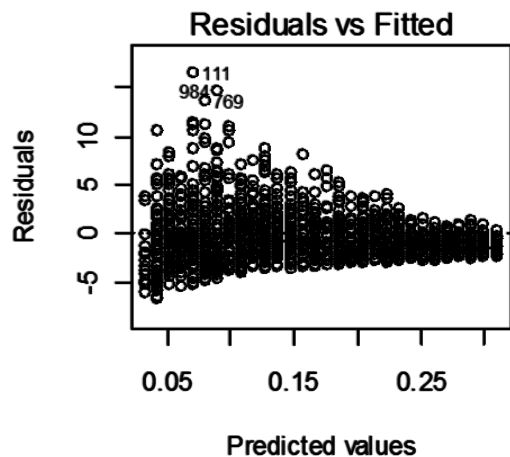
- Případně, pokud je i ϕ odhadované,

$$\frac{1}{\hat{\phi}} \frac{D_{\omega} - D_{\Omega}}{df_{\omega} - df_{\Omega}} \sim F_{df_{\omega} - df_{\Omega}, df_{\Omega}}, \quad df_{\omega} > df_{\Omega} \quad \hat{\phi} = \frac{X^2}{df} \quad \hat{\phi} = \frac{D}{df}$$

- Porovnávání různých modelů – Akaikeho informační kritérium

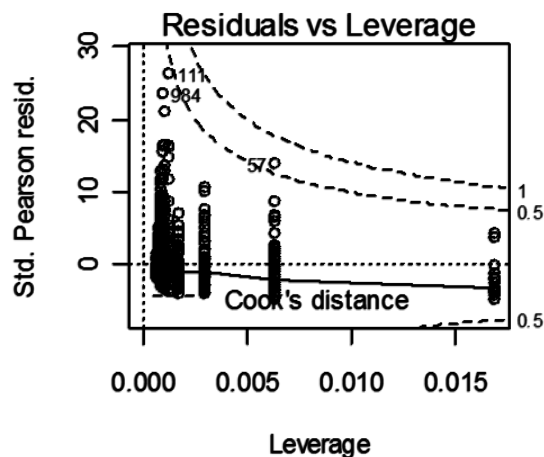
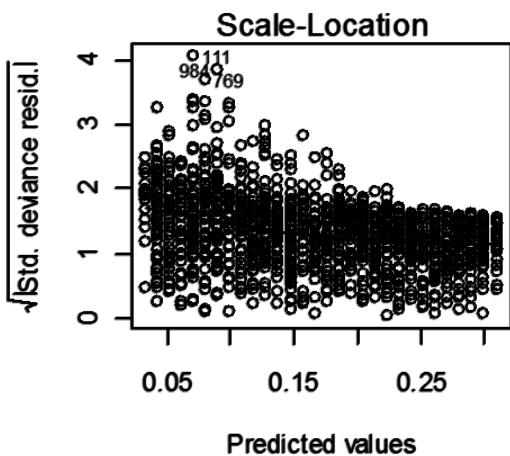
$$AIC = 2[-\ell(\hat{\beta}) + p] \vee 2[-\ell(\hat{\beta}) + p + 1]$$

Diagnostika modelu – obecně



Levý horní graf zobrazuje závislost reziduí na hodnotách lineárního prediktoru. Tento graf pomáhá vyhodnotit adekvátnost předpokladů o rozdělení, spojovací funkci, vysvětlujících proměnných aj.

Pravý horní graf zobrazuje QQ diagram, který porovnává rozdělení standardizovaných deviančních reziduí se standardním normálním rozdělením.



Levý dolní graf slouží k lepší identifikaci případných významných závislostí v rozptylu reziduí.

Pravý dolní graf, zobrazuje standardizovaná devianční rezidua a diagonální prvky projekční matice. Slouží k identifikaci měření s příliš velkým vlivem na odhad parametrů modelu. Zobrazeny jsou rovněž kontury Cookovy vzdálenosti.

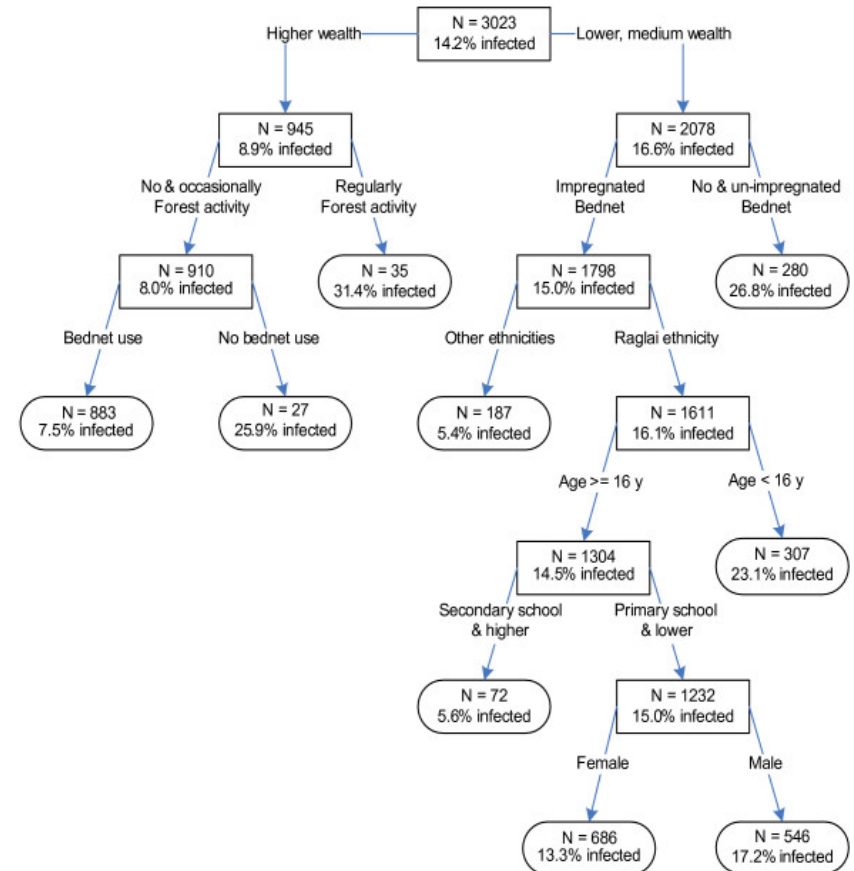


**Další možnosti
zohlednění opcí**

Jiné metody pro zohlednění opcí

Klasifikační stromy

- Chci rozdělit zkoumané osoby do skupin stejných charakteristik podle pravidel, která vycházejí ze vztahu storna k charakteristice
- Lze nastavit asymetrické penalizace za špatné zařazení
- Lze budovat zespona (postupná agregace listů) nebo seshora (větvení při dostatečném zlepšování modelu)
- Neviděl jsem zatím uplatněné v praxi pojišťovny



„Na základě těchto skutečností musíme dovodit, že snahy o zabránění časných storen přinesly malý efekt, ačkoliv víme výrazně více o charakteru kvalitního obchodu.

- a) Tomáš Síkora, výkonný ředitel ČAP
- b) Petr Borkovec, generální ředitel Partners Financial Services, a.s.
- c) Jaroslav Mlynář, generální ředitel České pojišťovny, a.s.
- d) nejmenovaný pojistný matematik

Již v roce 1951

Mohu vhodnou akcí (dopis, telefonát) odradit klienta od stornování?

- Na některé skupiny má akce nebude mít vliv – problém zbytečných nákladů
- U určitého typu klientů mohu svou akcí naopak storno způsobit

Jak vybrat vhodnou skupinu klientů? Dva možné přístupy:

- Zkoumám, která skupina osob je nejnáchylnější ke stornování (reálně pomíjí první problém, tedy zbytečné náklady u klientů, které neovlivním)

$$E[y_i | x_i; t = 0] \quad (y_i = 1 \text{ pro storno, } 0 \text{ jinak; } x_i - \text{ vysvětlující proměnné, } t = 0 \text{ „nevolám“})$$

- Zkoumám, u které skupiny „nejvíce získám“, tedy u které „v průměru“ nejvíce snížím pravděpodobnost stornování, pokud zavolám (UPLIFT)

$$E[y_i | x_i; t = 0] - E[y_i | x_i; t = 1] \quad (y_i \text{ a } x_i \text{ jako výše, } t = 0 \text{ „nevolám“, } t = 1 \text{ „volám“})$$

První možný přístup

- dva modely – jeden modelující pravděpodobnost storna v situaci, kdy nezavolám, druhý v opačném případě.
- Problém: rozdíl nezávislých „přesných“ modelů nemusí být „přesným“ modelem.

Druhý možný přístup:

- určovat přímo rozdíl mezi pravděpodobnostmi storna v obou případech
 - Model založený na logistické regresi. Vysvětlující proměnné jsou interakce mezi jednotlivými charakteristikami klienta a faktem, zda jsem zatelefonoval či nikoliv.
 - Do modelu musím přidat interakci volání se všemi ostatními proměnnými – složitý model, snižuje počet stupňů volnosti
 - Regresní model založený na předcházejícím posouzení relevance vysvětlujících proměnných
 - Modely založené na klasifikačních stromech: postupně přidávám kritéria rozdělující portfolio na skupiny podle toho, které je dle stanoveného klíče „nejvýznamnější“

Problém klasifikačního stromu: chyby v horních patrech se propagují do hloubky → vysoká variance výsledku.

- Možné řešení: generování více náhodných stromů z náhodně vybraných vzorků

Možné schéma modelu založeného na náhodně generovaných klasifikačních stromech.

- a) Vezmi náhodný vzorek velikosti M z tréninkových dat (bootstrap)
- b) Sestroj klasifikační strom. V každém uzlu
 - i. Vyber náhodnou skupinu z vysvětlujících znaků
 - ii. Z této skupiny vyber znak, který poskytuje „nejlepší split“ na dvě větve (jedna znak mající, druhá nikoliv) dle předem stanoveného kritéria
 - iii. Pokračuj dále, dokud velikost zbývajícího vzorku v uzlu není menší než stanovená mez
- c) Dle bodů a) a b) sestroj B „náhodných“ klasifikačních stromů. Předpovídaný „zisk“ v pravděpodobnosti se stanoví jako průměr ze „zisku“ vyplývající z jednotlivých stromů.

GLMM

- Zobecněný lineární model s náhodnými efekty

$$\eta_i = \sum_{j=1}^p x_{ij} \beta_j + U_i + W_{ij}; i = 1, \dots, n$$

- U_i, W_{ij} se uvažuje jako náhodná veličina, obvykle s normálním rozdělením. U_i reprezentuje odchylku výběru kategorie, W_{ij} individuální odchylku v rámci kategorie.

$$\eta_i = \sum_{j=1}^p x_{ij} \beta_j + \sum_{k=1}^q Z_{ik} u_{ik}, u_{ik} \sim N(0, \sigma_k^2)$$

- u_{ik} jsou parametry náhodných efektů a Z_{ik} jejich úrovně. η_i jsou potom korelované a je nutné tyto korelace odhadovat.
- Náhodné efekty kvantifikují variabilitu regresních proměnných mezi jednotlivými kategoriálními úrovněmi vybrané proměnné pomocí pravděpodobnostního rozdělení místo odhadování pevných regresních koeficientů. V GLMM je toto rozdělení normální s nulovou střední hodnotou a rozptylem, který je nutné odhadnout.

GLMM - pokračování

- Jednotlivé úrovně proměnné, kterou používám jako náhodný efekt, by měla představovat náhodný výběr z větší skupiny, tzn. jedinci vybraní z populace, a variabilita mezi úrovněmi je pro nás významnější než efekt jednotlivé úrovně.

Ocenění bez predikce

- Oceňování amerických opcí
- Zahrnutí iracionálního elementu
 - (např. *De Giovanni D., Working paper F-2007-03; Finance Research Group, University of Aarhus*)

Existuje široká škála možných přístupů k chování pojistníků, od nejjednodušších k v současnosti spíše teoretickým konceptům

- lze např. ~ využití teorie extrémních hodnot – rozdělení podmíněné tím, že jiná proměnná překročí kritickou hranici (délka období bez krize, nezaměstnanost)

Každý model nám dá nějaký výsledek

- ne nutně správný

U všech modelů narážíme na nutnost kvalitních podkladových dat

Pokud se zdá, že něco funguje a jde o market practice, proč to **NEzměnit? ...**

- Lehman Brothers

Kent, J. and Morgan, E.,; Staple Inn Actuarial Society 18 November 2008, *Dynamic Policyholder Behaviour*

Renshaw, A. E. and Haberman, S.J.; Inst. Act.; 113 (1986), 459-497 *Statistical Analysis of Life Assurance Lapses*

Richardson, C. B. F. and Hartwell, J. M.; Transactions of Soc. of Act. Vol. 3 No. 7 (1951); 338-396

Cox, S.H. and Lin, Y.; *Annuity Lapse Rate modelling: Tobit or not tobit*

Briere-Giroux G., Huet J.-F., Spaul R., Staudt A., Weinsner D.; Towers Watson (2010): *Predictive Modeling for Life Insurers (Application of Predictive Modeling Techniques in Measuring Policyholder Behavior in Variable Annuity Contracts)*

Milhaud X., Loisel S., Maume-Deschamps V.; *Surrender Triggers in Life Insurance: Classification and Risk Predictions*

Guelman L., Guillén M., Pérez-Marín A. M. MS 2012, LNBIP 115, pp. 123–133, 2012.: *Random Forests for Uplift Modeling: An Insurance Customer Retention Case.*

Rzepakowski P, Jaroszewicz S., published online <http://link.springer.com/content/pdf/10.1007%2Fs10115-011-0434-0.pdf>, 2011: *Decision trees for uplift modeling with single and multiple treatments.*

Prostor pro otázky

Presentation by Zdeněk Roubal





cutting through complexity

© 2013 Central and Eastern Europe Ltd., a Cyprus limited liability company and a member firm of the KPMG network of independent member firms affiliated with KPMG International Cooperative (“KPMG International”), a Swiss entity. All rights reserved. Printed in the Czech Republic.

The KPMG name, logo and ‘cutting through complexity’ are registered trademarks or trademarks of KPMG International Cooperative (KPMG International).