

OPONENTSKÝ POSUDEK HABILITAČNÍ PRÁCE

Název práce:

Modely přežití ve finanční a pojistné praxi

Autor: Mgr. David Zapletal, Ph.D.

Oponent: prof. Ing. Radim Briš, CSc., Fakulta elektrotechniky a informatiky, Kat.
aplikované matematiky, Vysoká škola báňská - Technická univerzita
Ostrava

Analýza přežití je v dnešní době rozvinutá matematicko-statistická disciplína, která pokrývá široké spektrum aplikací, nejčastěji v oblasti medicíny, biologie, ale také ekonomie, průmyslu, či jinde. Studium a výzkum náhodné veličiny, kterou lze interpretovat jako „dobu do výskytu nějaké události“, někdy také životnost zkoumaného objektu, pokud onou událostí rozumíme ukončení života objektu, je velmi atraktivní disciplína zejména v kontextu masivního rozvoje vědy a techniky na prahu třetího tisíciletí, kdy výrobci se snaží tuto náhodnou veličinu všemožně prodlužovat, protože je nezbytné ji umět správně a efektivně monitorovat. Tato habilitační práce přináší vyspělé metody užívané v rámci klasické analýzy přežití. Téma habilitační práce tedy považují za velmi aktuální.

Předložena habilitační práce nemá strukturu standardní habilitační práce, nýbrž je to spíše monografie, tedy podle § 72, odst. 3c Zákona č. 111/1998 Sb. o vysokých školách se jedná o „tiskem vydanou monografii, která přináší nové vědecké poznatky“. V tomto smyslu práce splňuje legislativní požadavek na habilitační práci.

Práce má celkem 331 stran, strukturovaných do 11 kapitol plus závěr, dále velmi podrobnou pasáž o dostupných literárních zdrojích na dané téma, 2 přílohy (informace o analyzovaných datech a příklady použitých kódů v softwaru R) a rejstřík. V literárních zdrojích nejsou vyčleněny vlastní publikace autora, které se tudíž těžko dohledávají. Nicméně publikace uchazeče lze dohledat na vysoce respektovaném serveru vědeckých poznatků Web of Science, kde uchazeč registruje 23 publikací, z čehož je většina 17 časopiseckých (7 krát Q1, 3 krát Q2, 4 krát Q3, atd.), 56 citací bez auto-citací, *h*-index 6. Tedy publikační činnost uchazeče je vyhovující a nad rámec požadavků pro habilitační řízení.

Cíle habilitační práce nejsou explicitně formulovány s ohledem na formu monografie, nicméně lze je implicitně vyčíst z úvodní části, která přináší komentáře k jednotlivým kapitolám, z nichž je patrné, které kapitoly mají popisný charakter (1,2,3,7 a 8) a metodologický charakter (4,5,6). Diagnostika regresních modelů včetně posouzení kvality a předpovědních schopností jsou prezentovány v kapitole 9. Nejvíce kreativní z hlediska nových poznatků je kapitola 10, která přináší možné modifikace a rozšíření Coxova modelu, a kapitola 11, přinášející konstrukci a aplikace pokročilých modelů analýzy přežití v kontextu s různými aspekty datových souborů úvěrových dat (P2P).

Po krátkém představení charakteru dat v analýze přežití (kap. 1) autor přináší v kapitole 2 zevrubný popis použitých dvou datových souborů (úvěrová a pojistná data) s vysvětlujícími a vysvětlovanými proměnnými včetně úpravy dat, zahrnující generaci výběrových souborů s upravenými rozsahy dat. Výklad je jasný, srozumitelný a nemám k němu žádné výhrady. Kapitola 3 přináší elementární matematiku nutnou pro analýzu přežití, kde se vyskytuje zásadní chyba na str. 37, kde integrál z hustoty samozřejmě nemůže být 0, což by se u statistické habilitační práce nemělo vyskytnout. Kapitola 4 přináší neparametrické metody pro odhady funkce přežití, kumulativní rizikové funkce, jakož i střední doby a mediánu přežití včetně intervalových odhadů, kde oceňuji velmi sofistikovaný a ucelený přístup výkladu zejména v kontextu se současným stavem problematiky. Totéž lze konstatovat o kapitole 5, která přehledně a vyčerpávajícím způsobem rekapituluje neparametrické metody pro porovnání funkcí přežití pro různé skupiny subjektů, jejich rozmanité modifikace při nesplněných předpokladech o proporcionalitě rizik apod. V podobném duchu je vedena i kapitola 6, přinášející Coxův model proporcionálních rizik, kde autor přijatelně jednoduchým způsobem vysvětluje čtenáři poměrně složitou konstrukci modelu včetně parciální věrohodnostní funkce, sloužící pro odhady parametrů modelu, případně jejich rozptylů, jakož i testování statistické významnosti parametrů modelu, což pak implikuje různé strategie a výběr proměnných do modelu. Diskuse k jednotlivým strategiím je dostatečně podrobná a její závěry jsou srozumitelně interpretovány na použitých datových souborech. Kratičká kapitola 7 dává do kontextu doposud použitý matematický aparát s teorií čítacích procesů. Podobně krátká kapitola 8 představuje platformu (různé typy reziduí) pro následnou diagnostiku modelu v kapitole 9. Zde autor rekapituluje různé postupy, jak ověřit předpoklady Coxova modelu (zejména lineární vztah numerických proměnných k logaritmu rizikové funkce a proporcionalita rizik) a jak postupovat v různých případech jejich porušení, např. zahrnutí nelinearity metrických proměnných do modelu pomocí frakcionálních polynomů, atd. Z hlediska diagnostiky modelu jsou cenné i statistiky a příslušné komentáře k identifikaci vlivných pozorování a také kvality modelu. Kapitola 10 je vlastně pokračování předešlé kapitoly, tedy dává instrukce, jak postupovat v případě nesplnění základních předpokladů Coxova modelu, přináší modifikace Coxova modelu a odpovídající vyčerpávající interpretaci výsledků, která vůbec není triviální zejména při nesplněných předpokladech modelu a která je obzvláště cenná pro experty z aplikační oblasti. Tuto interpretaci považuji za jeden z hlavních tvůrčích přínosů autora. Závěrečná kapitola 11 přináší alternativy ke Coxově modelu, tzv. mixture cure model, který zohledňuje situaci, kdy některé subjekty buď jsou nebo nejsou citlivé k nastání sledované události (složen ze 2 submodelů – latence a incidence), kde model je rovněž sestaven, naplněn parametry (pomocí EM algoritmu), diagnostikován a interpretován. Podobně jsou pak stručně nastíněny dva z modelů konkurujících rizik, tzv. model s rizikovou funkcí specifickou pro konkrétní událost a model se sub-distribuční rizikovou funkcí, oba mají svá specifika a opodstatnění. Obě alternativy jsou nakonec propojeny v posledně představeném modelu PHMC (proportional hazard mixture cure model), který byl autorem práce kreativně algoritmován, jelikož k němu neexistuje žádný balíček v softwaru R. Ukázalo se, že všechny představené modely mají poměrně nízké prediktivní schopnosti, což ostatně autor v závěru přiznává. Je zřejmé, že je to dáno nevhodným typem dat pro účely predikce.

Hlavní cíl této monografie vidím ve snaze aplikovat klasické, jakož i vyspělé metody analýzy přežití na úvěrová a pojistná data. Třebaže nejsem expertem z oblasti finanční praxe, z výsledků habilitační práce soudím, že tento cíl byl naplněn. Z práce není na první pohled zřejmé, zda autor má nějaký svůj originální metodický přínos k analýze přežití v kontextu s uvedenými daty. Práce však pokrývá široké spektrum témat klasické analýzy přežití, a tudíž může posloužit každému badateli v akademické sféře (studenti, akademičtí pracovníci, pracovníci vědecko-výzkumných institucí), který zpracovává a vyhodnocuje data o přežívání v kontextu svého výzkumu. Přidanou hodnotou pak jsou algoritmy, či kódy pro statistické a grafické

zpracování těchto dat v prostředí volně dostupného statistického programovacího jazyka R, přičemž oceňuji zejména to, že jeden z představených pokročilých modelů analýzy přežití byl programově zpracován přímo autorem.

Kritické připomínky:

- str.110: „...lineární prediktor $\eta = \exp(\dots)$ “ ...není lineární prediktor!
- Počet gramatických chyb je dosti velký, avšak je ještě přijatelný. Většinou je dán nedokončenými slovy nebo špatnou formou přivlastnění dané metody k autorovi (např. „Andersenovi grafy“).

Otázky k obhajobě:

1. Str. 33, poslední odstavec: Jakou metodou proběhla generace 10 000 výběrových souborů s rozsahem 636 (necenzorovaných) z původního rozsahu 19 856, resp. 16 745? (Zřejmě podobně se to pak provádělo s cenzorovanými subjekty s rozsahy 1904 a 546).
2. Prosím autora o stručnou formulaci toho, které pasáže v práci, zejména v metodice, považuje za svůj vlastní originální přínos.

Závěr:

Nutno konstatovat, že práce je přehledná, srozumitelná a přínosná. Většina předložených poznatků je rozpracována pro další použití. Autor prokázal schopnost kreativně vědecky pracovat, nové poznatky umí dovést až ke konkrétnímu praktickému využití. Autor prokázal, že jeho vědecké výsledky jsou konkurenceschopné, akceptovatelné ve vysoce uznávaných světových vědeckých časopisech.

Habilitační práce splňuje všechny zákonné požadavky stanovené v §72 odst. 3c Zákona č. 111/1998 Sb. o vysokých školách a tudíž

doporučuji

habilitační komisi její přijetí a po úspěšné obhajobě před vědeckou radou

následné jmenování pana Mgr. Davida Zapletala, Ph.D. docentem.

V Ostravě dne 14.8. 2024

podpis

