

## Posudek na doktorskou disertační práci Mgr. Miriam Marušiakové

### Tests for multiple changes in linear regression models

Předložená disertační práce se zabývá v současné době velmi aktuální a dynamicky se rozvíjející problematikou detekce změn ve statistických modelech. Autorka se ve své práci soustředila na úlohu odhalování změn v posloupnosti časově uspořádaných pozorování, která lze vysvětlit lineární regresí, zejména na problém, zda lze pozorování vysvětlit jediným modelem s konstantními parametry nebo zda data lze rozdělit na časové segmenty, ve kterých se řídí odlišnými regresními modely. Jde tedy o detekci vícenásobných změn v lineárním regresním modelu. Okamžiky změny a také jejich počty jsou obvykle neznámé a je třeba je odhadnout. Rovněž tak je třeba odhadnout parametry jednotlivých regresních modelů. V této práci se předpokládá, že ke změnám dochází jen v regresních parametrech, parametry procesu chyb se nemění.

První kapitola uvádí do problematiky a shrnuje současný stav. V práci jsou dále studovány 4 problémové okruhy týkající se detekce vícenásobných změn, jsou obsahem kapitol 2-5. Nejdříve se vyšetřuje lineární regresní model s náhodnými regresory (nontrending regressors), chyby a regresory jsou vzájemně nezávislé. Odhady regresních parametrů a okamžiky změn jsou založeny na metodě nejmenších čtverců, testová statistika je odvozena z reziduálního součtu čtverců a globalizuje statistiky typu F. Jsou zde prezentovány výsledky o konzistenci odhadů a jejich asymptotickém rozdělení, odvozeno asymptotické chování testové statistiky za nulové hypotézy (nedochází ke změně) i za alternativy a zmíněna konzistence odhadu rozptylu chyb. Tyto výsledky jsou převzaty z literatury. Nově je vypracována teorie permutačních testů pro detekci vícenásobných změn. Výhodou tohoto přístupu je, že asymptotické rozdělení je invariantní vůči volbě alternativy. Kritické hodnoty permutačního testu mohou být spočteny nikoliv z asymptotického rozdělení, což předpokládá velký rozsah datového souboru, ale přímo z daných pozorování jejich permutováním pomocí jednoduchého výpočetního algoritmu.

V další (třetí) kapitole je metodika předchozí kapitoly použita pro detekci vícenásobných změn v autoregresním modelu obecného řádu s konstantním členem (interceptem).

Ve čtvrté kapitole se uvažuje lineární regresní model, ve kterém regresory jsou spojitě diferencovatelné funkce rostoucího argumentu (trending regressors). Odhady metodou nejmenších čtverců jsou nahrazeny robustními M-odhady. Pro odvození výsledků (odhady parametrů a bodů změny, odhad rozptylu, odvození testové statistiky a její asymptotické chování, permutační test) je použito stejného scénáře jako v předchozích kapitolách.

Ve všech těchto situacích se předpokládá, že posloupnost chyb je tvořena nezávislými stejně rozdělenými náhodnými veličinami. Pátá kapitola se zabývá detekcí vícenásobných změn v modelu polohy, ve kterém chyby tvoří stacionární posloupnost typu strong ( $\alpha$ )-mixing. Opět jsou uvažovány M-odhady pro parametry posunutí a body změny a odvozeno limitní chování testové statistiky za nulové hypotézy. Dále je zde navržen odhad asymptotického rozptylu a dokázána jeho slabá konzistence.

Všechny teoretické výsledky dosažené v kapitolách 3-5 lze považovat za původní. Kapitola 6 shrnuje výsledky simulačních experimentů, které se vztahují ke kapitolám 2-4. Jsou zde prezentovány kvantilové grafy, grafy ilustrující sílu testů a tabelovány kritické hodnoty permutačních testů. V sedmé kapitole je uvedena aplikace výsledků ze 4. kapitoly na reálná klimatologická data. V poslední kapitole jsou zmíněny některé neřešené a otevřené problémy. V dodatcích A-D jsou uvedena podpůrná tvrzení z literatury, na které se v práci odkazuje.

Základními matematickými prostředky jsou pokročilé limitní věty teorie pravděpodobnosti jako funkcionální centrální limitní věta, věta o spojitém zobrazení, věty o supremu Wienerova procesu, limitní věty a nerovnosti týkající se posloupností strong mixing, dále též věty o asymptotické linearitě robustních odhadů. Z tohoto hlediska má práce vysokou teoretickou úroveň.

Podrobnější komentář:

- Mnoho důkazů je psáno zkratkovitě nebo jen v základních krocích, neboť úplná prezentace mnohdy velmi technických postupů by neúměrně rozšířila obsah práce, někdy je to však na úkor srozumitelnosti. Poněkud roztržitěně působí vkládání pomocných vět a jejich důkazů do důkazů hlavních vět. Orientaci v textu by také usnadnilo označení některých základních formulí referenčním číslem (např. na str. 11, ř. 14 a na str. 14, ř. 10), některé vzorce by se naopak nemusely opisovat vícekrát.
- Předpoklad (A5) pro důkaz lemmatu 2.3.2 se zdá být zbytečný.
- Formulace na str. 32 a 33 týkající se autoregresního procesu jsou poněkud nepřesné: předpoklad (B3) jen říká, že počáteční pozorování jsou hodnoty nějakého lineárního procesu. Za  $H_0$  a předpokladu (B3) je proces stacionární (nejen asymptoticky).
- Na str. 40 poslední dva řádky- musí se předpokládat i nulová střední hodnota.
- Důkazová technika v kapitole 4, zejména v odst. 4.3, spoléhá a odkazuje na práci Bai (1998), kde se studuje regresní model s vícenásobnými

změnami a náhodnými regresory a s  $L_1$  odhady regresních parametrů. To však není v práci zdůrazněno, rovněž není zmíněno, za jakých předpokladů lemmata a tvrzení v Bai (1998) dokázaná platí. Mělo by být ukázáno, a bylo by to mnohem korektnější, že z předpokladů (C1)-(C6) v kapitole 4 plynou předpoklady B1-B3 v práci Bai (1998) a důkazy se provedou s funkcí  $\rho$ , která nahradí absolutní hodnotu; potom stačí ukázat jen odlišnosti.

- Důkaz asymptotické linearity gradientu v (4.4.7) obsahuje příliš velké skoky a odkaz na použití Hájkovy a Rényiovy nerovnosti a na práci Hušková a Pícek (2005) je jen velmi obecný. Totéž se týká důkazu lemmatu 4.4.1, kde by měly být uvedeny aspoň základní kroky.
- Dále by se mělo ukázat, jak (4.4.10) plyne z (4.1.5) aspoň pro jeden případ.
- V práci se často užívá obratu, že nějaká nerovnost platí v pravděpodobnosti, aniž by to bylo blíže specifikováno, např. na str. 26, v tvrzení (iii) nebo v úvahách na str. 63.
- Proč v simulacích nebylo použito autoregrese s interceptem jako v teoretické části? Je rovněž škoda, že simulace se netýkají závislých pozorování a chování Bartlettova odhadu.

Přesto je nesporné, že předložená dizertace přináší nové významné teoretické výsledky, které lze aplikovat v praxi, doktorandka v ní prokázala schopnost samostatné tvůrčí činnosti, a proto doporučuji práci k obhajobě v doktorském studijním programu 4 m-5 na MFF UK.

V Praze 5. listopadu 2009

Doc. RNDr. Zuzana Prášková, CSc.