

POSUDEK OPONENTA DIPLOMOVÉ PRÁCE

Název: Modely multiplikatívnych chyb

Autor: Bc. Matyáš Krahulík

SHRNUTÍ OBSAHU PRÁCE

Práce je věnována tzv. multiplikatívním chybovým modelům (MEM). V první kapitole se řešitel snaží zavést GARCH modely a uvést nějaké jejich vlastnosti. Následně jsou definovány MEM modely, a to jak jednorozměrné, tak i vícerozměrné. Určitá část práce je věnována i modelům s chybami, které nabývají nulové hodnoty s kladnou pravděpodobností. V části 2.4. se autor věnuje vícerozměrným MEM modelům, v části 2.5 pak tzv. semi-parametrickému přístupu. V kapitole 3 je následně provedena analýza reálných dat, a to konkrétně časových řad nahlášených škod.

CELKOVÉ HODNOCENÍ PRÁCE

Práce je kompilát několika článků, jejichž vybrané části autor přeložil a přepsal z nich jednotlivé matematické formule. Žádné vysvětlující komentáře, rozpisy nebo zdůvodnění rovností neposkytuje. Text je proto obtížně čitelný a na mnoha místech působí velmi zmateným dojmem. Navíc se z tohoto důvodu práce pohybuje, podle mého názoru, na hranici plagiátu.

Téma práce je podle mého názoru pro diplomovou práci přiměřené a přirozeným způsobem navazuje na látku vyučovanou v rámci povinného předmětu Časové řady.

Vlastní příspěvek. Vlastní příspěvek autora v teoretické části práce (kapitoly 1 a 2) považuji za minimální. Neprovedl téměř žádné logické propojení jednotlivých článků, nedoplnil žádné rozpisy ani komentáře. Značení také není zcela sjednocené. Vlastní je pouze praktická část, kde popsání modely odhadl vždy pro jeden datový soubor.

Matematická úroveň. Autor žádné vlastní matematické výpočty neprovádí. Pokud uvádí nějakou větu, pak bez důkazu a jedná se o přesný překlad z předlohy. Na několika místech se řešitel pokusil o vlastní definice, tam ovšem pak požaduje poněkud nesmyslné předpoklady (viz připomínka 2 níže).

Práce se zdroji. Autor v práci vždy řádně uvádí, ze kterého zdroje čerpá. Následně jeho text dodržuje strukturu předlohového téměř dokonale. U dvou ze článků uvedených v seznamu použité literatury chybí jakékoliv informace o vydání. Uvedení jsou jen autoři, název a rok, jedná se o Brownless a kol. (2012) a Cipollini a Gallo (2022). Program R ani využití knihovny řádně citovány nejsou.

Formální úroveň. Po vzhledové stránce text vypadá rozumně, až na to, že obsahuje velmi často jednopísmenné předložky na konci řádku a zalamování řádků je občas ad hoc a text není pak dobře zarovnaný. Použité matematické značení není vždy zcela vysvětleno a lze se jen domnívat, co jím autor myslí. Navíc není značení ani mezi jednotlivými částmi (čerpajícími z jiných zdrojů) dostatečně sjednoceno. Dále pak text obsahuje určité množství dalších drobnějších formálních nedostatků týkajících se zápisu matematického textu.

Jazyková úroveň. Po jazykové stránce je text na rozumné úrovni, zvláštní mi přijde pouze používání a skloňování slova *kopula* namísto *kopule* v části 2.4.2.

Podle mého názoru si na sebe autor uložil příliš velký úkol a snažil se jít příliš do šířky, aniž by dostatečně pochopil a popsal zcela základní modely. Jako největší nedostatek práce vidím absenci vlastní iniciativy při prezentaci dané látky a formu textu na hranici plagiátu. Níže uvedené připomínky jsou proto pouze doplňující.

1. Kapitola 1 si klade za cíl představení GARCH modelů. Bohužel hned úvodní část je poněkud nesrozumitelná. V celé práci se pracuje se sigma algebrou $\Psi_t = \sigma\{Y_s, s \leq t\}$. V Definicí 1 se specifikuje podmíněná střední hodnota $E[Y_t|\Psi_t] = \mu_t$, ale v Poznámce za defínicí je uvažováno μ_t jako lineární funkce nějakého \mathbf{X}_t , přičemž vztah \mathbf{X}_t k $\{Y_t\}$ (a tedy Ψ_t) je nejasný. Totéž se vyskytuje za Definicí 2. Dále je dosti omezující, že se Defínice 1, 2 a 3 omezují na podmíněné normální rozdělení. Toto pojetí také do značné míry znemožňuje využívat vztah mezi MEM a GARCH modely.
2. V práci se v řadě defínic vyskytuje předpoklad, že *veličiny* $\{\varepsilon_t, t \in \mathbb{Z}\}$ jsou *podmíněně nezávislé* (viz např. str. 18, kde se pro to zavádí i matematické značení). Prosím o přesné matematické vyjádření toho, co je tímto myšleno.
3. Na str. 8 je zavedena podmíněná logaritmičká věrohodnost, používá se značení $\ell_n(\boldsymbol{\theta}|\Psi_{n-1})$. Prosím o přesné vysvětlení, čím je přesně daná věrohodnost podmíněná a jak byla odvozena.
4. Na str. 9 je uvedeno tvrzení týkající se vlastností odhadů lineárního MEM(1,1) modelu. Nesprávně je zde uvedeno, že specifikace μ_t odpovídá GARCH(1,1) modelu (podobná poznámka je i na str. 31). Mezi GARCH a MEM modely jistě vztah je, ale ne takový jako autor uvádí a navíc ne přímo za předpokladů uvažovaných v práci (viz i poznámka 1).
5. Str. 10: Co přesně je myšleno tvrzením, že rozdělení je nespojitě v nule a o jakou hustotu f_X se jedná (vzhledem k čemu)?
6. Jak přesně je odvozena log věrohodnost na str. 10? Jedná se o podmíněnou nebo nepodmíněnou věrohodnost? Dále pak v defínici 6 se požaduje nějaké podmíněné rozdělení ε_t při daném Ψ_{t-1} . Jakou závislost ε_t a Ψ_{t-1} tento předpoklad implikuje?
7. Proč se v části 2.3 uvažuje σ -algebra \mathcal{F}_t namísto Ψ_t ?
8. Str. 13: Popis vícerozměrného gama rozdělení formou *rozdělení má složitý tvar obsahující integrál* mi přijde poněkud úsměvné. Jedná se samozřejmě o překlad textu z předlohy, kde je ale v příloze toto rozdělení skutečně formálně zavedeno (pomocí vhodné konstrukce).
9. V Defínici 10 se připouští jakékoliv vícerozměrné normální rozdělení? Nebo musí toto rozdělení něco splňovat?
10. Co přesně znamená aproximační znak na str. 15?
11. Model uvedený na str.15 padne jen tak z nebe, není jasné, proč je jeho struktura právě taktová a jak souvisí veličiny r_t se sigma algebrou Ψ_t , kterou se podmiňuje v celé práci. Prosím tedy o vysvětlení, co v této fázi práce přesně μ_t vyjadřuje. Podobné je to u modelu na str. 19 - tam opět není jasné, jak souvisí R_t s Ψ_t a proč se model zavádí právě takto. Totéž u vícerozměrné verze.
12. Na str. 16, 6. řádek zdola je chyba.
13. Do části 2.4 je věrohodnost značená s argumentem (někde navíc s jakousi podmínkou), od 2.4 pak je zjevně argument vypuštěn. Co přesně jsou zde parametry, není jasné.

14. Text na str.17 a 18 je zjevně jen přepsaný a přeložený a autor nad ním zřejmě asi nepřemýšlel, ale zeptám se na intuici za zavedením odhadu \mathbf{R} na str. 18. Proč lze tento odhad pak dosadit do věrohodnosti?
15. V části 2.5. chybí explicitní předpoklady na jednotlivé složky. Co přesně se předpokládá o $\{\xi_t\}$ aby platila rovnost na posledním řádku na str. 18? Nebo to platí vždy?
16. V čem přesně se liší momentová metoda popsaná v části 2.5.1 a QML metoda popsaná v části 2.5.2? Jaký je rozdíl mezi $\nabla_{\theta}\xi_t$ a $\partial\xi_t/\partial\theta$? Proč se metoda z 2.5.2 nazývá kvazi-maximálně věrohodná, když se vychází z parametrického předpokladu o rozdělení ε_t ?
17. Proč predikce v modelu probíhá právě podle vzorců na str. 23? Jak se k těmto vztahům dojde?
18. Praktická část je sepsaná spíše jednoduchým jazykem s naivními formulacemi o tom, co, jak a kde se spustí v programu R. Autor se neodkazuje na obrázky a tabulky číselně, ale vkládá je přímo na potřebné místo do textu. Rozlišení některých obrázků je nedostatečné. Celkově se jedná spíše o jednodušší aplikaci, jediný zajímavý moment pak shledávám na str. 40–41.
 - Na str. 31 tvrdíte, že nemusí být $\alpha + \beta < 1$. Jaké omezení na parametry tedy bylo voleno v numerické optimalizaci a proč?
 - Na str. 32 (a posléze i jinde) je na řadu reziduí použit jednovýběrový t-test. Proč jej zde lze použít?
 - Volba parametru zpoždění pro Ljungův Boxův test ad hoc na základě naměřených dat je krajně nevhodná. Navíc mi není jasné, jak přesně byl tento test proveden pro rezidua z MEM modelu. Jaký byl uvažován kritický obor a co přesně je nulová hypotéza?
 - Z kolika pozorování byl odhadován model v části 3.3? Je to dostatečný počet pro tolik parametrů? Proč nebyla uvažována i jiná struktura modelu?
 - Proč by v části 3.3. mělo být ω omezené? Obecně daný přístup ke konstrukci omezení pro parametry mi přijde hodně nestandardní.
 - Jaká jádrová funkce a proč byla použita pro odhad τ_t ? A jak přesně byla spočtena šířka pásma?
 - Vložení použitého kódu přímo do pdf, bez dodatečných komentářů, mi přijde spíše nestandardní.

ZÁVĚR

Předloženou práci považuji za podprůměrnou. Moje hodnocení bude záviset na průběhu obhajoby a odpovědích na připomínky a otázky položené výše.

RNDr. Šárka Hudecová, Ph.D.
 KPMS MFF UK
 V Praze dne 27. května 2024