Posudek oponenta diplomové práce

Název: Regresní hloubka a podobné metody
Autor: Bc. Denisa Dočekalová

Shrnutí obsahu práce

Předložená práce se zabývá regresním hloubkou a porovnává tuto metodu s metodou nejmenších čtverců (LS) a metodou nejmenších absolutních odchylek (LAD).

Téma práce. Téma odpovídá znalostem a možnostem příslušného oboru magisterského studia a autorka jej naplnila. I když ze zadání by asi člověk čekal, že v množině dalších robustních regresních odhadů nebude pouze metoda LAD.

Vlastní příspěvek. Vlastní příspěvek autora vidím zejména v kapitole 5, ve které se autorka zabývá regresní hloubkou. Dále pak v simulační studii v kapitole 6.

Matematická úroveň. Po matematické stránce je úroveň práce slušná. Práce obsahuje rigorózně a korektně zformulovaný matematický text. Je však nutné říct, že úroveň práce je poněkud nevyrovnaná. Vypadá to, že autorku především zaujaly geometrické aspekty regresní hloubky popisované v kapitole 5. Zde bych mimo jiné rád ocenil ilustrační obrázky. Méně pozornosti a pečlivosti však autorka dle mého názoru věnovala asymptotických vlastnostem. To se projevuje v úvodních kapitolách 1 až 4, kde autorka stručně pojednává o lineárním regresním modelu a metodě LS a LAD. V těchto kapitolách je až příliš mnoho nejasností a nepřesností. A občas i závažnější matematické prohřešky.

Práce se zdroji. Na kolik dokáži posoudit, tak práce obsahuje vlastní text autorky a zdroje jsou řádně citovány. Citace by však bylo na některých místech záhodno upřesnit. Občas by bylo také lépe citovat původní zdroj a ne pouze článek, který sám výsledek pouze přebral. Jistými nedokonalostmi také trpí seznam literatury (např. u článků občas nejsou uvedeny stránky a u diplomové práce J. Rábka chybí instituce).

Formální úprava. Formální úroveň práce je dobrá, množství překlepů není nijak zásadně velké. Značení je konzistentní, ale občas ne dobře zavedené/vysvětlené.

Celkové hodnocení práce

Dle mého názoru se jedná o solidní práci, na které autorka očividně usilovně pracovala. Dojem však kazí řada menších i větších chyb a matematicky nejasných vyjádření.

Připomínky a dotazy

1. str. **4 Předpoklad 1**: Není jasné, v čem jsou toto předpoklady **lineárního** modelu. Dále by bylo asi potřeba v (2) zohlednit, že $rank(\widetilde{\mathbb{X}}) = p$ je náhodný jev. V (4) mně přijde matoucí používat slovo *ordinální*. V (5) pak není jasné, co myslí autorka tím, že podmíněné rozdělení $Y|\boldsymbol{X}$ je spojité.

2. str. **5 Definice 2**: Nějak nerozumím této definici. Pokud bych definoval $\varepsilon_i = Y_i - \beta_0 - \boldsymbol{X}_i^\top \boldsymbol{\beta}_X$, tak je vždy (1.1) splněno.

3. str. **8 Definice 8**: Tato definice je vskutku značně neformální, protože není vysvětleno, co přesně matematicky znamená *kontaminovaný výběr*.

4. str. **11** první *Pozorování* na této straně: Jelikož se **Definice 10** zabývá funkcionály, tak není jasné, co je zde myšleno slovem *odhad*.

5. str. **11** *Poznámka*: Ze zápisu funkce ρ^* to spíše vypadá, že je definována na $\mathbb{R}$ a nikoliv na $\mathbb{R} \times \mathbb{R}^p$.

6. str. **12 Tvrzení 12**: Vzorec (2.9) není správně.

7. str. **12**: I na intuitivní rovině by šla lépe čtenáři vysvětlit aproximace $\mathbf{T}(\boldsymbol{Y}, \mathbb{X}))$ získaná pomocí von Misesova rozvoje.

8. str. **14 Tvrzení 15**: Pokud se nepletu, tak by v předpokladech měl být spíše **předpoklad 2** (než definice 2).

9. str. **14 Tvrzení 16**: Používat na důkaz asymptotické normality LS odhadu obecnou větu o M-odhadech a ještě přitom předpokládat konzistenci mně nepřijde v diplomové práci právě nejvhodnější.

10. str. **17 Tvrzení 20**: Takto zformulované tvrzení za daných předpokladů neplatí. Obzvláště pokud se v *Pozorování I* na následující straně tvrdí, že *hustota chybových členů f_ε stále může záviset na regresorech.* Podobně k porovnání (4.4) nestačí předpoklad homoskedasticity.

11. str. **21 Definice 23**: Nešla by tato definice zjednodušit/zkrátit tím, že by se uvažovalo pouze (5.3) nebo (5.4)?

12. str. **28**$_4$: Nerozumím tomu, co přesně má v daném kontextu znamenat předpoklad, že *rozdělení distribuční funkce H je diskrétní a rovnoměrné.*

13. str. **30**^{7}: Není jasné, co znamená symbol P_F.

14. str. **32**^{8}: Nemělo by zde být spíše *kterou jsme při rotaci prošli je rovna* ***nejvýše*** *1/2*?

15. str. **32**: Nejsem schopen sledovat druhou část důkazu **lemmatu 35**. Proč je dobré *předpokládat pro spor, že vektor koeficientů $\boldsymbol{\beta}$ nabývá regresní hloubky v libovolném bodě $v^* \in \mathbb{R}$ takovém, že $v^* \neq v$*?

16. str. **37 Předpoklad 4**: Je škoda, že autorka v podstatě pouze opsala podmínky z článku Bai and He (1999) a přidala tam *s.j.* Pro případ náhodných regresorů, kdy pozorujeme nezávislé stejné rozdělené náhodné vektory $\binom{Y_1}{\boldsymbol{X}_1}, \ldots, \binom{Y_n}{\boldsymbol{X}_n}$, by se zejména podmínky (2), (3) a (4) daly zjednodušit. Navíc autorka opomněla vysvětlit, co je $\mathbf{w}_i$ v definici funkce $Q_n(c)$.

17. str. **38**: V druhém odstavci kapitoly 5.5.3 je z mého pohledu řada nejasností. Zaprvé nerozumím, co znamenají níže uvedené relativní eficience pro absolutní člen a směrnici v případě dvourozměrného normálního rozdělení. Dále nerozumím, proč by hodnota

relativní eficience měla v případě asymptoticky normálního rozdělení vycházet pouze cca 75 %.

18. str. $\mathbf{44}_{10}$: Nerozumím tomu, že by funkce g měla být tvaru

$$g(u) := g(x^2 + y^2) = \dots$$

19. str. $\mathbf{47}^{9}$: Nerozumím větě, že *…můžeme zajistit například tak, že pro všechna* $i \in \{1,\dots,n\}$ *zvolíme* $F_i = F$. V následujícím odstavci pak není jasné, co by mělo mít *normální* rozdělení a co *Studentovo* t-rozdělení.

20. str. **49**: Pokud se nepletu, tak nejdříve se generuje odezva (bod 6) a pak se teprve se kontaminují hodnoty regresoru.

Otázky k obhajobě

Připravte si odpovědi na připomínky č. 1, 6, 10, 11, 15 a 17.

Závěr

I přes výše uvedené výtky pokládám považuji práci za solidní. Domnívám se, že práce splňuje požadavky kladené na diplomovou práci na oboru Pravděpodobnost, matematická statistika a ekonometrie a doporučuji ji za ni uznat.

doc. Ing. Marek Omelka, Ph.D.
KPMS MFF UK
31. srpna 2022