

POSUDEK OPONENTA DIPLOMOVÉ PRÁCE

Název: Kauzální usuzování a klasická ekonometrie

Autor: Bc. Filip Olejko

SHRNUTÍ OBSAHU PRÁCE

Práce se zabývá problémem vyhodnocování kauzality. V první kapitole je představen model potenciálních výsledků. V druhé kapitole jsou podrobně představeny orientované acyklické grafy a jejich vybrané vlastnosti. Třetí kapitola popisuje metodu odhadu regresních parametrů pomocí metody instrumentálních proměnných. Na konci této kapitoly autor též uvádí krátkou simulační studii pro tři jednoduché situace.

CELKOVÉ HODNOCENÍ PRÁCE

Téma práce je zajímavé a pro diplomovou práci přiměřené. Přijde mi ale, že mohlo být zpracováno podstatně lépe, aby bylo zadání práce více naplněno.

Vlastní příspěvek. Autor poskytuje úvod do problému vyhodnocování kauzality pro případ nerandomizovaných studií. Předlohový text doplnil o nějaké svoje příklady a interpretace a drobná rozepsání některých rovností. Cílem práce zjevně mělo být i propojení teorie kauzality s odhady pomocí IV odhadů. To se autorovi podařilo, podle mého názoru, jen částečně. Vlastním přínosem je jistě krátká simulační studie.

Matematická úroveň. Matematická tvrzení jsou formulována většinou korektně, ale někde mohly být lépe specifikovány předpoklady (např. věta 8) a formulace mohla být jasnější (např. věta 7). V běžném textu práce není vždy jasné, s jakými předpoklady autor v jednotlivých částech pracuje, viz připomínka 2 níže.

Práce se zdroji. Práce čerpá z několika různých zdrojů a autorovi se bohužel nepodařilo sjednotit značení, což hodně znesnadňuje čtení práce. V kapitole 3 mi přijde, že autor čerpá i ze zdrojů, které necituje. Dále mi také přišlo dost nešťastné, že autor neuvádí, které příklady přebírá z literatury a které jsou jeho vlastní. Osobně mi příliš nevyhovuje, že seznam literatury není řazen podle abecedy. Některé jeho položky pak nejsou zadány formálně správně.

Formální úroveň. Práce obsahuje větší množství drobnějších chyb, a to zejména ve značení, ale i ve formátu (chyby ve fontu, nejednotné značení operace transpozic apod.). Popisky obrázků na str. 16 mi přijdou zvláště zarovnané, popisky tabulek naopak velmi dlouhé. Jinak jsou formální úprava a množství překlepů rozumné.

Jazyková úroveň. Práce je psaná slovenským jazykem, takže jazykovou úroveň práce nejsem schopna zcela zhodnotit, nicméně je patrné, že autor by mohl vylepšit psaní čárek ve složitějších souvětích. Slovenskou odbornou terminologií si také nedovoluji posuzovat, nicméně by mě zajímalo, zda toto autor odněkud čerpal, nebo jsou překlady jeho vlastní.

Připomínky.

1. První kapitola práce je zpracována pečlivěji než následující pasáže. Autor se zde problematice zjevně snažil porozumět. Ale bohužel ani zde není výklad vždy zcela srozumitelný a vhodně

řazený. Část kapitoly 1 a celá kapitola 2 velmi úzce sledují strukturu z [1] a přijde mi, že autor se nad samotným obsahem zamýšlel méně, než by bylo vhodné. V kapitole 2 je zbytečně moc prostoru věnováno markovským rozdělením na grafech, přestože se toto nikde dále nevyužívá. Není tedy vůbec jasné, proč tuto pasáž autor v práci zahrnul a jak souvisí se zbytkem práce. Kapitola 3 pak působí dojmem, že byla psána spíše narychlo a její struktura mi přijde dost nešťastná. Strídají se zde různá značení a celý text je hodně zmatený, viz další připomínky níže.

V první polovině práce je patrná snaha o sjednocení značení, ale bohužel na mnoha místech se to nepodařilo zcela a výsledek je tak místy zmatený. V poslední kapitole je značení v části 3.4 naprosto odlišné od předchozí části této kapitoly. Celkově mi jednotlivé části (kapitoly) práce přijdou málo propojené, zejména pak části 3.1 až 3.3 se zbytkem práce.

2. Na str. 22 nahoře se najednou objeví předpoklad o X a není mi vůbec jasné, kdy přesně byl před tím specifikován. Navíc mám pocit, že zajímavé jsou právě situace, kdy X náhodně přiřazeno není. Podobná situace se opakuje v celé práci vícekrát (např. strana 24, 32 aj.).
3. V tabule na str. 9 je evidentně chyba v jednom z řádků.
4. Na mnoha místech práce (např. str. 13) se diskutuje nějaký odhad, aniž by bylo jasné, jaká data přesně máme k dispozici a jak by se příslušný odhad konstruoval.
5. V části 1.3 je zcela pomícháno značení U a Z .
6. Část 2.4 působí jako zcela vytržená z kontextu: chybí úvod do problematiky a vysvětlení, co jsou veličiny s hvězdičkou.
7. Při vysvětlování Simpsonova paradoxu bych čekala, že autor představí danou situaci matematicky pomocí sdruženého rozdělení a následně bude odkazovat na marginální rozdělení.
8. V důkazu třetího tvrzení z věty 4 se v druhé rovnosti podmíněná nezávislost X a Y při daném (U, Z) , což je ale něco, co se právě snažíme dokázat. Důkaz tedy není v pořádku.
9. Některé pasáže a vyjádření působí poněkud „naivním dojmem“, jako třeba výraz, že „veličina je z dvou světů“ (na str. 21). Jiné pasáže (např. poslední odstavec v části 3.1, zmínka o hodnotnosti podmínce na str. 39 nahoře apod.) jsou v kontextu práce zbytečné a nesrozumitelné.
10. Předpoklad homoskedasticity náhodných chyb na str. 34 musí být formulován pomocí podmíněného momentu. Totéž na str. 39 nahoře.
11. Není mi jasné, proč autor nejprve představuje IV odhad bez odvození a teprve v další části představuje ideu 2SLS odhadu. Na str. 37 se bez vysvětlení najednou objeví vzorec, aniž by bylo zavedeno potřebné značení. V této pasáži se také poměrně náhodně střídá, zda je β psáno tlustým nebo obyčejným fontem. Teprve v další sekci je zaveden odhad 2SLS, a to vše opět s jiným značením. Celé to působí dojmem, že autor problematice velmi nerozumí a jednotlivé části slepil na poslední chvíli.
12. Ze zápisu předpokladů v příkladu 3.3. není jasné, jaký je vztah U a Z . Předpoklady o iid by měly být formulovány pro celý vektor. Stejná připomínka pak platí ke všem následujícím příkladům.
13. Část 3.4 působí v kontextu kapitoly 3 podivně: Najednou je použito úplně jiné značení, přechází se zjevně na binární odezvu a binární regresory a celý text je velmi obtížně srozumitelný. Vlivem neúspěšné změny značení se místy v textu vyskytuje veličina A , u níž se mi nepodařilo rozklíčovat, zda má být X , U , Z nebo ještě něco jiného.

14. Simulační část: V popisu 3.5.1 chybí zmínka implicitního předpokladu, že U a Z jsou nezávislé (podobně i další situace). V popisu studie je pak další řada nepřesností ve formulacích. V celé simulační studii mi chybí porovnání odhadů pro různé úrovně γ a různé síly vztahu X a U a Z a také nižší rozsahy výběru n .

Otázky.

1. Na str. 15 se představuje $C(x)$. Jaké jsou předpoklady, které bychom měli klást na zobrazení C (definiční obor, obor hodnot, vlastnosti. apod.)?
2. Na str. 26 je v Definicí 5 definována nezávislost dvou veličin pomocí sdružené hustoty. Tato část ale není správně, prosím tedy o správnou formulaci. Totéž pro defínici 6.
3. Defínice 7 je celá naprosto zmatená: Není jasné, v jakém smysli je f hustota, jak je možné, že f vystupuje na levé i pravé straně rovnosti, a naprosto chybí potřebný kvantifikátor. Prosím tedy o korektní formulaci, kde bude i jasně uvedeno, jakou množinu probíhá v .
4. Na str. 34 se uvedeno, že výše uvedené předpoklady zaručují konzistenci odhadu metodou nejmenších čtverců. To ale není úplně pravda: Jaké předpoklady tedy potřebujeme?
5. Na str. 35 jsou formulovány podmínky (3.1) a (3.2). Jaké postačující momentové předpoklady zaručují jejich platnost (v kontextu iid veličin)?
6. V příkladě 3.3. je uvedena řada předpokladů o konvergenci postupných průměrů. Tyto však platí za předpokladu jednoduchých momentových předpokladů. Jak lze formulovat postačující podmínky, aby konvergence specifikované na str. 40 platily?

ZÁVĚR

Předloženou práci považuji spíše za podprůměrnou. Za předpokladu, že student uspokojivě zodpoví výše uvedené otázky, doporučuji práci uznat jako diplomovou.

Reference

- [1] Wasserman, Larry (2013): All of statistics: A concise course in statistical inference. Springer Science& Business Media.

RNDr. Šárka Hudecová, Ph.D.
KPMS MFF UK
V Hannoveru dne 17. srpna 2024