

**MATEMATICKO-FYZIKÁLNÍ
FAKULTA**
Univerzita Karlova

DIPLOMOVÁ PRÁCE

Filip Olejko

**Kauzální usuzování a klasická
ekonometrie**

Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: prof. RNDr. Ivan Mizera, CSc.

Studijní program: Pravděpodobnost, matematická
statistika a ekonometrie

Praha 2024

Prohlašuji, že jsem tuto diplomovou práci vypracoval(a) samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů. Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V dne

Podpis autora

Chcel by som veľmi poďakovať všetkým, ktorí mi pomáhali a podporovali ma behom môjho štúdia. Predovšetkým mojim rodičom, starým rodičom, bratom a kamarátom.

Vďaka tiež patrí vedúcemu mojej diplomovej práce, prof. Ivanovi Mizerovi, za odborné, profesionálne a ľudské vedenie tejto práce. Chcel by som poďakovať aj MUDr. Zuzane Nedelskej, za konzultáciu niektorých príkladov.

Název práce: Kauzální usuzování a klasická ekonometrie

Autor: Filip Olejko

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: prof. RNDr. Ivan Mizera, CSc., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: Diplomová práce poskytuje základný náhľad na teóriu používanú pri kauzálnom usudzovaní. Zameriava sa pritom na model potenciálnych výsledkov a použitie usmernených acyklických grafov (DAGs). Tiež prezentuje jednu z klasických metód používaných v ekonometrii, ktorá využíva inštrumentálne premenné a dvojstupňový odhad pomocou metódy najmenších štvorcov. Táto metóda je daná do kontextu so situáciou, kedy dochádza k zastieraniu a identifikácia kauzálnych vzťahov s použitím väčšiny klasických metód zlyháva. Vo vhodných situáciach potom z pohľadu modelu potenciálnych výsledkov prezentuje predpoklady potrebné pre kauzálne usudzovanie. Nakoniec v simulačnej štúdií pre vybrané situácie, kedy dochádza k zastieraniu, demonštruje použitie metódy používajúcej inštrumentálne premenné. V celej práci sa zdôrazňuje dôležitosť uvádzaných predpokladov potrebných pre vyvodenie platných záverov o kauzalite.

Klíčová slova: Kauzálne usudzovanie, Model potenciálnych výsledkov, Inštrumentálne premenné

Title: Causal inference and classical econometrics

Author: Filip Olejko

Department: Department of Probability and Mathematical Statistics

Supervisor: prof. RNDr. Ivan Mizera, CSc., Department of Probability and Mathematical Statistics

Abstract: The thesis attempts to provide a basic insight into the theory used in causal inference. In this regard, it focuses on the potential outcomes model and the use of directed acyclic graphs (DAGs). It also presents one of the classical methods used in econometrics, which uses instrumental variables and two-stage estimation using the least squares method. This method is put in the context of a situation where there is confounding and the identification of causal relationships using more conventional methods therefore fails. In suitable situations, the assumptions necessary for causal inference from the perspective of the potential outcomes model are presented. Finally, for selected situations where confounding occurs, the simulation study demonstrates the use of a method using instrumental variables. Throughout the thesis, the importance of the stated assumptions needed to draw valid inferences about causality is emphasized.

Keywords: Causal inference, Potential outcomes model, Instrumental variables

Obsah

Úvod	6
1 Model potenciálnych výsledkov	8
1.1 Binárne ošetrenie	8
1.2 Mnohopočetné ošetrenie	15
1.3 Observačné štúdie a zastrenie	16
1.4 Simpsonov paradox	18
2 Orientované acyklické grafy	26
2.1 Úvod	26
2.2 DAGs	28
2.3 Pravdepodobnosť a DAGy	29
2.4 Vychýlenie meraní	31
2.5 DAGy a SWIGy	33
3 Metóda inštrumentálnych premenných	34
3.1 Úvod a motivácia	34
3.2 Odhady založené na inštrumentálnych premenných	35
3.3 Dvojstupňový odhad metódou najmenších štvorcov	38
3.4 Metódy inštrumentálnych premenných pre kauzálne usudzovanie	42
3.5 Simulačná štúdia	47
3.5.1 Kauzálny inštrument	48
3.5.2 Náhradný inštrument	48
3.5.3 Náhradný inštrument vo vybranej populácii	50
3.5.4 Zhodnotenie simulačnej štúdie	51
Záver	54
Literatúra	55
Zoznam obrázkov	57
Zoznam tabuliek	58

Úvod

Ukazovanie kauzálnych vzťahov je dôležité v mnohých oblastiach, predovšetkým vo vedeckom výskume a zdravotníctve. Existuje viacero moderných matematických metód, ktoré skúmajú kauzalitu. Medzi najznámejšie patria prístupy, ktoré vypracovali Donald Rubin, Judea Pearl, a Guido Imbens.

Donald Rubin je známy najmä vďaka vývoju Rubinovho kauzálneho modelu a modelu potenciálnych výsledkov (*potential outcomes framework*). Rubinov kauzálny model rieši problém zastieracích premenných (*confounding variables*) a výberového vychýlenia v observačných štúdiách, v ktorých dochádza k odhadu stredného účinku ošetrovania, prostredníctvom porovnania výsledkov v ošetrovaných a neošetrovaných skupinách. Model potenciálnych výsledkov poskytuje formálny rámec na definovanie kauzálneho efektu a jeho odhadovanie pomocou pozorovaných údajov.

Judea Pearl je priekopníkom v oblasti kauzálneho usudzovania. V roku 2011 získal Turingovu cenu za zásadný prínos k umelej inteligencii prostredníctvom vývoja kalkulu pre pravdepodobnostné a kauzálne usudzovanie. Vyvinul rámec štrukturálneho kauzálneho modelu (*structural causal model, SCM*). Rámec SCM umožňuje reprezentovať kauzálne vzťahy medzi premennými pomocou orientovaných acyklických grafov (*directed acyclic graph, DAG*). Tento rámec umožňuje vyhodnocovať účinok intervencií na výsledok a identifikovať kauzálne vzťahy medzi premennými v komplexnom kauzálnom systéme.

Guido Imbens je popredným odborníkom v oblasti kauzálneho usudzovania. V roku 2021 bol jedným z nositeľov Nobelovej ceny za prínos k analýze kauzálnych vzťahov v ekonometrii. Významne prispel k rozvoju metód porovnávania (*matching methods*) a inštrumentálnych premenných (*instrumental variables*). Metódy porovnávania sa zameriavajú na zníženie vychýlenia v kauzálnom usudzovaní porovnávaním ošetrovaných a kontrolných jednotlivcov na základe ich pozorovaných charakteristík, zatiaľ čo inštrumentálne premenné sa používajú na odhad kauzálneho efektu v prípade nepozorovaného zastierania (*unmeasured confounding*).

Je všeobecne známe, že prostriedkom pre získanie kauzálnych záverov sú randomizované kontrolované experimenty. Niektorí štatistickí [1] neuznávajú kauzálne závery získané za iných okolností. V praxi však nie je vždy možné realizovať randomizovaný kontrolovaný experiment: napríklad v niektorých zdravotníckych štúdiách (fajčenie) alebo v ekonometrii. Moderná metodológia kauzálneho usudzovania sa zaoberá práve takýmito situáciami; randomizovaných experimenty v tomto kontexte predstavujú "zlatý štandard".

Prvým cieľom tejto diplomovej práce je zoznámiť čitateľa s pojmovým aparátom moderných teórií kauzálneho usudzovania. Sú nimi model potenciálnych výsledkov a DAGov. Prvá kapitola sa zaoberá modelom potenciálnych výsledkov. Prezentovaný je základný rámec tohto modelu, sú preň uvedené rôzne príklady a vety, ktoré je možné použiť pri odhadovaní kauzálneho efektu pomocou pozorovaných údajov. Dôraz je tiež kladený na splnenie predpokladov týchto viet. Model potenciálnych výsledkov poskytuje formálny rámec na definovanie kauzálneho efektu a pracuje s potenciálnymi výsledkami (*counterfactuals, potential outcomes*).

Druhá kapitola sa venuje orientovaným acyklickým grafom (*directed acyclic graph*, *DAG*), ktoré sú vizuálnym nástrojom pre znázornenie kauzálnych vzťahov. DAGy pomáhajú identifikovať premenné, ktoré spôsobujú zastieranie a uľahčujú stanovenie súvislostí medzi premennými aj v zložitejších situáciách. Sú prezentované rôzne základné vlastnosti DAGov a ilustrované na konkrétnych príkladoch. Na záver kapitoly je diskutované možné spojenie modelu potenciálnych výsledkov a DAGov.

Druhým cieľom tejto diplomovej práce je štúdium vzťahu medzi kauzálnym usudzovaním a klasickou ekonometriou. Z tohto dôvodu sa v tretej kapitole venuje inštrumentálnym premenným, ktoré zohrávajú v ekonometrii kľúčovú úlohu, pretože umožňujú riešiť problémy v situáciách, kedy dochádza k zastieraniu. To je častým problémom pri kauzálnom usudzovaní a väčšina tradičných metód kvôli nemu zlyháva. Preto sú metódy založené na inštrumentálnych premenných dôležitým matematickým konceptom, ktorý môžeme použiť pri kauzálnom usudzovaní. Zjednodušenie si môžeme inštrumentálne predstaviť tak, že súvisia s ošetrovaním, ale nie s odozvou, ktorú skúmame. Vďaka tomu sú užitočné na identifikáciu kauzálnych účinkov v prítomnosti zastieracích premenných. My sa zamierame predovšetkým na metódu dvojstupňového odhadu pomocou najmenších štvorcov. Popíšeme konkrétne situácie, kedy môžeme použiť inštrumentálne premenné a uvedieme predpoklady, na základe ktorých môžeme robiť kauzálne závery. Nakoniec je uvedená simulačná štúdia pre vybrané situácie.

1 Model potenciálnych výsledkov

V tejto kapitole predstavíme jeden z moderných prístupov používaných pri kauzálnom usudzovaní. Je ním model kauzálného usudzovania založený na potenciálnych výsledkoch (*counterfactual model*). Tento model prvýkrát navrhol Neyman [2] a neskôr sa mu venoval až Rubin [3]. Pri popisovaní tohto modelu sa budeme opierať predovšetkým o znalosti z knihy Wassermana [4] a knihy Hernána a Robinsa [5]. Bolo pridaných niekoľko vlastných príkladov a dôkazov niektorých viet. Ďalej boli oproti [4] doplnené niektoré vzťahy, ktoré prispievajú k celkovému pochopeniu modelu potenciálnych výsledkov. Podkapitola týkajúca sa Simpsonovho paradoxu navyše obsahuje detailne rozobrané niektoré z možných situácií pri práci s tromi premennými. Je poskytnutý pohľad na tieto situácie pomocou modelu potenciálnych výsledkov. Oporou pri ukazovaní vzťahov bola predovšetkým [5].

1.1 Binárne ošetrovanie

Predstavme si, že pracujeme s dvomi náhodnými veličinami X a Y . Pri skúmaní kauzálnych vzťahov medzi nimi je dôležité rozlíšiť prípad, kedy X je príčinou Y , od prípadu, kedy X a Y sú asociované (*associated*). Vieme, že ak X je príčinou Y , tak nutne X a Y sú asociované (X a Y nie sú nezávislé). Opačná implikácia však nemusí platiť. Pre skúmanie kauzálnych súvislostí v tejto kapitole predstavíme model potenciálnych výsledkov (*counterfactual model*). Tento model prvýkrát navrhol [2] a neskôr sa mu venoval až [3].

Pre jednoduchosť zo začiatku uvažujme X ako binárnu premennú, ktorá napríklad reprezentuje ošetrovanie. X má hodnotu 1, ak dochádza k ošetrovaniu a X má hodnotu 0, ak k ošetrovaniu nedochádza. Pod ošetrovaním si môžeme predstaviť podávanie liečiva, vakcíny alebo napríklad fajčenie. Ak nedochádza k ošetrovaniu znamená to, že napríklad bolo podané placebo alebo že skúmaný subjekt nefajčí.

Po podaní ošetrovania nás väčšinou zaujíma, či malo nejaký vplyv na zdravie pacienta. Premennú, ktorá v sebe obsahuje výsledok/odozvu experimentu budeme označovať Y . Pre rozlíšenie situácie, kedy X je príčinou Y a situácie, kedy X a Y sú asociované, budeme pracovať s potenciálnymi výsledkami (C_0, C_1). Výsledok pre situáciu, kedy nedochádza k ošetrovaniu subjektu, s ktorým pracujeme ($X = 0$), označíme C_0 (napr. zdravotný stav, ak pacient dostal placebo). Výsledok pre situáciu, kedy dochádza k ošetrovaniu subjektu, s ktorým pracujeme ($X = 1$), označíme C_1 (napr. zdravotný stav, ak pacient dostal ošetrovanie). Teda

$$Y = \begin{cases} C_0, & \text{ak pacient nedostal ošetrovanie } (X = 0), \\ C_1, & \text{ak pacient dostal ošetrovanie } (X = 1). \end{cases}$$

To môžeme vyjadriť aj vzťahom

$$Y = C_X.$$

Tento vzťah nazývame tiež vzťah konzistencie. Inými slovami každý pacient má dva potenciálne výsledky C_0 a C_1 . To platí za predpokladu SUTVA (Stable

Unit Treatment Value Assumption, [6]). Ten hovorí, že potenciálny výsledok pre pacienta v štúdií nie je ovplyvnený priradením ošetrovania iným pacientom v štúdií a že sa uvažuje len jedna verzia ošetrovania.

Inými slovami, SUTVA vyžaduje, aby medzi pacientami nedochádzalo k vzájomnému ovplyvňovaniu a aby neexistovali žiadne rozdiely v ošetrovaní. Tento predpoklad je rozhodujúci pre správne vyhodnotenie kauzálnych záverov a budeme v celom texte predpokladať, že je tento predpoklad splnený.

Príklad 1.1. Pre ujasnenie uvažujme nasledujúcu populáciu:

X	Y	C_0	C_1
0	1	1	*
0	1	1	*
1	0	*	0
1	1	*	1

Potenciálne výsledky, ktoré nepozorujeme označujeme znakom *. Ak je pre konkrétny subjekt hodnota $X = 0$, tak pozorujeme C_0 , avšak nepozorujeme C_1 , ktoré predstavuje potenciálny výsledok, ktorý by sme pozorovali, ak by sme sa rozhodli subjekt ošetriť ($X = 1$). Tiež hovoríme, že C_1 je kontrafaktuál (*contrafactual outcome*). Potenciálne výsledky (C_0, C_1) môžeme vnímať ako premenné, ktoré v sebe uchovávajú všetku relevantnú informáciu o danom subjekte.

Podľa toho, aké sú hodnoty potenciálnych výsledkov rozlišujeme štyri typy subjektov:

Typ	C_0	C_1
<i>Prežívajúci</i>	1	1
<i>Respondenti</i>	0	0
<i>Anti-Respondenti</i>	1	0
<i>Zomierajúci</i>	0	0

Vidíme, že pre prežívajúcich (*survivors*) a zomierajúcich (*doomed*), hodnota odozvy nezávisí na tom, či im bolo podané ošetrovanie alebo nie. V tejto skupine teda je kauzálny efekt ošetrovania nulový, pretože $C_1 - C_0 = 0$. Naopak v skupine respondentom (*responders*) a anti-respondentov (*anti-responders*) je nenulový kauzálny efekt medzi ošetrovaním a odozvou, pretože $C_1 - C_0 \neq 0$. Je však dobré si uvedomiť, že v praxi nevieme tieto skupiny rozlíšiť. Ak sa pozrieme na populáciu v Príklade 1.1, tak nevieme či je prvý subjekt prežívajúci alebo anti-respondent. Podobne nevieme, či je druhý subjekt zomierajúci alebo respondent.

Definícia 1. *Stredný kauzálny efekt ošetrovania (X) na odozvu (Y) definujeme rozdielom*

$$\theta = E[C_1] - E[C_0].$$

Sú aj iné spôsoby, ako merať kauzálny efekt. Napríklad pre C_0, C_1 binárne, definujeme kauzálny pomer šancí ako

$$\frac{P(C_1 = 1)}{P(C_1 = 0)} \div \frac{P(C_0 = 1)}{P(C_0 = 0)}.$$

a kauzálné relatívne riziko ako

$$\frac{P(C_1 = 1)}{P(C_0 = 1)}.$$

Pre jednoduchosť budeme pracovať len s stredným kauzálnym efektom θ .

Definícia 2. *Asociáciu ošetrenia X s odozvou Y definujeme rozdielom*

$$\alpha = E[Y|X = 1] - E[Y|X = 0].$$

Všeobecne neplatí rovnosť $\alpha = \theta$. Ak by populácia, na ktorej skúmame odozvu Y na ošetrenie X obsahovala len subjekty, ktoré prežívajú alebo zomierajú, tak by stredný kauzálny efekt X na Y bol nulový ($\theta = 0$). Avšak asociácia α nemusí byť nulová. Môžeme to vidieť na nasledujúcom príklade.

Príklad 1.2. Uvažujme populáciu:

X	Y	C_0	C_1
0	0	0	0*
0	0	0	0*
0	0	0	0*
0	0	0	0*
1	1	1*	1
1	1	1*	1
1	1	1*	1
1	1	1*	1

Vidíme, že $C_0 = C_1$ pre všetky subjekty, takže ošetrenie nemá žiadny efekt na odozvu.

$$\begin{aligned} \theta &= E[C_1] - E[C_0] = \frac{1}{8} \sum_{i=1}^8 C_{1i} - \frac{1}{8} \sum_{i=1}^8 C_{0i} \\ &= \frac{0 + 0 + 0 + 0 + 1 + 1 + 1 + 1}{8} - \frac{0 + 0 + 0 + 0 + 1 + 1 + 1 + 1}{8} \\ &= 0. \end{aligned}$$

Stredný kauzálny efekt je nulový ($\theta = 0$). Pre hodnotu asociácie α medzi ošetrením X a odvozvou Y , dostávame

$$\begin{aligned} \alpha &= E[Y|X = 1] - E[Y|X = 0] \\ &= \frac{1 + 1 + 1 + 1}{4} - \frac{0 + 0 + 0 + 0}{4} = 1. \end{aligned}$$

Hodnota asociácie α sa nerovná hodnote stredného kauzálného efektu θ . Tiež vidíme, že θ nezávisí na tom, ako zvolíme hodnoty ošetrenia pre jednotlivé subjekty, avšak α na tom závisí.

Ak označíme $\delta = C_1 - C_0$, potom môžeme stredný kauzálny efekt X na Y rozpísať ako

$$\theta = E[C_1] - E[C_0] = E[C_1 - C_0] = E[\delta].$$

Ak podmienime priradením subjektu do skupiny, ktorá je ošetroená ($X = 1$) alebo priradením do kontrolnej skupiny ($X = 0$), dostávame

$$\begin{aligned} E[\delta|X = 1] &= E[C_1|X = 1] - E[C_0|X = 1], \\ E[\delta|X = 0] &= E[C_1|X = 0] - E[C_0|X = 0]. \end{aligned}$$

Ak ďalej označíme $\pi = P(X = 1)$, môžeme $E[\delta]$ rozpísať z vety o úplnej pravdepodobnosti ako

$$\begin{aligned} E[\delta] &= \pi E[\delta|X = 1] + (1 - \pi)E[\delta|X = 0] \\ &= E[\delta|X = 1] - (1 - \pi)E[\delta|X = 1] + (1 - \pi)E[\delta|X = 0] \\ &= E[\delta|X = 1] - (1 - \pi)(E[\delta|X = 1] - E[\delta|X = 0]) \\ &= E[C_1|X = 1] - E[C_0|X = 1] - (1 - \pi)(E[\delta|X = 1] - E[\delta|X = 0]). \end{aligned}$$

Následne pre α s využitím vzťahu $Y = C_X$ dostávame

$$\alpha = E[C_1|X = 1] - E[C_0|X = 0] = E[Y|X = 1] - E[Y|X = 0].$$

Ak teda pokračujeme v rozpise $E[\delta]$ dostávame

$$\alpha = E[\delta] + E[C_0|X = 1] - E[C_0|X = 0] + (1 - \pi)(E[\delta|X = 1] - E[\delta|X = 0]). \quad (1.1)$$

Pripomeňme, že $E[\delta] = \theta$ označuje stredný kauzálny efekt X na Y . Ak by sme teda namiesto θ pracovali s α , vidíme dva možné zdroje vychýlenia. Prvým zdrojom je základné vychýlenie (*baseline bias*), ktoré má tvar $E[C_0|X = 1] - E[C_0|X = 0]$. Druhým zdrojom je vychýlenie dané rozdielnym efektom ošetroenia (*differential treatment effect bias*). Má tvar

$$(1 - \pi)(E[\delta|X = 1] - E[\delta|X = 0]).$$

Tieto zdroje vychýlenia sú viac diskutované v [7]. Ich význam si priblížime v nasledujúcom príklade.

Príklad 1.3. Uvažujme populáciu z Príkladu [1.2]:

X	Y	C_0	C_1
0	0	0	0*
0	0	0	0*
0	0	0	0*
0	0	0	0*
1	1	1*	1
1	1	1*	1
1	1	1*	1
1	1	1*	1

Potom z Príkladu [1.2] vieme, že $\theta = 0$ a $\alpha = 1$. Základné vychýlenie v tomto prípade bude

$$E[C_0|X = 1] - E[C_0|X = 0] = \frac{1 + 1 + 1 + 1}{4} - \frac{0 + 0 + 0 + 0}{4} = 1.$$

Vychýlenie dané rozdielnym efektom ošetrovania v tomto prípade bude nulové (populáciu tvoria len subjekty, ktoré zomierajú alebo prežívajú, $\delta = 0$ pre každý subjekt). Dostávame

$$(1 - \pi)(E[\delta|X = 1] - E[\delta|X = 0]) = (1 - \frac{1}{2})(\frac{0 + 0 + 0 + 0}{4} - \frac{0 + 0 + 0 + 0}{4}) = 0.$$

V tomto príklade môžeme pod ošetrovaním rozumieť napríklad podanie liečiva a výsledok by mohol označovať zlepšenie zdravotného stavu pacienta ($Y = 1$). Potom by sa nám zdalo, že dochádza k pozitívnemu účinku liečiva, aj keď by stredný kauzálny efekt bol nulový.

Ak by sme sa rozhodli zvýšiť počet ošetrovaných pacientov, situácia by vyzerala nasledovne:

X	Y	C_0	C_1
0	0	0	0*
0	0	0	0*
1	0	0*	0
1	0	0*	0
1	1	1*	1
1	1	1*	1
1	1	1*	1
1	1	1*	1

V tomto prípade by sa hodnota $\alpha = \frac{4}{6} - \frac{0}{2} = \frac{4}{6}$ znížila z 1 na $\frac{4}{6}$. Kauzálny efekt by bol rovnaký, ale experimentátor, ktorý nerozlišuje asociáciou a kauzalitou, by mohol ostať zmätený.

Príklad 1.4. Uvedieme ešte jeden príklad. Ak by naša populácia mala tvar

X	Y	C_0	C_1
0	1	1	1*
0	1	1	1*
1	1	0*	1
1	1	0*	1

Asociácia medzi X a Y by bola $\alpha = \frac{2}{2} - \frac{2}{2} = 0$, aj keď $\theta = \frac{4}{4} - \frac{2}{4} = \frac{1}{2}$. Ak by sme sa teda v tomto prípade riadili iba na základe asociácie α , chybné by sme mohli urobiť záver, že kauzálny efekt je nulový. Základné vychýlenie by v tomto prípade bude

$$E[C_0|X = 1] - E[C_0|X = 0] = \frac{0 + 0}{2} - \frac{1 + 1}{2} = -1.$$

Vychýlenie dané rozdielnym efektom ošetrovania v tomto prípade bude

$$(1 - \pi)(E[\delta|X = 1] - E[\delta|X = 0]) = (1 - \frac{1}{2})(\frac{1 + 1}{2} - \frac{0 + 0}{2}) = \frac{1}{2}.$$

Ak by sme s nenulovou pravdepodobnosťou ($P(X = 0) > 0$ a $P(X = 1) > 0$) každému subjektu priradili ošetrovanie náhodne, ošetrovanie X je nezávislé s (C_0, C_1) a teda platí

$$\begin{aligned} E[C_0|X = 1] &= E[C_0|X = 0], \\ E[C_1|X = 1] &= E[C_1|X = 0]. \end{aligned}$$

V tom prípade by hodnota základného vychýlenia bola nulová

$$E[C_0|X = 1] - E[C_0|X = 0] = 0.$$

Hodnota vychýlenia spôsobeného rozdielnym efektom ošetrovania by bola tiež nulová, pretože

$$\begin{aligned} (1 - \pi)(E[\delta|X = 1] - E[\delta|X = 0]) &= \\ (1 - \pi)(E[C_1|X = 1] - E[C_0|X = 1] - E[C_1|X = 0] + E[C_0|X = 0]) &= 0. \end{aligned}$$

Z formuly (1.1) teda dostávame $\alpha = E[\delta] = \theta$. Tento výsledok zhrnieme v nasledujúcej vete.

Veta 1. *Ak každému subjektu s nenulovou pravdepodobnosťou ($P(X = 0) > 0$ a $P(X = 1) > 0$) priradíme ošetrovanie náhodne, potom je stredný kauzálny efekt rovný asociácii, $\alpha = \theta$.*

Dôkaz. Keďže predpokladáme, že ošetrovanie je priradené náhodne, X je nezávislé s (C_0, C_1) . Teda dostávame

$$\begin{aligned} E[C_0|X = 1] &= E[C_0|X = 0], \\ E[C_1|X = 1] &= E[C_1|X = 0]. \end{aligned}$$

Vieme, že pre stredný kauzálny efekt platí

$$\begin{aligned} \theta &= E[C_1] - E[C_0] \\ &= E[C_1|X_1] - E[C_0|X_0] && \text{pretože } X \text{ je nezávislé s } (C_0, C_1) \\ &= E[Y|X_1] - E[Y|X_0] && \text{pretože } Y = C_X \\ &= \alpha. \end{aligned} \quad \square$$

Z Vety 1 teda dostávame, že ak je splnený jej predpoklad, každý konzistentný odhad α je konzistentným odhadom θ .

V prípade, že ošetrovanie nie je priradené náhodne (predpoklad Vety 1 je porušený), môžeme pre stredný kauzálny efekt nájsť dolnú a hornú hranicu. Jej tvar dostávame po rozpise stredného kauzálneho efektu s použitím vety o úplnej pravdepodobnosti

$$\begin{aligned} \theta_z &= E[C_1] - E[C_0] \\ &= \pi E[C_1|X = 1] + (1 - \pi)E[C_1|X = 0] - \pi E[C_0|X = 1] - (1 - \pi)E[C_0|X = 0]. \end{aligned}$$

Vieme, že hodnoty π , $E[C_1|X = 1]$ a $E[C_0|X = 0]$ vieme konzistentne odhadnúť. Hodnoty $E[C_1|X = 0]$ a $E[C_0|X = 1]$ sa pohybujú v intervale $[0, 1]$. Dolná hranica preto bude mať tvar

$$\pi E[C_1|X = 1] - \pi - (1 - \pi)E[C_0|X = 0].$$

Horná hranica bude mať tvar

$$\pi E[C_1|X=1] + (1-\pi) - (1-\pi)E[C_0|X=0].$$

Šírka hranice bude

$$\pi E[C_1|X=1] + (1-\pi) - (1-\pi)E[C_0|X=0] - (\pi E[C_1|X=1] - \pi - (1-\pi)E[C_0|X=0]) = 1.$$

V situáciách, keď máme k dispozícii dodatočné informácie o subjektoch, môže byť užitočné skúmať kauzálny efekt pre konkrétnu skupinu subjektov.

Definícia 3. Pre regresor Z definujeme podmienený kauzálny efekt ako

$$\theta_z = E[C_1|Z=z] - E[C_0|Z=z].$$

Definícia 4. Pre regresor Z definujeme podmienenú asociáciu ako

$$\alpha_z = E[Y|X=1, Z=z] - E[Y|X=0, Z=z].$$

Príklad 1.5. Ak by naša populácia mala tvar

Z	X	Y	C_0	C_1
0	0	1	1	1*
0	1	1	0*	1
0	1	0	0*	0
0	1	0	0*	0
1	0	0	0	0*
1	0	0	0	0*
1	1	1	1*	1
1	1	0	1*	0

hodnoty podmieneného kauzálneho efektu by boli

$$\begin{aligned} \theta_0 &= E[C_1|Z=0] - E[C_0|Z=0] = \frac{2}{4} - \frac{1}{4} = \frac{1}{4}, \\ \theta_1 &= E[C_1|Z=1] - E[C_0|Z=1] = \frac{1}{4} - \frac{2}{4} = -\frac{1}{4}. \end{aligned}$$

Hodnoty podmienenej asociácie by boli

$$\begin{aligned} \alpha_0 &= E[Y|X=1, Z=0] - E[Y|X=0, Z=0] = \frac{1}{3} - 1 = -\frac{2}{3}, \\ \alpha_1 &= E[Y|X=1, Z=1] - E[Y|X=0, Z=1] = \frac{1}{2} - \frac{0}{2} = \frac{1}{2}. \end{aligned}$$

Vidíme teda, že $\theta_0 > 0$, ale $\alpha_0 < 0$ a podobne $\theta_1 < 0$, ale $\alpha_1 > 0$. Pri odhadovaní kauzálneho efektu pomocou asociácie teda musíme byť opatrní, pretože ich znamienka sa môžu líšiť.

Nech sme v situácii randomizovaného experimentu, v ktorom X je podmienene nezávislé s (C_0, C_1) pri danom $Z = z$ pre všetky z . Pre podmienený kauzálny efekt $\theta_z = E[Y|X=1, Z=z] - E[Y|X=0, Z=z]$ potom platí.

$$\begin{aligned} \theta_z &= E[C_1|Z=z] - E[C_0|Z=z] \\ &= E[C_1|X_1, Z=z] - E[C_0|X_0, Z=z] \\ &= E[Y|X_1, Z=z] - E[Y|X_0, Z=z] \quad \text{pretože } Y = C_X \end{aligned} \quad (1.2)$$

$$\theta_z = \alpha_z. \quad (1.3)$$

Teda podmienený kauzálny efekt je rovný podmienenej asociácii.

1.2 Mnohopočetné ošetrenie

Prípád dichotómie pri ošetrení môžeme zovšeobecniť. Uvažujme teraz $X \in \mathcal{X}$. Môžeme si predstaviť, že ošetrenie bude predstavovať skúmanie väčšieho počtu liečiv ($X \in \mathbf{N}$) alebo podávanie určitého objemu liečiva ($X \in \mathbf{R}$). Namiesto kontrafaktuálov C_0 a C_1 , budeme pracovať funkciou kontrafaktuálov $C(x)$, kde $C(x)$ predstavuje výsledok subjektu, ktorý dostal dávku liečiva o objeme x . Pre pozorovanú odozvu teda platí vzťah konzistencie

$$Y = C(X).$$

Namiesto stredného kauzálneho efektu θ , budeme pracovať s kauzálnou regresnou funkciou

$$\theta(x) = E[C(x)].$$

Uvedomme si, že pokiaľ je hodnota $\theta(x)$ konštantná, kauzálny efekt je nulový. Ako analógiu k asociácií α môžeme zaviesť regresnú funkciu asociácie

$$r(x) = E[Y|X = x].$$

Opäť môžeme konštatovať, že ak je $r(x)$ konštantná, je asociácia medzi X a Y nulová.

Veta 2. *Nech je ošetrenie X náhodne priradené náhodne. Potom $\theta(x) = r(x)$ pre všetky $x \in \mathcal{X}$.*

Dôkaz. Keďže predpokladáme, že ošetrenie je priradené náhodne, X je nezávislé s $C(x)$ pre všetky $x \in \mathcal{X}$. Teda dostávame

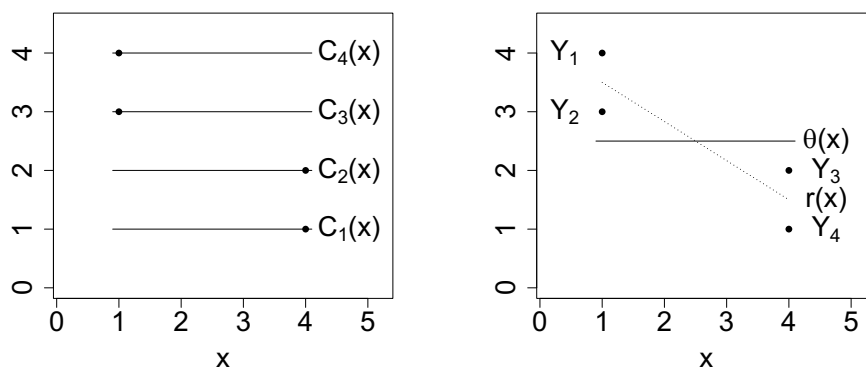
$$\begin{aligned} \theta(x) &= E[C(x)] = E[C(x)|X = x] && \text{pretože } X \text{ je nezávislé s } C(x) \\ &= E[Y|X = x] && \text{pretože } Y = C_X \\ &= r(x). \end{aligned} \quad \square$$

Príklad 1.6. Uvažujme populáciu na obr. 1.1. Hodnoty potenciálnych výsledkov sú konštantné pre všetky x . Kauzálna regresná funkcia bude mať tvar

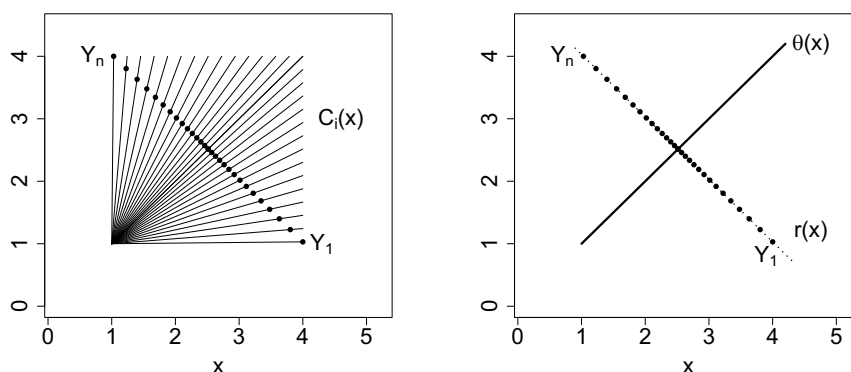
$$\theta(x) = E[C(x)] = \frac{C_1(x) + C_2(x) + C_3(x) + C_4(x)}{4}.$$

Ak teda pre daný subjekt zmeníme množstvo podávaného liečiva, výsledok sa nezmení. Avšak funkcia regresnej asociácie $r(x) = E[Y|X = x]$, vytvorená z pozorovaných hodnôt $Y_1 = C(X_1)$, $Y_2 = C(X_2)$, $Y_3 = C(X_3)$, $Y_4 = C(X_4)$, nie je konštantná. Vidíme, že hodnoty funkcie asociácie a kauzality sa môžu výrazne líšiť.

Príklad 1.7. Ďalšiu zaujímavú situáciu vidíme na obr. 1.2. Uvažujme, že pre ošetrenie platí $X \in \mathbf{R}$. Potenciálne výsledky $C_i(x)$ sú rastúce. Teda aj kauzálna regresná funkcia $\theta(x)$ bude rastúca. Avšak regresná funkcia asociácie $r(x) = E[Y|X = x]$ bude klesajúca. Vidíme teda, že ak hodnoty ošetrenia nie sú priradené náhodne, trendy kauzálnej regresnej funkcie a regresnej funkcie asociácie môžu byť opačné.



Obr. 1.1 Vľavo: Potenciálne výsledky. Vpravo: Pozorované výsledky. Zobrazený je tvar regresnej funkcie kauzality a regresnej funkcie asociácie pre mnohopočetné ošetrovanie pre Príklad 1.6. Vidíme, že sa môžu ich hodnoty výrazne líšiť.



Obr. 1.2 Vľavo: Potenciálne výsledky. Vpravo: Pozorované výsledky. Zobrazený je tvar regresnej funkcie kauzality a regresnej funkcie asociácie pre mnohopočetné ošetrovanie pre Príklad 1.7. Vidíme, že sa môžu ich hodnoty výrazne líšiť.

1.3 Observačné štúdie a zastrenie

Štúdia, v ktorej nie sú hodnoty ošetrovania priradené náhodne sa nazýva observačná štúdia. V týchto štúdiách veľa krát nie je možné subjektu priradiť hodnotu ošetrovania. Predstavme si, že by sme napríklad chceli skúmať vplyv fajčenia na rozvoj rakoviny pľúc (viz. napr. [8]). Ak by sme chceli urobiť randomizovanú štúdiu, o každom subjekte by sme museli náhodne rozhodnúť, či bude fajčiť alebo nie a skúmať vplyv fajčenia po dlhší časový úsek (typicky niekoľko rokov). To je v praxi takmer nedosiahnuteľné, preto si vyberáme subjekty podľa toho, či fajčia alebo nie. Vtedy už ale potenciálny výsledok nemusí byť nezávislý na hodnote ošetrovania. Avšak predpokladajme, že by sme vedeli nájsť podskupiny, v ktorých by X a $\{C(x), x \in \mathcal{X}\}$ boli nezávislé. Napríklad by sme mohli nájsť ľudí, ktorí sú v rovnakom veku, majú rovnaké pohlavie a sociálne zázemie. V rámci tejto podskupiny by nám mohlo pripadať rozumné predpokladať, že je výber ošetrovania (X) náhodný. Premenné, ktoré určujú túto podskupiny budeme nazývať zastieracie

(*confounding*) premenné (označíme ich U). Potom môžeme situáciu vyjadriť ako

$$\{C(x), x \in \mathcal{X}\} \text{ je podmienenene nezávislé s } X \text{ pri danom } U. \quad (1.4)$$

Inými slovami v rámci skupiny U , voľba ošetrenia (X) nezávisí na potenciálnom výsledku. Ak pozorujeme U a (1.4) platí, hovoríme, že nie je žiadne ďalšie nepozorované zastieranie. Tento predpoklad sa tiež nazýva predpoklad nezastierania (*unconfoundness assumption, ignorability, 9*).

Veta 3. *Predpokladajme, že (1.4) platí a nech U pochádza z rozdelenia s distribučnou funkciou F_U vzhľadom k Lebesgueovej miere. Potom*

$$\theta(x) = \int E[Y|X = x, U = u] dF_U(u). \quad (1.5)$$

Dôkaz. Pre kauzálnu regresnú funkciu platí

$$\begin{aligned} \theta(x) &= E[C(x)] = \int E[C(x)|U = u] dF_U(u) \\ &= \int E[C(x)|X = x, U = u] dF_U(u) \\ &= \int E[Y|X = x, U = u] dF_U(u) \\ &= \int r(x, u) dF_U(u). \end{aligned}$$

Prvú rovnosť dostávame z vety o úplnej pravdepodobnosti. Keďže predpokladáme, že ošetrenie je v rámci skupiny Z priradené náhodne, $\{C(x), x \in \mathcal{X}\}$ je podmienne nezávislé s X pri danom U . Z toho dostávame druhú rovnosť. Tretiu rovnosť dostávame, pretože $Y = C_X$. \square

Z Vety 3 teda dostávame, že ak $\hat{r}(x, z)$ je konzistentný odhad regresnej funkcie $r(x, z) = E[Y|X = x, U = u]$, potom konzistentný odhad $\theta(x)$ je

$$\hat{\theta}(x) = \frac{1}{n} \sum_{i=1}^n \hat{r}(x, U_i).$$

Teda, ak $r(x, u) = \beta_0 + \beta_1 x + \beta_2 u$ je lineárna, potom konzistentný odhad $\theta(x)$ dostávame ako

$$\hat{\theta}(x) = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 \bar{U}_n,$$

kde $\bar{U}_n = \frac{1}{n} \sum_{i=1}^n U_i$ a $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)$ sú konzistentné odhady $(\beta_0, \beta_1, \beta_2)$.

Po rozpísaní $r(x) = E[Y|X = x] = \int E[Y|X = x, U = u] dF_{U|X}(u|x)$ ju môžeme porovnať s (1.5). Rozdiel je v tom, že v $r(x)$ pracujeme s podmienenou distribučnou funkciou $F_{U|X}$.

Postup, pri ktorom počítame upravený efekt ošetrenia (1.5) pre dané skupiny môžeme nazvať korekciou na zastieranie. Premenné, ktoré spôsobujú zastieranie identifikujeme pomocou vedeckých poznatkov a konzultáciou s odborníkmi z oblasti, ktorú skúmame. Aj po korekcií na premenné, o ktorých si myslíme, že spôsobujú zmätok si nemôžeme byť istý, že sme na nejaké premenné nezabudli. Taktiež môžu nastať situácie, kedy premenné, ktoré prispievajú k zmätku nie sme

schopný zmerať. Preto by sme mali k observačným štúdiám pristupovať opatrne. Výsledky z observačných štúdií sa stávajú dôveryhodnými, ak dostávame podobné výsledky z viacerých štúdií, každá zo štúdií kontroluje pravdepodobné zastieracie faktory a existuje dôveryhodné vedecké vysvetlenie pre existenciu kauzálneho vzťahu. Dobrým nástrojom ako si uvedomiť, ktoré premenné máme kontrolovať je vytváranie DAGov (*directed acyclic graphs, orientované acyklické grafy*). Avšak aj pri nich si nie vždy môžeme byť istý ich správnosťou. DAGy budeme viac rozoberať v ďalšej kapitole.

Než prejdeme k ďalšej časti uvedieme zaujímavý postreh z článku [10], v ktorom sa pozeráme v modeli potenciálnych výsledkov na problém kauzálneho usudzovania ako na problém chýbajúcich pozorovaní. Môžeme potom urobiť paralelu medzi

- úplne randomizovanými kontrolovanými experimentmi a úplne náhodne chýbajúcimi pozorovaniami (MCAR, missing completely at random)
- observačnými štúdiami, kde nie je žiadne nepozorované zmätenie a náhodne chýbajúcimi pozorovaniami (MAR, missing at random)
- observačnými štúdiami, kde je nejaké nepozorované zmätenie a pozorovaniami chýbajúcimi nie náhodne (MNAR, missing not at random).

1.4 Simpsonov paradox

Model s potenciálnymi výsledkami je užitočný nástroj pre vysvetlenie Simpsonovho paradoxu, pri ktorom dochádza k zdanlivému rozporu medzi podmienenými a marginálnymi pravdepodobnosťami. Tento paradox prvýkrát v článku popísal [11], avšak už predtým podobný efekt spomenul [12] a [13]. V tejto podkapitole tiež detailne rozoberieme niektoré z možných situácií pri práci s tromi premennými, pri ktorých môže k Simpsonovmu paradoxu dochádzať. Bude poskytnutý pohľad na tieto situácie pomocou modelu potenciálnych výsledkov. Oporou pri ukazovaní vzťahov bola predovšetkým [5].

Nech X, Y, Z sú dichotomické (binárne) premenné pozorované v populácii 100 subjektov. Počty subjektov pre jednotlivé podskupiny našej populácie budú

	$Y = 1$	$Y = 0$	$Y = 1$	$Y = 0$
$X = 1$	4	16	3	12
$X = 0$	6	24	7	28
	$Z = 1$		$Z = 0$	

Ak sa pozrieme na podskupiny pre X a Y , dostávame

	$Y = 1$	$Y = 0$	
$X = 1$	7	28	35
$X = 0$	13	52	65
	20	80	100

Ak spočítame marginálny pomer šancí toho, že $Y = 1$, keď porovnáваме prípade s ošetrením ($X = 1$) s prípadom bez ošetrenia ($X = 0$) dostávame

$$\frac{P(Y = 1|X = 1)}{P(Y = 0|X = 1)} \div \frac{P(Y = 1|X = 0)}{P(Y = 0|X = 0)} = \frac{7}{28} \div \frac{13}{52} = 1$$

Avšak, ak spočítame pomer šancí podmienene pre skupinu Z dostávame

$$\begin{aligned} \frac{P(Y = 1|X = 1, Z = 1)}{P(Y = 0|X = 1, Z = 1)} \div \frac{P(Y = 1|X = 0, Z = 1)}{P(Y = 0|X = 0, Z = 1)} &= \frac{4}{16} \div \frac{6}{24} = \frac{1}{4} \\ \frac{P(Y = 1|X = 1, Z = 0)}{P(Y = 0|X = 1, Z = 0)} \div \frac{P(Y = 1|X = 0, Z = 0)}{P(Y = 0|X = 0, Z = 0)} &= \frac{3}{12} \div \frac{7}{28} = \frac{1}{4} \end{aligned}$$

Mohli by sme si preto myslieť, že ošetrenie X nemá efekt na odozvu Y , pretože pomer šancí pre X a Y je 1. Avšak, keď podmienime hodnotou $Z = 0,1$, pomer šancí pre X a Y je $\frac{1}{4}$, teda by sa nám zdalo, že v rámci skupiny ošetrenie X znižuje hodnotu odozvy Y .

Ak sa na problém pozrieme v reči potenciálnych výsledkov uvidíme, že k rozporu v skutočnosti nedochádza. Ak budeme predpokladať, že ošetrenie X znižuje hodnotu odozvy Y malo by platiť

$$P(C_1 = 1|Z = z) < P(C_0 = 1|Z = z),$$

pre všetky z . Ďalej platí

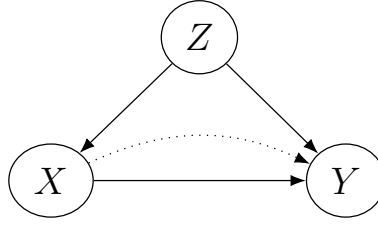
$$\begin{aligned} P(C_1 = 1) &= \sum_z P(C_1 = 1|Z = z)P(Z = z) \\ &< \sum_z P(C_0 = 1|Z = z)P(Z = z) \\ &= P(C_0 = 1). \end{aligned}$$

Dostávame, že $P(C_1 = 1) < P(C_0 = 1)$, teda ošetrenie marginálne neznižuje hodnotu odozvy.

Dôležité je si uvedomiť, aká je štruktúra medzi premennými X, Y, Z . Pre určenie jej tvaru nám však samotné pozorovania nestačia. Záleží na tom, aký majú jednotlivé premenné význam a konzultovať ich vzájomné prepojenia s odborníkmi v danej oblasti.

Jedným z možných scenárov je, že premenná Z spôsobuje zastieranie (confounding). Situáciu môžeme znázorniť pomocou orientovaného acyklického grafu (DAG), ktorý vidíme na obr. 1.3. Bodkovanou čiarou znázorňujeme, ktoré kauzálne prepojenie nás zaujíma. V našom prípade chceme vyhodnotiť, či ošetrenie X má kauzálny vplyv na Y . Ak by sme sa v tejto situácii obmedzili na skúmanie marginálneho kauzálneho efektu, dostali by sme, že nie je kauzálny vzťah medzi X a Y . To by však nebol správny postup. Keďže dochádza k zastieraniu, musíme vyhodnocovať kauzálny efekt pre jednotlivé skupiny Z (podmienene na $Z = z$). Tým kontrolujeme vplyv premennej Z .

Príklad 1.8. Tiež môžeme uviesť konkrétny príklad. Predstavme si, že by sme skúmali vplyv žltej farby nechtov (X) na rozvoj rakovinu pľúc (Y). V tomto



Obr. 1.3 DAG v ktorom dochádza k zastieraniu (*confounding*), medzi ošetrením X a odozvou Y , spôsobenému premennou Z .

prípade by sa nám po skúmaní populácie marginálne mohlo zdať, že žlté nechty spôsobujú rakovinu pľúc. To samozrejme nie je pravda. Vieme, že fajčenie cigariet (Z) zrejme spôsobuje žltnutie nechtov a taktiež rakovinu pľúc. Fajčenie cigariet (Z) je teda spoločnou príčinou pre X a Y . Ak budeme skúmať kauzálny efekt podmienene v skupine fajčiarov a nefajčiarov, kauzálny efekt X na Y sa vytratí. Ak si teda nie sme vedomí toho, že dochádza k zastieraniu a pri našich subjektoch nepozorujeme informáciu o tom, či fajčia alebo nie, dopúšťame sa chyby a chybne by sme mohli vyhodnotiť, že žlté nechty spôsobujú rakovinu pľúc.

Ak sa na situáciu pozrieme pomocou modelu potenciálnych výsledkov, chceme vyhodnotiť kauzálny efekt X na Y . To môžeme urobiť pomocou stredného kauzálneho efektu, ktorý má tvar

$$\theta = E[C_1] - E[C_0].$$

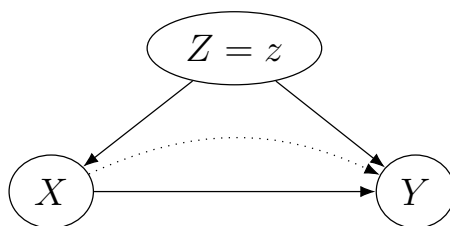
Na jeho výpočet potrebujeme hodnoty potenciálnych výsledkov. K dispozícii však máme len hodnoty pre výpočet asociácie

$$\alpha = E[Y|X = 1] - E[Y|X = 0].$$

Ak by platilo, že hodnoty (C_0, C_1) sú nezávislé s hodnotami ošetrenia X , tak by sme mohli použiť Vetu 1 a namiesto θ by sme mohli pracovať s α . V tejto situácii však nie sme, pretože hodnoty X aj Y sú závislé na Z (viz. obr. 1.3). Teda hodnoty X nie sú priradené náhodne. Pripomeňme tiež, že $Y = C_X$, teda potenciálne výsledky sú závislé na Z . Teda všeobecne neplatí rovnosť $\theta = \alpha$. Ak však podmienime hodnotou $Z = z$, veličina Z prestane byť náhodná a dostaneme sa do situácie na obr. 1.4. Teda hodnota ošetrenia X bude podmienene nezávislá na ostatným veličinách, pri danom $Z = z$. Potom z (1.3) dostávame, že $\theta_z = \alpha_z$. Teda ak pracujeme marginálne, asociácia nemusí odpovedať strednému kauzálnemu efektu. Avšak keď pracujeme s podmienene pre $Z = z$, asociácia odpovedá strednému kauzálnemu efektu.

Ďalšia možná situácia je znázornená na obr. 1.5. V tomto prípade dochádza k mediácii (*mediation*) kauzálneho efektu X na Y , pomocou premennej Z . Ak by sme teda skúmali len kauzálny efekt X na Y podmienene na Z , dostali by sme hodnotu kauzálneho efektu, ktorý nie je sprostredkovaný cez Z . Ten môže byť zanedbateľný. Nás ale zaujíma celkový kauzálny efekt X na Y , preto správny prístup je skúmať marginálny kauzálny efekt X na Y .

Príklad 1.9. Ak by sme napríklad u fajčiarov skúmali vplyv odvykania od fajčenia (hodnota X by bola priradená náhodne ($P(X = 0), P(X = 1) > 0$), $X = 0$,



Obr. 1.4 DAG v ktorom dochádza ku kontrole zastierania, medzi ošetrovaním X a odozvou Y , spôsobenému premennou Z , podmienením hodnotou $Z = z$.

ak jednotlivec prestal fajčiť, $X = 1$, ak jednotlivec pokračoval fajčiť a uvažujeme, že všetci jednotlivci dodržali nariadenie o fajčení) na riziko infarktu myokardu po jednom roku (Y). Mediátorom by bola napríklad znížená hodnota hypertenzie (Z). Zaujímalo by nás, či je znížené riziko infarktu myokardu spôsobené efektom odvykania na hypertenziu.

Celkový stredný kauzálny efekt X na Y môžeme pomocou potenciálnych výsledkov prepísať ako

$$\begin{aligned} \theta &= E[C_1] - E[C_0] = E[C_{1,Z^1}] - E[C_{0,Z^0}] \\ &= E[C_{1,Z^1}] - E[C_{1,Z^0}] + E[C_{1,Z^0}] - E[C_{0,Z^0}] \end{aligned}$$

kde Z^x pre daný subjekt predstavuje potenciálnu hodnotu Z , ak by hodnota ošetrovania bola x ($X = x$, v našom prípade $x \in \{0,1\}$). C_{k,Z^x} potom predstavuje potenciálnu hodnotu výsledky Y , ak by hodnota ošetrovania X bola k a potenciálna hodnota Z by bola Z^x . Potom definujeme priamy kauzálny efekt X na Y , ktorý nie je sprostredkovaný cez Z , ako

$$E[C_{1,Z^0}] - E[C_{0,Z^0}].$$

Ide teda o stredný kauzálny efekt X na Y , ak by pre každého jednotlivca bola hodnota Z nastavená na hodnotu, ktorá by nastala, ak by sa $X = 0$, teda ak by sa Z nastavilo na hodnotu $Z_{x=0}$ (čo môže byť 1 pre niektorých jednotlivcov a 0 pre ostatných).

V našom príklade by teda priamy kauzálny efekt odvykania od fajčenia (X) bola hodnota, ktorú dostaneme ak by sme mohli každej osobe priradiť hodnotu hypertenzie Z (1 alebo 0), ktorú by mala, keby prestala fajčiť ($X = 0$). Táto hodnota hypertenzie Z je známa pre osoby, ktoré skutočne prestali fajčiť, ale zostáva neznáma pre osoby, ktoré naďalej fajčili. Preto vidíme, že $E[C_{x=1,Z^x=0}]$ a teda aj priamy kauzálny efekt, je veličina z dvoch svetov. Dôvodom je, že zahŕňa kontrafaktuálny výsledok indexovaný dvoma hodnotami ošetrovania, $x = 1$ a $x = 0$, ktoré nemôžu nastať súčasne pre toho istého jednotlivca v tom istom svete.

Nepriamy kauzálny efekt X na Y definujeme ako

$$E[C_{1,Z^1}] - E[C_{1,Z^0}].$$

Vidíme, že aj nepriamy kauzálny efekt obsahuje veličinu $E[C_{x=1,Z^x=0}]$, ktorá je indexovaná oboma hodnotami ošetrovania. Vidíme teda, že celkový kauzálny efekt vieme rozložiť ako súčet priameho a nepriameho kauzálneho efektu.

Ak skúmame celkový (marginálny) kauzálny efekt X na Y , tak môžeme použiť Vetu [1](#), pretože hodnoty X sú priradené náhodne a predpokladáme, že $P(X = 0) > 0$, $P(X = 1) > 0$.

Hodnotu $E[C_{x=1,Z^x=0}]$ môžeme z vety o úplnej pravdepodobnosti rozpisat ako

$$\begin{aligned}
E[C_{x=1,Z^x=0}] &= \sum_z E[C_{x=1,z}|Z^{x=0} = z]P(Z^{x=0} = z) \\
&= \sum_z E[C_{x=1,z}]P(Z^{x=0} = z) \\
&= \sum_z E[C_{x=1,z}|X = 1,Z = z]P(Z^{x=0} = z) \\
&= \sum_z E[C_{x=1,z}|X = 1,Z = z]P(Z^{x=0} = z|X = 0) \\
&= \sum_z E[Y|X = 1,Z = z]P(Z = z|X = 0).
\end{aligned}$$

Kde sme v druhej rovnosti použili, že $C_{x=1,z} \perp\!\!\!\perp Z^{x=0}$. V tretej rovnosti sme využili, že $C_{x=1,z} \perp\!\!\!\perp \{X,Z\}$ a $Z^x \perp\!\!\!\perp X$. Pri poslednej rovnosti používame vzťahy konzistencie $Y = C_{X,Z}$ a $Z^X = Z$. Tento rozpis je pomerne technický rozpis a spája v sebe model potenciálnych výsledkov a DAGy. Pre hlbšie nahliadnutie môžu slúžiť napríklad [5](#).

Podobne pre $E[C_{x=0,Z^x=0}]$ a $E[C_{x=1,Z^x=1}]$ dostávame

$$\begin{aligned}
E[C_{x=0,Z^x=0}] &= \sum_z E[Y|X = 0,Z = z]P(Z = z|X = 0), \\
E[C_{x=1,Z^x=1}] &= \sum_z E[Y|X = 1,Z = z]P(Z = z|X = 1).
\end{aligned}$$

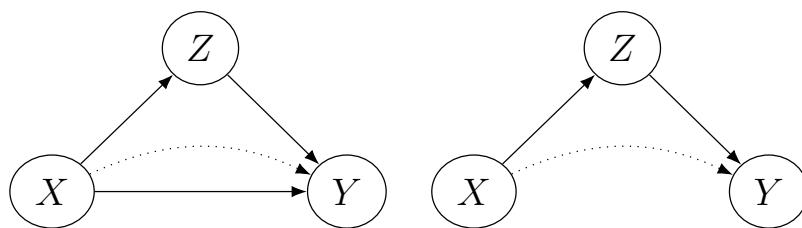
Ak by nás teda zaujímal podmienený kauzálny efekt X na Y pri danom $Z = z$, vypočítame ho súčtom podmieneného nepriameho a priameho kauzálneho efektu. Nepriamy kauzálny efekt bude nulový

$$\begin{aligned}
&E[C_{x=1,Z^x=1}|Z = z] - E[C_{x=1,Z^x=0}|Z = z] \\
&= E[Y|X = 1,Z = z] - E[Y|X = 1,Z = z] = 0.
\end{aligned}$$

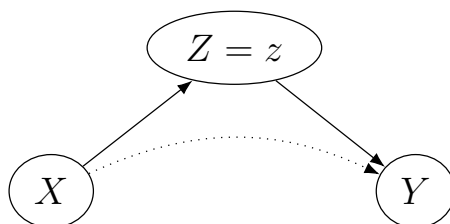
Tento záver je intuitívne z obr. [1.4](#) a obr. [1.6](#) zrejmý. Pri podmienení hodnotou $Z = z$ sa preruší kauzálna previazanosť medzi X,Z a medzi Z,Y . Ak je navyše priamy kauzálny efekt X na Y nulový, je podmienený efekt X na Y pri danom $Z = z$ tiež nulový (viz. obr. [1.6](#)). Ak je však priamy kauzálny efekt X na Y nenulový (viz. obr. [1.4](#)), pri podmienení hodnotou $Z = z$ dostávame

$$\begin{aligned}
&E[C_{x=1,Z^x=0}|Z = z] - E[C_{x=0,Z^x=0}|Z = z] \\
&= E[Y|X = 1,Z = z] - E[Y|X = 0,Z = z] = \theta_z.
\end{aligned}$$

Preto, ak chceme vypočítat celkový kauzálny efekt X na Y , správny prístup je skúmať marginálny kauzálny efekt X na Y . Výpočet podmieneného kauzálneho efektu môže byť užitočný, ak nás zaujíma priamy kauzálny efekt X na Y . Kombináciou týchto prístupov môžeme skúmať nepriamy kauzálny efekt X na Y . Pri skúmaní priameho a nepriameho kauzálneho efektu však nemusí byť štruktúra podobná obr. [1.5](#). Niektoré zložitejšie situácie sú rozobrané napríklad v [14](#) a [5](#).



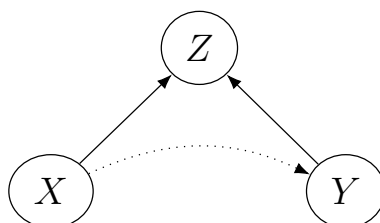
Obr. 1.5 Vľavo: Mediácia efektu X na Y prostredníctvom Z . Navyiac je nenulový priamy efekt X na Y . Vpravo: Mediácia bez priameho efektu X na Y .



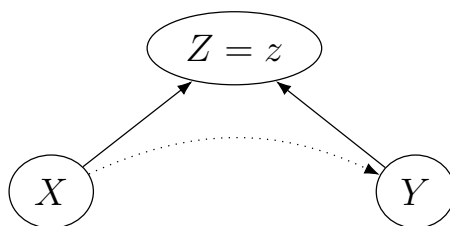
Obr. 1.6 Mediácia efektu X na Y prostredníctvom Z , bez priameho efektu X na Y , pričom dochádza k podmieneniu hodnotou $Z = z$.

Môže tiež nastať situácia, kedy X a Y sú spoločnou príčinou pre Z . Túto situáciu vidíme na obr. 1.7. Vtedy hovoríme, že dochádza ku kolízii a Z predstavuje kolidujúcu premennú (*collider*). Podmienенý kauzálny efekt X na Y by v tomto prípade dával nesprávny výsledok, že existuje kauzálny vzťah medzi X a Y , aj keď by v skutočnosti žiaden vzťah neexistoval. Správne by sme v tejto situácii mali použiť marginálny kauzálny efekt X na Y .

Príklad 1.10. Príkladom situácie z obr. 1.7 by mohlo byť skúmanie vzťahov medzi genetickým faktorom (X), znečistením ovzdušia (Y) a rakovinou (Z). Znečistené ovzdušie aj genetické faktory zrejme sú príčinou rakoviny. Súčasne sa nezdá byť rozumné predpokladať kauzálny vzťah medzi genetickými faktormi a znečisteným ovzduším. Avšak, ak by sme podmienene skúmali populáciu ľudí, ktorí majú rakovinu, mohli by sme pozorovať inverzný vzťah medzi genetickými faktormi a znečisteným ovzduším. Ak totiž skúmane subjekt, ktorý má rakovinu a nežil v znečistenom ovzduší, pravdepodobnou príčinou rakoviny by mohli byť genetické faktory. Chybne by sa nám teda mohlo zdať, že existuje kauzálny efekt medzi genetickými faktormi a znečistením životného prostredia. Vychýlenie kauzálne efektu, ktoré nastáva v situácii kedy podmienime spoločným efektom nazývame výberovým vychýlením (*selection bias*).



Obr. 1.7 Kolízia X a Y v premennej Z .



Obr. 1.8 Kolízia X a Y v premennej Z , pri ktorej podmieniujeme hodnotou $Z = z$.

Ak sa na situáciu pozrieme pomocou modelu s potenciálnymi výsledkami, môžeme pre výpočet marginálneho kauzálneho efektu X na Y použiť Vetu [1](#), pretože hodnota X je nezávislá na ostatných veličinách ($P(X = 0), P(X = 1) > 0$). Dostávame teda, že $\theta = \alpha$ a teda marginálny kauzálny efekt odpovedá strednému kauzálnemu efektu X na Y . V našej situácii (viz. obr. [1.7](#)) by teda kauzálny efekt X na Y bol nulový. Ak by sme podmienili hodnotu $Z = z$, podmienený kauzálny efekt X na Y by v niektorej podskupine z mohol byť nenulový, aj keď X a Y nie sú kauzálne previazané. Pripomeňme, že pre θ platí

$$\theta = E[C_1 - C_0] = \sum_z E[C_1 - C_0 | Z = z] = \sum_z \theta_z.$$

Ak by sme napríklad uvažovali populáciu

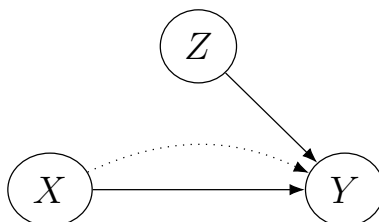
Z	X	Y	C_0	C_1
0	0	0	0	0*
0	0	1	1	0*
1	1	0	0*	0
1	1	1	0*	1

Stredný kauzálny efektu X na Y je pre túto populáciu nulový ($\theta = \frac{1}{4} - \frac{1}{4} = 0$), avšak hodnota podmieneného kauzálneho efektu X na Y je nenulová.

$$\theta_0 = E[C_1 - C_0 | Z = 0] = -\frac{1}{2},$$

$$\theta_1 = E[C_1 - C_0 | Z = 1] = \frac{1}{2}.$$

Vidíme teda, že nie nutne platí rovnosť $\theta = \theta_z$. Pre vyhodnotenie stredného kauzálneho efektu X na Y teda použijeme marginálny kauzálny efekt X na Y .



Obr. 1.9 Príklad DAGu pre randomizovanú štúdiu.

Príklad štruktúry pre randomizovanú štúdiu môžeme vidieť na obr. [1.9](#). Vidíme, že hodnota ošetrenia X je nezávislá s ostatnými veličinami. Pre odhad stredného

kauzálneho efektu X na Y teda môžeme použiť Vetu [1](#) ($P(X = 0), P(X = 1) > 0$). Veličina Z , ktorá má kauzálny vplyv na výsledok Y nemusí byť pozorovaná, pretože je nezávislá s X . Oproti tomu príklad štruktúry pre observačnú štúdiu predstavuje obr. [1.3](#). V tomto prípade je pre správne vyhodnotenie kauzálne efektu X na Y nutné pozorovať veličinu Z .

Vo všeobecnosti môžu mať kauzálne diagramy rozmanitú štruktúru. Viac sa im budeme venovať v ďalšej kapitole. Z uvedených situácií na obr. [1.3](#) až obr. [1.9](#) však vidíme, že analýza kauzálnych vzťahov sa môže v konkrétnej situácii výrazne odlišovať.

2 Orientované acyklické grafy

V tejto kapitole zavedieme orientované acyklické grafy (*directed acyclic graphs*, DAGS) a uvedieme niektoré ich základné vlastnosti. DAGy sú oproti modelu s potenciálnymi výsledkami ďalším nástrojom používaným pri kauzálnom usudzovaní. Pri ich popise budeme opäť vychádzať predovšetkým z knihy Wassermana [4] a knihy Hernána a Robinsa [5]. Prínosom autora je podrobnejšie rozpísanie niektorých situácií a pridanie vlastných príkladov, avšak snažíme sa zavedenie nových pojmov, preto sa budeme držať [4]. V podkapitole *vychýlenie meraním* sú uvedené DAGy, ktoré vychádzajú z [5]. Na záver kapitoly uvádzame možné rozšírenie DAGov a modelu potenciálnych výsledkov, ktorým sú SWIGy (*Single-World Intervention Graphs*, SWIGs, viz. [15]).

2.1 Úvod

V minulej kapitole sme videli príklady orientovaných grafov (viz. obr. 1.3 - obr. 1.7). Tie sú súborom uzlov so šípkami medzi niektorými uzlami. Predstavujú tak intuitívny spôsob ako skúmať kauzálne vzťahy. Z matematického pohľadu pomocou DAGov reprezentujeme vzťahy nezávislosti medzi skúmanými premennými. Neskôr uvidíme, že táto reprezentácia nemusí byť jednoznačná. Môže sa teda stať, že rôzne DAGy prezentujú tie isté vzťahy nezávislosti.

Pripomeňme definíciu nezávislosti a podmienenej nezávislosti.

Definícia 5. Uvažujme dve náhodne veličiny X a Y definované na (Ω, \mathcal{A}, P) . Predpokladajme, že X , resp. Y pochádza z rozdelenia s distribučnou funkciou $F_X(x)$, resp. $F_Y(y)$. Náhodné veličiny X a Y sú nezávislé práve vtedy keď, pre združenú distribučnú funkciu náhodnej veličiny (X, Y) platí

$$F_{X,Y}(x,y) = F_X(x)F_Y(y) \text{ pre všetky } x,y.$$

Ekvivalentne, ak existujú príslušné hustoty $f_X(x)$ a $f_Y(y)$ a existuje príslušná združená hustota $f_{X,Y}(x,y)$ (vzhľadom k Lebesgueovej miere),

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) \text{ pre všetky } x,y.$$

Budeme používať značenie $X \perp Y$.

Definícia 6. Uvažujme náhodne veličiny X , Y a Z definované na (Ω, \mathcal{A}, P) . Predpokladajme, že X , resp. Y podmienené pri danom Z , pochádza z rozdelenia s distribučnou funkciou $F_{X|Z}(x|z)$, resp. $F_{Y|Z}(y|z)$. Náhodné veličiny X a Y sú podmienené nezávislé pri danom Z práve vtedy keď, pre podmienenú združenú distribučnú funkciu náhodnej veličiny (X, Y) pri danom Z , platí

$$F_{X,Y|Z}(x,y|z) = F_{X|Z}(x|z)F_{Y|Z}(y|z) \text{ pre všetky } x,y \text{ a } z.$$

Ekvivalentne, ak existujú príslušné podmienené hustoty $f_{X|Z}(x|z)$ a $f_{Y|Z}(y|z)$ a existuje príslušná podmienená združená hustota $f_{X,Y|Z}(x,y|z)$ pri danom Z (vzhľadom k Lebesgueovej miere)

$$f_{X,Y|Z}(x,y|z) = f_{X|Z}(x|z)f_{Y|Z}(y|z) \text{ pre všetky } x,y \text{ a } z.$$

Z definície podmienenej hustoty, pre spojitú náhodnú veličinu, dostávame

$$f_{X,Y|Z}(x,y|z) = f_{X|Y,Z}(x|y,z)f_{Y|Z}(y|z) \text{ pre všetky } x,y \text{ a } z.$$

Po dosadení do $f_{X,Y|Z}(x,y|z)$ teda môžeme podmienenú nezávislosť ekvivalentne definovať pomocou vzťahu

$$f_{X|Y,Z}(x|y,z) = f_{X|Z}(x|z) \text{ pre všetky } x,y \text{ a } z.$$

Budeme používať značenie $X \perp\!\!\!\perp Y \mid Z$.

Veta 4. *Pre podmienenú nezávislosť platia implikácie*

$$\begin{aligned} X \perp\!\!\!\perp Y \mid Z &\implies Y \perp\!\!\!\perp X \mid Z \\ X \perp\!\!\!\perp Y \mid Z \text{ a } U = h(X) &\implies U \perp\!\!\!\perp Y \mid Z \\ X \perp\!\!\!\perp Y \mid Z \text{ a } U = h(X) &\implies X \perp\!\!\!\perp Y \mid (U,Z) \\ X \perp\!\!\!\perp Y \mid Z \text{ a } X \perp\!\!\!\perp W \mid (Y,Z) &\implies X \perp\!\!\!\perp (W,Y) \mid Z \\ X \perp\!\!\!\perp Y \mid Z \text{ a } X \perp\!\!\!\perp Z &\implies X \perp\!\!\!\perp (Y,Z). \end{aligned}$$

Dôkaz. Budeme uvažovať spojitú rozdelenia. Ak platí, že $X \perp\!\!\!\perp Y \mid Z$, z definície podmienenej nezávislosti dostávame

$$\begin{aligned} f_{X,Y|Z}(x,y|z) &= f_{X|Z}(x|z)f_{Y|Z}(y|z) \\ &= f_{Y|Z}(y|z)f_{X|Z}(x|z) = f_{Y,X|Z}(y,x|z), \text{ pre všetky } x,y \text{ a } z. \end{aligned}$$

Teda $Y \perp\!\!\!\perp X \mid Z$.

Pre $X \perp\!\!\!\perp Y \mid Z$ a $U = h(x)$, z definície podmienenej nezávislosti dostávame

$$\begin{aligned} f_{X,Y|Z}(x,y|z) &= f_{X|Z}(x|z)f_{Y|Z}(y|z) \\ &= g_{h(X)|Z}(h(x)|z)f_{Y|Z}(y|z), \text{ kde } g(h(x)) = f(x), \text{ pre všetky } x \\ &= g_{u|z}(u|z)f_{Y|Z}(y|z) = f_{U,Y|Z}(u,y|z), \text{ pre všetky } x,u,y \text{ a } z. \end{aligned}$$

Teda $U \perp\!\!\!\perp Y \mid Z$.

Chceme ukázať, že $X \perp\!\!\!\perp Y \mid (U,Z)$. Z definície podmienenej hustoty pre spojitú náhodnú veličinu máme

$$\begin{aligned} f_{X,Y|U,Z}(x,y|u,z) &= f_{X|Y,U,Z}(x|y,u,z)f_{Y|U,Z}(y|u,z) \\ &= f_{X|U,Z}(x|u,z)f_{Y|U,Z}(y|u,z), \text{ pretože } X \perp\!\!\!\perp Y \mid Z, \text{ pre všetky } x,u,y \text{ a } z. \end{aligned}$$

Teda $X \perp\!\!\!\perp Y \mid (U,Z)$.

Chceme ukázať, že $X \perp\!\!\!\perp (W,Y) \mid Z$. Z definície podmienenej hustoty pre spojitú náhodnú veličinu máme

$$\begin{aligned} f_{X,W,Y|Z}(x,y,w|z) &= f_{X|W,Y,Z}(x|w,y,z)f_{W,Y|Z}(w,y|z) \\ &= f_{X|Y,Z}(x|y,z)f_{W,Y|Z}(w,y|z), \text{ pretože } X \perp\!\!\!\perp W \mid (Y,Z) \\ &= f_{X|Z}(x|z)f_{W,Y|Z}(w,y|z), \text{ pretože } X \perp\!\!\!\perp W \mid (Y,Z), \text{ pre všetky } x,w,y \text{ a } z. \end{aligned}$$

Teda $X \perp\!\!\!\perp (W,Y) \mid Z$.

Chceme ukázať, že $X \perp\!\!\!\perp (Y,Z)$. Z definície podmienenej hustoty pre spojitú náhodnú veličinu máme

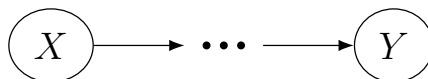
$$\begin{aligned} f_{X,Y,Z}(x,y,z) &= f_{X|Y,Z}(x|y,z)f_{Y,Z}(y,z) \\ &= f_{X|Z}(x|z)f_{Y,Z}(y,z), \text{ pretože } X \perp\!\!\!\perp Y \mid Z \\ &= f_X(x)f_{Y,Z}(y,z), \text{ pretože } X \perp\!\!\!\perp Z \text{ pre všetky } x,y \text{ a } z. \end{aligned}$$

Teda $X \perp\!\!\!\perp (Y,Z)$.

□

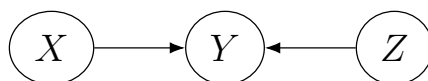
2.2 DAGs

Najskôr zavedieme základné značenie pre prácu s grafmi a postupne sa dostaneme až k definícii DAGu. Orientovaný graf sa skladá z množiny vrcholov V a množiny hrán E , ktorú tvoria usporiadané dvojice vrcholov. V našom prípade budú vrcholy zodpovedať náhodným veličinám a orientované hrany budú zodpovedať kauzálnym vzťahom medzi náhodnými veličinami. Ak napríklad $(X,Y) \in E$, znamená to, že existuje šípka z X do Y . Ak existuje šípka, ktorá spája X a Y v ľubovoľnom smere, nazveme X a Y susednými. Ak existuje šípka z X do Y hovoríme, že X je rodičom Y a Y je dieťaťom X . Množinu všetkých rodičov X označíme $\pi(X)$ (prípadne π_X). Orientovaná spojnica medzi dvomi premennými je súbor rovnako orientovaných šípok, ktorá spájajú jednu a druhú premennú (viz. obr. 2.1).

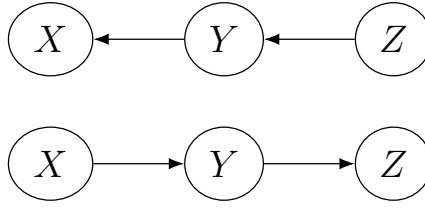


Obr. 2.1 Orientovaná spojnica medzi X a Y .

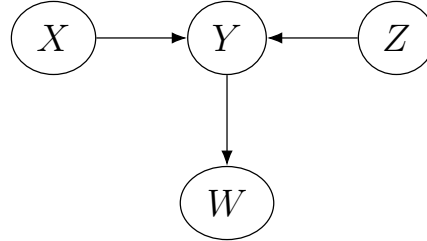
Postupnosť susedných vrcholov, ktoré začínajú v X a končia v Y (bez ohľadu na orientáciu) nazývame neorientovanou spojnicou. Postupnosť $\{X,Y,Z\}$ na obr. 2.3 tvoria neorientovanú spojnicu. X je predkom Y , ak existuje orientovaná spojnica medzi X a Y . Štruktúru, pri ktorej Y je predkom X aj predkom Z (viz. obr. 2.2) nazveme kolízorom v Y na spojnici $\{X,Y,Z\}$. Ak je štruktúra odlišná (viz. obr. 2.3), nazveme ju nekolízorom v Y na spojnici $\{X,Y,Z\}$. Vlastnosť kolízie závisí na spojnici. Ak budeme uvažovať štruktúru na obr. 2.4, dostávame kolízor v Y na spojnici $\{X,Y,Z\}$, avšak nekolízor v Y na spojnici $\{X,Y,W\}$. Ak premenné, ktoré smerujú do kolízoru nie sú susediace, hovoríme, že kolízor je netienený (unshielded). Orientovaná spojnica, ktorá začína končí v rovnakej premennej sa nazýva cyklus. Orientovaný graf nazveme acyklický, ak neobsahuje žiadny cyklus. Orientovaný acyklický graf budeme nazývať DAG.



Obr. 2.2 Kolízor v Y .



Obr. 2.3 Nekolízor v Y .



Obr. 2.4 Kolízor v Y na spojnici $\{X, Y, Z\}$.

2.3 Pravdepodobnosť a DAGy

Nech \mathcal{G} je DAG s vrcholmi $V = (X_1, X_2, \dots, X_k)$.

Definícia 7. Nech P je pravdepodobnostné rozdelenie pre V s hustotou f . Hovoríme, že P je Markovské ku \mathcal{G} , alebo \mathcal{G} reprezentuje P , ak

$$f(v) = \prod_{i=1}^k f(x_i | \pi_i),$$

kde π_i sú rodičia X_i . Množinu rozdelení reprezentovaných \mathcal{G} označíme $M(\mathcal{G})$.

Príklad 2.1. Význam definície môžeme ilustrovať na DAGu na obr. 2.4. Ak by napríklad X, Y, Z a W predstavovali genetický faktor, rakovinu pľúc, znečistenie ovzdušia a smrť, pre hustotu pravdepodobnosti by platilo

$$\begin{aligned} f(v) &= f(\text{genetický faktor}, \text{rakovina pľúc}, \text{znečistenie ovzdušia}, \text{smrť}) \\ &= f(\text{genetický faktor}) \cdot f(\text{znečistenie ovzdušia}) \\ &\quad \cdot f(\text{rakovina pľúc} \mid \text{genetický faktor}, \text{znečistenie ovzdušia}) \\ &\quad \cdot f(\text{smrť} \mid \text{rakovina pľúc}) \end{aligned}$$

Veta 5. Rozdelenie $P \in M(\mathcal{G})$ práve vtedy keď, platí Markovská podmienka

$$W \perp \tilde{W} \mid \pi_W \text{ pre všetky } W,$$

kde \tilde{W} označuje všetky ostatné veličiny, okrem rodičov a potomkov. Pripomeňme, že π_W označuje množinu všetkých rodičov W .

Markovská podmienka teda zhruba hovorí, že každá premenná W je nezávislá na minulosti, pri daných rodičoch.

Príklad 2.2. Napríklad pre štruktúru na obr. 2.4, Markovská podmienka implikuje vzťahy nezávislosti

$$X \perp Z \text{ a } W \perp \{X, Z\} \mid Y.$$

Príklad 2.3. Uvažujme štruktúru na obr. 2.6. Markovská podmienka implikuje vzťahy nezávislosti

$$U_X \perp\!\!\!\perp \{X, Y, Y^*, U_Y\}, \quad X^* \perp\!\!\!\perp \{Y, Y^*, U_Y\} | \{U_X, X\}, \quad Y \perp\!\!\!\perp \{U_X, X^*, U_Y\} | X, \\ X \perp\!\!\!\perp \{U_X, U_Y\}, \quad Y^* \perp\!\!\!\perp \{X, X^*, U_X\} | \{U_Y, Y\} \text{ a } U_Y \perp\!\!\!\perp \{U_X, X^*, X, Y\}.$$

Pri práci s kolízormi a potomkami kolízorov môžeme používať d-separačné (directed separation) pravidlá. Ak budeme uvažovať štruktúru na obr. 2.4 a na obr. 2.5, platia pravidlá d-separácie.

1. Ak Y nie je kolízor, potom X a Z sú d-prepojené, ale sú d-separované pri danom $Y = y$.
2. Ak Y je kolízor na spojnici $\{X, Y, Z\}$, potom X a Z sú d-separované, avšak sú d-prepojené pri danom $Y = y$.
3. Ak podmienime potomkom kolízoru, má to rovnaký účinok, ako keby sme podmienili kolízorom. Teda na obr. 2.4 sú X a Z sú d-separované, ale sú d-prepojené pri danom $W = w$.

Formálne môžeme d-separáciu a d-prepojenie definovať nasledovne.

Definícia 8. *Nech X a Y sú rôzne a nech W je množina vrcholov, ktorá neobsahuje X alebo Y . Potom X a Y sú d-separované vzhľadom na W , ak neexistuje žiadna neorientovaná spojnica U medzi X a Y taká, že*

1. *každý kolízor na U má potomka vo W ,*
2. *žiadny iný vrchol na U nie je vo W .*

Ak A , B a W sú rôzne množiny vrcholov a A a B nie sú prázdne, potom A a B sú d-separované vzhľadom na W , ak pre každé $X \in A$ a $Y \in B$, X a Y sú d-separované vzhľadom na W . Množiny vrcholov, ktoré nie sú d-separované, sú d-prepojené.

Príklad 2.4. Ak budeme uvažovať situáciu na obr. 2.6, z pravidiel d-separácie dostaneme

1. U_X a U_Y sú d-separované,
2. U_X a U_Y sú d-prepojené, pri danom $\{X^*, Y^*\}$,
3. U_X a U_Y sú d-separované, pri danom $\{X^*, Y^*, X\}$.

Veta 6. *Nech P je reprezentované DAGom \mathcal{G} . Nech A, B a C sú oddelené množiny vrcholov tohto \mathcal{G} . Potom $A \perp\!\!\!\perp B \mid C$ práve vtedy keď, A a B sú d-separované, pri danom C .*

Príklad 2.5. S d-separáciou a d-prepojením sme už intuitívne pracovali v minulej kapitole, keď sme skúmali situáciu na obr. 1.7. V príklade, ktorý sme k tomuto obrázku uvažovali by pre nás Veta 6 znamenala, že ak genetické faktory X a znečistenie ovzdušia Y sú d-separované, teda $X \perp\!\!\!\perp Y$. Ak však, keď podmienime rakovinou pľúc Z , sú X a Y d-prepojené a pri danom $Z = z$, už teda nemusia byť nezávislé. K podobnému záveru sme dospeli aj v minulej kapitole.

Aj pre grafy, ktoré sa zdajú byť rozdielne môže platiť, že implikujú rovnaké vzťahy nezávislosti.

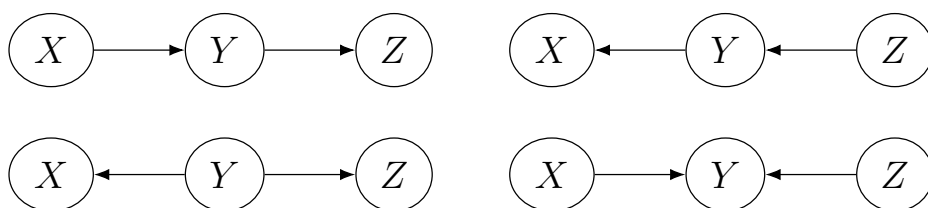
Definícia 9. Nech \mathcal{G} je DAG a $\mathcal{I}(\mathcal{G})$ nech sú všetky vzťahy nezávislosti implikované \mathcal{G} . Dva DAGy \mathcal{G}_1 a \mathcal{G}_2 sú Markovsky ekvivalentné, ak $\mathcal{I}(\mathcal{G}_1) = \mathcal{I}(\mathcal{G}_2)$.

Pre DAG \mathcal{G} , označíme kostrou (\mathcal{G}) neorientovaný graf, ktorý dostaneme, ak v grafe \mathcal{G} nahradíme všetky šípky neorientovanými hranami. Pripomeňme, že ak premenné, ktoré smerujú do kolízoru nie sú susediace, hovoríme, že kolízor je netienený (unshielded). Potom nám nasledujúca veta dáva nástroj na skúmanie ekvivalencie DAGov.

Veta 7. Dva DAGy \mathcal{G}_1 a \mathcal{G}_2 sú Markovsky ekvivalentné práve vtedy keď,

1. $kostra(\mathcal{G}_1) = kostra(\mathcal{G}_2)$,
2. \mathcal{G}_1 a \mathcal{G}_2 majú rovnaké netienené kolízory.

Príklad 2.6. Ak sa pozrieme na DAGy na obr. 2.5, vidíme, že majú rovnaké kostry. Okrem DAGu vpravo dole, ostatné DAGy nemajú kolízor. Z Vety 7 sú teda tieto tri DAGy Markovsky ekvivalentné.



Obr. 2.5 DAG vpravo dole má kolízor Y na spojnici $\{X, Y, Z\}$. Ostatné tri DAGy nemajú kolízor.

2.4 Vychýlenie meraní

Štruktúra na obr. 2.6 sa používa na vyjadrenie chyby merania ($X^* - X \neq 0$, $Y^* - Y \neq 0$). V tomto prípade sú veličiny, ktoré kauzálne ovplyvňujú pozorované veličiny (X^*, Y^*) vzájomne nezávislé ($U_X \perp U_Y$). Bodkovanou čiarou vyjadrujeme kauzálny efekt, ktorý v skutočnosti môžeme skúmať (X^* na Y^*), aj keď by sme chceli skúmať kauzálny efekt X na Y . Tieto kauzálne efekty sa však môžu líšiť. Ak k tomu dochádza hovoríme, že dochádza k vychýleniu meraní.

Príklad 2.7. Môžeme si predstaviť situáciu, kedy chceme skúmať efekt aspirínu (X) na zápal pečene (Y). K dispozícii by sme ale mali údaje o užívaní aspirínu a o zápalových hodnotách pečene z databázy (X^*, Y^*). Pri vyplňaní však mohlo dôjsť k chybám pri vkladaní údajov. Veličiny, ktoré kauzálne ovplyvňujú chybovosť pre X, Y označíme U_X, U_Y . Predpokladáme, že údaje boli vkladane rôznymi ľuďmi, v rôznych časoch, preto sú U_X a U_Y nezávislé.

Ak budeme chcieť vyjadriť situáciu pomocou modelu s potenciálnymi výsledkami a opäť sa obmedzíme na prácu s binárnymi premennými X a Y , dostávame

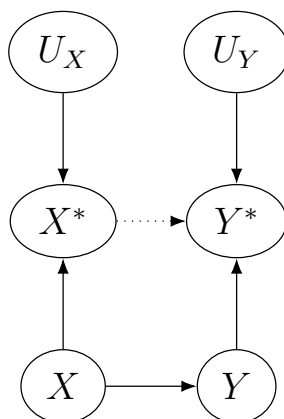
$$\theta = E[C_1] - E[C_0] = E[Y|X = 1] - E[Y|X = 0] = \alpha.$$

Platí totiž, že ošetrenie X je nezávislé na ostatných veličinách a teda môžeme použiť Vetu [1](#) ($P(X = 0), P(X = 1) > 0$). Sme však v situácii, kedy máme k dispozícii len X^* a Y^* .

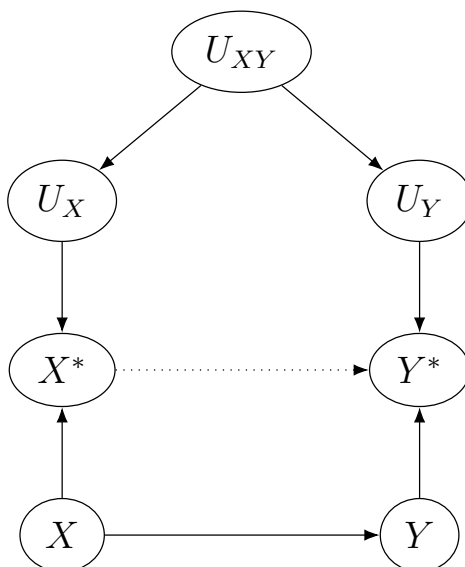
$$\theta = E[C_1] - E[C_0] \neq E[Y^*|X^* = 1] - E[Y^*|X^* = 0].$$

Teda vo všeobecnosti nemôžeme bez ďalších predpokladov skúmať kauzálny efekt X na Y , ak máme k dispozícii len hodnoty X^* a Y^* .

Na obr. [2.6](#) a [2.7](#) môžeme vidieť dva prípady situácií, ktoré môžu nastať, ak dochádza k nepresnému meraniu veličín X a Y . Podobne môže tiež dochádzať k nepresnému meraniu hodnôt kolízorov, mätúcich veličín či mediátorov. Tiež je možné zostaviť štruktúry DAGov, ktoré uvažujú meranie veličín v čase. Analýza takýchto DAGov je však nad rámec tejto diplomovej práce. Viac o týchto témach môžeme nájsť napríklad v [5](#).



Obr. 2.6 Chyba pri meraní X a Y . Faktory U_X a U_Y sú nezávislé.

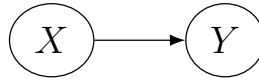


Obr. 2.7 Chyba pri meraní X a Y . Faktory U_X a U_Y sú závislé.

2.5 DAGy a SWIGy

V minulých podkapitolách sme videli, že DAGy poskytujú relatívne jednoduchý spôsob, ako reprezentovať kauzálne vzťahy medzi skúmanými veličinami. Avšak táto reprezentácia má svoje nedostatky. Napríklad nevieme vyjadriť veľkosť kauzálneho efektu X na Y . V DAGoch tiež zobrazujeme náhodné veličiny, nie kontrafaktuály. Podrobnejšie niektoré ďalšie úskalia DAGov rozoberá napríklad [16]. Spôsobov ako zovšeobecniť DAGy bolo navrhnutých niekoľko viz. napríklad prístup štruktúry dvojčiat (*twin network*, [17]) alebo kontrafaktuálne grafy (*contrafactual graphs*, [18]). Nie všetky tieto rozšírenia však splňujú požadované teoretické vlastnosti. Ako vhodné spojenie DAGov a modelu potenciálnych výsledkov sa ukazujú byť SWIGy (*Single-World Intervention Graphs*, *SWIGs*, viz. [15]), pretože v nich funguje veľa dôležitých teoretických výsledky odvodených pre DAGy a model potenciálnych výsledkov (napríklad Markovské vlastnosti a rozšírená g-formula zavedená v [19]). Vrcholy vo SWIGoch budú odpovedať kontrafaktuálom. Inými slovami budeme môcť v diagramoch pracovať s potenciálnymi výsledkami (C_0, C_1). S niečim podobným sme sa už stretli napr. pri analýze mediátora na obr. [1.5]. V tomto prípade sme pracovali s DAGom a danú situáciu sme sa tiež snažili priblížiť pomocou práce s potenciálnymi výsledkami.

Pre jednoduchosť skúmame kauzálny efekt ošetrenia X na výsledok Y , pričom uvažujeme X a Y nadobúdajú hodnoty $\{0,1\}$. DAG pre túto situáciu vidíme na obr. [2.8]. Pri vytváraní SWIGu dochádza ku rozdeľovaniu vrcholov. Namiesto práce s náhodnými veličinami teda pracujeme s kontrafaktuálmi (viz. obr. [2.9]). Situáciu môžeme tiež vyjadriť pomocou obr. [2.10].



Obr. 2.8 DAG \mathcal{G} pre skúmanie kauzálneho efektu X na Y .

$$X \mid x = 0 \longrightarrow Y^{x=0}$$

$$X \mid x = 1 \longrightarrow Y^{x=1}$$

Obr. 2.9 SWIGy $\mathcal{G}_{(x=0)}$ a $\mathcal{G}_{(x=1)}$ pre skúmanie kauzálneho efektu X na Y . Horný SWIG vyjadruje svet, v ktorom všetky subjekty nedostali ošetrovanie a dolný SWIG vyjadruje svet, v ktorom všetky subjekty dostali ošetrovanie.

$$X \mid x \longrightarrow Y^x$$

Obr. 2.10 SWIG $\mathcal{G}_{(x)}$ pre skúmanie kauzálneho efektu X na Y .

Analýza vlastností SWIGov a ich dôkladnejšie skúmanie je však nad rámec tejto diplomovej práce a uvádzame ich preto, že poskytujú rozšírenie DAGov a modelu potenciálnych výsledkov. Viac o tejto téme môžeme nájsť v [15].

3 Metóda inštrumentálnych premenných

Cieľom tejto kapitoly je poskytnúť náhľad na to, ako môžu byť metódy používané v klasickej ekonometrii použité pre kauzálne usudzovanie. Budeme sa zaoberať predovšetkým inštrumentálnymi premennými (*instrumental variables, IV*), ktoré sú bežnou súčasťou ekonometrických modelov. Ukážeme, že odhad pomocou metódy najmenších štvorcov (*ordinary least squares*) vo všeobecnosti nie je nestranný ani konzistentný. Zameriame sa na dvojstupňový odhad metódou najmenších štvorcov (*two-stage least squares, 2SLS*), ktorá používa inštrumentálne premenné. Budeme sa pri tom opierať hlavne o Ciprovu knihu [20] a Ciprove skriptá [21]. Uvedieme predpoklady, pri ktorých môžeme metódu 2SLS použiť na kauzálne usudzovanie. Budeme pritom vychádzať z knihy Hernána a Robinsa [5] a z článku [22]. Prínosom autora je predovšetkým simulačná štúdia uvedená na konci tejto kapitoly, v ktorej sú analyzované vybrané situácie, pre ktoré môžeme použiť inštrumentálne premenné.

3.1 Úvod a motivácia

V klasickej lineárnej regresii pracujeme so situáciou kedy máme n nezávislých kópií náhodných vektorov $(Y_i, \mathbf{X}_i^T)^T, i = 1, \dots, n$. Každé X_i má $p < n$ komponentov $\mathbf{X}_i = (X_{i1}, \dots, X_{ip})$. Pre odozvu \mathbf{Y} a maticu regresorov \mathbb{X} zavádzame značenie

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \text{ a } \mathbb{X} = \begin{pmatrix} \mathbf{X}_1^T \\ \vdots \\ \mathbf{X}_n^T \end{pmatrix}.$$

Ďalej predpokladáme, že $P(\text{rank}(\mathbb{X}) = p) = 1$.

Nezávislé (Y_i, \mathbf{X}_i) splňujú lineárny regresný model, ak odozvu Y_i vieme zapísať v tvare

$$Y_i = \mathbf{X}_i^T \boldsymbol{\beta}_0 + \varepsilon_i,$$

kde $\boldsymbol{\beta}_0 = (\beta_{01}, \dots, \beta_{0p})^T$ je vektor neznámych regresných koeficientov a $\varepsilon_1, \dots, \varepsilon_n$ sú nezávislé náhodné veličiny také, že $E[\varepsilon_i | \mathbf{X}_i] = 0$ a $\text{var}[\varepsilon_i] = \sigma^2$.

Z predpokladu $E[\varepsilon_i | \mathbf{X}_i] = \mathbf{0}$ dostávame, že $E[\varepsilon_i \mathbf{X}_i] = \mathbf{0}$. Teda ε_i a \mathbf{X}_i s nekorelované. Tento predpoklad zaručuje konzistenciu a nestrannosť klasického odhadu pomocou metódy najmenších štvorcov (*ordinary least squares, OLS odhad*). Ten má tvar

$$\hat{\boldsymbol{\beta}}_{OLS} = (\mathbb{X}^T \mathbb{X})^{-1} \mathbb{X}^T \mathbf{Y}.$$

Ak by tento predpoklad nebol splnený, teda $E[\varepsilon_i \mathbf{X}_i] \neq \mathbf{0}$, dostávame po dosadení za \mathbf{Y}

$$\hat{\boldsymbol{\beta}}_{OLS} = (\mathbb{X}^T \mathbb{X})^{-1} \mathbb{X}^T (\mathbb{X} \boldsymbol{\beta}_0 + \boldsymbol{\varepsilon}) = \boldsymbol{\beta}_0 + \left(\frac{1}{n} \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^T \right)^{-1} \frac{1}{n} \sum_{i=1}^n X_i \varepsilon_i.$$

Vo všeobecnosti $E[\hat{\beta}_{OLS}] \neq \beta_0$ a teda odhad nie je nestranný. Vo všeobecnosti neplatí ani konzistencia odhadu. Situáciu môžeme ilustrovať na nasledujúcom príklade.

Príklad 3.1. Uvažujme model regresnej priamky prechádzajúcej počiatkom

$$y_i = \beta x_i + \varepsilon_i, \quad x_i \sim iid(0, q), \quad \varepsilon_i \sim iid(0, \sigma^2), \quad \text{pre } i = 1, \dots, n,$$

kde značením $iid(\mu, \sigma^2)$ rozumieme nezávislé a rovnako rozdelené náhodné veličiny so strednou hodnotou μ a rozptylom σ^2 . Prirodzene uvažujeme $q > 0$ a $\sigma^2 > 0$. Nech $E[x_i \varepsilon_i] = \gamma \neq 0$ a γ je konečná. Potom zo slabého zákona veľkých čísel

$$\hat{\beta}_{OLS} = \beta + \left(\frac{1}{n} \sum_{i=1}^n x_i x_i^T \right)^{-1} \frac{1}{n} \sum_{i=1}^n x_i \varepsilon_i \xrightarrow{P} \beta + \frac{\gamma}{q}, \quad \text{pre } n \rightarrow \infty,$$

teda klasický OLS odhad nie je konzistentný.

Predpoklad nekorelovanosti ε_i a \mathbf{X}_i často nebýva splnený v modeloch s chýbami merania (*measurement error models*), kde tieto chyby bývajú korelované s reziduálnou zložkou. Predpoklad je tiež porušený vo viacúrovňových modeloch, kde pre odhadovanú rovnicu existuje iná rovnica vysvetľujúca určitý regresor pomocou vysvetľujúcej premennej odhadovanej rovnice. Porušenie tiež nastáva pri dynamických modeloch s korelovanými zložkami, kde sa na pozície regresorov môžu dostať oneskorené hodnoty vysvetľujúcej premennej. Nás bude zaujímať predovšetkým situácia, kedy dochádza k porušeniu kvôli nepozorovaniu dôležitej zastieracej premennej (viz. obr. [3.1](#)).

3.2 Odhady založené na inštrumentálnych premenných

Metóda inštrumentálnych premenných (IV metóda) sa používa pre získanie konzistentných odhadov parametrov v situáciách, kedy je porušený predpoklad nekorelovanosti ε_i a \mathbf{X}_i . Nás bude zaujímať predovšetkým situácia, kedy je dôležitá zastieracia premenná nepozorovaná. Aby sme mohli túto metódu použiť, musíme mať k dispozícii okrem pozorovaných hodnôt vysvetľovanej premennej y_i a vysvetľujúcich premenných \mathbf{x}_i tiež pozorované hodnoty inštrumentálnych premenných \mathbf{z}_i , pričom každá má $p < n$ komponentov z_{i1}, \dots, z_{ip} (tieto hodnoty usporiadame do matice \mathbf{Z} , ktorá má n riadkov a p stĺpcov). Na inštrumentálne premenné kladieme nasledujúce požiadavky

$$\frac{1}{n} \sum_{i=1}^n \mathbf{Z}_i \varepsilon_i \xrightarrow{P} \mathbf{0}, \quad \text{pre } n \rightarrow \infty, \quad (3.1)$$

$$\frac{1}{n} \sum_{i=1}^n \mathbf{Z}_i \mathbf{X}_i^T \xrightarrow{P} \Sigma_{ZX}, \quad \text{pre } n \rightarrow \infty, \quad (3.2)$$

kde Σ_{ZX} je regulárna matica. Volne môžeme tieto predpoklady formulovať tak, že inštrumentálne premenné majú byť nekorelované s náhodnými zložkami a významné korelované s pôvodnými vysvetľujúcimi premennými.

Príklad 3.2. Pre jednoduchosť uvažujme situáciu, kedy máme k dispozícii n nezávislých, rovnako rozdelených dvojíc (Y_i, X_i) , kde

$$Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i, \quad i = 1, \dots, n,$$

kde $E[\varepsilon_i] = 0$. Ak $\text{cov}(X_i, \varepsilon_i) = 0$, potom

$$\begin{aligned} \text{cov}(X_i, Y_i) &= \beta_2 \text{var}(X_i) + 0, \\ \beta_2 &= \frac{\text{cov}(X_i, Y_i)}{\text{var}(X_i)}, \\ E[Y_i] &= \beta_1 + \beta_2 E[X_i] + \varepsilon_i, \\ \beta_1 &= E[Y_i] - \beta_2 E[X_i]. \end{aligned}$$

Konzistentný OLS odhad teda dostávame ako

$$\begin{aligned} \hat{\beta}_{2,OLS} &= \frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2}, \\ \hat{\beta}_{1,OLS} &= \bar{Y} - \hat{\beta}_2 \bar{X}, \end{aligned}$$

kde $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$ a $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$.

Ak $\text{cov}(X_i, \varepsilon_i) \neq 0$, potom

$$\text{cov}(X_i, Y_i) = \beta_2 \text{var}(X_i) + \text{cov}(X_i, \varepsilon_i),$$

pričom $\text{cov}(X_i, \varepsilon_i)$ nevieme odhadnúť z pozorovaní. Ak budeme predpokladať, že máme k dispozícii inštrumentálnu premennú Z_i pre ktorú platí

$$\begin{aligned} \text{cov}(Z_i, \varepsilon_i) &= 0, \\ \text{cov}(Z_i, X_i) &\neq 0. \end{aligned}$$

Potom dostávame

$$\begin{aligned} \text{cov}(Z_i, Y_i) &= \beta_2 \text{cov}(Z_i, X_i) + 0, \\ \beta_2 &= \frac{\text{cov}(Z_i, Y_i)}{\text{cov}(Z_i, X_i)}. \end{aligned}$$

Potom pre nezávislé a rovnako rozdelené (Y_i, X_i, Z_i) , dostávame zo zákona veľkých čísel konzistentný odhad

$$\hat{\beta}_{2,IV} = \frac{\frac{1}{n} \sum_{i=1}^n (Z_i - \bar{Z})(Y_i - \bar{Y})}{\frac{1}{n} \sum_{i=1}^n (Z_i - \bar{Z})(X_i - \bar{X})},$$

kde $\bar{Z} = \frac{1}{n} \sum_{i=1}^n Z_i$. Rovnaký odhad dostávame, ak budeme postupovať v dvoch krokoch. Veľmi voľne povedané budeme v prvom kroku robiť OLS regresiu $X_i \sim Z_i$

a dostaneme predikované hodnoty \hat{X}_i . V druhom kroku budeme robiť OLS regresiu $Y_i \sim \hat{X}_i$. Ak postup rozpíšeme tak pre prvý krok dostávame

$$\begin{aligned} X_i &= \gamma_1 + \gamma_2 Z_i + u_i, \quad E[u_i] = 0, \\ \hat{X}_i &= \hat{\gamma}_1 + \hat{\gamma}_2 Z_i, \\ \overline{\hat{X}} &= \hat{\gamma}_1 + \hat{\gamma}_2 \overline{Z}, \\ \hat{X}_i - \overline{\hat{X}} &= \hat{\gamma}_2 (Z_i - \overline{Z}), \\ \hat{\gamma}_2 &= \frac{\frac{1}{n} \sum_{i=1}^n (Z_i - \overline{Z})(X_i - \overline{X})}{\frac{1}{n} \sum_{i=1}^n (Z_i - \overline{Z})^2}. \end{aligned}$$

Pre druhý krok dostávame

$$\begin{aligned} Y_i &= \beta_1 + \beta_2 \hat{X}_i + \varepsilon_i, \\ \hat{\beta}_2 &= \frac{\frac{1}{n} \sum_{i=1}^n (\hat{X}_i - \overline{\hat{X}})(Y_i - \overline{Y})}{\frac{1}{n} \sum_{i=1}^n (\hat{X}_i - \overline{\hat{X}})^2}, \\ &= \frac{\hat{\gamma}_2 \sum_{i=1}^n (Z_i - \overline{Z})(Y_i - \overline{Y})}{\hat{\gamma}_2^2 \sum_{i=1}^n (Z_i - \overline{Z})^2} = \frac{\frac{1}{n} \sum_{i=1}^n (Z_i - \overline{Z})(Y_i - \overline{Y})}{\frac{1}{n} \sum_{i=1}^n (Z_i - \overline{Z})(X_i - \overline{X})}. \end{aligned}$$

Dostávame teda, že $\hat{\beta}_2 = \hat{\beta}_{2,IV}$. Dvojstupňový odhad, ktorý sme použili sa tiež spolieha na existenciu inštrumentálnych premenných. Tomuto odhadu sa tiež hovorí dvojstupňový odhad metódou najmenších štvorcov (*two stage least squares*, 2SLS odhad). Viac ho budeme diskutovať v ďalšej podkapitole.

V Prípade 3.2 sme pre jednoduchú situáciu ilustrovali, že pri porušení predpokladu nekorelovanosti X_i a ε_i je klasický OLS odhad nekonzistentný. Ukázali sme, že je možné použiť IV odhad, ktorý pre túto jednoduchú situáciu odpovedá 2SLS odhadu.

Odhadom parametrov β_0 metódou inštrumentálnych premenných rozumieme odhad

$$\hat{\beta}_{IV} = (\mathbf{Z}^T \mathbf{X})^{-1} \mathbf{Z}^T \mathbf{Y}.$$

Veta 8. Odhad $\hat{\beta}_{IV}$ metódou inštrumentálnych premenných je konzistentný odhadom parametru β_0 .

Dôkaz. Postupnými úpravami dostávame

$$\begin{aligned} \hat{\beta}_{IV} &= (\mathbf{Z}^T \mathbf{X})^{-1} \mathbf{Z}^T (\mathbf{X} \beta_0 + \boldsymbol{\varepsilon}) = \beta_0 + (\mathbf{Z}^T \mathbf{X})^{-1} \mathbf{Z}^T \boldsymbol{\varepsilon} \\ &= \beta_0 + \left(\frac{1}{n} \sum_{i=1}^n Z_i X_i^T \right)^{-1} \frac{1}{n} \sum_{i=1}^n Z_i \varepsilon_i, \end{aligned}$$

takže z predpokladov [3.1](#) a [3.2](#) platí

$$\hat{\beta}_{IV} \xrightarrow{P} \beta_0 + \Sigma_{ZX}^{-1}0 = \beta_0, \text{ pre } n \rightarrow \infty.$$

□

Ďalej je možné za veľmi všeobecných predpokladov dokázať, že S_{IV} je konzistentným odhadom asymptotickej rozptylovej matice odhadu $\hat{\beta}_{IV}$

$$S_{IV} = s_{IV}^2 (\mathbf{Z}^T \mathbf{X})^{-1} \mathbf{Z}^T \mathbf{Z} (\mathbf{X}^T \mathbf{Z})^{-1},$$

kde s_{IV}^2 je konzistentný odhad reziduálnej zložky σ^2

$$s_{IV}^2 = \frac{(\mathbf{Y} - \mathbf{X}\hat{\beta}_{IV})^T (\mathbf{Y} - \mathbf{X}\hat{\beta}_{IV})}{n - p}.$$

Všimnime si, že odhad $\hat{\beta}_{IV}$ môžeme použiť asymptociky, avšak vo všeobecnosti nie je nestranným odhadom β_0

$$E[\hat{\beta}_{IV}] = E[(\mathbf{Z}^T \mathbf{X})^{-1} \mathbf{Z}^T (\mathbf{X}\beta_0 + \boldsymbol{\varepsilon})] = \beta_0 + E[(\mathbf{Z}^T \mathbf{X})^{-1} \mathbf{Z}^T \boldsymbol{\varepsilon}].$$

Pri praktickom použití je výhodou metódy inštrumentálnych premenných, že nemusíme pozorovať všetky dôležité zastieracie premenné. Na druhej strane problematický môže byť výber inštrumentálnych premenných splňujúcich predpoklady [3.1](#) a [3.2](#). Ak je závislosť medzi inštrumentálnymi a vysvetľujúcimi premennými malá (ako je tomu pri voľbe inštrumentálnych premenných rovných 1, absolútny člen), získame síce konzistentné, ale málo eficientné odhady s veľkými rozptylmi. Aj pri vysokej závislosti medzi inštrumentálnymi a vysvetľujúcimi premennými môže nastať problém, ak nebude splnený predpoklad [3.1](#) (ako je tomu pri extrémnom prípade, kedy $\mathbf{X} = \mathbf{Z}$), dostaneme odhady s malými rozptylmi, ktoré sú nekonzistentné. V praxi je náročné overiť, či je zvolená inštrumentálna premenná naozaj dostatočne nekorelovaná s náhodnou zložkou modelu.

3.3 Dvojstupňový odhad metódou najmenších štvorcov

Dvojstupňový odhad metódou najmenších štvorcov bol navrhnutý tak, aby zostal konzistentný pri porušení predpokladu nekorelovanosti \mathbf{X}_i a ε_i . K jeho konštrukcii však potrebujeme inštrumentálne premenné. Pracujeme v situácii, kedy máme n nezávislých kopíí náhodného vektorku $(y_i, \mathbf{X}_i^T, \mathbf{Z}_i^T)^T, i = 1, \dots, n$. Každé X_i má $p < n$ komponentov $\mathbf{X}_i = (X_{i1}, \dots, X_{ip})^T$ a každé Z_i má $q < n$ komponentov $\mathbf{Z}_i = (Z_{i1}, \dots, Z_{iq})^T$. Ďalej používame značenie

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \mathbf{X} = \begin{pmatrix} \mathbf{X}_1^T \\ \vdots \\ \mathbf{X}_n^T \end{pmatrix} \text{ a } \mathbf{Z} = \begin{pmatrix} \mathbf{Z}_1^T \\ \vdots \\ \mathbf{Z}_n^T \end{pmatrix}.$$

Naším cieľom bude odhadnúť hodnotu neznámeho koeficientov β , ak pre Y_i platí

$$y_i = \mathbf{X}_i^T \beta + \varepsilon_i,$$

kde $\beta = (\beta_1, \dots, \beta_p)^T$ je vektor neznámych koeficientov a $\varepsilon_1, \dots, \varepsilon_n$ sú nezávislé náhodné veličiny také, že $\text{var}[\varepsilon_i] = \sigma^2$. Tiež predpokladáme, že $E[\varepsilon_i X_i] \neq \mathbf{0}$. Nech ďalej platia predpoklady [3.1], [3.2] a $q \geq p$. Posledná podmienka sa nazýva rozmerová podmienka identifikácie (môže byť tiež nahradená hodnotnou podmienkou identifikácie). Ak chceme zaručiť existenciu rozptylovej matice 2SLS odhadu aj v prípade malého počtu pozorovaní, kedy nemôžeme používať asymptotické výsledky, malo by dokonca platiť $q \geq p + 2$. Inštrumentálne premenné hľadáme aj medzi regresormi \mathbf{X}_i . Ak platí, že $E[X_{ij}\varepsilon_i] = 0$, zaradíme tento $j - t$ regresor k Z_i . Špeciálne medzi inštrumentálne premenné zaradujeme vždy absolútny člen. Niekedy k regresorom pridávame rovnice, ktoré ich vysvetľujú, to však spadá pod sústavy simultánnych rovníc.

Budeme predpokladať, že $P(\text{rank}(\mathbf{Z} = q)) = 1$ a $P(\text{rank}(\mathbf{Z}^T \mathbf{X} = p)) = 1$. Konštrukcia 2SLS odhadu sa vykonáva v dvoch stupňoch. V prvom stupni získame vypočítané OLS odhady pomocou lineárnej regresie pôvodných vysvetľujúcich premenných \mathbf{X} na inštrumentálnej premennej \mathbf{Z}

$$\hat{\mathbf{X}} = \mathbf{Z}(\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{X}.$$

V druhom stupni sa finálny odhad $\hat{\beta}_{2SLS}$ vypočíta tým spôsobom, že sa pre pôvodný model nahradia vysvetľujúce premenné \mathbf{X} vypočítanými hodnotami $\hat{\mathbf{X}}$ a opäť sa skonštruuje OLS odhad.

$$\hat{\beta}_{2SLS} = \hat{\mathbf{X}}(\hat{\mathbf{X}}^T \hat{\mathbf{X}})^{-1} \hat{\mathbf{X}}^T \mathbf{Y} = (\mathbf{X}^T \mathbf{Z}(\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Z}(\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{y}.$$

Ak je počet premenných \mathbf{Z} a počet premenných \mathbf{X} rovnaký ($p = q$) dostávame z $\hat{\beta}_{2SLS}$ jeho redukovanú podobu $\hat{\beta}_{IV}$

$$\hat{\beta}_{2SLS} = (\mathbf{Z}^T \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{Z}(\mathbf{Z}^T \mathbf{Z})^{-1})^{-1} \mathbf{X}^T \mathbf{Z}(\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{y} = (\mathbf{Z}^T \mathbf{X})^{-1} \mathbf{Z}^T \mathbf{y} = \hat{\beta}_{IV},$$

pričom z Vety [8] vieme, že odhad $\hat{\beta}_{IV}$ je konzistentný. Pre $q \geq p$ je odhad $\hat{\beta}_{2SLS}$ tiež konzistentný (viz. [21]). Odhad asymptotickej rozptylovej matice dostávame ako

$$S_{2SLS} = s_{2SLS}^2 (\mathbf{X}^T \mathbf{Z}(\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{X})^{-1}, \quad (3.3)$$

kde s_{2SLS}^2 je odhad reziduálnej zložky σ^2

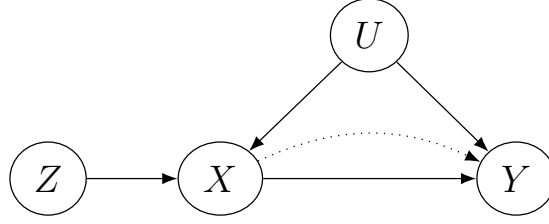
$$s_{2SLS}^2 = \frac{(\mathbf{y} - \mathbf{X}\hat{\beta}_{2SLS})^T (\mathbf{y} - \mathbf{X}\hat{\beta}_{2SLS})}{n - p}.$$

V praxi sa tiež často používa Hausmanov test špecifikácie modelu, ktorý testuje, či je naozaj nutné používať inštrumentálne premenné (alternatívna hypotéza H1) voči klasickému OLS postupu (nulová hypotéza odpovedá predpokladu $E[\varepsilon_i X_i] = 0$) viz. [20].

Nakoniec podkapitoly uvedieme niekoľko jednoduchých príkladov, na ktorých ilustrujeme použitie $\hat{\beta}_{2SLS}$ a ukážeme, že použitie klasického $\hat{\beta}_{OLS}$ nie je vhodné. Prezentované príklady budú odpovedať situáciám na obr. [3.1], obr. [3.2] a obr. [3.3].

Príklad 3.3. Uvažujme situáciu zodpovedajúcu obr. [3.1]. Nech pre jednoduchosť platí, že

$$\begin{aligned} Y_i &= X_i \beta + U_i \gamma + \varepsilon_i, \quad \varepsilon_i \sim iid(0,1), \\ U_i &\sim iid(0,1), \quad Z_i \sim iid(0,1), \\ X_i | (U_i, Z_i) &\sim Alt(\pi_i), \\ \text{kde } \pi_i &= f(U_i, Z_i), \quad \text{pre } i = 1, \dots, n. \end{aligned}$$



Obr. 3.1 Typický príklad situácie, kedy Z je inštrumentálnou premennou. Vplyv X na Y je zastrený premennou U .

Hodnoty zastieracej premennej U_i nie sú pozorované a β, γ sú neznáme nenulové parametre. Nech $\mathbf{X} = (X_1, \dots, X_n)^T$, $\mathbf{Z} = (Z_1, \dots, Z_n)^T$, $\mathbf{Y} = (Y_1, \dots, Y_n)^T$ a nech tvar funkcie f je napríklad $f(U_i, Z_i) = \frac{1}{1 + \exp(-U_i - Z_i)}$. Ďalej predpokladáme, že platia vzťahy $\frac{1}{n} \sum_{i=1}^n X_i U_i \xrightarrow{P} \theta \neq 0$, $\frac{1}{n} \sum_{i=1}^n X_i^2 \xrightarrow{P} \eta \neq 0$, $\frac{1}{n} \sum_{i=1}^n X_i \varepsilon_i \xrightarrow{P} 0$, $\frac{1}{n} \sum_{i=1}^n Z_i (U_i \gamma + \varepsilon_i) \xrightarrow{P} 0$ a $\frac{1}{n} \sum_{i=1}^n Z_i X_i \xrightarrow{P} \Sigma_{ZX} \neq 0$, pre $n \rightarrow \infty$. Časť týchto predpokladov je ilustračných pre ukázanie nekonzistencie OLS odhadu. Ďalšia časť prirodzene vyplýva zo situácie na obr. 3.1. Pre klasický OLS odhad platí

$$\hat{\beta}_{OLS} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} = \beta + \left(\frac{1}{n} \sum_{i=1}^n X_i^2 \right)^{-1} \frac{1}{n} \sum_{i=1}^n X_i (U_i \gamma + \varepsilon_i),$$

$$\hat{\beta}_{OLS} \xrightarrow{P} \beta + \gamma \frac{\theta}{\eta}, \text{ pre } n \rightarrow \infty.$$

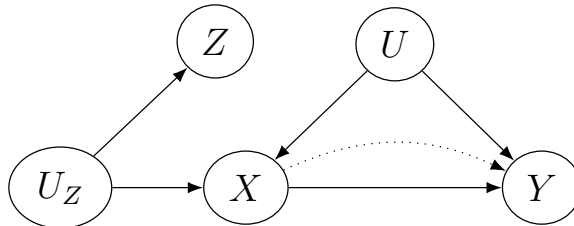
Vidíme, že $\theta \neq 0$, teda $\hat{\beta}_{OLS}$ nie je konzistentný (prirodzene predpokladáme, že $\gamma \neq 0$). Pre použitie $\hat{\beta}_{2SLS}$ sú splnené podmienky 3.1, 3.2 a počet vysvetľujúcich premenných je rovný počtu inštrumentálnych premenných. Teda s využitím týchto predpokladov dostávame

$$\hat{\beta}_{2SLS} = \hat{\beta}_{IV} = (\mathbf{Z}^T \mathbf{X})^{-1} \mathbf{Z}^T \mathbf{Y} = \beta + \left(\frac{1}{n} \sum_{i=1}^n Z_i X_i \right)^{-1} \frac{1}{n} \sum_{i=1}^n Z_i (U_i \gamma + \varepsilon_i),$$

$$\hat{\beta}_{2SLS} \xrightarrow{P} \beta + \gamma \frac{0}{\Sigma_{ZX}} = \beta, \text{ pre } n \rightarrow \infty.$$

Odhad $\hat{\beta}_{2SLS}$ je teda konzistentný.

Príklad 3.4. Uvažujme situáciu zodpovedajúcu obr. 3.2. Oproti neprechádzajúcemu prípadu pozorujeme náhradný inštrument Z (*surrogate instrument*). Nech



Obr. 3.2 Situácia, pre ktorú nie je kauzálny inštrument U_Z porovaný. Namiesto toho pozorujeme náhradný inštrument Z .

pre jednoduchost platí, že

$$\begin{aligned} Y_i &= X_i\beta + U_i\gamma + \varepsilon_i, \quad \varepsilon_i \sim iid(0,1), \\ U_i &\sim iid(0,1), \quad U_{Zi} \sim iid(0,1), \\ Z_i|U_{Zi} &\sim iid(U_{Zi},1), \quad X_i|(U_i, U_{Zi}) \sim Alt(\pi_i), \\ \text{kde } \pi_i &= f(U_i, U_{Zi}), \quad \text{pre } i = 1, \dots, n. \end{aligned}$$

Hodnoty zastieracích premenných U_i, U_{Zi} nie sú pozorované a β, γ sú neznáme nenulové parametre. Budeme používať značenie $\mathbf{X} = (X_1, \dots, X_n)^T$, $\mathbf{Z} = (Z_1, \dots, Z_n)^T$, $\mathbf{Y} = (Y_1, \dots, Y_n)^T$ a nech tvar funkcie f je napríklad $f(U_i, U_{Zi}) = \frac{1}{1+\exp(-U_i-U_{Zi})}$. Nech ďalej platí, že $\frac{1}{n} \sum_{i=1}^n X_i U_i \xrightarrow{P} \theta \neq 0$, $\frac{1}{n} \sum_{i=1}^n X_i^2 \xrightarrow{P} \eta \neq 0$, $\frac{1}{n} \sum_{i=1}^n X_i \varepsilon_i \xrightarrow{P} 0$, $\frac{1}{n} \sum_{i=1}^n Z_i (U_i \gamma + \varepsilon_i) \xrightarrow{P} 0$ a $\frac{1}{n} \sum_{i=1}^n Z_i X_i \xrightarrow{P} \Sigma_{ZX} \neq 0$, pre $n \rightarrow \infty$. Potom pre klasický OLS odhad podobne ako v Príklade 3.3 platí

$$\begin{aligned} \hat{\beta}_{OLS} &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} = \beta + \left(\frac{1}{n} \sum_{i=1}^n X_i^2 \right)^{-1} \frac{1}{n} \sum_{i=1}^n X_i (U_i \gamma + \varepsilon_i), \\ \hat{\beta}_{OLS} &\xrightarrow{P} \beta + \gamma \frac{\theta}{\eta}, \quad \text{pre } n \rightarrow \infty. \end{aligned}$$

Opäť vidíme, že $\theta \neq 0$, teda $\hat{\beta}_{OLS}$ nie je konzistentný (prirodzene predpokladáme, že $\gamma \neq 0$). Pre použitie $\hat{\beta}_{2SLS}$ sú splnené podmienky 3.1, 3.2. To čiastočne vidíme z Dagu na obr. 3.2. Vidíme, že nie je žiadna zastieracia premenná, ktorá pôsobí medzi Z a Y , navyše nie je žiaden priamy kauzálny vzťah medzi Z a Y . Preto predpokladáme, že $\frac{1}{n} \sum_{i=1}^n Z_i (U_i \gamma + \varepsilon_i) \xrightarrow{P} 0$, pre $n \rightarrow \infty$ (predpoklad 3.1). Tiež vidíme, že na X a Z pôsobí rovnaká nepozorovaná zastieracia premenná U_Z . To je dôvod, pre ktorý je prirodzené predpokladať, že $\frac{1}{n} \sum_{i=1}^n Z_i X_i \xrightarrow{P} \Sigma_{ZX} \neq 0$, pre $n \rightarrow \infty$ (predpoklad 3.2). S využitím týchto predpokladov podobne ako v Príklade 3.3 dostávame

$$\begin{aligned} \hat{\beta}_{2SLS} &= \hat{\beta}_{IV} = (\mathbf{Z}^T \mathbf{X})^{-1} \mathbf{Z}^T \mathbf{Y} = \beta + \left(\frac{1}{n} \sum_{i=1}^n Z_i X_i \right)^{-1} \frac{1}{n} \sum_{i=1}^n Z_i (U_i \gamma + \varepsilon_i), \\ \hat{\beta}_{2SLS} &\xrightarrow{P} \beta + \gamma \frac{0}{\Sigma_{ZX}} = \beta, \quad \text{pre } n \rightarrow \infty. \end{aligned}$$

Odhad $\hat{\beta}_{2SLS}$ je teda konzistentný.

Uvedieme ešte jeden príklad, v ktorom opäť môžeme použiť inštrumentálne premenné.

Príklad 3.5. Uvažujme situáciu zodpovedajúcu obr. 3.3. Opäť pozorujeme náhodný inštrument Z . Ten však môžeme vhodne použiť, len ak do 2SLS odhadí podmienime hodnotou $S = s$. Nech pre jednoduchost platí, že

$$\begin{aligned} Y_i &= X_i\beta + U_i\gamma + \varepsilon_i, \quad \varepsilon_i \sim iid(0,1), \\ U_i &\sim iid(0,1), \quad U_{Zi} \sim iid(0,1), \\ Z_i &\sim iid(0,1), \quad S_i|(U_i, U_{Zi}) \sim iid(U_{Zi} + Z_i, 1), \\ X_i|(U_i, U_{Zi}) &\sim Alt(\pi_i), \\ \text{kde } \pi_i &= f(U_i, U_{Zi}), \quad \text{pre } i = 1, \dots, n. \end{aligned}$$

Hodnoty zastieracích premenných U_i, U_{Zi} nie sú pozorované a β, γ sú neznáme nenulové parametre. Oproti prechádzajúcim prípadom však pozorujeme premennú S , ktorá pôsobí ako kolízor medzi premennými U_Z a Z . Nech $\mathbb{X} = (X_1, \dots, X_n)^T$, $\mathbb{Z} = (Z_1, \dots, Z_n)^T$, $\mathbb{Y} = (Y_1, \dots, Y_n)^T$ a nech tvar funkcie f je napríklad $f(U_i, U_{Zi}) = \frac{1}{1 + \exp(-U_i - U_{Zi})}$. Nech ďalej platí, že $\frac{1}{n} \sum_{i=1}^n X_i U_i \xrightarrow{P} \theta \neq 0$, $\frac{1}{n} \sum_{i=1}^n X_i^2 \xrightarrow{P} \eta \neq 0$, $\frac{1}{n} \sum_{i=1}^n X_i \varepsilon_i \xrightarrow{P} 0$, pre $n \rightarrow \infty$. Potom podobne ako v Príkladoch 3.3 a 3.4 platí, že klasický OLS odhad nie je konzistentný. Pre 2SLS odhad budeme postupovať podobne ako v prechádzajúcom príklade, avšak podmienime hodnotou $S = s$. Našu populáciu teda rozdelíme do skupín podľa premennej S . Pod označením $Y_i^s, X_i^s, Z_i^s, U_i^s, \varepsilon_i^s$ budeme rozumieť hodnoty daných veličín, ak $S = s$. Nech ďalej v zhode s obr. 3.3 platí $\frac{1}{n} \sum_{i=1}^n Z_i^s (U_i^s \gamma + \varepsilon_i^s) \xrightarrow{P} 0$ a $\frac{1}{n} \sum_{i=1}^n Z_i^s X_i^s \xrightarrow{P} \Sigma_{ZX^s} \neq 0$, pre $n \rightarrow \infty$. Pre 2SLS odhad v prvom stupni získame vypočítané OLS odhady pomocou lineárnej regresie pôvodných vysvetľujúcich premenných \mathbb{X} na inštrumentálne premenné \mathbb{Z} , pričom pracujeme len s vybranou populáciou v ktorej $S = s$.

$$\hat{\mathbb{X}} = \mathbb{Z}(\mathbb{Z}^T \mathbb{Z})^{-1} \mathbb{Z}^T \mathbb{X}.$$

V druhom stupni sa finálny odhad $\hat{\beta}_{2SLS}$ vypočíta tým spôsobom, že sa pre pôvodný model nahradia vysvetľujúce premenné \mathbb{X} vypočítanými hodnotami $\hat{\mathbb{X}}$ a opäť sa skonštruuje OLS odhad (vo vybranej populácii pre ktorú $S = s$). Po úprave dostávame

$$\hat{\beta}_{2SLS} = \hat{\beta}_{IV} = (\mathbb{Z}^T \mathbb{X})^{-1} \mathbb{Z}^T \mathbb{Y} = \beta + \left(\frac{1}{n} \sum_{i=1}^n Z_i^s X_i^s \right)^{-1} \frac{1}{n} \sum_{i=1}^n Z_i^s (U_i^s \gamma + \varepsilon_i^s).$$

Pretože pracujeme vo vybranej populácii pre ktorú $S = s$, $\frac{1}{n} \sum_{i=1}^n Z_i^s \varepsilon_i^s \xrightarrow{P} 0$ a $\frac{1}{n} \sum_{i=1}^n Z_i^s X_i^s \xrightarrow{P} \Sigma_{ZX^s} \neq 0$, pre $n \rightarrow \infty$. Dostávame teda

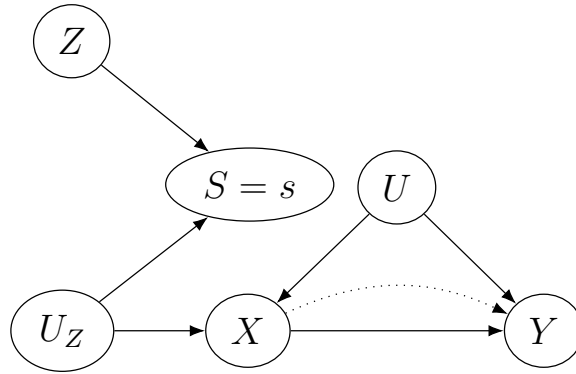
$$\hat{\beta}_{2SLS} \xrightarrow{P} \beta + \gamma \frac{0}{\Sigma_{ZX}} = \beta, \text{ pre } n \rightarrow \infty.$$

Všimnime si, že zásadný vplyv toho, že pracujeme vo vybranej populácii v ktorej $S = s$ je predovšetkým na predpoklad $\frac{1}{n} \sum_{i=1}^n Z_i^s X_i^s \xrightarrow{P} \Sigma_{ZX^s} \neq 0$, pre $n \rightarrow \infty$. Predpoklad $\frac{1}{n} \sum_{i=1}^n Z_i^s (U_i^s \gamma + \varepsilon_i^s) \xrightarrow{P} 0$, pre $n \rightarrow \infty$, by mal platiť v celej populácii.

Uvedené odhady asymptotických rozptylových matíc pre $\hat{\beta}_{2SLS}$ a $\hat{\beta}_{IV}$ sa spoliehajú na predpoklad homoskedasticity. Ak je tento predpoklad porušený, je možné ich nahradiť zovšeobecnenými odhadmi metódou najmenších štvorcov s využitím inštrumentálnych premenných (*generalized least squares instrumental variable estimators*, viz. [23]). Je dobré si uvedomiť, že metóda 2SLS funguje len v prípade, že sú vzťah medzi odozvou a regresormi lineárne. Ak sú tieto vzťahy nelineárne je vhodné použiť iné metódy (napríklad 2SRI viz. [24] a [25]).

3.4 Metódy inštrumentálnych premenných pre kauzálne usudzovanie

Metódy, ktoré používajú inštrumentálne premenné sa používajú aj pri skúmaní kauzálnych vzťahov. Ich výhodou je, že ak máme k dispozícii inštrumentálne pre-



Obr. 3.3 Situácia, pre ktorú je náhradný inštrumentálna premenná Z asociovaná s X vo vybranej populácii, v ktorej $S = s$.

mené, môžeme skúmať kauzálne vzťahy aj pre prípad nepozorovaného zastierania. Pri ich analýze budeme vychádzať predovšetkým z [5] a [22].

Situáciu na obr. 3.1 si môžeme predstaviť tak, že v randomizovanom experimente je Z znáhodnený indikátor pre priradenie ošetrovania (1 : áno, 0 : nie), X je indikátor dodržania ošetrovania pacientom. Y je odozva a U sú všetky (niektoré nenamerané) faktory, ktoré ovplyvňujú aj odozvu aj dodržanie ošetrovania pacientom. Ak by sme chceli konzistentne odhadnúť stredný kauzálny efekt A na Y , museli by sme väčšine metód používaných pre odhadnutie kauzálneho efektu (inverzné váženie pravdepodobnosti, štandardizácia, g-odhad, stratifikácia alebo párovanie) pozorovať všetky faktory U a správne ich zahrnúť do použitých modelov. Inak by naše odhady kauzálneho efektu mohli byť vychýlené.

Pre použitie IV metód na kauzálne usudzovanie musí byť splnených niekoľko predpokladov:

A1 : SUTVA. Tento predpoklad hovorí, že ošetrovanie ovplyvňuje len jednotlivca, ktorý ošetrovanie dostal, a že neexistujú rôzne verzie ošetrovania, ktoré by mali odlišné účinky.

A2 : Z a X nie sú nezávislé.

A3 : Z neovplyvňuje Y inak ako svojím vplyvom na X .

A4 : Z a Y nemajú spoločné príčinu.

Predpoklad A1 sme diskutovali už v prvej kapitole. V klasickej štatistike ho môžeme pripodobniť k predpokladu nezávislosti. Navyše vyžadujeme, aby existovala len jedna verzia ošetrovania. Ak napríklad dochádza k vzájomnému ovplyvňovaniu medzi jednotlivcami v experimente (t. j. ošetrovanie, ktoré jednotlivec dostal, ovplyvňuje výsledky iných jednotlivcov), tak by bol tento predpoklad porušený. My však v celej práci predpokladáme, že je tento predpoklad splnený. Predpoklad A2 je podobný predpokladu 3.2 a hovorí, že je nenulová asociácia medzi Z a X . Tento predpoklad je v randomizovaných experimentoch prirodzený, pretože očakávame, že priradenie ošetrovania bude mať vplyv nato, či pacient ošetrovanie skutočne prijme. Tomuto predpokladu tiež hovoríme podmienka relevantnosti (*relevance condition*). Predpokladu A3 sa hovorí obmedzenie vylúčenia (*exclusion restriction*) a znamená, že nie je žiadny priamy vplyv Z na Y . Pomocou potenciálnych výsledkov

ho môžeme tiež vyjadriť tak, že $Y_i^{z',x} = Y_i^{z,x} = Y_i^x$ pre všetky x a i . Ak budeme uvažovať randomizovanú štúdiu, ktorá je dvojito zaslepená (*double-blind*, ani pacient ani doktor nevedia, či dostávajú resp. podávajú ošetrovanie alebo placebo), tak by tento predpoklad mal byť z dizajnu experimentu splnený. Predpoklad A4 môžeme vyjadriť vo forme marginálnej zameniteľnosti (*marginal exchangeability*), ktorá znamená, že $Y^{x,z}$ a Z sú nezávislé, pre všetky x a z . Môžeme to chápať tak, že medzi Y a Z nie je žiadna premenná, ktorý by spôsobovala zastieranie.

Situácia na obr. 3.1 zobrazuje inštrument Z , ktorý má priamy vplyv na ošetrovanie X . V tejto situácii ho tiež nazývame kauzálnym inštrumentom (*causal instrument*). V tejto situácii je predpoklad A2 splnený, pretože Z ma kauzálny vplyv na X . Predpoklad A3 je tiež splnený, pretože Z nemá žiaden priamy kauzálny vplyv (okrem vplyvu na X). Naviac neexistuje premenná, ktorá by pôsobila medzi Z a Y , teda je splnený tiež predpoklad A4. Niekedy však kauzálny inštrument U_Z nie je pozorovaný, preto používame náhradný inštrument Z (*surrogate instrument*, situácia na obr. 3.2). Ten splňuje predpoklad A2, pretože medzi Z a X pôsobí U_Z ako zastieracia premenná. Predpoklad A3 je triviálne splnený, pretože Z nemá žiaden kauzálny vplyv a predpoklad A4 je modifikovaný do podoby, že Z a Y nemajú spoločnú príčinu, až na U_Z . Na obr. 3.3 je zobrazená situácia nevšedného náhradného inštrumentu. Asociácia medzi Z a X vzniká podmienením hodnotou spoločného dôsledku Z a nepozorovaného U_z . Týmto spoločným dôsledkom je S . Vo vybranej populácii, ktorú dostaneme podmienením $S = s$ sú teda Z a X asociované. Je tiež splnený predpoklad A2. Rovnako vidíme, že nie je žiadny priamy kauzálny vzťah medzi Z a Y , teda je splnený tiež predpoklad A3. Predpoklad A4 je upravený podobne ako pre situácia na obr. 3.2, Z a Y nemajú spoločnú príčinu, až na U_Z (pracujeme vo vybranej populácii, kedy $S = s$). Všimnime si však, že predpoklady A3 a A4 sú v praxi neoveriteľné. Ak sa napríklad nachádzame v situácii na obr. 3.1, tak ak by sme napríklad chceli overiť podmienku A3 tým, že podmienime hodnotou $X = x$, tak by ošetrovanie X pôsobilo ako kolízor medzi Z a U a keďže U má kauzálny vplyv na Y mohli by sme podmienene pozorovať vplyv Z na Y , aj keď by v skutočnosti žiaden nebol. Podmienku A4 nemôžeme overiť, pretože si nikdy nemôžeme byť istý, či nedochádza ku zastieraniu (okrem prípadu, kedy vhodne používame randomizáciu).

Pripomeňme, že Z a X sú binárne. Potom môžeme stredný kauzálny efekt $E[Y^{x=1} - Y^{x=0}]$ prepísať do tvaru

$$\frac{E[Y|Z = 1] - E[Y|Z = 0]}{E[X|Z = 1] - E[X|Z = 0]}.$$

Tento tvar tiež odpovedá klasickému Waldovmu odhadu pre binárny inštrument ([26]). Okrem predpokladov A1-A4 však musí byť splnený ďalší dodatočný predpoklad. Ten má rôzne podoby. Situácia vysvetlíme na nasledujúcom príklade. Uvažujme saturovaný, aditívny model strednej hodnoty pre binárne ošetrovanie X a binárny inštrument Z , pričom uvažujeme situáciu na obr. 3.1, 3.2 a 3.3.

$$E[Y^{x=1} - Y^{x=0}|X = 1, Z] = \beta_0 + \beta_1 Z.$$

Pretože platí, že $Y = XY^{x=1} + (1 - X)Y^{x=0}$, tak model môžeme prepísať do tvaru $E[Y - Y^{x=0}|X, Z] = X(\beta_0 + \beta_1 Z)$. Parameter β_0 zodpovedá strednému kauzálnemu efektu medzi ošetrovanými so $Z = 0$. Súčet parametrov $\beta_0 + \beta_1$ môžeme interpretovať ako kauzálny efekt medzi ošetrovanými so $Z = 1$. Parameter β_1 teda zodpovedá rozdielu v strednom kauzálnom efekte medzi ošetrovanými

pri zmene $Z = 1$ na $Z = 0$. Ak je Z inštrument, tak z podmienky A3 máme, že $E[Y^{x=0}|Z = 1] = E[Y^{x=0}|Z = 0]$ a z tvaru modelu dostávame, že

$$E[Y - X(\beta_0 + \beta_1 Z)|Z = 1] = E[Y - X(\beta_0)|Z = 0].$$

Ak budeme predpokladať, že $\beta_1 = 0$, tak platí, že

$$\beta_0 = \frac{E[Y|Z = 1] - E[Y|Z = 0]}{E[X|Z = 1] - E[X|Z = 0]}.$$

Dôvod prečo volíme hodnotu $\beta_1 = 0$ je, že máme rovnicu s dvoma neznámymi (β_0 a β_1) a táto rovnica vyčerpáva obmedzenia na rozdelenie dát vyplývajúce z inštrumentálnych podmienok. Potrebovali sme ďalšie obmedzenie, ktoré je však neoveriteľné. Dostávame, že

$$\beta_0 = E[Y^{x=1} - Y^{x=0}|X = 1, Z = z] = E[Y^{x=1} - Y^{x=0}|X = 1].$$

Parameter β_0 teda odpovedá kauzálnemu efektu v skupine ošetrovaných pre ľubovoľnú hodnotu $Z = z$ (za predpokladu, že $\beta_1 = 0$). Ak chceme ukázať, že parameter β_0 odpovedá kauzálnemu efektu v celej populácii ($E[Y^{x=1} - Y^{x=0}]$), a teda odpovedá klasickému Waldovmu odhadu (resp. IV odhadu) musíme pridať ďalší (neoveriteľný) predpoklad, že efekt ošetrovania je rovnaký v skupine ošetrovaných aj neošetrovaných (teda, že parameter pre Z je tiež nulový pre model v ktorom $A = 0$).

Vidíme teda, že okrem predpokladov A1-A4 potrebujeme dodatočný predpoklad, aby sme dokázali nájsť odhad pre stredný kauzálny efekt. Uvažujme teraz, že tento predpoklad má podobu všeobecnej podmienky homogenity (*general homogeneity condition*). Tá má tvar

$$A5 : \text{cov}(e(U), t(U)) = 0.$$

Kde $e(U) = E[Y^{x=1} - Y^{x=0}|U]$ a $t(U) = E[X|Z = 1, U] - E[X|Z = 0, U]$. Podmienku si môžeme predstaviť tak, že zmena účinku ošetrovania X na výsledok Y prostredníctvom U je nekorelovaná so zmenou asociácie $Z - X$ pomocou U na aditívnej úrovni. Všimnime si, že podmienka je splnená, ak je aspoň jeden z výrazov $e(U)$ a $f(U)$ konštantný. Ak by sme chceli za platnosti predpokladov A1 – A5 pre situáciu na obr. 3.1 ukázať, že

$$E[Y^{x=1} - Y^{x=0}] = \frac{E[Y|Z = 1] - E[Y|Z = 0]}{E[X|Z = 1] - E[X|Z = 0]},$$

postupovali by sme nasledovne. Vieme, že $E[e(U)] = E[Y^{x=1} - Y^{x=0}]$ a $E[t(U)] = E[X|Z = 1] - E[X|Z = 0]$, pretože U a Z sú nezávislé. Potom z podmienky A5 dostávame, že

$$\begin{aligned} \text{cov}(e(U), t(U)) &= E[e(U)t(U)] - E[e(U)]E[t(U)] = 0, \\ E[e(U)t(U)] &= E[e(U)]E[t(U)], \\ \frac{E[e(U)t(U)]}{E[X|Z = 1] - E[X|Z = 0]} &= E[Y^{x=1} - Y^{x=0}]. \end{aligned}$$

Ostáva ukázať, že $E[e(U)t(U)] = E[Y|Z = 1] - E[Y|Z = 0]$. Vieme, že $Y = XY^{x=1} + (1 - X)Y^{x=0} = X(Y^{x=1} - Y^{x=0}) + Y^{x=0}$ a že $X = P(X = 1)$. Tiež

vieme, že pre situáciu na obr. 3.1 platí, že Y^x a (X, Z) sú podmienene nezávislé pri danom U . Potom môžeme $E[Y|Z]$ rozpísať ako

$$\begin{aligned} E[Y|Z] &= \sum_u \sum_{x=\{0,1\}} E[Y|X=x, Z, U=u] P(X=x|Z, U=u) P(U=u|Z) \\ &= \sum_u \{E[Y^{x=1} - Y^{x=0}|U=u] P(X=1|Z, U=u) + E[Y^{x=0}|U=u]\} P(U=u). \end{aligned}$$

Teda pre $E[Y|Z=1] - E[Y|Z=0]$ dostávame

$$E[\{E[Y^{x=1} - Y^{x=0}|U]\} \{P(X=1|Z=1, U=u) - P(X=1|Z=0, U=u)\}].$$

Takže platí, že $E[Y|Z=1] - E[Y|Z=0] = E[e(U)t(U)]$, ako sme chceli ukázať.

Alternatívne môžeme podmienku A5 formulovať iným spôsobom. Tejto podmienke sa hovorí monotónnosť a označíme ju A5b. Má tvar $X^{z=1} \geq X^{z=0}$ pre všetkých pacientov. Kontrafaktuálna premenná $X^{z=1}$ označuje či by bol pacient ošetrovaný, ak by mu bolo priradené ošetrovanie (analogicky $X^{z=0}$). Pacientov potom môžeme rozdeliť do štyroch skupín. Pre vždy primajúcich (*always-takers*) platí, že $X^{z=1} = 1$ a $X^{z=0} = 1$. Pre vždy neprijímajúcich (*never-takers*) platí, že $X^{z=1} = 0$ a $X^{z=0} = 0$. Pre spolupracujúcich (*compliers, cooperative*) platí, že $X^{z=1} = 1$ a $X^{z=0} = 0$ a pre kontraspolpracujúcich (*defiers, contrarians*) platí, že $X^{z=1} = 0$ a $X^{z=0} = 1$. Pre podskupiny, v ktorých sa subjekty správajú ako spolupracujúci, však prakticky nevieme identifikovať, či sa jedná o skutočne spolupracujúcich, alebo sú to vždy prijímajúci, resp. vždy neprijímajúci (ak subjektu priradíme, aby bol ošetrovaný a on ošetrovanie príjme, nevieme, či je vždy prijímajúci alebo spolupracujúci). Ak v štúdií nie sú kontraspolpracujúce subjekty, platí alternatívna podmienka monotónnosti A5b. Ak podmienku homogenity A5 nahradíme podmienkou monotónnosti A5b, IV odhad už nebude identifikovať stredný kauzálny efekt ošetrovania na odozvu v celej populácii, ale len s populáciou spolupracujúcich subjektov, teda

$$E[Y^{x=1} - Y^{x=0}|X^{z=1} = 1, X^{z=0} = 0].$$

Uvažujme pre Z, X jednoduchú dichotómiu, situáciu na obr. 3.1 (teda Z je kauzálny inštrument a sú splnené predpoklady A1 – A4). Ďalej predpokladajme, že Z bolo efektívne náhodne pridelené (to implikuje, že Z a $\{Y^{x,z}, X^z; z=0,1; a=0,1\}$ sú nezávislé). Potom bude náznak dôkazu vyzeráť takto. Stredný efekt zámeru ošetrovania (*intention-to-treat*) môžeme rozpísať ako

$$\begin{aligned} E[Y^{z=1} - Y^{z=0}] &= E[Y^{z=1} - Y^{z=0}|X^{z=1} = 1, X^{z=0} = 1] P(X^{z=1} = 1, X^{z=0} = 1) \\ &+ E[Y^{z=1} - Y^{z=0}|X^{z=1} = 0, X^{z=0} = 0] P(X^{z=1} = 0, X^{z=0} = 0) \\ &+ E[Y^{z=1} - Y^{z=0}|X^{z=1} = 1, X^{z=0} = 0] P(X^{z=1} = 1, X^{z=0} = 0) \\ &+ E[Y^{z=1} - Y^{z=0}|X^{z=1} = 0, X^{z=0} = 1] P(X^{z=1} = 0, X^{z=0} = 1). \end{aligned}$$

V skupine vždy prijímajúcich a vždy neprijímajúcich budú stredné hodnoty nulové, pretože Z a X sú nezávislé a z predpokladu A3, Z nemá priamy vplyv na Y . Ak budeme ďalej predpokladať, že v skupine nemáme kontraspolpracujúcich, dostávame

$$E[Y^{z=1} - Y^{z=0}] = E[Y^{z=1} - Y^{z=0}|X^{z=1} = 1, X^{z=0} = 0] P(X^{z=1} = 1, X^{z=0} = 0).$$

V skupine spolupracujúcich však platí, že $X = Z$, teda

$$E[Y^{z=1} - Y^{z=0} | X^{z=1} = 1, X^{z=0} = 0] = E[Y^{x=1} - Y^{x=0} | X^{z=1} = 1, X^{z=0} = 0].$$

Teda pre efekt ošetrenia v skupine spolupracujúcich platí

$$E[Y^{x=1} - Y^{x=0} | X^{z=1} = 1, X^{z=0} = 0] = \frac{E[Y^{z=1} - Y^{z=0}]}{P(X^{z=1} = 1, X^{z=0} = 0)}.$$

Pretože sme predpokladali, že Z bolo pridelené náhodne dostávame, že čitateľ môžeme upraviť do tvaru $E[Y^{z=1} - Y^{z=0}] = E[Y|Z = 1] - E[Y|Z = 0]$. Pretože predpokladáme, že uvažujeme experiment, ktorý neobsahuje skupinu kontraspolpracujúcich, tak pre pravdepodobnosť, že bude subjekt v skupine vždy spolupracujúcich platí $P(X^{z=0} = 1) = P(X = 1|Z = 0)$ a pre pravdepodobnosť, že bude subjekt v skupine vždy nespolpracujúcich platí $P(X^{z=1} = 0) = P(X = 0|Z = 1) = 1 - P(X = 1|Z = 1)$. Potom pravdepodobnosť, že bude subjekt v skupine spolupracujúcich dostávame ako

$$\begin{aligned} P(X^{z=1} = 1, X^{z=0} = 0) &= 1 - (P(X = 1|Z = 0) + P(X = 1|Z = 0)) \\ &= 1 - (P(X = 1|Z = 0) + 1 - P(X = 1|Z = 1)) \\ &= P(X = 1|Z = 1) - P(X = 1|Z = 0). \end{aligned}$$

Spolu dostávame

$$E[Y^{x=1} - Y^{x=0} | X^{z=1} = 1, X^{z=0} = 0] = \frac{E[Y|Z = 1] - E[Y|Z = 0]}{P(X = 1|Z = 1) - P(X = 1|Z = 0)}.$$

Avšak vidíme, že sme museli predpokladať, že inštrument Z je priradený náhodne. Teda v observačných štúdiách môže byť veľmi náročné nájsť inštrumenty, ktoré túto podmienku budú spĺňať.

Vidíme, že pre získania kauzálnych záverov potrebujeme množstvo predpokladov, ktoré v mnohých prípadoch nemôžeme overiť. Najdetailnejšie sme skúmali situáciu, ktorá sa týkala obr. 3.1. Pre situácie na 3.2 a 3.3 sme diskutovali predpoklady potrebné na kauzálne usudzovanie, no konkrétne dôkazy, by sa pri tieto situácie líšili. Je tiež treba dodať, že sme sa obmedzili na situáciu jednoduchej dichotómie a pre mnohopočetné ošetrenie je potrebné naše predpoklady upraviť. Viac o predpokladoch pre všeobecnejšie prípady môžeme nájsť napríklad v [27]. Preto aj keď prístup pomocou inštrumentálnych premenných ponúka možnosť skúmanie kauzálnych efektu aj v prípade, kedy dochádza k zastieraniu, je potrebné si byť vedomí uvedených predpokladov.

3.5 Simulačná štúdia

V tejto podkapitole bude urobená simulačná štúdia pre situácie na obr. 3.1, obr. 3.2 a obr. 3.3. Vyhodou pri simulácií je, že poznáme skutočnú hodnotu kauzálného efektu, ktorý chceme skúmať. Tiež predpokladáme, že sú splnené predpoklady $A1 - A5$, ktoré pre kauzálne usudzovanie potrebujeme. Opodstatnenosť týchto predpokladov sme diskutovali v minulej podkapitole.

3.5.1 Kauzálny inštrument

V Príkľade 3.3 sme pre konkrétnu voľbu parametrov ukázali konzistenciu 2SLS odhadu pre situáciu na obr. 3.1. Z predchádzajúcej podkapitoly vieme, že pre túto situáciu sú tiež splnené predpoklady, ktoré potrebujeme pre kauzálne usudzovanie. Teraz pre túto situáciu vygenerujeme súbor pozorovaní a urobíme simulačnú analýzu. Naším cieľom bude skúmať, či pomocou 2SLS metódy vieme konzistentne odhadnúť hodnotu stredného kauzálneho efektu X na Y . Pre rôzne hodnoty počtu pozorovaní n budeme generovať $i = 1, \dots, n$ nezávislých hodnôt kauzálnej inštrumentálnej premennej Z_i z alternatívneho rozdelenia s pravdepodobnosťou úspechu $\psi = \frac{1}{2}$ a nepozorovanej zastieracej U_i z normovaného normálneho rozdelenia. Ošetrovanie X_i bude generované z alternatívneho rozdelenia s pravdepodobnosťou

$$\pi_i = \frac{1}{1 + \exp(-U_i - Z_i)}.$$

Odozva Y_i bude generovaná z normálneho rozdelenia s rozptylom 1 a so strednou hodnotou μ_i , kde

$$\mu_i = \alpha + X_i\beta + U_i\gamma.$$

Táto situácia môže odpovedať randomizovanému experimentu s jednou hodnotou ošetrovania, kde Z_i je indikátor, či bude pacientovi priradené ošetrovanie, X_i označuje, či bol pacient skutočne ošetrovaný, U_i predstavuje hodnotu nepozorovanej zastieracej premennej a Y_i hodnotu odozvy. Voľme hodnoty koeficientov $(\alpha, \beta, \gamma) = (-2, \frac{1}{2}, 1)$. Bude nás zaujímať predovšetkým hodnota parametru β , ktorý v tomto prípade odpovedá kauzálnemu efektu X na Y . Po vygenerovaní dát odhadneme hodnotu β pomocou OLS odhadu (regresiou Y na X). Hodnoty 2SLS odhadu dostávame v prvom stupni vypočítaním OLS odhadov \hat{X} pomocou lineárnej regresie X na Z . V druhom stupni vypočítame finálnu hodnotu 2SLS odhadu parametru β pomocou lineárnej regresie Y na odhadnuté hodnoty \hat{X} . Hodnoty priemernej smerodajnej odchýlky dostávame sprimerovaním odhadov vypočítaných na základe 3.3. Empirické hodnoty smerodajnej odchýlky predstavujú smerodajnú odchýlku získanú z 1000 vypočítaných odhadov (počet simulačných opakovaní je 1000). Výsledky simulačnej štúdie vidíme v tab. 3.1. Podľa očakávaní môžeme konštatovať, že klasický OLS odhad sa nezdá byť konzistentným odhadom skúmaného parametru $\beta = 0.5$. Tiež vidíme, že priemerné aj empirické hodnoty smerodajnej odchýlky sa príliš nelíšia. Hodnoty smerodajnej odchýlky pre OLS odhad sú menšie než hodnoty smerodajnej odchýlky pre 2SLS odhad. Hodnoty odhadov pre 2SLS odhad sa zdajú byť konzistentné.

Vidíme teda, že ak platia vzťahy znázornené na DAGu, ktorý je reprezentovaný na obr. 3.1 a je splnený predpoklad A5, tak sa zdá, že pomocou 2SLS metódy vieme konzistentne odhadnúť hodnotu kauzálneho efektu X na Y . V praxi si však takmer nikdy nemôže byť istý, že DAG s ktorým pracujeme je skutočne správny. Ak budeme alternatívne predpokladať, že je splnený predpoklad A5b, β identifikuje stredný kauzálny efekt X na Y v populácií spolupracujúcich.

3.5.2 Náhradný inštrument

Pre obr. 3.2 budeme postupovať podobne ako v Príkľade 3.4, kde sme teoreticky ukázali konzistenciu odhadu pomocou 2SLS metódy a vychýlenie klasického OLS

Počet pozorovaní	Priemer $\hat{\beta}_{OLS}$	Empirická SD_{OLS}	Priemerná SD_{OLS}	Priemer $\hat{\beta}_{2SLS}$	Empirická SD_{2SLS}	Priemerná SD_{2SLS}
1000	1.30	0.09	0.09	0.51	0.46	0.46
5000	1.30	0.04	0.04	0.48	0.21	0.20
10000	1.30	0.03	0.03	0.50	0.14	0.14
100000	1.30	0.01	0.01	0.50	0.05	0.05

Tabuľka 3.1 Výsledky simulačnej analýzy pre kauzálny inštrument. Skutočná hodnota odhadovaného parametru $\beta = 0.5$. Pre každý počet pozorovaní, je počet simulačných opakovaní 1000. Uvedené sú priemerné hodnoty odhadov, empirické hodnoty smerodajnej odchýlky, ktorú dostaneme jej výpočtom z odhadov a priemerné hodnoty smerodajnej odchýlky odhadov, ktoré sme dostali jej priemerovaním vždy z 1000 simulačných opakovaní. Je uvedené porovnanie odhadov klasického OLS a odhadov pomocou metódy 2SLS. Podľa očakávaní vidíme, že klasický OLS je nekonzistentný pri odhadovaní parametru β . Odhad pomocou metódy 2SLS sa zdá byť blízky skutočnej hodnote parametru β . Vidíme tiež, že hodnoty smerodajnej odchýlky sú pre metódy 2SLS oveľa väčšie a je potrebný veľký počet pozorovaní, aby začali byť zmysluplné (vidíme napríklad, že 1000 pozorovaní rozhodne nestačí). Toto je dobré brať do úvahy pri používaní metódy 2SLS. Vidíme tiež, že empirické a priemerné SD a veľmi nelíšia.

odhadu. V predchádzajúcej kapitole sme tiež pre túto situáciu diskutovali opodstatnenosť predpokladov, ktoré potrebujeme pre kauzálne usudzovanie. Teraz pre túto situáciu vygenerujeme súbor pozorovaní a urobíme simulačnú analýzu. Naším cieľom bude skúmať, či pomocou 2SLS metódy vieme konzistentne odhadnúť hodnotu stredného kauzálneho efektu X na Y . Pre skúmanie našich záverov, budeme pre rôzne hodnoty počtu pozorovaní n generovať $i = 1, \dots, n$ nezávislých hodnôt náhradnej inštrumentálnej premennej Z_i z alternatívneho rozdelenia s pravdepodobnosťou úspechu $\psi = \frac{1}{1 + \exp(-U_{Z_i})}$, nepozorovanej premennej U_{Z_i} z normovaného normálneho rozdelenia a nepozorovanej zastieracej U_i z normovaného normálneho rozdelenia. Ošetrenie X_i bude generované z alternatívneho rozdelenia s pravdepodobnosťou

$$\pi_i = \frac{1}{1 + \exp(-U_i - Z_i)}.$$

Odozva Y_i bude generovaná z normálneho rozdelenia s rozptylom 1 a so strednou hodnotou μ_i , kde

$$\mu_i = \alpha + X_i\beta + U_i\gamma.$$

Táto situácia sa od predchádzajúcej líši, pretože nepozorujeme priamy kauzálny inštrument, ale náhradný, ktorý je hodnotou ošetrenia asociovaný vďaka premennej U_Z , ktorá pôsobí ako zastieracia premenná. Podobne ako v predchádzajúcej situácii voľme hodnoty koeficientov $(\alpha, \beta, \gamma) = (-2, \frac{1}{2}, 1)$. Bude opäť nás zaujímať hodnota parametru β , ktorý v tomto prípade odpovedá vplyvu X na Y . Po vygenerovaní dát odhadneme hodnotu β pomocou OLS odhadu (regresiou Y na X). Hodnoty 2SLS odhadu dostávame v prvom stupni vypočítaním OLS odhadov \hat{X} pomocou lineárnej regresie X na Z . V druhom stupni vypočítame finálnu hodnotu 2SLS odhadu parametru β pomocou lineárnej regresie Y na odhadnuté hodnoty \hat{X} . Výsledky simulačnej štúdie vidíme v tab. 3.2. Hodnoty priemernej a empirickej smerodajnej odchýlky dostávame podobne ako v predchádzajúcom prípade kauzálneho inštrumentu. Počet simulačných opakovaní bol opäť 1000. Opäť konštatujeme,

že klasický OLS odhad sa nezdá byť konzistentným odhadom parametru $\beta = 0.5$. Oproti tomu 2SLS odhad sa zdá byť konzistentný. Pri porovnaní [3.1](#) a [3.2](#) vidíme, že rýchlosť konvergencie oboch odhadov sa výrazne nelíši (hodnoty smerodajných odchýlok sa zdajú byť podobné).

Počet pozorovaní	Priemer $\hat{\beta}_{OLS}$	Empirická SD_{OLS}	Priemerná SD_{OLS}	Priemer $\hat{\beta}_{2SLS}$	Empirická SD_{2SLS}	Priemerná SD_{2SLS}
1000	0.24	0.09	0.09	0.51	0.43	0.41
5000	0.24	0.04	0.04	0.51	0.13	0.13
10000	0.24	0.03	0.03	0.49	0.09	0.09
100000	0.24	0.01	0.01	0.50	0.03	0.03

Tabuľka 3.2 Výsledky simulačnej analýzy pre náhradný inštrument. Skutočná hodnota odhadovaného parametru $\beta = 0.5$. Pre každý počet pozorovaní, je počet simulačných opakovaní 1000. Uvedené sú priemerné hodnoty odhadov, empirické hodnoty smerodajnej odchýlky a priemerné hodnoty smerodajnej odchýlky. Hodnoty empirickej smerodajnej odchýlky dostávame výpočtom smerodajnej odchýlky z príslušných 1000 odhadov (uvažujeme 1000 simulačných opakovaní). Priemerné hodnoty smerodajnej odchýlky dostávame spriemerovaním 1000 hodnôt smerodajnej odchýlky príslušných odhadov. Je uvedené porovnanie odhadov klasického OLS a odhadov pomocou metódy 2SLS. Podľa očakávaní vidíme, že klasický OLS je nekonzistentný pri odhadovaní parametru β . Odhad pomocou metódy 2SLS sa zdá byť blízky skutočnej hodnote parametru β . Vidíme tiež, že hodnoty smerodajnej odchýlky sú pre metódy 2SLS oveľa väčšie a je potrebný veľký počet pozorovaní, aby začali byť zmysluplné (vidíme napríklad, že 5000 pozorovaní nestačí). Toto je dobré brať do úvahy pri používaní metódy 2SLS. Vidíme tiež, že empirické a priemerné SD a veľmi nelíšia. Oproti situácií v tab. [3.1](#), sú smerodajné odchýlky porovnateľné.

Dostávame teda, že ak platia vzťahy znázornené v DAGu na obr. [3.2](#) a je splnený predpoklad A5, tak sa zdá, že pomocou 2SLS metódy vieme konzistentne odhadnúť hodnotu kauzálneho efektu X na Y . Ak však chceme tento postup použiť v praktickej situácii musíme byť opatrní, pretože málokedy si môžeme byť istí, že DAG s ktorým pracujeme je skutočne správny. Môže dôjsť k napríklad v porušení predpokladov A2, A3. Tie sú neoveriteľné, takže porušenie nezaznamenáme. Tiež vidíme, že pre prípad kedy uvažujeme náhradný inštrument namiesto kauzálneho inštrumentu, došlo k miernej zmene uvažovaných predpokladov, ale praktický postup pri odhadovaní kauzálneho efektu pomocou 2SLS metódy sa nezmenil.

3.5.3 Náhradný inštrument vo vybranej populácii

Na obr. [3.3](#) vidíme prípad neobvyklého náhradného inštrumentu Z , ktorý je s premennou X asociovaný len ak podmienime hodnotou $S = s$. Situáciu si môžeme predstaviť tak, že Z môžeme použiť ako inštrument, len ak skúmame vybranú populáciu v ktorej $S = s$. V Príklade [3.3](#) sme teoreticky ukázali konzistenciu odhadu pomocou 2SLS metódy a nekonzistenciu klasického OLS odhadu. V predchádzajúcej kapitole sme tiež pre túto situáciu diskutovali predpoklady potrebné pre kauzálne usudzovanie. Teraz pre túto situáciu vygenerujeme súbor pozorovaní a urobíme simulačnú analýzu. Naším cieľom bude skúmať, či pomocou 2SLS metódy vieme konzistentne odhadnúť hodnotu stredného kauzálneho efektu X na Y . Pre rôzne hodnoty počtu pozorovaní n budeme generovať $i = 1, \dots, n$

nezávislých hodnôt náhradnej inštrumentálnej premennej Z_i z alternatívneho rozdelenia s pravdepodobnosťou úspechu $\psi = \frac{1}{2}$, nepozorovanej premennej U_{Z_i} z normovaného normálneho rozdelenia, pozorovanej premennej S z alternatívneho rozdelenia s pravdepodobnosťou úspechu $\eta = \frac{1}{1 + \exp(-U_{Z_i} - Z_i)}$ a nepozorovanej zástieracej U_i z normovaného normálneho rozdelenia. Ošetrenie X_i bude generované z alternatívneho rozdelenia s pravdepodobnosťou

$$\pi_i = \frac{1}{1 + \exp(-U_i - Z_i)}.$$

Odozva Y_i bude generovaná z normálneho rozdelenia s rozptylom 1 a so strednou hodnotou μ_i , kde

$$\mu_i = \alpha + X_i\beta + U_i\gamma.$$

V tomto prípade budeme postupovať mierne odlišne od prechádzajúcej situácie, pretože výsledky budeme vyhodnocovať podmienene na $S = s$. Voľme hodnoty koeficientov $(\alpha, \beta, \gamma) = (-2, \frac{1}{2}, 1)$. Bude nás zaujímať predovšetkým hodnota parametru β . S je binárna, preto budeme mať dva 2SLS odhady parametru β . Očakávame, že OLS odhad bude nekonzistentný vo vybranej populácii pre voľbu $S = s$ aj v celej populácii. 2SLS odhad by mal byť v celej populácii nekonzistentný, avšak ak podmienime hodnotou $S = s$, mali by sme dostať konzistentný odhad. Odhady vo vybranej populácii, kedy $S = s$ dostávame analogicky ako v prechádzajúcej situácii. Výsledky simulačnej štúdie v celej populácii vidíme v tab. [3.3](#). Výsledky vo vybranej populácii pre hodnotu $S = 0,1$ vidíme v [3.4](#). OLS odhady vo vybranej populácii sa ukázali byť nekonzistentné (neuvádzame ich). Hodnoty priemernej a empirickej smerodajnej odchýlky dostávame podobne ako v predchádzajúcom prípade náhradného inštrumentu. Počet simulačných opakovaní bol opäť 1000. Na základne výsledkov konštatujeme, že klasický OLS odhad sa nezdá byť konzistentným odhadom parametru $\beta = 0.5$. 2SLS odhad sa zdá byť v celej populácii tiež nekonzistentný, avšak vo vybranej populácii pre ktorú $S = s$ sa zdá byť konzistentný. Pri porovnaní výsledkov s [3.1](#) a [3.2](#) vidíme, že rýchlosť konvergencie oboch odhadov sa zdá byť v tomto prípade pomalšia (hodnoty smerodajných odchýlok sa pre kauzálny inštrument zdajú byť pre rovnaký počet pozorovaní výrazne menšie).

Opäť sa zdá, že pre situáciu na obr. [3.3](#) môžeme použiť metódu 2SLS na odhadnutie kauzálneho efektu X na Y . Na druhej strane podobne ako v predchádzajúcich dvoch príkladoch musíme konštatovať, že v praktickej situácii sú predpoklady $A2$, $A3$ neovetielné. Oproti predchádzajúcemu prípadu vyhodnocujeme kauzálny efekt vo vybranej populácii ($S = s$), pretože v celej populácii nemôžeme použiť Z ako inštrumentálnu premennú (nie je splnený predpoklad $A2$). Preto sa praktický postup vyhodnocovania pomocou 2SLS odhadu oproti prechádzajúcej situácii mierne líši.

3.5.4 Zhodnotenie simulačnej štúdie

V simulačnej časti sme skúmali tri rôzne situácie, pri ktorých je možné použiť prístup pomocou metódy inštrumentálnych premenných pre kauzálny usudzovanie. Pre kauzálny, náhradný a podmienene náhradný inštrument sme použili metódu 2SLS, pre odhad kauzálneho efektu X na Y . Zvolený počet simulačných opakovaní

Počet pozorovaní	Priemer $\hat{\beta}_{OLS}$	Empirická SD_{OLS}	Priemerná SD_{OLS}	Priemer $\hat{\beta}_{2SLS}$	Empirická SD_{2SLS}	Priemerná SD_{2SLS}
1000	1.25	0.09	0.09	1.36	136.15	6.97
10000	1.24	0.03	0.03	1.15	100.35	6.15
100000	1.24	0.01	0.01	1.35	114.82	6.81
500000	1.24	0.00	0.00	1.36	49.00	5.89
1000000	1.24	0.00	0.00	1.25	166.34	6.13

Tabuľka 3.3 Výsledky simulačnej analýzy pre podmienený náhradný inštrument v celej populácii (nepodmieňujeme hodnotou $S = s$). Skutočná hodnota odhadovaného parametru $\beta = 0.5$. Pre každý počet pozorovaní, je počet simulačných opakovaní 1000. Uvedené sú priemerné hodnoty odhadov, empirické hodnoty smerodajnej odchýlky, ktorú dostaneme výpočtom smerodajnej odchýlky z odhadov a priemerné hodnoty smerodajnej odchýlky odhadov, ktoré sme dostali spriemerovaním hodnôt smerodajných odchýlok odhadov. Je uvedené porovnanie klasických OLS odhadov a odhadov pomocou metódy 2SLS. Podľa očakávaní vidíme, že klasický OLS je nekonzistentný pri odhadovaní parametru β . Odhad pomocou metódy 2SLS je v celej populácii taktiež nekonzistentný. Empirická SD je pre 2SLS výrazne odlišná od priemernej SD. Len z toho by sme mohli usudzovať, že 2SLS odhad sa nezdá byť v poriadku.

pre všetky situácie je 1000. Bol volený tiež menší počet simulačných opakovaní (100), pričom výsledky sa príliš nelíšili. 2SLS odhad sa zdá byť vo všetkých troch prípadoch konzistentný. Oproti tomu klasický OLS odhad sa ukazuje byť nekonzistentný pre odhad parametru β . To je dané porušením jedného zo základných predpokladov klasického lineárneho modelu (porušenie nekorelovanosti náhodného člena a vysvetľujúcich premenných). Príkladom takéhoto porušenie je situácia, kedy medzi odozvou a vysvetľujúcimi premennými existuje premenná, ktorá spôsobuje zastieranie. Táto situácia nastáva vo všetkých skúmaných prípadoch. Vidíme, že počet pozorovaní potrebný na dosiahnutie konzistencie 2SLS odhadu sa naprieč prezentovanými situáciami výrazne líši. Pre 2SLS odhad využívajúci podmienené náhradný inštrument je aj pre počet pozorovaní 1000000 stále relatívne veľký odhad smerodajnej odchýlky odhadu. To môže byť problémom zvlášť v praktických aplikáciách. Tiež môže byť problém, že v praxi veľa krát nemáme možnosť overiť, či situácia, ktorú skúmame zodpovedá nami navrhnutému DAGu. Lahko sa teda môžeme ocitnúť v situácií, kedy sú predpoklady, ktoré vyžadujeme (zvlášť pre skúmanie kauzálnych vzťahov) porušené a nemáme možnosť to zistiť. Preto by sme mali byť pri vyhodnocovaní opatrní. Možným návrhom ako sa vyhnúť nesprávnym záverom je uvažovať viacero modelov a skúmať ako sa výsledky líšia naprieč týmito modelmi. Zložitejšie štruktúry DAGov a ich skúmanie pomocou metódy využívajúcich inštrumentálne premenné môžeme nájsť napríklad v [28]. Prístup pomocou metódy 2SLS tiež spoľahlivo funguje len ak je vzťah medzi odozvou a vysvetľujúcimi premennými lineárny. Pre zložitejšie situácie je preto vhodnejšie použiť iné metódy.

Počet pozorovaní	Priemer $\hat{\beta}_{2SLS s=0}$	Empirická $SD_{2SLS s=0}$	Priemerná $SD_{2SLS s=0}$	Priemer $\hat{\beta}_{2SLS s=1}$	Empirická $SD_{2SLS s=1}$	Priemerná $SD_{2SLS s=1}$
1000	0.85	313.06	3.43	0.91	61.59	3.37
10000	0.51	1.17	0.97	0.49	2.44	0.93
100000	0.50	0.30	0.30	0.49	0.29	0.29
500000	0.50	0.14	0.13	0.50	0.13	0.13
1000000	0.50	0.10	0.10	0.49	0.09	0.09

Tabuľka 3.4 Výsledky simulačnej analýzy pre podmienený náhradný inštrument v oboch vybraných populáciách (podmieňujeme hodnotou $S = 0,1$). Skutočná hodnota odhadovaného parametru $\beta = 0.5$. Pre každý počet pozorovaní, je počet simulačných opakovaní 1000. Uvedené sú priemerné hodnoty odhadov, empirické hodnoty smerodajnej odchýlky, ktorú dostaneme výpočtom smerodajnej odchýlky z odhadov a priemerné hodnoty smerodajnej odchýlky odhadov, ktoré sme dostali spriemerovaním hodnôt smerodajných odchýlok odhadov. Je uvedené porovnanie odhadov pomocou metódy 2SLS pre situácie kedy $s = 0$ a $s = 1$. Odhad pomocou metódy 2SLS sa zdá byť blízky skutočnej hodnote parametru β , avšak až pre veľmi veľký počet pozorovaní (pri porovnaní s tab. 3.1 a tab. 3.2). Vidíme tiež, že hodnoty smerodajnej odchýlky sú pre metódy 2SLS oveľa väčšie a je potrebný veľký počet pozorovaní, aby začali byť zmysluplné (vidíme napríklad, že aj pre 1000000 pozorovaní sú hodnoty smerodajnej odchýlky stále relatívne veľké). Toto je dobré brať do úvahy pri používaní metódy 2SLS, zvlášť pre podmienený náhradný inštrument. Vidíme tiež, že empirické a priemerné SD a sú pre malý počet pozorovaní (1000) výrazne odlišné. Avšak s rastúcim počtom pozorovaní sa od seba takmer nelíšia.

Záver

Na záver môžeme konštatovať, že metódy kauzálneho usudzovania uvedené v tejto práci, vrátane modelu potenciálnych výsledkov, DAGov a inštrumentálnych premenných, zohrávajú kľúčovú úlohu pri pochopení kauzálnych vzťahov v rôznych oblastiach, ako je vedecký výskum a zdravotníctvo. Hoci tieto metódy poskytujú cenné nástroje na odhadovanie kauzálnych efektov, je dôležité si uvedomiť obmedzenia a výzvy s nimi spojené.

V prvej kapitole sme ukázali, že model potenciálnych výsledkov ponúka formálny rámec na definovanie kauzálneho efektu a jeho odhadovanie s využitím údajov z pozorovaní, ale opiera sa o predpoklady, ktoré môžu byť v praxi neoveriteľné. Podobne sme v druhej kapitole predstavili DAGy, ktoré poskytujú vizuálne znázornenie kauzálnych vzťahov, čo pomáha pri identifikácii zastieracích premenných, avšak nemôžu preukázať kauzalitu. Praxi len málokedy môžeme mať istotu, že sú používané DAGy správne. V tretej kapitole sme sa venovali vzťahu klasickej ekonometrie a kauzality. Sústredili sme sa na inštrumentálne premenné, ktoré môžu byť užitočné pri riešení zastierania a odhadovaní kauzálneho efektu. Ich použitie si však vyžaduje starostlivé zváženie ich predpokladov. Tie sú však opäť v mnohých prípadoch neoveriteľné. Na záver kapitoly bola urobená simulačná štúdia pre vybrané prípady DAGov, v ktorých je možné použiť prístup pomocou inštrumentálnych premenných. Výhodou simulačnej štúdie je, že poznáme skutočnú hodnotu kauzálneho efektu. Ukázali sme, že ak je 2SLS metóda vhodne použitá, dokážeme pomocou nej úspešne odhadovať hodnotu kauzálneho efektu. Oproti tomu v uvedených situáciách klasický OLS odhad zlyháva.

V celej tejto práci sme zdôrazňovali dôležitosť kritického hodnotenia predpokladov a používania vhodných štatistických metód pri odhadovaní kauzálneho efektu. Ideálna situácia nastáva, ak kauzálne vzťahy skúmame pre prípad randomizovaného experimentu, pretože ošetrenie je priradené náhodne. V mnohých prípadoch však realizácia randomizovaného experimentu nie je realistická. V týchto prípadoch je treba byť pri skúmaní kauzálnych vzťahov obzvlášť opatrný, pretože požadované predpoklady pre jeho odhadnutie sú často neoveriteľné. Je potom otázne, akú váhu prikladať záverom o kauzalite, pri tomto type štúdií. Je zrejmé, že odhadovanie kauzálnych vzťahov je zložitá a náročná úloha. V tejto práci bola prezentovaná len malá časť metód, ktoré sa pre skúmanie kauzálnych vzťahov používajú. Viac o tejto problematike môžeme nájsť napríklad v knihe Hernána a Robinsa [5].

Literatúra

1. FISHER, Ronald Aylmer. *The Design Of Experiments*. Oxford University Press, 1935.
2. NEYMAN, Jerzy. On the application of probability theory to agricultural experiments. *Annals of Agricultural Sciences*. 1923, s. 1–51.
3. RUBIN, Donald B. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*. 1974, roč. 66, č. 5, s. 688.
4. WASSERMAN, Larry. *All of statistics: A concise course in statistical inference*. Springer Science & Business Media, 2013.
5. HERNÁN, Miguel A; ROBINS, James M. *Causal inference: What if*. Boca Raton: Chapman Hall/CRC, 2020.
6. RUBIN, Donald B. Randomization analysis of experimental data: The Fisher randomization test comment. *Journal of the American statistical association*. 1980, roč. 75, č. 371, s. 591–593.
7. MORGAN, Stephen L; WINSHIP, Christopher. *Counterfactuals and causal inference*. Cambridge University Press, 2015.
8. BURCH, PRJ. Smoking and lung cancer. Tests of a causal hypothesis. *Journal of Chronic Diseases*. 1980, roč. 33, č. 4, s. 221–238.
9. RUBIN, Donald B. Inference and missing data. *Biometrika*. 1976, roč. 63, č. 3, s. 581–592.
10. DING, Peng; LI, Fan. Causal inference. *Statistical Science*. 2018, roč. 33, č. 2, s. 214–237.
11. SIMPSON, Edward H. The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1951, roč. 13, č. 2, s. 238–241.
12. PEARSON KARL, LA; LESLIE, BM. Genetic (reproductive) selection: inheritance of fertility in man, and of fecundity in thoroughbred racehorses. *Philosophical Transactions of Royal Society of London, Series A*. 1899, roč. 192, s. 257–330.
13. YULE, G Udny. Notes on the theory of association of attributes in statistics. *Biometrika*. 1903, roč. 2, č. 2, s. 121–134.
14. GENELETTI, Sara. Identifying direct and indirect effects in a non-counterfactual framework. *Journal of the Royal Statistical Society Series B: Statistical Methodology*. 2007, roč. 69, č. 2, s. 199–215.
15. RICHARDSON, Thomas S; ROBINS, James M. Single world intervention graphs (SWIGs): A unification of the counterfactual and graphical approaches to causality. *Center for the Statistics and the Social Sciences, University of Washington Series. Working Paper*. 2013, roč. 128, č. 30, s. 2013.
16. SUZUKI, Etsuji; SHINOZAKI, Tomohiro; YAMAMOTO, Eiji. Causal diagrams: Pitfalls and tips. *Journal of Epidemiology*. 2020, roč. 30, č. 4, s. 153–162.

17. BALKE, A; PEARL, J. Bounds on probabilistically evaluated counterfactual queries. *Technical Report R-213-B, UCLA Cognitive Systems Lab*. 1994.
18. SHPITSER, I; PEARL, J. What counterfactuals can be tested. *In Proceedings of the 23rd Annual Conference on Uncertainty in Artificial Intelligence, Vancouver, BC Canada*. 2007.
19. ROBINS, James M; HERNÁN, Miguel A; SIEBERT, Uwe. Effects of multiple interventions. *Comparative Quantification of Health Risks: Global and Regional Burden of Disease Attributable to Selected Major Risk Factors*. 2004, roč. 1, s. 2191–2230.
20. CIPRA, Tomáš. *Finanční ekonometrie*. Zv. 30. Ekopress Praha, Czech Republic, 2008.
21. CIPRA, Tomáš. *Ekonometrie*. SPN, 1984.
22. BAIOCCHI, Michael; CHENG, Jing; SMALL, Dylan S. Instrumental variable methods for causal inference. *Statistics in Medicine*. 2014, roč. 33, č. 13, s. 2297–2340.
23. MCFADDEN, Daniel. *Econometrics 240B*. Lecture notes, University of California Berkeley, 1999.
24. MULLAHY, John. Instrumental-variable estimation of count data models: Applications to models of cigarette smoking behavior. *Review of Economics and Statistics*. 1997, roč. 79, č. 4, s. 586–593.
25. TERZA, Joseph V; BASU, Anirban; RATHOUZ, Paul J. Two-stage residual inclusion estimation: addressing endogeneity in health econometric modeling. *Journal of Health Economics*. 2008, roč. 27, č. 3, s. 531–543.
26. WALD, Abraham. The fitting of straight lines if both variables are subject to error. *The Annals of Mathematical Statistics*. 1940, roč. 11, č. 3, s. 284–300.
27. ANGRIST, Joshua D; PISCHKE, Jörn-Steffen. *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press, 2009.
28. CHALAK, Karim; WHITE, Halbert. An Extended Class of Instrumental Variables for the Estimation of Causal Effects. UCSD Department of Economics Discussion Paper. 2009.

Zoznam obrázkov

1.1 Prvý príklad tvaru regresnej funkcie kauzality a asociácie.	16
1.2 Druhý príklad tvaru regresnej funkcie kauzality a asociácie.	16
1.3 Zastieranie.	20
1.4 Kontrolovanie na zastieranie	21
1.5 Mediácia.	23
1.6 Mediácia pri podmienení.	23
1.7 Kolízor.	23
1.8 Kolízia pri podmienení.	24
1.9 Randomizovaná štúdia.	24
2.1 Orientovaná spojnica medzi X a Y	28
2.2 Kolízor v Y	28
2.3 Nekolízor v Y	29
2.4 Kolízor v Y na spojnici $\{X, Y, Z\}$	29
2.5 Dagy pre $\{X, Y, Z\}$	31
2.6 Chyba pri meraní X a Y . Faktory U_X a U_Y sú nezávislé.	32
2.7 Chyba pri meraní X a Y . Faktory U_X a U_Y sú závislé.	32
2.8 DAG \mathcal{G} pre skúmanie kauzálneho efektu X na Y	33
2.9 SWIGy $\mathcal{G}_{(x=0)}$ a $\mathcal{G}_{(x=1)}$	33
2.10 SWIG $\mathcal{G}_{(x)}$	33
3.1 Inštrumentálna premenná Z	40
3.2 Náhradná inštrumentálna premenná Z	40
3.3 Podmienene náhradná inštrumentálna premenná Z	43

Zoznam tabuliek

3.1	Výsledky simulačnej analýzy pre kauzálny inštrument.	49
3.2	Výsledky simulačnej analýzy pre náhradný inštrument.	50
3.3	Výsledky simulačnej analýzy pre podmienený náhradný inštrument.	52
3.4	Výsledky simulačnej analýzy pre podmienený náhradný inštrument.	53