

Tento projekt bol spolufinancovaný Európskou úniou. Investícia do Vašej budúcnosti.

Kontrafaktuálne metódy hodnotenia dopadov

ePublikácia

Prehľad a aplikácia konceptov hodnotenia
dopadov intervencií

2015



APTP	Aktívna politika trhu práce
ŠF	Štrukturálne fondy
KF	Kohézny fond
ATT	Priemerný účinok podpory na podporených
ATE	Priemerný účinok podpory
TOT	Účinok podpory na podporených
PSM	Propensity score matching
DiD	Difference in differences
2SLS	Dvojstupňová metóda najmenších štvorcov
RDD	Regression discontinuity design
TBIE	Hodnotenie dopadov založené na teórii
IV	Inštrumentálne premenné
LATE	Lokálny priemerný účinok podpory
LLM	Lokálne lineárne párovanie

Porovnávacía skupina	Pojem sa používa na označenie nepodporených jednotiek ktoré slúžia ako porovnávacía skupina pre jednotky vystavené intervencii.
Kontrolná skupina	Kontrolná skupina je ekvivalentom porovnávacej skupiny. V literatúre často dochádza k zamieňaniu týchto pojmov. Tento pojem sa správne používa v kontexte náhodných experimentov (<i>random experiments</i>), zatiaľ čo porovnávacía skupina sa používa v kontexte kvázi- experimentálnych metód.
Podporená skupina	Jednotky populácie vystavené pôsobeniu intervencie (podpory).
Intervencia	Opatrenie resp. politika, ktorá sa pôsobením na cieľovú populáciu snaží dosiahnuť určitú zmenu.
F- štatistika	Test významnosti rozdielu medzi dvoma rozptylmi.
Výsledková premenná	Veličina, prostredníctvom ktorej sa sleduje dosiahnutie výsledku intervencie.
Cieľová populácia	Skupina jednotiek (obyvateľstva, podnikov a pod.) na ktorú je intervencia namierená.
Program	Programom v kontexte hodnotení dopadov označujeme opatrenie, alebo súbor opatrení, ktorých cieľom je vyvolať určité správanie alebo zmenu.
Náhodný experiment (<i>Random experiment</i>)	Postup označujúci náhodný výber jednotiek, ktoré budú podporené. Všetky jednotky v cieľovej populácii majú rovnakú pravdepodobnosť byť podporené, výber je vykonaný náhodným spôsobom.
Pozorovateľné charakteristiky	Vlastnosti jednotiek, ktorých hodnoty poznáme, sú zachytené v dátach, alebo sa dajú inak kvantifikovať. Príkladom je napríklad: u jednotlivcov - pohlavie, vek, výška, vzdelanie, mzda; u spoločností – tržby, náklady, počet zamestnancov.
Nepozorovateľné charakteristiky	Vlastnosti jednotiek, ktoré nie sú známe, resp. sú ťažko kvantifikovateľné. Príkladom je napríklad: u jednotlivcov – motivácie, talent, skúsenosti, vrodené schopnosti; u spoločností – goodwill, know-how, kontakty.

Táto publikácia sa člení na tri hlavné kapitoly:

Východiská a koncepty hodnotenia dopadov – kapitola poskytuje základný rámec, teoretické východiská, podmienky a predpoklady aplikácie kontrafaktuálnych metód hodnotenia dopadov.

Kontrafaktuálne metódy hodnotenia dopadov – v tejto kapitole sú podrobne rozpracované jednotlivé metódy, ich využitie a postup aplikácie.

Otázky požiadaviek na dáta pre realizáciu hodnotení dopadov – v tejto kapitole popisujeme všeobecné požiadavky na údaje potrebné pre realizáciu kvantitatívnych hodnotení dopadov.

	Strana
Východiská a koncepty hodnotenia dopadov	5
■ Intervencie v kontexte verejných politík	7
■ Základné koncepty	9
■ Podmienky hodnotenia kauzálnych účinkov verejných politík	15
■ Kvalitatívne a kvantitatívne hodnotenia dopadov	16
■ Hodnotenie dopadov založené na teórii	17
Kontrafaktuálne metódy hodnotenia dopadov	19
■ Prehľad použitých metód	20
■ Difference-in-Differences	21
■ Propensity score matching	32
■ Regression discontinuity design	50
■ Instrumental variables	57
Otázky požiadaviek na dáta pre realizáciu hodnotení dopadov	65
■ Požiadavky na dáta pre realizáciu hodnotení dopadov	66
■ Scenáre postupov hodnotení s obmedzeniami času, rozpočtu a dát	72

Medzi hlavné zdroje literatúry použitej pri tvorbe tejto publikácie patria uvedené tituly:

	<p>A note on the impact evaluation of public policies: the counterfactual analysis European Commission Joint Research Centre; Institute for the Protection and Security of the Citizen JRC Scientific and policy reports (JRC74778), 2012 Massimo Loi a Margarida Rodrigues ISBN 978-92-79-26425-2</p>		<p>Impact Evaluations and Development NONIE Guidance on Impact Evaluation World Bank, Washington, D.C. Leeuw, F. Vaessen, J. ISBN-10: 1-60244-120-0 ISBN-13: 978-1-60244-120-0</p>
	<p>Evalued Sourcebook: Method and Techniques European Commission Regional and Urban policies</p>		<p>Design and Commissioning of Counterfactual Impact Evaluations: A practical Guidance for ESF Managing Authorities European Commission, 2013 Directorate-General for Employment, Social Affairs and Inclusion, Unit A3 ISBN: 978-92-79-28238-6 DOI: 10.2767/94454</p>
	<p>Handbook on Impact Evaluation; Qualitative Methods and Practices The World Bank, Washington, D.C.; 2010 Khandker, S. Koolwal, G. Samad, H. ISBN: 978-0-8213-8028-4 eISBN: 978-0-8213-8029-1 DOI: 10.1596/978-0-8213-8028-4</p>		<p>Quantitative and Qualitative Methods in Impact Evaluation and Measuring Results (Issue Paper) Governance and Social Development Resource Centre Garbarino, S. Holland, J. Social Development Direct, 2009 ISBN: 0 7044 27567 / 9780704427563</p>

Východiská a koncepty hodnotenia dopadov

Východiská

Intervencie v kontexte verejných politík

Základné koncepty

Podmienky hodnotenia kauzálnych účinkov verejných politík

Kvalitatívne vs kvantitatívne hodnotenia

Hodnotenie dopadov založené na teórii

Táto publikácia je súčasťou projektu „Hodnotenie vybraných intervencií ŠF a KF využitím metód Counterfactual Impact Evaluation“. V rámci uvedeného projektu sa vykonali hodnotenia dvoch intervencií spolufinancovaných zo zdrojov EÚ.

Prvá intervencia bola realizovaná v rámci Operačného programu Konkurencieschopnosť a hospodársky rast – podopatrenie 1.1.1. Podpora zavádzania inovácií a technologických transferov. Druhá intervencia bola realizovaná v rámci Operačného programu Zamestnanosť a sociálna inklúzia – opatrenie §46 Vzdelávanie a príprava pre trh práce. Vypracovanie tejto publikácie je súčasťou uvedeného hodnotenia.

Účelom publikácie je predstavenie metód hodnotenia dopadov a zvýšenie povedomia o využívaní kontrafaktuálnych metód hodnotenia verejných politík na Slovensku.

Intervencie realizované v rámci verejných politík sú implementované za účelom vyvolania určitej zmeny, resp. zlepšenia situácie jednotlivcov, skupín obyvateľstva, firiem, alebo iných subjektov ktoré sú nimi zasiahnuté. Aby bolo možné zistiť do akej miery dosiahli tieto intervencie svoj cieľ, je potrebné vykonať adekvátne hodnotenie týchto politík.

Hodnotenia verejných politík môžu byť kvalitatívneho alebo kvantitatívneho charakteru. Ako príklad môžeme uviesť programy verejne poskytovaného doplnkového vzdelávania pre nezamestnaných, ktorých účelom je zvýšiť vzdelanostnú úroveň a zručnosti týchto skupín občanov s cieľom zvýšiť tak pravdepodobnosť ich zamestnania sa. Hodnotenie vplyvov týchto programov na úrovni jednotlivcov a ich schopnosti nájsť si vďaka nim zamestnanie sa môže na prvý pohľad javiť jednoducho avšak v skutočnosti je to komplexnejšia úloha. Mali by byť výsledky týchto uchádzačov o zamestnanie porovnané so situáciou v ktorej sa nachádzali pred implementáciou programu? Alebo by mali byť porovnané so skupinou nezamestnaných, ktorí sa programu nezúčastnili? Čo ak ani jedna z týchto alternatív nie je správna? Čo je najvhodnejším prístupom hodnotenia účinkov verejných politík na výsledky účastníkov v nich zapojených? Táto publikácia si dáva za cieľ predstaviť rámec hodnotenia verejných politík, objasniť prečo sú dva vyššie uvedené prístupy zväčša nesprávne a popísať metódy kontrafaktuálneho hodnotenia dopadov ktoré sa v súčasnosti používajú na meranie účinkov politík.

Potreba kvantifikácie účinkov politík je v dnešnej dobe nepopierateľná, nakoľko umožňuje stanoviť ich skutočný vplyv a porovnanie s očakávaniami. Hodnotenie politík zároveň poskytuje kľúčový a nenahraditeľný zdroj informácií a dôkazov pre budúce politiky v danej oblasti. Ako také, je dôležitým nástrojom ex-ante hodnotenia politík, pod podmienkou že minulé intervencie boli podobného kontextu. Dôsledkom tohto je rastúci význam literatúry zaoberajúcej sa hodnotením politík a vznik nových metód identifikácie kauzálnych účinkov politík.

Cieľom hodnotenia politík je určiť kauzálny účinok určitej politiky na oblasť jej záujmu, na ktorú sa očakáva že bude mať vplyv. Kauzálny účinok politiky môžeme definovať ako rozdiel medzi výsledkom jednotiek ovplyvnených politikou a výsledkom, ktoré by tie isté jednotky dosiahli bez existencie politiky. Základným problémom hodnotení je nemožnosť simultánneho sledovania tých istých jednotiek v dvoch rozdielnych situáciách, t.j. za účinnosti politiky a v situácii kedy politika nie je implementovaná – kontrafaktuálna situácia – ťažko simulovateľná a zachytiteľná.

Na identifikáciu a porovnanie rozdielov vo výsledkoch je potrebná adekvátna porovnávací skupina (*comparison group*), ktorá je čo možno najviac podobná skupine zasiahnutej intervenciou. Účinok politiky je následne určený na základe rozdielov medzi kontrolnou skupinou a podporenou skupinou (*treatment group*). Identifikácia a výber kontrolnej skupiny je komplexnou úlohou. Často nás môže zvädzať použitie jedného z dvoch chybných prístupov:

- i) **porovnanie výsledkov podporenej skupiny pred a po intervencii, alebo**
- ii) **porovnanie jednotiek zasiahnutých intervenciou s tými, ktoré ňou zasiahnuté neboli.**

Ani jeden z týchto prístupov neodhalí kauzálne účinky politiky. V prvom prípade môže byť výsledok záujmu ovplyvnený aj inými faktormi, mimo pôsobenia politiky, ktoré sa mohli v čase zmeniť. V takomto prípade je možné že politika mala na výsledok záujmu účinok aj napriek tomu, že sa jeho hodnota nezmenila, alebo naopak, politika nemala žiadny účinok no výsledok záujmu sa v sledovanom období zmenil. Druhý prístup je vo všeobecnosti tiež nevhodný pretože politikou zasiahnuté a nezasiahnuté jednotky sú za normálnych okolností odlišné aj bez prítomnosti politiky. Je to obzvlášť znateľné v prípadoch, kedy je politika namierená voči špecifickej časti populácie, alebo keď si jednotlivci/jednotlivé subjekty sami volia účasť na programe/politike.

Kontrafaktuálne metódy hodnotenia si dávajú za cieľ identifikovať adekvátnu kontrolnú skupinu, následne kontrafaktuálny výsledok a účinky politiky. Tieto metódy sa stali štandardným prístupom identifikácie kauzálnych účinkov politík vo väčšine inštitúcií a medzinárodných organizáciách v posledných desaťročiach, na čele so skupinou Svetovej banky.

Táto publikácia opisuje rámce hodnotenia politík a jednotlivé metódy kontrafaktuálnej analýzy:

- ***Propensity score matching***
- ***Regression discontinuity design***
- ***Difference-in-Differences***
- ***Instrumental variables***

Intervencia

Intervencia je politickým nástrojom namiereným voči špecifickej skupine populácie za účelom vyvolania zmeny definovaného stavu alebo správania (Loi a Rodrigues, 2012). V tomto kontexte je potrebné zdôrazniť tri za sebou nasledujúce elementy politickej intervencie:

- Cieľová populácia – dôkladne definovaný súbor jednotiek (napr. osôb, skupín obyvateľstva, domácností, firiem, oblastí/regiónov) na ktoré bude počas určitého obdobia opatrenie pôsobiť.
- Intervencia – akcia, alebo súbor akcií (*treatment*), ktorých účinkov na výsledok záujmu (*outcome*) sa snažíme posúdiť voči situácii ktorá by nastala bez pôsobenia intervencie. Jednotlivci populácie, ktorí sú vystavení intervencii sa označujú ako podporení (*treated*), jednotlivci bez vystavenia alebo účasti na intervencii ako nepodporení (*non-treated*).
- Výsledková premenná (*outcome variable*) – pozorovateľná a merateľná charakteristika jednotiek populácie na ktorú má intervencia pôsobiť (ovplyvniť).

Kauzalita

Vo svojej podstate je hodnotenie dopadov štúdiom **vzťahov medzi príčinami a dôsledkami**. Jeho cieľom je poskytnúť odpoveď na otázku: či účasť na programe ovplyvňuje sledovanú výsledkovú premennú. Inak povedané, do akej miery môžeme variácie výsledkovej premennej pripísať intervencii, za predpokladu zachovania ostatných vplyvov konštantnými. Odpoveď na túto otázku vieme získať odčítaním hodnoty výsledkovej premennej po intervencii od hodnoty ktorú by táto premenná nadobudla pri absencii intervencie (čistý rozdiel, *net difference*). V tomto kontexte sa **kauzalita vzťahuje na čistý prírastok alebo úbytok vo výsledku (*outcome*) podporených jednotiek (*treated units*) ktorý možno pripísať intervencii**.

Náhodný výber a kvázi-experiment

Najspoľahlivejšou metódou posúdenia účinkov intervencie je náhodný výber (*random experiment*), často označovaný ako zlatý štandard medzi metódami hodnotenia politik. V rámci tejto metódy sú jednotlivci cieľovej populácie (*target population*) náhodne rozdelení do intervenčnej (podporenej) a kontrolnej (nepodporenej) skupiny, tzn. majú rovnakú pravdepodobnosť ocitnúť sa v ktorejkoľvek z týchto skupín. Dosiahnuté výsledky (*outcomes*) oboch skupín sa sledujú a ich rozdiely sú pripísané pôsobeniu intervencie.

Najväčšou prekážkou náhodného výberu je jeho finančná a časová náročnosť a (hlavne v Európe) býva považovaný za neetický. Je to kvôli odopretiu možnosti jednotlivcom náhodne zaradeným do kontrolnej skupiny zúčastniť sa na programe, ktorý by im v princípe mal pomôcť. V prípadoch kedy randomizácia nie je možná sú k dispozícii alternatívne spôsoby hodnotenia. Ide o kvázi-experimentálne metódy, resp. štúdie založené na pozorovaní, ktoré porovnávajú jednotlivcov cieľovej populácie (*target group*) zúčastnených na programe (intervencii) s kontrolnou skupinou nie-náhodne zvolených jednotlivcov cieľovej populácie, ktorí sa na programe nezúčastnili. Ak je kontrolná skupina v relevantných charakteristikách podobná podporenej skupine, účinky programu vieme vyhodnotiť s relatívnou presnosťou.

Zásadné problémy pri hodnotení

Hlavnou oblasťou záujmu pri hodnotení politik je otázka či je výsledková premenná jednotlivcov v cieľovej populácii ovplyvnená účasťou na programe. Ako príklad môžeme uviesť *APTP* (aktívnu politiku trhu práce), kde by hlavným výsledkom (*outcomes*) účastníkov programu mohla byť vyššia pravdepodobnosť zamestnania sa alebo vyššie príjmy po určitom čase od absolvovania programu. V tomto kontexte by sme chceli zistiť hodnotu výsledku účastníka programu po jeho absolvovaní a hodnotu jeho výsledku ak by sa programu nebol zúčastnil. V tomto momente narážame na zásadný problém evaluácií, a to nemožnosť sledovania toho istého jednotlivca populácie v oboch situáciách (tzn. za účasti aj neúčasti na intervencii) v rovnakom čase. Z tohto dôvodu vyvodzovanie dopadov politiky na výsledok jednotlivca cieľovej populácie zahŕňa úvahu o tom, ako by si počínal ak by nebol politikou zasiahnutý. Štandardným rámcom formalizácie tohto problému je kontrafaktuálny prístup (tiež nazývaný „prístup potenciálneho výsledku“ alebo „*Roy-Rubinov model*“). Kontrafaktuálnym výsledkom podporenej jednotky je jej potenciálny výsledok pri absencii podpory a naopak kontrafaktuálnym výsledkom nepodporenej jednotky je jej potenciálny výsledok dosiahnutý po obdržaní podpory. Kľúčovým predpokladom kontrafaktuálneho prístupu je, že každá jednotka cieľovej populácie dosahuje potenciálny výsledok v každom stave (podporená/nepodporená), avšak v každom okamihu môže byť sledovaná iba v jednom stave.

Formalizácia kontrafaktuálneho prístupu

Pri využívaní kontrafaktuálneho prístupu sledujeme súbor premenných (Y, X, D) pre každú jednotku populácie: (Y_i, X_i, D_i) , kde $i = 1, \dots, N$. D je umelá premenná (*dummy variable*) indikujúca či bola jednotka podporená ($D = 1$) alebo nepodporená ($D = 0$). Y je výsledková premenná, tzn. premenná ktorá má byť intervenciou ovplyvnená. X je súbor pozorovateľných charakteristík jednotky populácie, resp. na vyššej úrovni domácností, firiem, regionálnych charakteristík (X je súbor pozorovateľných charakteristík, zatiaľ čo x je jedna pozorovateľná charakteristika). Tieto premenné sa zvyčajne označujú ako vysvetľujúce premenné alebo kovarianty (*covariates*). Príkladmi pozorovateľných charakteristík pri skúmaní trhu práce (a jednotlivcov na trhu práce) sú napríklad vek, pohlavie, vzdelanie, príjem, bydlisko a iné. Pri skúmaní politik na podporu MSP to môže byť napr. počet zamestnancov podniku, lokalita, tržby, právna forma podniku, dĺžka pôsobenia na trhu a iné.).

Nech $(Y_1, Y_0)_i$ sú dva potenciálne výsledky i -tej jednotky populácie, Y_1 v podporenom stave a Y_0 v nepodporenom stave. Ak jednotka cieľovej populácie obdrží podporu, tak potom Y_1 je pozorovateľná (\equiv faktúrna) premenná a Y_0 je naopak nepozorovateľná a korešponduje so situáciou ktorú by sme pozorovali ak by jednotka populácie nebola podporená (\equiv kontrafaktuálna premenná). Podobne, ak určitá jednotka cieľovej populácie nie je podporená môžeme pozorovať iba Y_0 (\equiv faktúrna premenná), čiže Y_1 je v tomto prípade výsledok tejto jednotky za situácie že by bola obdržala podporu (\equiv kontrafaktuálna premenná).

Rovnicu vyjadrujúcu vzťah Y_i , reálneho výsledku pozorovateľného na i -tej jednotke cieľovej populácie, k potenciálnemu výsledku tej istej jednotky zapíšeme nasledovne:

$$Y_i = Y_{1i}D_i + Y_{0i}(1 - D_i) = \begin{cases} Y_{1i} & \text{if } D_i = 1 \\ Y_{0i} & \text{if } D_i = 0 \end{cases}$$

D_i vyjadruje stav podpory (*treatment status*) i -tej jednotky populácie. $D_i = 1$ v prípade podpory a $D_i = 0$ ak nebola podporená. Táto rovnica indikuje, ktorý z dvoch výsledkov bude v databáze sledovaný (Y_{1i} alebo Y_{0i}) v závislosti od stavu podpory ($D_i = 1$ alebo $D_i = 0$). Hlavným cieľom rovnice je poukázať na to, že na vyjadrenie kauzálneho vzťahu medzi D_i (príčinou) a Y_i (výsledkom) nemôžeme dať priamo do spojitosti Y_{1i} a D_i pod podmienkou $D_i = 1$, ale namiesto toho, musíme overiť výsledok Y_{0i} pod podmienkou $D_i = 0$ a porovnať Y_{0i} s Y_{1i} .

Odhad čo možno najpresnejšej hodnoty kontrafaktuálneho výsledku Y_{0i} je hlavným cieľom kvantitatívnych metód hodnotenia dopadov.

Priemerný účinok podpory

Keďže vyrátať kauzálny účinok podpory na jednotlivca (jednotku populácie) je v podstate nemožné (a z pohľadu tvorcov politik aj nezaujímavé), pozornosť sa sústreďuje predovšetkým na odhad agregovaných kauzálnych účinkov. Dvomi najčastejšie používanými spôsobmi sú: i) priemerný účinok podpory na celú populáciu (*ATE* – *average treatment effect*) a ii) priemerný účinok na jednotky v cieľovej populácii ktorým bola pridelená podpora (priemerný účinok na podporené jednotky, *ATT* – *average treatment effect on the treated*). Tieto dve veličiny možno vyjadriť nasledovne:

$$ATE = E(Y_{1i} - Y_{0i}) = E(Y_{1i}) - E(Y_{0i})$$

$$ATT = E(Y_{1i} - Y_{0i} | D_i = 1) = E(Y_{1i} | D_i = 1) - E(Y_{0i} | D_i = 1)$$

Priemerný účinok na celú populáciu (*ATE*) určuje priemerný účinok intervencie v prípade podpory všetkých jednotiek populácie. *ATE* je predmetom záujmu v prípade univerzálnej uplatniteľnosti hodnoteného opatrenia, tzn. všetky jednotky populácie sú vystavené intervencii.

Priemerný účinok na podporených (*ATT*) vyjadruje účinok podpory len na jednotky vystavené intervencii a je hlavným predmetom záujmu pri hodnotení politik. Prvá časť rovnice *ATT*, teda $E(Y_{1i} | D_i = 1)$ priemerný účinok na podporených je pozorovateľná premenná (faktuálny výsledok), zatiaľ čo druhá časť, $E(Y_{0i} | D_i = 1)$ priemerný účinok na podporených v prípade, že by neboli podporení, je nepozorovateľná premenná (kontrafaktuálny výsledok). Z tohto dôvodu nie je možné priamo identifikovať *ATT*.

Táto publikácia sa zameriava výlučne na identifikáciu *ATT* (nie *ATE*), keďže priemerný účinok na podporených je najdôležitejším parametrom pre tvorcov politik.

Výberové skreslenie 1/2

Keďže kontrafaktuálny výsledok nie je pozorovateľný, eventuálne by sme mohli použiť výsledok jednotiek nezúčastnených na opatrení ako vhodnú aproximáciu výsledku, ktorý by boli dosiahli zúčastnení pri absencii podpory. Uvedené by bolo správne len pod podmienkou veľmi podobných charakteristík zúčastnených a nezúčastnených jednotiek v období pred implementáciou programu (podpory). Vo všeobecnosti sa však zúčastnené a nezúčastnené jednotky odlišujú v kľúčových charakteristikách súvisiacich tak s účasťou na programe ako aj s výsledkom. Tento problém sa označuje ako výberové skreslenie.

Vhodným príkladom výberového skreslenia je napríklad prípad poskytovania tréningov na zlepšenie uplatniteľnosti na trhu práce, kedy majú šikovnejší a kvalifikovanejší jednotlivci vyššiu pravdepodobnosť zaradenia do programu a rovnako vyššiu pravdepodobnosť nájdenia si zamestnania (pozri Caliendo a Kopeinig, 2008).

Náhodné zaradenie do programu (randomizácia) rieši problém výberu keďže stav podpory ($D_i = 1$ alebo $D_i = 0$) je v tomto prípade nezávislý od potenciálneho výsledku (pozri Angrist a Pischke, 2009). Náhodný výber eliminuje všetky systematické skôr-existujúce skupinové rozdiely, keďže výber do podporenej a nepodporenej skupiny je založený výlučne na náhode. Z daného vyplýva, že každá skupina má rovnaké očakávané hodnoty všetkých charakteristík, pozorovateľných aj nepozorovateľných. Náhodný výber určitej vzorky môže vytvoriť skupiny, ktoré sú náhodou odlišné, avšak tieto odlišnosti sú náhodné chyby a nie odchýlky. Zákony pravdepodobnosti zaisťujú, že čím je počet jednotiek v cieľovej populácii väčší, tým je prítomnosť skôr-existujúcich skupinových rozdielov menšia.

Príklad: Predpokladajme, že máme údaje o mzdách a charakteristikách pracovníkov, ktoré zahŕňajú informáciu o ich členstve v odborovom združení. Naivný spôsob odhadu účinkov členstva v odboroch na mzdu je porovnanie priemerných miezd pracovníkov, ktorí sú zároveň členmi odborov ($D_i = 1$) a tých, ktorí členmi nie sú ($D_i = 0$). Problémom tohto prístupu je fakt, že členstvo v odboroch považuje za exogénny faktor, teda že je úplne nezávislé na iných osobných charakteristikách pracovníkov. V skutočnosti však existuje množstvo faktorov ovplyvňujúcich rozhodnutie pracovníkov vstúpiť do odborového združení, v dôsledku čoho je potrebné vziať do úvahy aj výberové skreslenie (pozri Guo a Fraser, 2010).

Náhodné zaradenie do programu rieši problém výberu keďže stav podpory ($D_i = 1$ alebo $D_i = 0$) je v tomto prípade nezávislý od potenciálneho výsledku (pozri Angrist a Pischke, 2009). Náhodný výber eliminuje všetky systematické skôr-existujúce skupinové rozdiely, keďže výber do podporenej a nepodporenej skupiny je založený výlučne na náhode. Z daného vyplýva, že každá skupina má rovnaké očakávané hodnoty všetkých charakteristík, pozorovateľných aj nepozorovateľných. Náhodný výber určitej vzorky môže vytvoriť skupiny, ktoré sú náhodou odlišné, avšak tieto odlišnosti sú náhodné chyby a nie odchýlky. Zákony pravdepodobnosti zaisťujú, že čím je počet jednotiek v cieľovej populácii väčší, tým je prítomnosť skôr-existujúcich skupinových rozdielov menšia.

Výberové skreslenie 2/2

Vezmime si pozorovateľný rozdiel medzi priemerným výsledkom podporených jednotiek a priemerným výsledkom nepodporených jednotiek:

$$E(Y_{1i}|D_i = 1) - E(Y_{0i}|D_i = 0)$$

ak od neho odčítame a pripočítame kontrafaktuálny výsledok podporených jednotiek $E(Y_{0i}|D_i = 1)$ dostaneme:

$$E(Y_{1i}|D_i = 1) - E(Y_{0i}|D_i = 0) =$$

$$\underbrace{[E(Y_{1i}|D_i = 1) - E(Y_{0i}|D_i = 1)]}_{\text{Výberové skreslenie}} + \underbrace{[E(Y_{0i}|D_i = 1) - E(Y_{0i}|D_i = 0)]}_{\text{rozdiel medzi nepodporenými jednotkami}}$$

Výberové skreslenie, tzn. rozdiel medzi podporenými a nepodporenými jednotkami, ktorý by sme boli pozorovali aj v prípade absencie podpory/programu. Závisí od pôvodne existujúcich rozdielov medzi týmito skupinami.

Táto rovnica poukazuje na fakt, že pozorovateľný rozdiel sa rovná priemernému účinku podpory na podporené jednotky jedine v prípade absencie výberového skreslenia. Či pozorovaný rozdiel stredných hodnôt medzi podporenými a nepodporenými jednotkami (rozdiel v stredných hodnotách skutočných výsledkov) korešponduje s priemerným účinkom podpory na podporených závisí od prítomnosti alebo neprítomnosti výberového skreslenia.

V odbornej literatúre sa odhad kontrafaktuálneho výsledku zhoduje s predchádzaním problému výberového skreslenia a preto sa všetky doteraz vyvinuté metódy hodnotenia politík zameriavajú na jeho riešenie.

Na odhalenie kauzálnych účinkov verejných politík je potrebné vychádzať z dvoch hlavných predpokladov:

1. *Predpoklad nemenného stavu podpory jednotky populácie (SUTVA – Stable unit treatment value assumption)*

Prvý predpoklad zahŕňa dve podmienky:

- I. pre každú jednotku existuje iba jedna forma podpory a jedna forma absencie podpory (tzn. podpora nesmie byť nejasná, musí byť jednoznačná) a
- II. medzi jednotkami neexistuje žiadna interakcia (spolupráca), tzn. výsledok dosiahnutý jednotkou i nie je ovplyvnený stavom podpory ani výsledkom žiadnej inej jednotky.

2. *Spoločná podmnožina definičného oboru (Common Support)*

Druhý predpoklad taktiež zahŕňa dve podmienky:

- i. podporené ako aj nepodporené jednotky sú pozorovateľné a
- ii. pre každú podporenú jednotku existuje aspoň jedna porovnateľná nepodporená jednotka, tzn. že existuje nepodporená jednotka s podobnými pozorovateľnými charakteristikami X a podobnou pravdepodobnosťou podpory. Matematicky vyjadrené:

$$\Pr(D_i = 1|X_i) < 1$$

čo znamená, že pravdepodobnosť vystavenia intervencii pre jednotku i , za predpokladu $X_i = x$ (tzn. pohlavie = žena) je menšia ako 1, čiže existuje aspoň jedna jednotka j s rovnakými charakteristikami (pohlavie = žena), ktorá nie je vystavená podpore ($D_j = 0$).

Ak spoločná podmnožina definičného oboru platí iba čiastočne, teda len pre podmnožinu hodnôt X (tzn. len pre ženy alebo len pre mužov), analýzu musíme obmedziť iba na túto podmnožinu.

Pri realizácii **kvantitatívnych hodnotení** je záujmom hodnotiteľov kvantifikovať veľkosť zmeny spôsobenej intervenciou. Výsledkom je vždy **číslo, ktoré vyjadruje rozdiel alebo zmenu ktorú nastala** medzi dvomi sledovanými obdobiami.

Na získanie odpovede prečo daná zmena nastala alebo nenastala, resp. či bola spôsobená len intervenciou alebo aj inými faktormi je potrebné využiť **metódy hodnotenia dopadov založené na teórii**. Významom hodnotení založených na teórii je popri kvantifikovateľných kauzálnych účinkov aj množstvo iných informácií a faktov, ktoré sú pre tvorcov politik dôležitými informáciami pri rozhodovaní. Otázky zaoberajúce sa príčinami účinkov intervencií, ich cílením a kontextom v ktorom sa implementujú sú rovnako dôležité ako kvantifikácia ich veľkosti.

Výsledkom kvalitatívnych hodnotení nie je konkrétne číslo ale naratívny popis situácie. Z tohto dôvodu sú ich závery ťažšie komunikovateľné a podložiteľné konkrétnymi číslami, čo môže vyvolávať dojem ich menšej vedeckej významnosti alebo objektívnosti. Ich sila však spočíva v možnostiach sprostredkovať hodnotiteľom a tvorcovi politik náhľad na skutočné príčiny fungovania intervencií v reálnom prostredí.

Kvalitatívna analýza	Kvantitatívna analýza
Začína všeobecnejšími otvorenými otázkami a s narastajúcim počtom informácií sa jej presnosť a precíznosť zvyšuje	Kľúčové vysvetľujúce premenné a výsledkové premenné sú identifikované vopred
Pred-definované premenné nie sú na začiatku identifikované	Kontextuálne premenné sú identifikované a kontrolované
Predbežná, resp. úvodná analýza je neoddeliteľnou súčasťou zberu dát	Zber dát a samotná analýza sú separátnymi fázami
	Analýza využíva formálne štatistické metódy

Zdroj: Nigatu, 2009

Hodnotenie dopadov založené na teórii (Theory based impact evaluation, TBIE) je prístupom, ktorý sa zameriava na súbor predpokladov a hypotéz aplikovaných tvorcami politik, programovými manažérmi a inými zainteresovanými osobami, ktoré sú logicky previazané a empiricky testovateľné (Leeuw, 2011).

Tieto teórie vyjadrujú intervenčnú logiku politiky (programu), teda jednotlivé opatrenia vedúce k alokácii finančných prostriedkov (vstupov) s cieľom dosiahnuť plánované výstupy, ktoré majú prispieť k zlepšeniu situácie v určitej oblasti. Úspešnosť týchto opatrení závisí na efektívite politiky, faktoroch, ktoré ju ovplyvňujú, ako aj na samotnom kontexte v ktorom sa politika aplikuje. Najdôležitejším elementom efektivity politiky je mechanizmus, prostredníctvom ktorého sa aplikuje. Mechanizmus aplikácie nemožno charakterizovať ako následnosť vstup-výstup-výsledok, logický model alebo štatistickú rovnicu. Mechanizmus zahŕňa presvedčenia, želania, poznania a ďalšie rozhodovacie procesy, ktoré ovplyvňujú správanie jednotlivcov a skupín.

Hodnotenia dopadov založené na teórii skúmajú mechanizmy aplikované tvorcami politik a porovnávajú ich s dôkazmi založenými na vedeckom skúmaní. TBIE sa zameriava na intervenčnú logiku, jej ciele a testovanie a v prípade potreby na jej zlepšenie.

Hodnotenia založené na teórii sú zložené z dvoch kľúčových elementov. Prvý je konceptuálny a druhý empirický. Z konceptuálneho hľadiska formulujú a vymedzujú hodnotenú politiku a jej teóriu, zatiaľ čo z empirického hľadiska sa snažia túto teóriu testovať a preskúmať či, a ako tieto politiky fungujú a či sa im darí dosahovať želané výsledky.

Keďže teórie súvisiace s politikou alebo programom nebývajú explicitne vyjadrené v oficiálnej dokumentácii alebo programových manuáloch, musia byť hodnotiteľmi vyskúmané a artikulované spôsobom, ktorý je následne možné otestovať. Testovanie teórií sa môže vykonať na základe existujúcich alebo nových údajov, tak kvantitatívnych ako kvalitatívnych. Nižšie uvádzame príklady troch v praxi najčastejšie využívaných prístupov.

- **Realistické hodnotenie** (*Realistic evaluation*; Pawson a Tilley, 1997) – tento prístup zdôrazňuje dôležitosť usporiadania „kontext-mechanizmus-výsledok“ pri politikách a programoch. Kontext v ktorom je program realizovaný ovplyvňuje výsledky, ktoré sa ním dosiahnu. Kontext programu zahŕňa aspekty ako sociálne, ekonomické a politické štruktúry, inštitucionálny usporiadanie, zainteresované osoby, geografické a historické aspekty a pod. Niektoré aspekty v rámci kontextu môžu podporovať mechanizmy realizácie, iné im na druhej strane brániť. Interakcia medzi kontextom a mechanizmom realizácie následne podnecuje a ovplyvňuje výsledok programu: kontext + mechanizmus = výsledok.
- **Teória zmeny** (*Theory of Change*; Weiss, 1995) – táto teória popisuje predpoklady vysvetľujúce tak jednotlivé kroky vedúce k dlhodobým cieľom, ako aj prepojenia a súvislosti medzi programovými aktivitami a výsledkami dosahovanými v každom kroku na ceste k dosiahnutiu cieľa.
- **Analýza prínosov** (*Contribution analysis*; 1990) – analýza prínosov je založená na vizuálnom zobrazení kauzálnych súvislostí a vysvetľujúcich úsudkov medzi pozorovanými zmenami a hodnotenými intervenciami. Cieľom analýzy prínosov nie je exaktné vymedzenie kauzálnych súvislostí medzi programom a želanými výsledkami, ale skôr poskytnutie vierohodných dôkazov umožňujúcich znížiť neistotu ohľadom zmeny sledovaných výsledkov spôsobených programom.

Kontrafaktuálne metódy hodnotenia dopadov

Prehľad kontrafaktuálnych metód

Metóda Difference-in-Differences

Metóda Propensity score matching

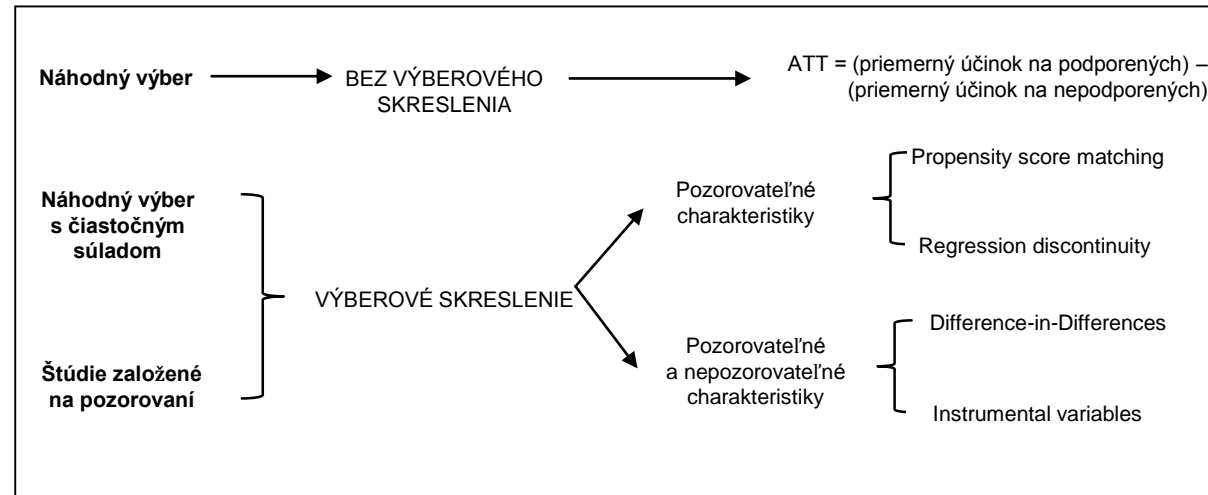
Metóda Regression discontinuity design

Metóda Instrumental variables

Táto časť popisuje jednotlivé kontrafaktuálne metódy na zníženie výberového skreslenia. Patria medzi ne:

- *Propensity score matching*
- *Regression discontinuity design*
- *Difference-in-differences*
- *Instrumental variables*

Prvé dve metódy – *Propensity score matching* a *Regression discontinuity design* – možno použiť, keď okrem podmienky SUTVA a spoločnej podmnožiny definičného oboru (*common support*), riešiteľ pozná a pozoruje v dátach všetky premenné, ktoré majú vplyv na vystavenie jednotky podpore a na potencionálne výsledky. Tento dodatočný predpoklad – známy ako **výber na základe pozorovateľných premenných** – je ťažké obhájiť v oblastiach ako vzdelávanie a trh práce, kde nepozorovateľné charakteristiky (ako osobné schopnosti, motivácia a inteligencia) budú s väčšou pravdepodobnosťou vplyvať na individuálne správanie. V prípadoch, keď má riešiteľ podozrenie, že výber do programu je ovplyvňovaný tak pozorovateľnými ako aj nepozorovateľnými faktormi/charakteristikami sú k dispozícii ďalšie dve metódy: *Difference-in-differences* alebo *Instrumental variables* (viď obrázok nižšie).



Kvôli prehľadnosti sú v nasledujúcej časti jednotlivé metódy prezentované samostatne. V praxi sú zväčša tieto metódy kombinované. Neexistuje presne stanovený postup, ako najlepšie tieto metódy kombinovať, ich použitie závisí od typu štúdie a prístupu k hodnotenému problému.

Zdroj: Loi a Rodrigues, 2012

Difference-in- Differences

Charakteristika metódy

Metóda Difference-in-Differences (DiD alebo DD – *double difference*) využíva časové rady údajov na určenie kontrafaktúálnej situácie. Jej aplikácia si vyžaduje údaje pre podporenú aj nepodporenú skupinu a pre obdobie pred a po realizácii intervencie. Priemerný účinok podpory na podporené jednotky (ATT - *average treatment effect on the treated*) sa určí porovnaním rozdielov výsledkov medzi podporenou a porovnávacou skupinou po určitom čase od ukončenia realizácie intervencie, v porovnaní s ich rozdielom v období pred intervenciou.

Metóda DiD berie na vedomie prítomnosť nepozorovateľnej heterogenity pri selekcii do programu, čo zaručuje odhad skutočného ATT v prípade v čase konštantného výberového skreslenia (tzn. výberové skreslenie je rovnaké v čase pred aj po intervencii a teda neovplyvní výsledný ATT). Na výpočet rozdielov sa využívajú časové rady údajov rovnakých jednotiek počas určitého obdobia.

Účinok podpory dosiahneme porovnaním priemerných hodnôt oboch skupín. V prvom kroku sa vypočíta priemerná hodnota rozdielu výsledkov pred a po implementácii u oboch skupín, matematicky vyjadrené:

$$E(Y_a^T - Y_b^T | T_i = 1) \text{ a } E(Y_a^C - Y_b^C | T_i = 0)$$

kde „Y“ označuje výsledkovú variabilnú premennú, dolný index „a“ obdobie po intervencii, dolný index „b“ obdobie pred intervenciou, horný index „T“ podporenú skupinu (*treatment group*) a „C“ porovnávaciu skupinu (*comparison group*), T_i označuje stav podpory ($T_i = 1$ pre podporených a $T_i = 0$ pre nepodporených).

Priemerný účinok podpory na podporené jednotky môžeme vyjadriť ako rozdiel týchto dvoch rozdielov:

$$ATT(DiD) = E(Y_a^T - Y_b^T | T_i = 1) - E(Y_a^C - Y_b^C | T_i = 0)$$

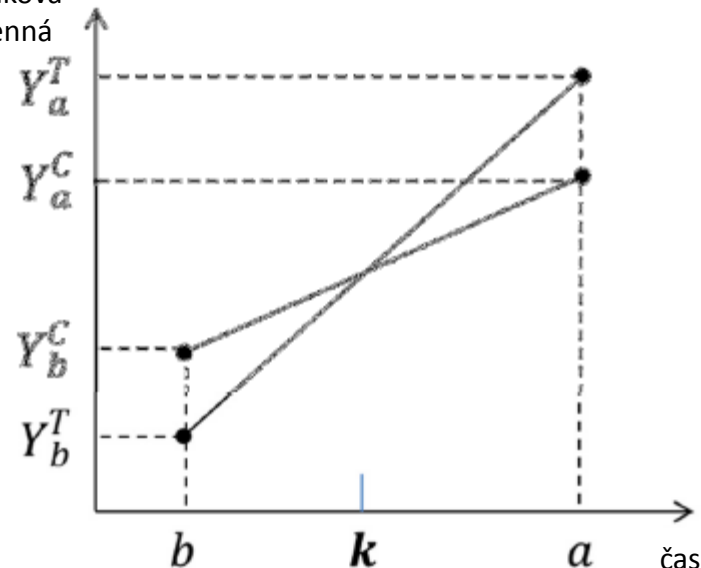
Predpoklady metódy

Metóda DiD závisí na dvoch predpokladoch:

- I. Nepozorovateľná heterogenita je v čase nemenná a teda pri porovnaní pred a po- intervenčného obdobia sa vynuluje,
- II. Spoločný trend (*common trend*): pri absencii podpory by podporené aj nepodporené jednotky prechádzali rovnakým trendom v rámci svojich výsledkových premenných. Z tohto dôvodu môže byť akákoľvek deviácia od trendu pozorovaná v skupine podporených jednotiek interpretovaná ako účinok dopady.

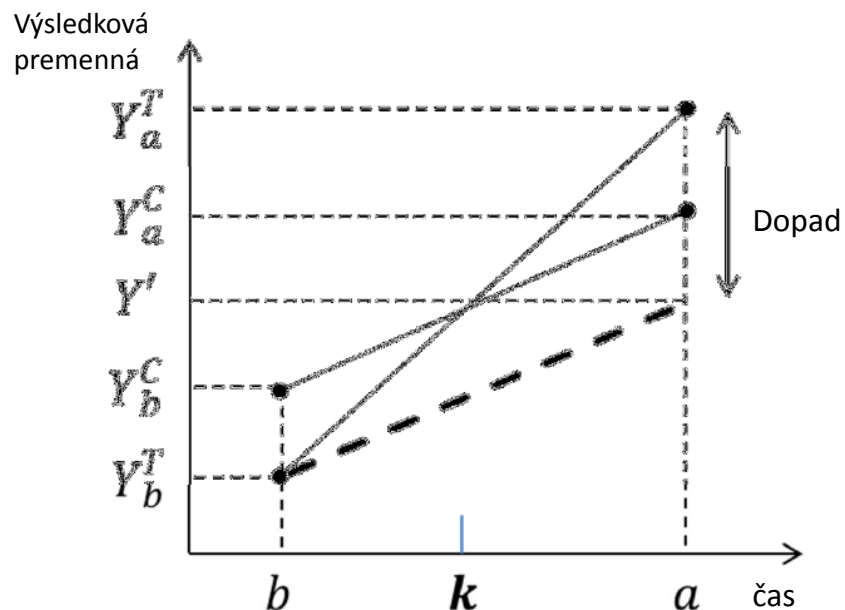
Grafické znázornenie metódy DiD, 1.časť

Výsledková
premená



Obdobie intervencie je na obrázku znázornené ako „k“, obdobie pred intervenciou ako „b“ (*before*) a obdobie po intervencii ako „a“ (*after*). Pre obe tieto obdobia je potrebné mať dostupné údaje. Hodnoty výsledkovej premennej pre podporenú skupinu sú Y_b^T v období pred intervenciou a Y_a^T v období po intervencii. Obdobne je to v prípade výsledkov porovnávacej skupiny, kde Y_b^C je hodnota pred a Y_a^C hodnota výsledkovej premennej po intervencii. Priemerný účinok podpory na podporených sa vypočíta podľa rovnice uvedenej vyššie. Na identifikáciu ATT na obrázku je potrebné aplikovať predpoklad spoločného trendu, ktorý hovorí že trend vývoja výsledku by bol pri absencii intervencie rovnaký pre podporenú aj nepodporenú skupinu (hrubá prerušovaná čiara na obrázku nižšie). Na základe uvedeného môžeme Y^I definovať ako výsledok podporenej skupiny pri absencii podpory a predpoklad spoločného trendu zapísať ako: $Y_a^C - Y_b^C = Y^I - Y_b^T$. ATT korešponduje s rozdielom medzi skutočným výsledkom podporenej skupiny a výsledkom, ktorý by bola dosiahla pri absencii podpory (intervencie): $Y_a^T - Y^I$, tzn. pozitívna zmena výsledku podporenej skupiny voči nepodporenej = dopad intervencie.

Grafické znázornenie metódy DiD, 2.časť



Metóda *Difference-in-Differences* (DiD, alebo aj *Double Difference* (DD)) umožňuje: porovnať vplyv intervencie za sledované obdobie, ako daná intervencia naplnila pôvodné očakávania, či bola úspešná alebo naopak, či nebola navrhnutá optimálne a mala len minimálny/žiadny účinok. Získané výsledky sa môžu následne použiť pri plánovaní intervencie v ďalšom programovom období, pomôžu navrhnúť zlepšenia pri poskytovaní podpory, resp. získané výsledky odhalia, že daná intervencia nenapĺňa požadované očakávania a v budúcnosti by sa už nemala realizovať, prípadne intervencia bude musieť byť preplánovaná, aby v budúcnosti spĺňala požiadavky.

Metóda pri hodnotení porovnáva závislé premenné dvoch skupín. Prvú skupinu tvoria podporení žiadatelia. Druhú skupinu tvoria nepodporení žiadatelia, ktorí nespĺnili požiadavky na získanie podpory. Pri porovnávaní metóda využíva obdobie pred intervenciou ako aj obdobie po intervencii.

Teoreticko – matematický základ metódy

Odhad metódy DiD rieši problém chýbajúcich údajov vzniknutých pri meraní výsledkov a kovariancie pri oboch skupinách, podporených ako aj nepodoprených žiadateľov, v období pred intervenciou ako aj po intervencii. Metóda DiD v podstate porovnáva obe skupiny, pokiaľ ide o zmeny v čase vzhľadom k výsledkom získaných z obdobia pred intervenciou. To znamená, že vzniknú dve obdobia (periódy), kde $t = 0$ je obdobie pred intervenciou a $t = 1$ je obdobie po intervencii. Nech Y_t^T a Y_t^C sú príslušné výsledky pre príjemcu programu (participujúcim na intervencii) a v čase t neošetrených jednotiek, potom bude mať táto metóda DiD tvar:

$$DiD = E(Y_1^T - Y_0^T | T_1 = 1) - E(Y_1^C - Y_0^C | T_1 = 0)$$

V rovnici predstavuje $T_1 = 1$ prítomnosť podporenej skupiny na programe v čase $t = 1$, naproti tomu $T_1 = 0$ predstavuje nepodoprenú skupinu.

Ak sú k dispozícii základné údaje, možno teda odhadnúť dopady intervencie za predpokladu, že nepozorovateľná heterogenita je časovo invariantná a nekorelovaná počas sledovaného obdobia. Tento predpoklad je slabší ako podmienka exogenity a vytvára výsledok zmeny pre porovnateľné skupiny z nepodoprených žiadateľov (to je, $E(Y_1^C - Y_0^C | T_1 = 0)$) ako vhodnú hypotézu rovnajúcu sa $E(Y_1^C - Y_0^C | T_1 = 1)$.

Pre lepšie pochopenie je možné rovnicu zapísať podrobne vo forme očakávania (pre tento moment sa bude počítat' s potlačením spodného indexu i):

$$E(Y_1^T - Y_0^T | T_1 = 1) = (\alpha + DD + \rho + \gamma) - (\alpha + \rho)$$

$$E(Y_1^C - Y_0^C | T_1 = 0) = (\alpha + \gamma) - \alpha$$

Je potrebné si uvedomiť, že výsledok rovnice DiD je objektívny len vtedy, ak potencionálny zdroj výberu odchýlok je aditívny a časovo nemenný. Rovnakým spôsobom, ako pre/proti vplyvu odhadu, sa vypočíta vzorka účastníkov (reflexná konštrukcia), vypočítaný vplyv programu bude $DiD + \gamma$ a zodpovedajúca odchýlka bude γ . Bez kontrolnej skupiny, ktorá by odôvodnila že ostatné faktory boli zodpovedné, ktoré by mohli ovplyvniť výstupy pre podporených žiadateľov, je zložitý potvrdiť dôveryhodnosť získaných výsledkov. Dal by sa tiež ešte porovnať rozdiel výsledkov v po-intervenčnom období celej podporenej a kontrolnej skupiny. Avšak v tomto prípade by bol odhadovaný vplyv politiky v tvare $DiD + \rho$, a odchýlka by bola ρ . Systematické nemerateľné rozdiely, ktoré by mohli ovplyvniť výsledky, nie je ľahké oddeliť.

Je potrebné mať na pamäti, že pre vyššie uvedenú rovnicu DiD musí byť správne prezentovaný výklad, ktorý musí spĺňať nasledovné:

- I. Model vo výslednej rovnici je zadaný správne, ak je aditívna štruktúra správne určená,
- II. Dátumová chyba je nekorelovaná s ostatnými premennými v rovnici:

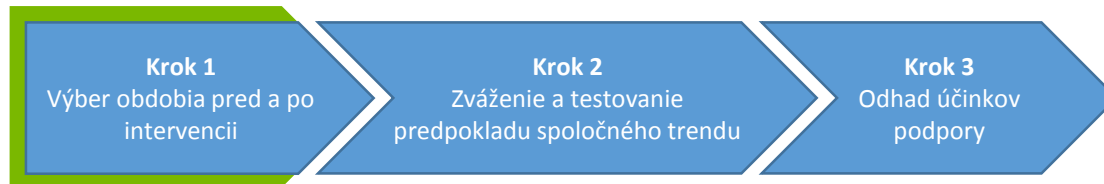
$$Cov(\varepsilon_{it}, T_{i1}) = 0$$

$$Cov(\varepsilon_{it}, t) = 0$$

$$Cov(\varepsilon_{it}, T_{i1}t) = 0.$$

Posledný z týchto predpokladov, tiež známi ako paralelný trend, patrí k najdôležitejším predpokladom. To znamená, že nepozorovateľné charakteristiky ovplyvňujú účasť na programe, ktorý sa mení v priebehu času spolu so stavom vyhodnocovania.

Postup výpočtu a kroky aplikácie metódy DiD



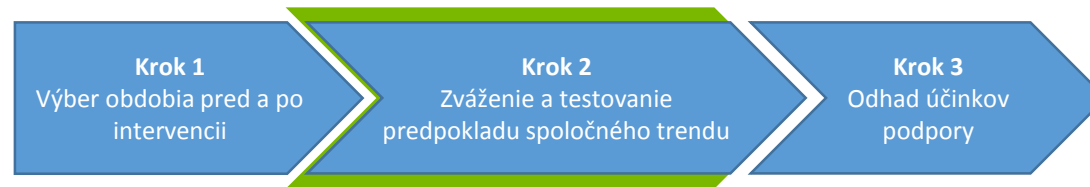
1. Výber obdobia pred a po intervencii

Výber pred a po intervenčného obdobia závisí od dostupnosti údajov za jednotlivé roky. Neexistuje univerzálne pravidlo na stanovenie presného počtu rokov ktoré je potrebné sledovať pred intervenciou, resp. po jej implementácii. Je potrebné zohľadniť charakter intervencie a dĺžku obdobia jej implementácie.

Pri jednorazových opatreniach, resp. opatreniach nenáročných na implementáciu (napr. nákup technológie, investícia do produkčných zariadení) je postačujúce sledovať obdobie dvoch alebo troch rokov pred realizáciou intervencie a obdobie po intervencii, kedy sa dá logicky predpokladať že jej účinky sa už stihli prejavíť.

Pri intervenciách majúcich za cieľ zmeny menej hmatateľných oblastí (ako napr. zvýšenie životnej úrovne, sociálneho a ekonomického postavenia, rodovej rovnosti, alebo zlepšenie životného prostredia) je potrebné počítať s dlhšími obdobiami počas ktorých sa bude pôsobenie intervencie sledovať a vyhodnocovať.

Postup výpočtu a kroky aplikácie metódy DiD



2. Zváženie a testovanie predpokladu spoločného trendu

V prípade dostupnosti údajov výsledkovej premennej oboch skupín v pred intervenčnom období (pozn. autora: avšak aj v iných rokoch než len v roku, ktorý bol použitý v rovnici pre výpočet ATT(DiD)), je možné testovať platnosť predpokladu spoločného trendu. V prípade aplikácie výpočtu ATT(DiD) iba pre pred intervenčné obdobie by mal byť výsledok nula, keďže ešte nedošlo k realizácii opatrenia a teda nemohlo dôjsť k rozdielom medzi jednotlivými skupinami.

Porušenie platnosti predpokladu spoločného trendu – metóda trojitého rozdielu Differences-in-Differences-in-Differences

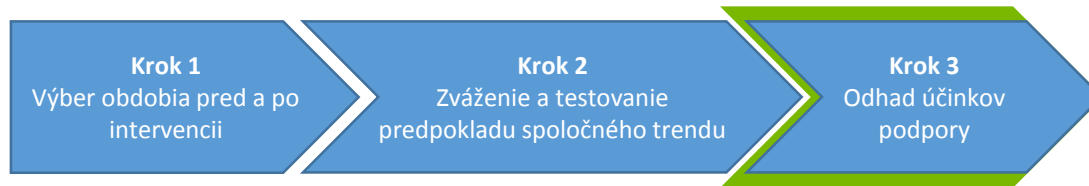
Ak z dôvodu určitých nepozorovateľných charakteristík reagujú podporená a porovnávací skupina na externý vývoj (napr. makroekonomický šok) odlišne, je predpoklad spoločného trendu porušený a použitie metódy DiD by podhodnotilo/nadhodnotilo hodnotu priemerného účinku podpory na podporených (ATT). Na riešenie tohto problému navrhujú Bell, Blundell a van Reenen (1999) techniku prispôbenia trendu (*trend adjustment technique*). Prispôbenie trendu pozostáva z aplikácie techniky trojitého rozdielu (Differences-in-Differences-in-Differences (DDD)) na zohľadnenie rozdielnych účinkov trendu vývoja.

Pre aplikáciu metódy trojitého rozdielu (DDD) sú potrebné údaje pre obe skupiny, pre obe obdobia (pred (*before*) a po (*after*)) a pre ďalšie dve obdobia, povedzme t a t' , kde: $(t' < t < b < k < a)$. Estimáciu trojitého rozdielu (DDD) môžeme matematicky vyjadriť ako:

$$DDD = [E(Y_a^T - Y_b^T | T_i = 1) - E(Y_a^C - Y_b^C | T_i = 0)] - [E(Y_t^T - Y_{t'}^T | T_i = 1) - E(Y_t^C - Y_{t'}^C | T_i = 0)]$$

kde je posledný výraz v hranatej zátvorke vyjadrením rozdielu vývoja trendu jednotlivých skupín sledovaného medzi obdobím t a t' .

Postup výpočtu a kroky aplikácie metódy DiD



3. Odhad účinkov podpory

Priemerný účinok podpory na podporených ATT(DiD) sa zvyčajne odhaduje prostredníctvom regresnej analýzy. Kľúčovými nezávislými premennými sú: indikátor jednotiek podporenej skupiny T_{it} ; indikátor po-intervenčného obdobia t ; a interakcia stavu podporenej skupiny s po-intervenčným obdobím $T_{it}t$:

$$Y_{it} = \alpha + \beta T_{i1}t + \gamma T_{i1} + \pi t + \varepsilon_{it}$$

Koeficient interakcie β predstavuje ATT(DiD), keďže meria rozdiel medzi oboma skupinami v po-intervenčnom období v porovnaní s pred-intervenčným obdobím. Tento parametrický prístup je vhodný z dvoch dôvodov:

- I. pre odhad štandardnej odchýlky a
- II. umožňuje zahrnúť iné vysvetľujúce premenné, konkrétne tie ktoré odrážajú rozdiely medzi počiatočnými podmienkami oboch skupín a tie ktoré by viedli k rozdielnym časovým trendom.

Heckman, Ichimura a Todd (1997) navrhujú kombináciu medzi metódou DiD a metódami párovania, aby sa zabezpečilo že na určenie účinku intervencie sa použijú iba podporené jednotky a jednotky porovnávacej skupiny ktoré sú porovnateľné v pred-intervenčnom období.

Smith a Todd (2005) poukazujú na to, že "podmienečná DiD" metóda je vhodnejšia ako tradičné prierezové párovanie (*cross-section matching*), pretože umožňuje výber na základe pozorovateľných charakteristík, ako aj na základe v čase nemenných nepozorovateľných charakteristík. Na realizáciu tohto kombinovaného prístupu je potrebné spárovať podporené a porovnávacie jednotky na základe ich pred-intervenčných charakteristík a na výpočet účinku podpory sa použijú jednotky v rámci spoločnej podmnožiny definičného oboru oboch skupín (*common support area*).

Zhrnutie metódy

Difference-in-Differences (DiD)			
Hlavná hypotéza	Potreba dát	Klady	Zápory
<p>Selekcia do podporenej skupiny je založená na v čase nemennej nepozorovateľnej heterogenite.</p> <p>Spoločný trend: pri absencii podpory by obe skupiny dosahovali podobný trend vývoja.</p>	Metóda skúma časový rozmer údajov, preto vyžaduje dlhodobé sledované alebo opakované prierezové údaje.	Zohľadňuje nepozorovateľnú, v čase nemennú heterogenitu skúmaných jednotiek.	Selekcia do podporenej skupiny môže byť ovplyvnená dočasnými nepozorovateľnými charakteristikami, ktoré v prípade nevydiferencovania vedú k skresleniu účinku podpory (ATT).
	Podpora (intervencia) musí nastať medzi dvomi pozorovanými časovými obdobiami.	Jednoduché testovanie predpokladu spoločného trendu, pokiaľ sú dostupné údaje z dvoch pred-intervenčných časových období (DiD medzi týmito dvomi pred-intervenčnými obdobiami by mal byť nula).	Predpoklad spoločného trendu nemusí byť zaručený, resp. môže byť netestovateľný.
		Umožňuje kombináciu s inými párovacími technikami pre zvýšenie porovnateľnosti medzi podporenou a porovnávacou skupinou.	Náročnosť na dostupné údaje: sú potrebné údaje za dve obdobia (dlhodobé časové rady alebo opakované prierezové údaje).
		Ide sa o flexibilný prístup, ktorý umožňuje ilustratívnu interpretáciu výsledkov.	

Zdroj: Loi, Rodrigues (2012)

Propensity score matching

Charakteristika metódy

Metóda Propensity score matching (PSM) je založená na štatistickej porovnávacej skupiny ktorá je vytvorená na základe pravdepodobnosti účasti na podpore (hodnoty propensity score) za využitia pozorovateľných charakteristík jednotiek populácie.

Podporené jednotky sú následne spárované s nepodporenými na základe hodnôt propensity score. Priemerný účinok podpory sa následne vypočíta ako priemer rozdielov vo výsledkoch naprieč týmito dvomi skupinami.

Platnosť PSM je závislá na dvoch podmienkach:

- I. podmiennečná nezávislosť (tzn. nepozorovateľné charakteristiky nemajú vplyv na účasť v programe),
- II. dostatočne veľká spoločná podmnožina definičného oboru (tzn. prekrývajúce sa hodnoty propensity score zúčastnených a nezúčastnených jednotiek).

Na párovanie podporených a nepodporených jednotiek na základe propensity score sa využívajú rozličné postupy. Patria medzi ne napr.:

- i. párovanie na základe najbližšieho suseda (nearest-neighbor matching),
- ii. caliper a radius matching,
- iii. stratifikácia a intervalové párovanie,
- iv. kernel matching,
- v. lokálne lineárne párovanie (LLM).

Metódy založené na regresii na vzorke podporených a nepodporených jednotiek využívajúce hodnoty propensity score ako váhy vedú k presnejším odhadom.

Metóda PSM je užitočným prístupom v prípadoch, kedy uvažujeme iba o pozorovateľných charakteristikách majúcich vplyv na účasť v programe. V prípadoch že je tento predpoklad dostatočne silný na určenie účasti v programe, údaje o širokej škále pred-programových charakteristík nám umožnia presnejšie špecifikovať pravdepodobnosť účasti v programe založenú na týchto pozorovateľných charakteristikách (Khadaker et al., 2010).

Využitie PSM v praxi

Vzhľadom na kontroverzný charakter implementácie randomizovaných (náhodných) hodnotení predstavuje tento prístup veľmi presné hodnotenie dopadu. Teda ak program nemôže byť implementovaný náhodne je simulácia náhodného experimentu kľúčovým krokom – tzn. pokus o pozorovateľnú analógiu náhodného experimentu.

Hodnotiteľ sa pomocou párovacích techník snaží vytvoriť kontrafaktúalnu alebo kontrolnú skupinu, ktorá je vzhľadom na pozorovateľné charakteristiky čo najpodobnejšia podporenej skupine. Podstatou je nájsť z čo najväčšej skupiny nezúčastnených jednotiek také jednotky, ktoré sú pozorovateľne podobné podporeným jednotkám v charakteristikách neovplyvnených účasťou v programe (tieto môžu zahŕňať pred-programové charakteristiky, keďže tieto s určitosťou nie sú ovplyvnené následnou účasťou v programe).

Každá podporená jednotka je spárovaná s pozorovateľne podobnou nepodporenou jednotkou a priemerný rozdiel vo výsledkoch naprieč týmito dvomi skupinami je účinkom programu (*programme treatment effect*).

Za predpokladu že rozdiely v účasti sú založené výlučne na pozorovateľných charakteristikách a ak máme k dispozícii dostatočne veľkú skupinu nepodporených jednotiek na spárovanie s podporenými, môže byť korešpondujúci účinok podpory odmeraný aj ak podpora nie je náhodná.

Problémom býva spoľahlivo identifikovať skupiny ktoré sú si podobné. Identifikácia je problémom pretože aj v prípade párovania jednotiek pozdĺž vektora X rozdielných charakteristík, je zriedka možné nájsť dve jednotky, ktoré sú si veľmi podobné v mnohých charakteristikách. Z dôvodu existencie množstva možných charakteristík je párovanie využitím PSM často využívanou technikou.

Pri aplikácii PSM je každá podporená jednotka spárovaná s nepodporenou na základe jej hodnoty propensity score, odrážajúcej pravdepodobnosť podpory podmienenú rozličnými pozorovateľnými charakteristikami X .

Metóda PSM predchádza tzv. „kľatbe dimenzionality“ ktorá je spojená so snahou párovania podporených a nepodporených jednotiek na základe každej možnej charakteristiky v prípade veľkej množiny X .

Úloha PSM

Pri aplikácii metódy PSM sa zostrojí štatistická porovnávacia skupina modelovaním pravdepodobnosti účasti v programe na základe pozorovateľných charakteristík neovplyvnených programom.

Podporení sú následne spárovaní s nepodporenými na základe tejto pravdepodobnosti, tzn. hodnoty *propensity score*. Priemerný účinok podpory (ATE) sa vyráta ako rozdiel priemerných hodnôt výsledkov týchto dvoch skupín.

Sama o sebe je metóda PSM užitočná v prípadoch ak sa má za to, že účasť na programe určujú iba pozorovateľné charakteristiky. Tento predpoklad závisí od pravidiel cielenia programu, ako aj od faktorov vedúcich k prihláseniu sa jednotiek do programu (*self-selection into programme*). V ideálnom prípade, ak sú pre podporené a nepodporené jednotky dostupné dáta z obdobia pred implementáciou programu, sa tieto môžu využiť na výpočet hodnôt *propensity score* a spárovanie týchto dvoch skupín na ich základe.

Teoretické východiská PSM

Metóda PSM sa snaží zachytiť účinky rozličných pozorovaných kovariantov X na účasť v programe pomocou hodnoty *propensity score*. Následne sú výsledky podporených a nepodporených jednotiek s podobnými hodnotami propensity score porovnané na určenie účinku programu. Jednotky pre ktoré nie je nájdená dvojica (jednotka s podobou hodnotou propensity score) sú z analýzy vylúčené pretože neexistuje základ pre ich porovnanie.

Pri aplikácii PSM je vytvorená štatistická porovnávacia skupina založená na pravdepodobnosti účasti v programe T (*treatment*), podmienenej pozorovanými charakteristikami X , teda propensity score: $P(X) = \Pr(T=1 | X)$. Rosenbaum a Rubin (1983) poukazujú na skutočnosť, že za určitých podmienok je párovanie na základe $P(X)$ rovnako vhodné ako párovanie na základe X .

Nevyhnutnými predpokladmi pre identifikáciu účinkov programu sú:

- I. podmiennečná nezávislosť a
- II. existencia spoločnej podmnožiny definičného oboru (spoločnej oblasti prelínania).

Predpoklad podmiennečnej nezávislosti

Predpoklad podmiennečnej nezávislosti hovorí, že ak existuje súbor pozorovateľných kovariantov X , ktoré nie sú ovplyvnené podporou, potenciálne výsledky Y sú potom nezávislé od účasti v programe T . Ak Y_i^T označuje výsledky zúčastnených a Y_i^C výsledky nezúčastnených jednotiek, potom podmiennečná nezávislosť implikuje:

$$(Y_i^T, Y_i^C) \perp T_i | X_i$$

Tento predpoklad je tiež označovaný ako predpoklad „nezameniteľnosť“ (*unconfoundedness*; Rosenbaum a Rubin 1983) a implikuje že účasť na programe je založená výlučne na pozorovaných charakteristikách.

Na odhad účinku podpory na podporených (TOT – treatment effect on the treated), na rozdiel od ATE, je postačujúci slabší predpoklad:

$$Y_i^C \perp T_i | X_i$$

Podmienečná nezávislosť je silným predpokladom a nie priamo testovateľným kritériom - závisí od špecifik programu ako takého. Ak účasť v programe určujú aj nepozorované charakteristiky, predpoklad podmienečnej nezávislosti nebude dodržaný a PSM v takomto prípade nie je vhodnou metódou.

Alternatívami v prípade existencie výberu na základe nepozorovaných charakteristík sú napríklad metódy difference-in-differences alebo inštrumentálne premenné, ktoré sú popísané v ďalších kapitolách tejto publikácie.

Predpoklad spoločnej podmnožiny definičného oboru

Druhým predpokladom je predpoklad spoločnej podmnožiny definičného oboru (*common support*), resp. podmienka prelínania (*overlap condition*), matematicky vyjadrená:

$$0 < P(T_i = 1 | X_i) < 1$$

Táto podmienka zaručuje že pozorovania v rámci podporenej skupiny majú k nim porovnateľné pozorovania čo najbližšie v rámci oblasti distribúcie hodnôt *propensity score* (Heckman, LaLonde a Smith 1999). Inak povedané, efektívnosť metódy PSM závisí na dostatočne veľkej a približne rovnako početnej vzorky pozorovaní podporenej a nepodporenej skupiny, aby bolo možné identifikovať oblasť prelínajúcich sa hodnôt *propensity score* týchto skupín.

Na určenie odhadu účinku podpory na podporených (TOT) môže byť tento predpoklad obmedzený iba na:

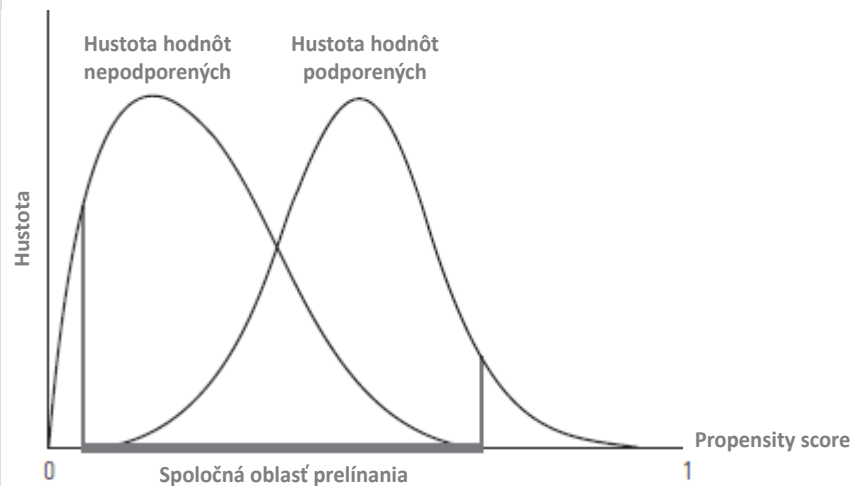
$$P(T_i = 1 | X_i) < 1.$$

Podporené jednotky by mali byť podobné nepodporeným jednotkám vzhľadom na pozorovateľné charakteristiky neovplyvnené účasťou v programe, z toho dôvodu môže dôjsť k vylúčeniu určitých nepodporených jednotiek aby sa zabezpečila porovnateľnosť.

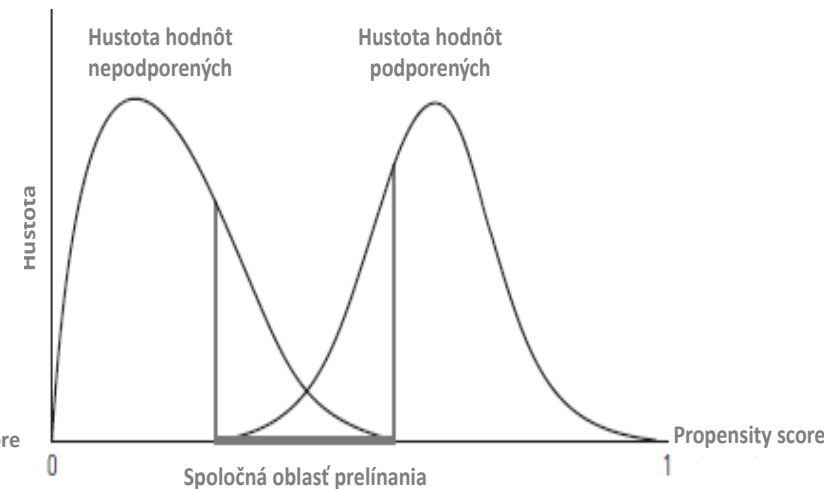
Niekedy tiež môže nastať situácia, kedy je potrebné vylúčiť určitú podskupinu podporených jednotiek, ak k nim neexistujú podobné nepodporené jednotky (Ravallion 2008). Takáto situácia môže byť problematickejšia, keďže môže potenciálne vyvolať skreslenie vzorky v účinkoch podpory. Analýza charakteristík vylúčených jednotiek môže byť užitočným nástrojom pri interpretácii potenciálneho skreslenia v odhadovaných účinkoch podpory.

Heckman, Ichimura a Todd (1997) odporúčajú vylúčiť pozorovania s nevýraznou oblasťou prelínania hodnôt propensity score. Logické úsudky o kauzalite môžu byť tvorené len pre oblasť spoločného prelínania, ako zobrazuje Graf č. 1. Graf č. 2 zobrazuje scenár s nevýraznou oblasťou prelínania hodnôt propensity score.

Graf č. 1: Príklad oblasti spoločného prelínania hodnôt propensity score



Graf č. 2: Príklad oblasti nevýrazného prelínania hodnôt propensity score



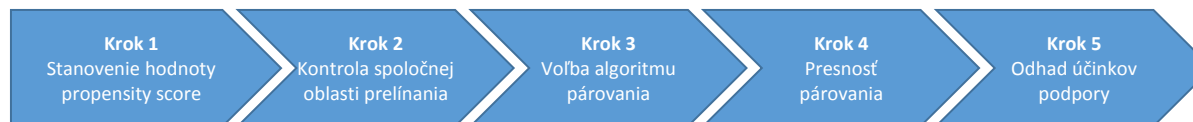
Postup výpočtu a kroky aplikácie metódy PSM

Na výpočet účinkov podpory (programu) je najprv potrebné vyrátať hodnoty propensity score $P(X)$ na základe všetkých pozorovateľných charakteristík X , ktoré ovplyvňujú tak účasť v programe ako aj očakávaný výsledok.

Cieľom párovania je vytvorenie čo možno najpodobnejšej porovnávacej skupiny nezúčastnených jednotiek k skupine podporených jednotiek. Podobnosť sa meria na základe pozorovateľných charakteristík neovplyvnených účasťou v programe (Khadaker, 2012).

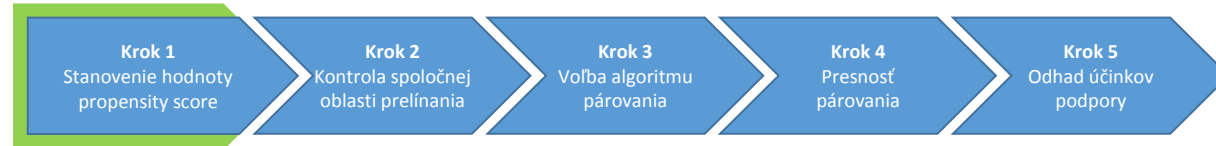
Aplikáciu metódy Propensity score matching môžeme rozdeliť do piatich základných krokov:

- I. Stanovenie hodnoty propensity score
- II. Kontrola spoločnej oblasti prelínania
- III. Voľba algoritmu párovania
- IV. Presnosť párovania
- V. Odhad účinkov podpory



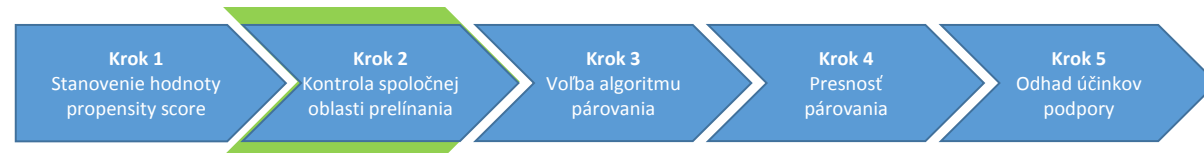
V nasledujúcej časti sú bližšie popísané jednotlivé kroky aplikácie metódy propensity score matching.

Postup výpočtu a kroky aplikácie metódy PSM



1. Stanovenie hodnoty propensity score

Hodnota propensity score určuje pravdepodobnosť jednotky v cieľovej populácii (podporené a porovnávacie jednotky) obdržať podporu (treatment) na základe jej pozorovateľných charakteristík X_i , matematicky vyjadrené: $Pr(D_i = 1|X_i)$. Ak majú dve jednotky podobnú hodnotu propensity score, znamená to že sú podobné aj v charakteristikách X použitých na jej určenie (Rosenbaum a Rubin, 1983).

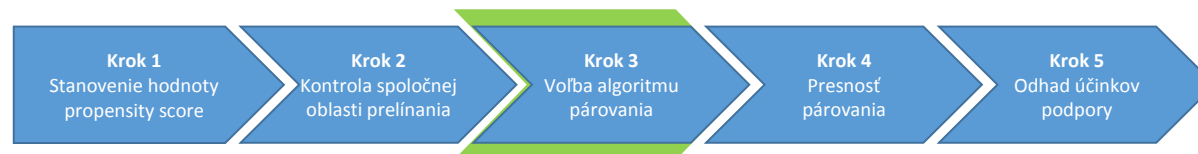


2. Krok – Kontrola spoločnej oblasti prelínania

Je potrebné zabrániť porovnávaniu neporovnateľného, tzn. že v analýze by sa mala použiť iba podmnožina porovnávacej skupiny ktorá je porovnateľná s podporenou skupinou. Je potrebné skontrolovať výskyt aspoň jednej podporenej a jednej nepodporenej jednotky pre každú z hodnôt propensity score. Existuje viacero metód kontroly avšak najjednoduchšou je vizuálna analýza hustoty rozloženia hodnôt propensity score.

Ďalšou možnosťou je porovnanie minim a maxim hodnôt propensity score v podporenej a nepodporenej skupine. Oba prístupy vyžadujú odstránenie všetkých pozorovaní pri ktorých je hodnota propensity score menšia ako minimum alebo väčšia ako maximum v opačnej skupine.

Postup výpočtu a kroky aplikácie metódy PSM



3. Voľba algoritmu párovania

Ďalší krok sa pozostáva zo spárovania podporených a nepodporených jednotiek, ktoré majú podobné hodnoty propensity score za použitia vhodného algoritmu párovania. Algoritmy párovania sa líšia v spôsobe merania podobnosti medzi podporenými a nepodporenými jednotkami (tzn. spôsobe akým sa tvoria dvojčky) a tiež v spôsobe pridelovania váh spárovaným jednotkám.

Algoritmy párovania využívané pri aplikácii Propensity score matching

- A. Párovanie na základe najbližšieho suseda (Nearest-neighbour matching)
- B. Calliper a Radius matching
- C. Kernel matching

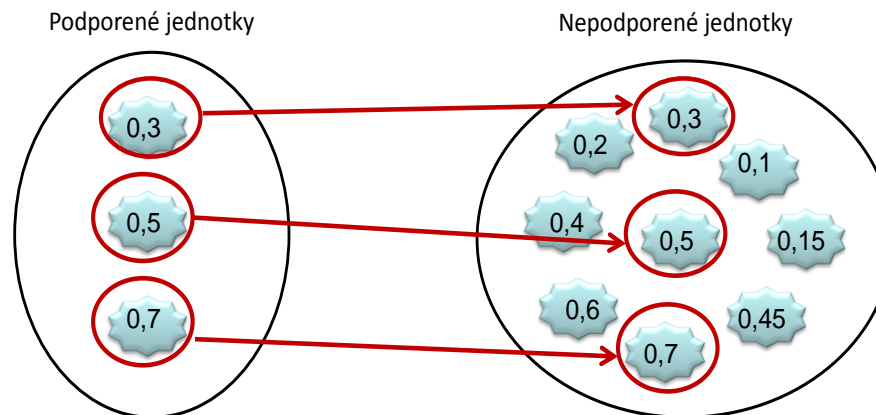
Algoritmy párovania

A. Párovanie na základe najbližšieho suseda (Nearest-neighbour matching) – Podporená jednotka je spárovaná s jednotkou z porovnávacej skupiny s najbližšou hodnotou propensity score. V rámci tejto techniky existujú dva varianty:

- i. párovanie s opakovaním – kedy môže byť nepodporená jednotka použitá ako dvojica k viacerým podporeným jednotkám, a
- ii. Párovanie bez opakovania – každá nepodporená jednotka sa použije iba raz.

Problémom párovania bez opakovania je závislosť odhadov účinkov od poradia v akom sa jednotlivé pozorovania párovali. Pri aplikácii tohto prístupu je preto potrebné zabezpečiť náhodnosť poradia v akom sú jednotky párované. (Khadaker, 2012).

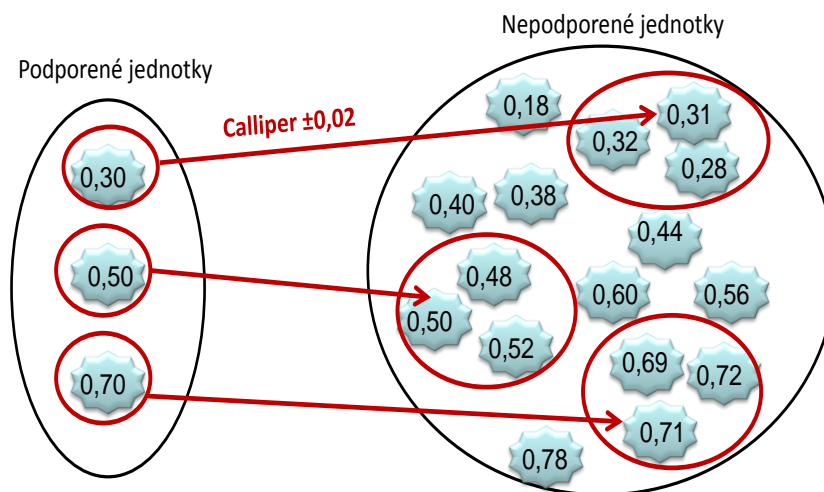
Obrázok 4: Párovanie na základe najbližšieho suseda



Algoritmy párovania

- B. Calliper a Radius matching** – Pri párovaní na základe najbližšieho suseda môže dôjsť k vytvoreniu zlých dvojčiek ak najbližší sused nie je dostatočne podobný. Tomuto sa dá zabrániť určením podmienky, na základe ktorej sa dvojčky môžu vytvoriť len z jednotiek, ktorých hodnoty propensity score sa nelíšia viac než o nami zvolenú toleranciu (napr. $\pm 2\%$). Táto tolerancia sa nazýva *calliper*.

Obrázok 5: Calliper matching



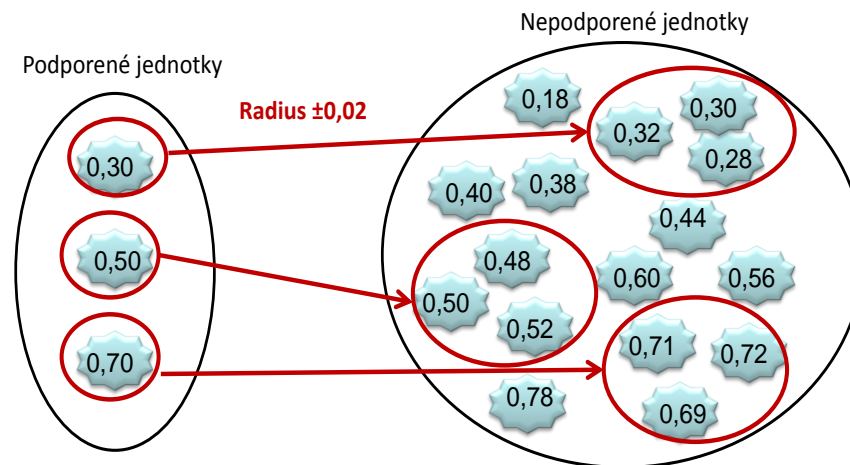
Algoritmy párovania

(pokračovanie)

B. Calliper a Radius matching – Pri párovaní na základe najbližšieho suseda môže dôjsť k vytvoreniu zlých dvojčiek ak najbližší sused nie je dostatočne podobný. Tomuto sa dá zabrániť určením podmienky, na základe ktorej sa dvojčky môžu vytvoriť len z jednotiek, ktorých hodnoty propensity score sa nelíšia viac než o nami zvolenú toleranciu (napr. $\pm 2\%$). Táto tolerancia sa nazýva calliper.

Radius matching je obdobou calliper matching, avšak jeho podstatou je využiť všetky jednotky spadajúce do nami zvolenej úrovne tolerancie, nie len najbližšieho suseda v každom callipry.

Obrázok 6: Radius matching



Algoritmy párovania

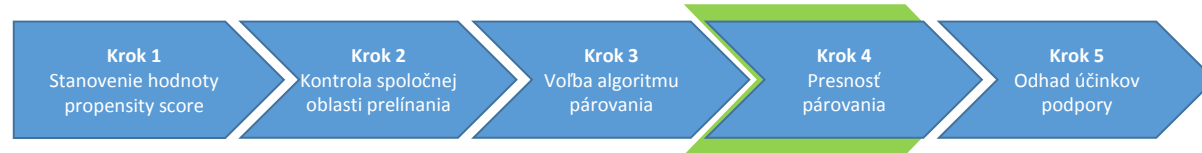
C. Kernel matching – Predchádzajúce dva algoritmy majú spoločné to, že len niektoré pozorovania z porovnávacej (nepodporenej) skupiny sa využijú na určenie kontrafaktúálneho výsledku podporenej jednotky.

Metóda *Kernel matching* používa na určenie kontrafaktúálneho výsledku vážené priemery všetkých jednotiek kontrolnej skupiny. Váhy závisia od vzdialenosti medzi každou jednotkou nepodporenej skupiny a jednotkou vystavenou intervencii pre ktorú sa určuje kontrafaktúálny výsledok. Funkcia Kernel priradzuje väčšiu váhu jednotkám porovnávacej skupiny ktoré sú v rámci hodnôt *propensity score* bližšie podporenej jednotke a nižšiu váhu hodnotám ktoré sú jej vzdialenejšie.

Voľba medzi rozličnými algoritmi párovania znamená kompromis medzi skreslením a znížením rozptylu. Párovanie na základe najbližšieho suseda využíva len podporenú jednotku a jej najbližšieho suseda z nepodporenej skupiny. Na jednej strane sa tým znižuje skreslenie, na strane druhej sa však znižuje aj efektívnosť, pretože sa neprihliada na viacero blízkych susedov.

Metóda Kernel matching využíva viacero (všetky) nepodporené jednotky pre každú podporenú jednotku, čím sa znižuje rozptyl ale potenciálne zvyšuje skreslenie. Využívanie tej iste nepodporenej jednotky viackrát (párovanie na základe najbližšieho suseda s opakovaním) môže do istej miery zlepšiť kvalitu párovania, avšak zvyšuje rozptyl (Caliendo et al, 2005).

Postup výpočtu a kroky aplikácie metódy PSM



4. Presnosť (kvalita) párovania

Kvalita párovania sa hodnotí na základe jeho schopnosti vyvážiť porovnávaciu a podporenú skupinu s ohľadom na charakteristiky použité pri určení hodnôt propensity score. Existuje viacero spôsobov hodnotenia kvality párovania. Základnou myšlienkou všetkých prístupov je porovnať rozloženie premenných (jednotlivých charakteristík) v oboch skupinách (podporenej a porovnávej) pred a po párovaní na základe propensity score. V prípade významných rozdielov po párovaní nebolo párovanie na základe hodnôt propensity score úplne úspešné a je potrebné vykonať určité nápravné opatrenia.

Rosenbaum a Rubin (1985) zdôrazňujú, že dobrá párovacia procedúra by mala znížiť štandardnú odchýlku (skreslenie) pre každú variabilnú premennú (pozorovateľnú charakteristiku) použitú pri určení propensity score. Tento prístup vyžaduje porovnanie štandardných odchýlok pre každú variabilnú premennú x pred a po párovaní.

Štandardnú odchýlku pred párovaním vypočítame ako:

$$\text{Štandardná odchýlka}_{pred} = \frac{\bar{x}_1 - \bar{x}_0}{\sqrt{0,5(V_1(x) - V_0(x))}}$$

Štandardnú odchýlku po párovaní vypočítame ako:

$$\text{Štandardná odchýlka}_{po} = \frac{\bar{x}_{1M} - \bar{x}_{0M}}{\sqrt{0,5(V_{1M}(x) - V_{0M}(x))}}$$

kde $\bar{x}_1(V_1)$ je priemerná hodnota (rozptyl) v podporenej skupine pred párovaním a $\bar{x}_0(V_0)$ je analogicky priemerná hodnota pre kontrolnú skupinu. $\bar{x}_{1M}(V_{1M})$ a $\bar{x}_{0M}(V_{0M})$ sú príslušné hodnoty pre spárované skupiny. Aj keď táto metóda jasne neprezentuje kritickú hranicu pre úspech procesu párovania, väčšina empirických štúdií považuje za takúto kritickú hranicu štandardnú odchýlku (skreslenie) pod úrovňou 3% alebo 5% po párovaní.

Podobný prístup sa využíva pri dvoj-vzorkovom *t*-teste (*two sample t-test*) pre kontrolu, či existujú významné rozdiely vo variabilných premenných pre obe skupiny. Po párovaní by mali byť variabilné premenné vyrovnané (vybilancované) pre obe skupiny, ako pre podporenú, tak aj pre nepodporenú skupinu, a žiadne významné rozdiely by nemali byť identifikované.

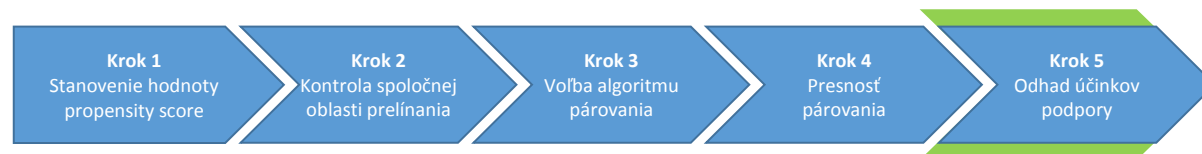
Sianesi (2004) navrhuje znovu odhadnúť *propensity score* v spárovanej vzorke a porovnať *pseudo-R²* pred a po párovaní. Po párovaní by nemali existovať žiadne systematické rozdiely v rozložení variabilných premenných medzi oboma skupinami. Z toho dôvodu by *pseudo-R²* po párovaní malo ostať nízke. Rovnakú kontrolu je možné vykonať aj pomocou *F*-štatistiky pred a po párovaní.

Výhody a nevýhody prístupov používaných pri hodnotení kvality párovania

	Výhody	Nevýhody
Štandardná odchýlka (skreslenie)	Nízka náročnosť výpočtu	Vykonáva sa pre každú variabilnú premennú použitú pri odhade <i>propensity score</i> . Bez objektívnej indikácie úspešnosti párovania
t-štatistika (t-test)	Nízka náročnosť výpočtu	Vykonáva sa pre každú variabilnú premennú použitú pri odhade <i>propensity score</i> . Zníženie odchýlky (skreslenia) nie je jednoznačne vidieť

Zdroj: Loi, Rodrigues (2012)

Postup výpočtu a kroky aplikácie metódy PSM



5. Odhad účinkov podpory

Po spárovaní podporených jednotiek s jednotkami kontrolnej skupiny je určenie účinkov podpory relatívne jednoduché – postačuje vyrátať priemernú hodnotu rozdielu medzi výsledkovou premennou v podporenej a nepodporenej skupine.

Poznámka:

Pred aplikáciou t-testu na overenie štatistickej významnosti účinkov je potrebné overiť prítomnosť štandardných chýb. Problémom je, že odhadovaná odchýlka účinku podpory zahŕňa aj odhad odchýlky hodnôt propensity score, spoločnej oblasti prelínania (common support) a prípadne aj poradia, v ktorom sa podporené jednotky párujú.

Jedným zo spôsobov, ako sa vysporiadať s týmto problémom je použitie metódy bootstrapping, ako navrhuje Lenchcer (2002). Na druhej strane Imbens (2004) konštatuje že metóda bootstrapping neposkytuje dostatočne formálny dôkaz, ktorý by podporoval jej aplikáciu, aj napriek jej rozšírenému využívaniu.

Zhrnutie metódy PSM

Propensity score matching			
Hlavná hypotéza	Potreba dát	Klady	Zápory
Selekcia do programu je daná len na základe pozorovateľných charakteristík	Údaje pre hodnotenie musia obsahovať všetky premenné, ktoré popisujú charakteristiky určujúce účasť v programe (pridelenie podpory)	Má dlhú tradíciu v mnohých odvetviach (napr. medicína, environmentálne štúdie, trh práce, vzdelávanie)	V oblastiach ako je vzdelávanie alebo trh práce je hypotéza výberu založeného na pozorovateľných charakteristikách ťažko obhájitelná
	Pre každú hodnotu propensity score musí existovať aspoň jedna podporená a jedna nepodporená jednotka	Má pod kontrolou súbor premenných (charakteristík) určujúcich účasť v programe (pridelenie podpory)	Metóda náročná na dáta; jej závery platia iba pre podmnožinu spárovaných jednotiek
	Prierezové dáta (cross-sectional data)	Nie je parametrickým prístupom preto je veľmi flexibilný	Externá validita (zovšeobecniteľnosť) jej výsledkov klesá so zvyšujúcim sa počtom nespárovaných jednotiek
			Odhadovaná odchýlka (variancia) účinku podpory by mala zahŕňať aj odchýlku hodnôt propensity score (neexistuje ustálená metóda na riešenie tohto problému)

Zdroj: Loi, Rodrigues (2012)

A blue trapezoidal graphic on the left side of the slide, tapering from left to right. The text is white and centered within the graphic.

Regression discontinuity design

Hraničnú hodnotu, resp. prah môže predstavovať jedna konkrétna charakteristika (napr. vek), alebo indikátor zložený z viacerých charakteristík (napr. počet bodov z odborného hodnotenia pozostávajúceho z viacerých hodnotených kategórií).

Charakteristika metódy

Metóda regression discontinuity design (RDD) má viacero prvkov náhodného experimentu a môže byť použitá v prípadoch, kedy náhodný výber podporených a nepodporených nie je možný. Ide o populárnu kvázi-experimentálnu metódu, ktorá je založená na poznaní presných pravidiel na základe ktorých je stanovená oprávnenosť zúčastniť sa intervencie.

V súlade s touto metódou je **zaradenie uchádzačov do oprávnenej skupiny založené výlučne na pozorovateľných pred-intervenčných premenných** a pravdepodobnosť podpory sa nespojitou mení ako funkcia týchto premenných. Pre objasnenie tejto myšlienky môžeme uviesť príklad, v ktorom je skupina žiadateľov o podporu rozdelená do dvoch skupín podľa hodnoty zvolenej premennej v pred-intervenčnom období (napr. vek, výška mzdy, počet bodov z odborného hodnotenia). Žiadatelia nachádzajúcim sa nad určitou hranicou sú vystavení intervencii, zatiaľ čo žiadatelia nachádzajúci sa pod touto hranicou sú zamietnutí (Battisin a Rettore, 2008). Nespojitosť v rámci RDD môže byť **sharp** (ostrá) alebo **fuzzy** („rozmazaná“).

Sharp regression discontinuity

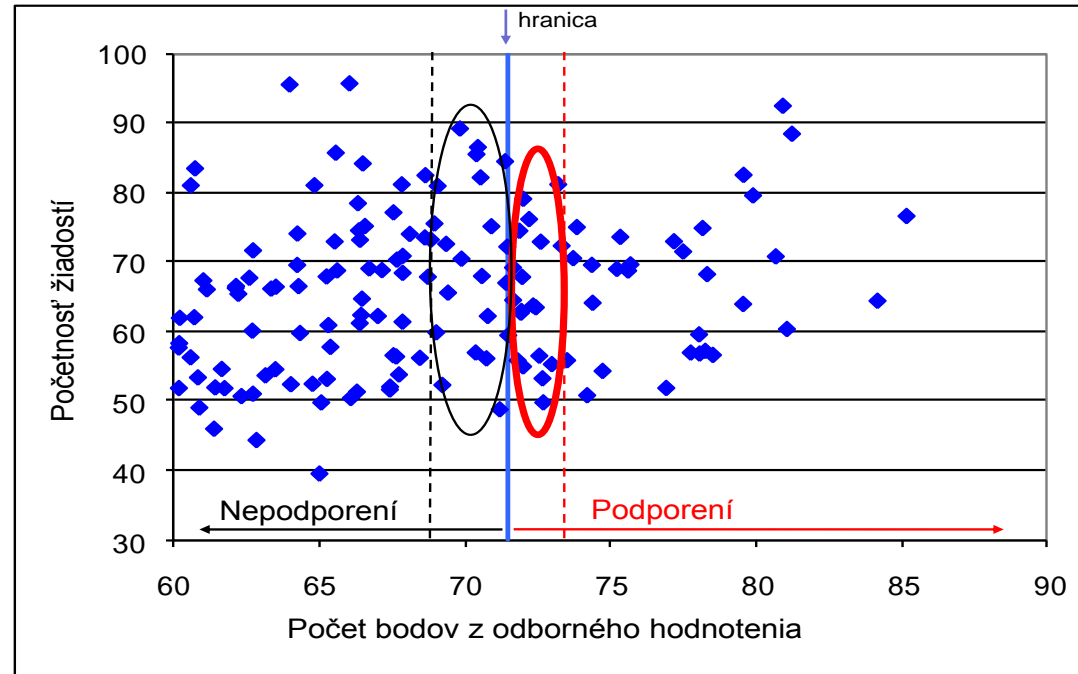
Sharp RDD sa používa ak je stav podpory nespojitou nemennou (deterministickou) funkciou kovariantu x (napr. vek, mzda, dĺžka nezamestnanosti, počet bodov z hodnotenia). Matematicky vyjadrené:

$$D_i = \begin{cases} 0 & \text{if } x_i < S \\ 1 & \text{if } x_i \geq S \end{cases}$$

kde S je **hraničná hodnota** (prah, *threshold*). Mechanizmus oprávnenosti je deterministickou funkciou x , pretože ak poznáme x_i vieme hneď určiť D_i . Podpora je nespojitou funkciou x , pretože bez ohľadu na to, ako blízko je x k S , stav podpory ostane nezmenený až pokiaľ $x = S$...z uvedeného vyplýva že neexistuje hodnota x_i pre ktorú môžeme sledovať oba stavy – podporený aj nepodporený (Angrist a Pischke, 2009).

Hlavnou myšlienkou tejto metódy je, že jednotky v cieľovej populácii pod stanovenou hranicou (ktorí neboli podporení) sú vhodnou porovnávacou skupinou k jednotkám nad touto hranicou (ktorí boli podporení). Účinok intervencie je možné určiť porovnaním priemerných výsledkov podporených žiadateľov nachádzajúcich sa tesne nad hranicou oprávnenosti so zamietnutými žiadateľmi nachádzajúcimi sa tesne pod hranicou oprávnenosti. Oblasť v okolí hranice sa dá za určitých podmienok porovnateľnosti označiť takmer ako náhodný experiment.

Ilustrácia konceptu metódy *Regression discontinuity design*



Z uvedeného vyplýva, že *sharp RDD* zachytáva účinok intervencie iba na podskupine populácie s hodnotami x blízko hranice. Tento účinok sa označuje ako lokálny priemerný účinok podpory (*Local Average Treatment Effect - LATE*). V prípade heterogénnych účinkov intervencie sa môže lokálny priemerný účinok podpory značne líšiť od účinkov podpory na jednotky nachádzajúce sa vo väčšej vzdialenosti od hranice.

Hoci je tento prístup charakterizovaný vysokou internou validitou, jeho výsledky nemôžu byť vždy generalizované na celú cieľovú populáciu.

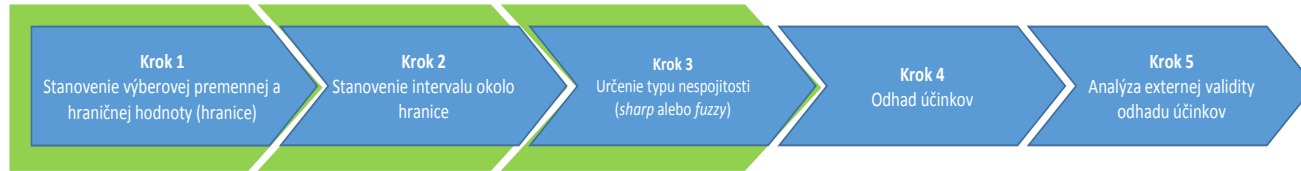
Fuzzy regression discontinuity design

Fuzzy RDD sa používa v prípade neúplného súladu s pravidlami oprávnenosti. V takomto prípade môžeme sledovať podporené jednotky aj ak $x_i < S$ alebo nepodporené jednotky ak $x_i \geq S$ (označované ako jednosmerný nesúlad) alebo zároveň podporené jednotky ak $x_i < S$ a nepodporené jednotky ak $x_i \geq S$ (označované ako obojsmerný nesúlad).

Metóda fuzzy RDD využíva nespojitosti v pravdepodobnosti podpory závislé na kovariante x . Nespojitosť je následne použitá ako inštrumentálna premenná.

Postup výpočtu a kroky aplikácie metódy RDD

Aplikácia metódy regression discontinuity pozostáva z 5 krokov.



1. Stanovenie výberovej premennej a hraničnej hodnoty

Podmienkou aplikácie metódy RDD je existencia spojitej výberovej premennej s hranicou rozdeľujúcou jednotky do podporenej a nepodporenej skupiny.

2. Stanovenie intervalu okolo hranice

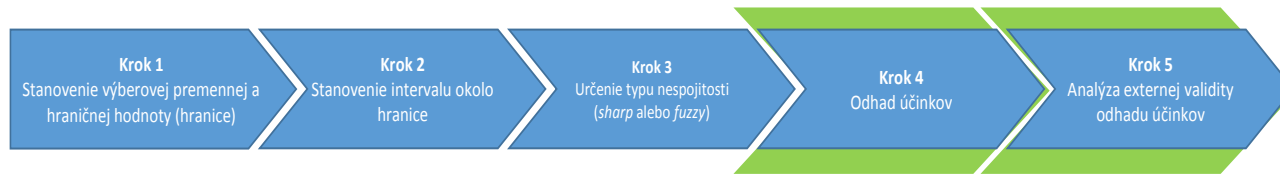
Stanovenie oblasti okolo hraničnej hodnoty závisí na hodnotiteľoch a početnosti pozorovaní. Zvyčajne sa aplikuje princíp viacerých intervalov, napr. veľký interval ($0\% < x < 50\%$), stredný interval ($10\% < x < 40\%$) a malý interval ($15\% < x < 35\%$). V tejto súvislosti je potrebné poznamenať, že čím je interval menší, tým viac sú podmienky podobné kvázi-experimentálnemu postupu. Na druhej strane však platí, že čím je interval menší, tým je menšia aj externá validita výsledkov analýzy, keďže sa porovnáva len časť populácie ktorá sa nachádza v blízkosti danej hranice.

3. Určenie typu nespojitosti (*sharp* alebo *fuzzy*)

Pri sharp regression discontinuity je podpora deterministickou funkciou kovariantu X. Pri fuzzy regression discontinuity sa využívajú nespojitosti v pravdepodobnosti podpory závislé na kovariante x. Nespojitosť je následne použitá ako inštrumentálna premenná.

Postup výpočtu a kroky aplikácie metódy RDD

Aplikácia metódy regression discontinuity pozostáva z 5 krokov.



4. Odhad účinkov

Odhad účinkov sa vykoná využitím regresného modelu. Regresia sa môže vykonať pre všetky zvolené intervaly.

5. Analýza externej validity odhadu účinkov

V prípade heterogénnych účinkov intervencie umožňuje metóda RDD iba identifikáciu priemerného dopadu v oblasti hranice výberu.

V reálnej situácii, pri výskyte heterogénnych účinkov naprieč podporenými jednotkami môže byť lokálny účinok podpory značne odlišný od účinku v oblasti vzdialenejšej od hranice výberu (Battisin a Rettore, 2008). Z tohto dôvodu by sa mali hodnotitelia uistiť, či a za akých podmienok sú výsledky v okolí hranice zovšeobecniteľné na celú populáciu.

Zhrnutie metódy RDD

Regression discontinuity design			
Hlavná hypotéza	Potreba dát	Klady	Zápory
Oprávnenosť účasti v programe je stanovená len na základe pozorovateľných pred-intervenčných premenných a pravdepodobnosť podpory sa nespojito mení ako funkcia týchto premenných.	Potrebné údaje zahŕňajú výberovú premennú a pozorovania oprávnených a neoprávnených jednotiek	V oblasti blízko hranice podpory má RDD charakteristiky náhodného experimentu	Obmedzená externá validita odhadov
	Prierezové dáta (cross-sectional data)	Metóda umožňuje identifikáciu kauzálnych účinkov programu bez nutnosti stanovenia svojvoľných obmedzení vylúčenia, predpokladov výberového procesu alebo distribúcie chýb	Na rozšírenie výsledkov v okolí hranice na celú populáciu je možné využiť iba neexperimentálne odhady, ktorých konzistencia je závislá na behaviorálnych (a netestovateľných) predpokladoch

Zdroj: Loi, Rodrigues (2012)

Instrumental variables

Charakteristika metódy

Metóda inštrumentálnych premenných sa priamo zaoberá **selekciami do programu na základe nepozorovateľných charakteristík**. Priemerný účinok podpory na podporených (ATT_{IV}) sa dá určiť ak je možné identifikovať premennú (inštrument), ktorá ovplyvňuje selekciami do programu ale nemá priamu súvislosť so sledovaným výsledkom (výsledkom záujmu) alebo nepozorovanými premennými ktoré ho určujú. **Inštrument je zdrojom exogénnej variácie, ktorý sa používa s cieľom aproximácie náhodného experimentu** (Blundell a Costa Dias, 2008).

Výber inštrumentu je fundamentálnym krokom aplikácie tejto metódy preto by mal byť dôkladne zvolený a odôvodniteľný ekonomickou intuíciou a teóriou. (Loi a Rodrigues, 2012).

Odhad ATT_{IV} sa vyráta pomocou lineárneho regresného modelu:

$$Y_i = \alpha + \beta D_i + \gamma X_i + \varepsilon_i$$

kde parameter záujmu β je: účinok podpory na výsledkovú premennú, ponechávajúci ostatné vopred stanovené premenné X_i konštantné.

Všeobecne platí, že v nie-experimentálnom prostredí, teda ak podpora nie je poskytovaná náhodne, je prítomné skreslenie pri výbere jednotiek do intervencie, konkrétne hovoríme o výberovom skreslení na základe nepozorovateľných charakteristík. Preto sú premenné, ktoré zároveň ovplyvňujú výsledok aj selekciami do intervencie často nepozorovateľné (napríklad vrodene schopnosti a zručnosti alebo motivácia). V prípade abstrahovania od týchto nepozorovateľných premenných bude priemerný výsledok podpory na podporených (ATT_{IV}) nesprávny (nahodnotený alebo podhodnotený).

Ako príklad môžeme uviesť vzdelávacie kurzy pre nezamestnaných na zlepšenie ich kvalifikácie. Predpokladajme že motivácia, teda premenná ktorú hodnotitelia nevedia sledovať, ovplyvňuje tak pravdepodobnosť jednotlivca prihlásiť sa do programu, ako aj pravdepodobnosť že si v budúcnosti nájde prácu (výsledok záujmu). Pri abstrahovaní od takýchto nepozorovateľných charakteristík by bol ATT nahodnotený, keďže by úloha motivácie pri hľadaní práce bola chybné pripísaná účasti na programe.

Kvôli jednoduchosti budeme uvažovať o prítomnosti iba jednej inštrumentálnej premennej, hoci v skutočnosti môže byť prítomných viacero inštrumentov.

*Metóda inštrumentálnych premenných sa zameriava na očistenie výberu na základe nepozorovateľných premenných použitím premennej (inštrumentu) Z , ktorá spĺňa **podmienky**:*

- I. Ovplyvňuje zapojenie sa do programu (intervencie): $cov(Z, D) \neq 0$,
- II. Nie je súčasťou modelu: $cov(Z, u) = 0$, táto podmienka je označovaná ako „obmedzenie vylúčenia“ (*exclusion restriction*). Podmienka obmedzenia vylúčenia zabraňuje akémukoľvek priamemu účinku inštrumentu na výsledok záujmu a akémukoľvek nepriamemu účinku prostredníctvom inej premennej okrem D .

V súvislosti s vyššie uvedeným príkladom vzdelávacích kurzov pre nezamestnaných musí hodnotiteľ **identifikovať takú pozorovateľnú premennú Z , ktorá:**

- i. má vplyv na rozhodnutie jednotlivcov zúčastniť sa programu,
- ii. nie je priamo spojená s pravdepodobnosťou nájdenia si zamestnania v budúcnosti,
- iii. nie je spojená s individuálnou motiváciou jednotlivcov.

Príkladom takejto premennej by mohla byť napríklad vzdialenosť bydliska uchádzača o zamestnanie od školiaceho strediska (pokiaľ je táto informácia hodnotiteľom známa). Neexistuje žiaden zrejmy dôvod domnievať sa, že táto vzdialenosť je spojená s motiváciou alebo pravdepodobnosťou uchádzača nájsť si zamestnanie. Túto domnienku je však potrebné odborne prediskutovať, vysvetliť a dôkladne overiť v dostupných údajoch.

Metóda inštrumentálnych premenných identifikuje priemerný účinok podpory na podporených využitím iba tej časti variácie stavu podpory (podporený/nepodporený), ktorá súvisí s premennou Z (inštrumentom).

Jedným z najčastejšie využívaných postupov je dvojstupňová metóda najmenších štvorcov (2SLS), v ktorej odhad parametra záujmu prebieha v dvoch fázach.

V prvej fáze sa izoluje časť premennej označujúcej stav podpory „D“, ktorá je nezávislá od nepozorovateľných charakteristík ovplyvňujúcich výsledok " ε_i ". Toto sa vykonáva v prvom stupni regresie, v ktorom je intervenčná premenná vyjadrená inštrumentom Z a ostatné exogénne premenné sú X:

$$D_i = \pi_0 + \pi_1 Z_i + \pi_2 X_i + \varepsilon_i$$

Zodpovedajúce hodnoty regresie \hat{D}_i odrážajú len exogénnu variáciu intervencie.

V druhom stupni sú tieto hodnoty začlenené do rovnice $cov(Z, u) = 0$ ako náhrada D_i . Odhad dvojstupňovej metódy najmenších štvorcov (2SLS) by nemal byť vykonávaný manuálne hodnotiteľom, pretože aj napriek tomu že koeficienty by boli správne, štandardná odchýlka by bola nesprávna. K chybe by došlo z dôvodu nezohľadnenia odhadov \hat{D} v druhom stupni regresie. Odhad by mal byť vykonaný pomocou štatistického softvéru a jeho príkazov (napríklad príkaz „ivregress“ v programe STATA), čo zaručí správny odhad štandardnej odchýlky.

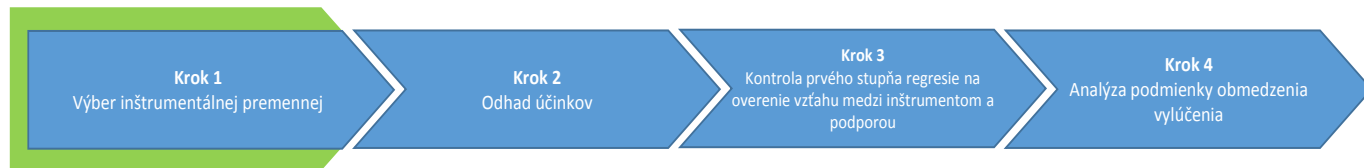
Je potrebné zdôrazniť že odhadovaný priemerný účinok podpory na podporených (*ATT*) bude závisieť na použitej inštrumentálnej premennej, keďže rôzne inštrumenty môžu vyvolať rôzne variácie v intervenčnej premennej.

Ak je vplyv intervencie rovnaký na všetky jednotky populácie, bude výsledkom metódy inštrumentálnych premenných *ATT*.

Ak však intervencia vykazuje heterogénny účinok na populáciu, bude výsledkom metódy IV lokálny účinok intervencie (*LATE - Local Average Treatment Effect*). Tento estimátor (odhadová funkcia) vyjadruje účinok podpory iba na jednotky, ktoré pri zmene inštrumentu zmenia svoj stav z nepodporených na podporených, ale neidentifikuje účinok podpory na jednotky, ktoré by boli podporené bez ohľadu na zmeny inštrumentu (Imbens a Angrist 1994). Aj keď je táto skutočnosť zvyčajne interpretovaná ako nevýhoda metódy inštrumentálnych premenných, odhadovaný parameter môže byť predmetom záujmu tvorcov politiky. Ak je inštrumentom diskretná premenná, napríklad zmena politiky, výsledkom *LATE* bude účinok intervencie na jednotky meniace svoj stav podpory v závislosti od zmeny politiky, čo môže byť významným ukazovateľom dopadu politiky (Blundell a Costa Dias, 2008).

Postup výpočtu a kroky aplikácie metódy IV

Aplikácia metódy inštrumentálnych premenných pozostáva zo štyroch krokov. V prvom kroku sa identifikuje vhodná premenná (inštrument), ktorá ovplyvňuje účasť v programe ale nemá priamy vplyv na sledovaný výsledok. V druhom kroku sa využije dvojstupňová metóda najmenších štvorcov (2SLS) na odhad premennej záujmu. Následne sa analýzou prvého stupňa regresie otestuje predpoklad vzťahu medzi inštrumentom a stavom podpory. V štvrtom kroku sa analyzuje platnosť podmienky obmedzenia vylúčenia, teda nezávislosti medzi inštrumentom a sledovanou výsledkovou premennou.



1. Výber inštrumentálnej premennej

Voľba inštrumentálnej premennej je kľúčovým krokom pre zabezpečenie odhadu kauzálneho účinku intervencie. Vo všeobecnosti možno povedať, že poznanie detailných informáciách o politike (programe), jej celení a implementácii môže odhaliť zdroje exogénnych rozdielov (variácií), ktoré by mohli byť použité ako inštrumenty.

Zvyčajné zdroje inštrumentov sú napríklad:

i. Politické a geografické rozdiely

Napríklad, ak sa z určitých dôvodov hodnotená politika realizuje len v niektorých regiónoch, vplyvu danej intervencie je vystavená len časť populácie krajiny.

i. Vonkajšie šoky ovplyvňujúce načasovanie politiky

Napríklad, ak sa kvôli vonkajším faktorom oneskorila realizácia politiky v jednom regióne (oblasti) alebo pre určitú skupinu obyvateľstva.

i. Pravidlá oprávnenosti podpory

Ak je politika navrhnutá spôsobom, že určité jednotky populácie sú oprávnené zúčastniť sa a iné nie.

Postup výpočtu a kroky aplikácie metódy IV



2. Odhad účinkov

Na aplikáciu dvojstupňovej metódy najmenších štvorcov sa zvyčajne využívajú špeciálne štatistické softvéry (napr. STATA, Eviews, SAS). Odhadovaná premenná záujmu ATT_{IV} sa matematicky vyjadrí:

$$ATT_{IV} = \frac{cov(Y, Z)}{cov(D, Z)} = ATT + \frac{cov(\varepsilon, Z)}{cov(D, Z)}$$

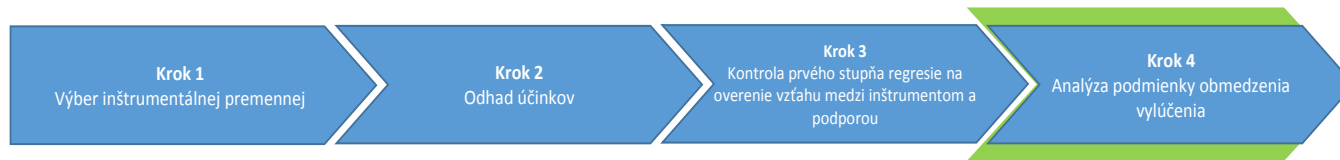
Táto rovnica zdôrazňuje kľúčovú úlohu dvoch vyššie uvedených predpokladov. Ak jeden z nich nie je splnený, odhadovaný ATT_{IV} sa bude líšiť od skutočného. Preto je dôležité, aby hodnotiteľ správne popísal spoľahlivosť týchto dvoch predpokladov.

3. Kontrola prvého stupňa regresie na overenie vzťahu medzi inštrumentom a podporou

Prvý predpoklad – súvislosť inštrumentálnej premennej s podporou je možné otestovať analýzou prvého stupňa regresie a posúdením sily vzťahu medzi intervenciou a inštrumentálnymi premennými. Ak je inštrument slabý na predpovedanie podpory, správnosť odhadu ATT môže byť narušená.

Zaužívaným pravidlom je, že F-štatistika, ktorá je spojená s inštrumentálnymi premennými by mala byť vyššia ako 10 (Stock a Yogo, 2005). V prípade slabých inštrumentálnych premenných môže byť skreslenie odhadovaného ATT väčšie, ako by bolo pri abstrahovaní od výberu založenom na nepozorovateľných charakteristikách.

Postup výpočtu a kroky aplikácie metódy IV



4. Analýza podmienky obmedzenia vylúčenia

Aj napriek nemožnosti priameho testovania predpokladu nezávislosti inštrumentálnej premennej a výrazu chyby, hodnotiteľ by mal dôkladne zvážiť a odôvodniť, prečo sa predpokladá, že inštrumentálna premenná a výraz chyby nie sú korelované.

V súvislosti s príkladom vyššie, v ktorom bola ako inštrument navrhnutá vzdialenosť bydliska uchádzača o zamestnanie od školiaceho strediska by sme mohli argumentovať, že dotýčny sa mohol presťahovať do miesta, kde sú vzdelávacie programy ponúkané častejšie. To by však narušilo predpoklad nezávislosti motivácie od vzdialenosti.

Ďalším príkladom môže byť systematické ponúkajúce vzdelávacích a tréningových programov v ekonomicky rozvinutejších oblastiach, kde je vyššia pravdepodobnosť nájdenia si práce. Ak by toto bolo prípadom, musel by existovať vzťah medzi vzdialenosťou a výsledkovou premennou.

Takéto dohady by mali byť dôkladne zvážené a vysvetlené a pomocou dostupných dát alebo iných dôkazov by mala byť potvrdená alebo zamietnutá ich platnosť.

Zhrnutie metódy IV

Instrumental variables			
Hlavná hypotéza	Potreba dát	Klady	Zápory
Inštrumentálna premenná Z musí spĺňať dve podmienky: 1. Oplyvňovať selekciu do programu, 2. Byť nezávislá na nepozorovateľných charakteristikách a nie priamo súvisiaca s výsledkom záujmu	Prierezoové dáta (cross-sectional data), časové rady údajov alebo opakované prierezoové dáta	Rieši priamo výber založený na nepozorovateľných charakteristikách	Je veľmi náročné nájsť vhodný inštrument, ktorý by spĺňal dve kľúčové podmienky
	Prierezoové dáta (cross-sectional data)	Ak je zvolený vhodný inštrument, je garantovaná kauzalita odhadovaného účinku podpory	V prípade heterogénnych účinkov podpory je výsledkom metódy IV lokálny priemerný účinok (LATE) a nie priemerný účinok podpory na podporených (ATT)
			Inštrument môže mať nedostatočnú variáciu pri selekcii do programu – slabý inštrument

Zdroj: Loi, Rodrigues (2012)

**Otázky požiadaviek
na dáta pre
realizáciu
hodnotení
dopadov**

Hoci sa otázky zberu dát, akými sú dostupnosť a kvalita môžu javiť ako bežné operatívne problémy ktoré je potrebné riešiť len na technickej úrovni, nemalo by sa zabúdať, že tieto aspekty majú kritický význam pre každé hodnotenie dopadov (a vo všeobecnosti akékoľvek hodnotenie). Dáta sú potrebné z dôvodu možnosti testovania zmien závislých premenných, resp. na určenie kontrafaktuálneho odhadu (estimácie), teda situácie ktorá by bola nastala ak by program/podpora nebola realizovaná.

Pre kontrafaktuálne hodnotenia dopadov sú potrebné dáta pre tri základné skupiny:

- I. Podporená skupina** (*treatment group*) – podporené jednotky (napr. jednotlivci, podniky)
- II. Porovnávací skupina** (*comparison group*) – skupina nepodporených jednotiek, ktoré sú vystavené rovnakým podmienkam a vplyvom ako podporené jednotky, okrem samotného pridelenia podpory. Jednotky porovnávací skupiny sú svojimi charakteristikami podobné podporenej skupine. Vhodnou porovnávacou skupinou bývajú zamietnutí žiadatelia. V literatúre sa môžeme stretnúť aj s pojmom kontrolná skupina (*control group*). Kontrolná skupina sa vo väčšine prípadov používa v kontexte náhodných experimentov (*randomised experiments*) a porovnávací skupina pri kontrafaktuálnych hodnoteniach dopadov. V praxi sa však tieto pojmy často zamieňajú a chápu sa ako synonymá.
- III. Kontrolná vzorka** (*control sample*) – skupina nepodporených jednotiek, ktorá nemusí byť nutne vystavená rovnakým podmienkam ako porovnávací skupina. Používa sa na kontrolu vývoja sledovaných premenných a zachytenie všeobecného trendu vývoja v ekonomike.

Pre uvedené skupiny sú potrebné tieto typy údajov:

i. Identifikačné údaje a charakteristika jednotiek danej skupiny

Pri jednotlivcoch to býva pohlavie, vek, vzdelanie, rodinný stav, kvalifikácia, bydlisko, dĺžka praxe (zamestnania), výška príjmu, dĺžka nezamestnanosti, jazykové znalosti, pc zručnosti, doplnkové vzdelávanie a rozšírenie kvalifikácie a iné, v závislosti od typu hodnotenia a sledovaných účinkov.

Pri podnikoch sa zvyčajne sleduje dátum založenia, právna forma, miesto sídla, miesto realizácie podnikateľskej činnosti, odvetvie, počet zamestnancov, obrat, výška príjmov, pridaná hodnota, národná/medzinárodná pôsobnosť, exportná/importná orientácia a iné, v závislosti od typu hodnotenia a sledovaných účinkov.

ii. Výsledkové údaje

Výsledkové údaje (hodnoty sledovaných premenných, na ktoré má intervencia pôsobiť). Výsledkové údaje je potrebné sledovať za určité časové obdobie pred a po realizácii intervencie, aby bolo možné vyhodnotiť ich vývoj a zmeny ktoré nastali/nenastali v dôsledku podpory.

iii. Kontrolné údaje

Kontrolné údaje sú potrebné z dôvodu možnosti sledovania a kontroly rozdielov medzi podporenou a nepodporenou skupinou počas analýzy. Je užitočné disponovať čo možno najväčším množstvom údajov vzťahujúcich sa tak k voľbe zúčastniť sa programu, ako aj potenciálnych výsledkov, teda výsledkových ukazovateľov ktoré sú k dispozícii aj pre pred-intervenčné obdobie.

Potrebné údaje je vhodné získavať z jedného zdroja a je potrebné zabezpečiť aby boli merané a zaznamenávané rovnakým spôsobom. V praxi sa však potrebné údaje získavajú z rôznych zdrojov (subjektov, inštitúcií).

V kontexte hodnotenia intervencií verejných politík môžu byť údaje získavané z databáz:

- Riadiacich orgánov a Sprostredkovateľských orgánov pod riadiacimi orgánmi,
- Štatistického úradu SR,
- Štatistického úradu EÚ (EUROSTAT),
- Sociálnej poisťovne SR,
- Finančnej správy SR,
- Ústredia práce, sociálnych vecí a rodiny SR,
- Národného cenzu a registrov,
- iných relevantných databáz.
- Z medzinárodných zdrojov sa obvykle využívajú databázy:
 - OECD Statistics,
 - European company data (portál Amadeus),
 - OSN (UNSD – United Nations Statistical Database).

Otázky ochrany osobných údajov

Pri získavaní údajov o jednotlivcoch alebo spoločnostiach sa hodnotitelia môžu stretnúť s problémami ochrany osobných a citlivých údajov. Riešením pri problémoch s poskytnutím takýchto informácií je ich anonymizácia poskytujúcou inštitúciou. Zaobchádzanie a spracovanie takýchto údajov je v rámci EÚ riešené Smernicou Európskeho parlamentu a Rady 95/46/EC z 24. októbra 1995 o ochrane jednotlivcov pri spracovaní osobných údajov a voľnom pohybe týchto údajov. Na národnej úrovni rieši otázky ochrany osobných údajov Zákon č. 122/2013 Z. z. o ochrane osobných údajov a o zmene a doplnení niektorých zákonov, ako vyplýva zo zmien a doplnení vykonaných zákonom č. 84/2014 Z. z.

Správne načasovanie zberu dát

V ideálnom prípade by mali byť hodnotenia dopadov založené na údajoch pokrývajúcich pred- a po-intervenčné obdobie. Dôležitým aspektom je správne posúdenie reprezentatívnosti východiskového a konečného roku (obdobia). Ak je niektorý z týchto dvoch hraničných časových bodov nereprezentatívny (vymykajúci sa z priemeru), bude sledovaná zmena počas tohto obdobia skreslená. Ak je napríklad východiskový rok ovplyvnený určitou špecifickou udalosťou (prírodná katastrofa, politický vývoj, zmeny ekonomického cyklu), ktorá spôsobí neprimerane vysoké alebo nízke hodnoty sledovaných veličín, potom budú zmeny počas sledovaného obdobia značne skreslené. Príkladom takejto situácie môže byť hodnotenie dopadov intervencií v sektore poľnohospodárstva, kedy sa za východiskový bod zvolí rok, ktorý nie je reprezentatívnym príkladom bežného vývoja v sektore (podpriemerne nízka úroda kvôli povodňam alebo naopak nedostatku zrážok a extrémnemu suchu). Rovnako to platí aj v prípade voľby po-implimentačného obdobia. Vo väčšine prípadov je stanovenie sledovaného časového rámca determinované načasovaním intervencie resp. jej hodnotenia. Hodnotitelia musia dôkladne zvážiť reprezentatívnosť údajov východiskového a konečného obdobia hodnotenej intervencie pred formuláciou akýchkoľvek záverov.

Nemenej dôležitým aspektom každého hodnotenia dopadov je správne posúdenie časového rámca očakávaných výsledkov intervencie, teda rozlíšenie krátkodobých a dlhodobých účinkov. V závislosti od druhu a kontextu intervencie sa v období zberu a analýzy dát niektoré účinky ešte nemuseli stihnúť prejaviť a iné už mohli naopak odznieť. Uvedené je potrebné zohľadniť pri formulácii záverov hodnotenia.

Dostupnosť dát

Hodnotenie dopadov sa v praxi začína posúdením existujúcich údajov, ktoré mohli byť zachytené pred realizáciou intervencie na vstupe, počas realizačnej fázy alebo na výstupe (ako výsledky). Takýto prehľad dostupných údajov je užitočný z nasledujúcich dôvodov:

- i. Dostupné údaje sú užitočnou pomôckou pri rekonštrukcii intervenčnej teórie, čo môže podnietiť ďalší zber a získavanie potrebných údajov,
- ii. Informácie o dostupnosti údajov majú vplyv na výber metód a návrh dizajnu hodnotenia, resp. môžu viesť k ich ďalšiemu spracovaniu a analýze (napr. ex ante a ex post databázy môžu byť doplnené ďalšími zdrojmi údajov na vytvorenie vhodnej kontrolnej skupiny, alebo iniciovať ich dodatočné získavanie),
- iii. Dostupné údaje z rôznych zdrojov umožňujú trianguláciu výsledkov.

Ďalšími zdrojmi údajov pre hodnotenia môžu byť:

- Údaje Štatistického úradu,
- Údaje zo sčítania obyvateľstva,
- Rôzne špecializované prieskumy,
- Administratívne údaje zbierané ministerstvami a inštitúciami štátnej správy,
- Štúdie realizované mimovládnyimi neziskovými organizáciami a univerzitami.

Užitočný prístup k posúdeniu dostupnosti dát môže byť charakterizovaný nasledujúcimi krokmi:

1. Tvorba zoznamu dostupných dát a posúdenie ich kvality. Niekedy môžu byť na vykonanie kompletného hodnotenia dopadov postačujúce aj údaje získané zo sekundárnych zdrojov (väčšinou pri národných resp. sektorových intervenciách). Väčšinou sú však sekundárne údaje využívané ako doplnkový zdroj informácií,

2. Analýza potreby dodatočných údajov z pohľadu intervenčnej teórie. Proces zberu dát musí byť založený na dizajne hodnotenia, ktorý je naopak z časti založený na intervenčnej teórii,
3. Porovnávací skupina by mala mať adekvátnu veľkosť a mala by byť predmetom rovnakého spôsobu zberu dát ako podporená skupina. Veľkosť porovnávací skupiny nie je niekde presne stanovená, závisí od veľkosti podporenej skupiny a dostupnosti dát. Z praxe a skúseností z iných hodnotení býva jej veľkosť niekoľkonásobne väčšia ako veľkosť podporenej skupiny, aby bola dosiahnutá istá štatistická významnosť.
4. Je potrebné skontrolovať či iné intervencie alebo neočakávané udalosti a procesy neovplyvnili porovnávaciu alebo podporenú skupinu (tzn. skontrolovať či je porovnávací skupina ovplyvňovaná inými procesmi/udalosťami ako podporená skupina),
5. Východiskové údaje musia pokrývať relevantné indikátory vývoja/pokroku, ako aj jeho hlavné determinanty, aby bolo v neskorších fázach možné určiť, či mali okrem intervencie vplyv na vývoj v sledovanom období aj iné faktory.

V prípade že za východiskové obdobie neexistujú dáta, je jednou z možností ich dodatočného získania dopytovanie sa (výskum) priamo v teréne. Hoci je tento prístup často kritizovaný, je taktiež relevantným zdrojom údajov a je na schopnostiach hodnotiteľov posúdiť ich váhu a kredibilitu.

Kvalita dát

Úspech každého hodnotenia závisí od kvality dostupných údajov. Na riešenie tohto problému sa využíva triangulácia a kombinácia metód. Ani tieto stratégie však samé o sebe nestačia na kontrolu kvality dát, ktorá je potrebná na zaistenie neskreslených výsledkov. Hodnotitelia by si mali položiť nasledovné otázky:

Aké techniky by bolo vhodné použiť na zlepšenie kvality dát (resp. zberu dát)?

Ako riešiť chýbajúce údaje (chýbajúce pozorovania v databázach, chýbajúce premenné)?

Ako riešiť chyby merania – Zodpovedá hodnota premennej alebo odpovedi na otázku jej reálnej hodnoty?

Ako riešiť chyby špecifikácie – Zodpovedá dopytovaná otázka alebo meraná premenná oblasti ktorú sa ňou snažíme pokryť?

Dovoľuje kvalita dát vykonať špecifické štatistické analýzy?

Nové postupy dátovej analýzy a rozšírenie kvázi-experimentálnych metód sú pri hodnotení dopadov veľmi perspektívne, hoci kvalita dát je často obmedzujúcim faktorom pri formulácii výsledkov a záverov (Deaton, 2005).

V prípade sekundárnych údajov je poznanie ich zdroja a procesu získavania faktorom, ktorý môže podporiť alebo oslabiť validitu zistení.

Riešenie obmedzení dostupnosti dát

Podľa Bamberger et al., finančné prostriedky na výkon hodnotení bývajú zriedka zahrnuté do pôvodného rozpočtu projektu/programu a preto musia byť hodnotenia vykonané s omnoho menším rozpočtom než by bol za normálnych okolností alokovaný na takéto hodnotenie. Dôsledkom toho nie je možné aplikovať želané spôsoby zberu dát (napr. retrospektívne analýzy, výskumy na vzorke), alebo aplikovať metódy na rekonštrukciu východiskových údajov a tvorbu kontrolných skupín. Problémy s dostupnosťou a kvalitou dát bývajú determinované časovými a rozpočtovými obmedzeniami.

V tabuľke na nasledujúcej strane sú uvedené rôzne scenáre postupov s obmedzeniami času, rozpočtu a dát.

Hodnotiace scenáre s časovými, dátovými a rozpočtovými obmedzeniami

Obmedzenie, s ktorým musí byť hodnotenie vykonané			Príklady scenárov
Čas	Rozpočet	Dáta	
x			Hodnotenie sa iniciuje v neskoršej fáze projektu/programu a je potrebné aby bolo ukončené do určitého dátumu, aby jeho zistenia a závery mohli byť použité v následnom rozhodovacom procese, nastavovaní obdobného programu resp. pre účely vykazovania pokroku. Aj v prípade adekvátneho nastavenia rozpočtu môžu nastať ťažkosti so zberom a analýzou dát z rôznych zdrojov z dôvodu časového obmedzenia trvania hodnotenia.
	x		Na hodnotenie sú alokované iba obmedzené finančné zdroje z rozpočtu. Aj v prípade dostatočného časového rámca na výkon hodnotenia môže byť z dôvodu nízkeho rozpočtu ťažké zabezpečiť všetky potrebné údaje.
		x	Hodnotitelia sú prizvaní až v čase kedy je už projekt/program v pokročilej fáze svojej realizácie. Z tohto dôvodu nemohla byť vykonaná analýza východiskového stavu podporenej alebo porovnávacej skupiny v potrebnom rozsahu. Takáto situácia je nepríjemná pri hodnoteniach intervencií v oblastiach, v ktorých sa systematicky nezberajú údaje, resp. je obtiažne ich merať (napr. domáce násilie, korupcia, otázky rodovej rovnosti a pod.).
x	x		Hodnotitelia musia pracovať pod časovým tlakom a s obmedzeným rozpočtom. Dáta získané zo sekundárnych zdrojov môžu byť k dispozícii, ale hodnotitelia majú málo času a zdrojov na ich analýzu.
x		x	Hodnotitelia majú málo času a obmedzený prístup k východiskovým údajom alebo kontrolnej skupine. Sú k dispozícii zdroje na zber dodatočných údajov potrebných k hodnoteniu, ale spôsob a metódy ich zberu sú obmedzené nedostatkom času.
	x	x	Hodnotitelia sú prizvaní v pokročilej fáze projektu/programu s obmedzeným prístupom k východiskovým údajom alebo kontrolnej skupine. Limitujúcim faktorom v tomto prípade nie je čas ale finančné možnosti.
x	x	x	Hodnotitelia sú prizvaní neskoro, majú obmedzený rozpočet, obmedzený alebo žiadny prístup k východiskovým údajom a nebola identifikovaná žiadna kontrolná skupina.

Zdroj: Bamberger et al. (2004)

Kľúčové body a odporúčania

- So zberom dát je potrebné začať čo najskôr,
- Spoľahlivé údaje za východiskové obdobie sú nevyhnutným predpokladom na porozumenie a odhad (estimáciu) dopadov,
- V závislosti od druhu intervencie si zber dát za východiskové obdobie, ako aj nastavenie iných aspektov hodnotenia vyžaduje efektívnu spoluprácu medzi hodnotiteľmi dopadov a implementátorom intervencie,
- Tvorcovia politiky by mali zapojiť odborníkov na hodnotenia dopadov v čo najskoršej fáze návrhu intervencie, aby boli schopní navrhnúť čo možno najkvalitnejšie hodnotenie,
- Zaistenie kvalitného zberu dát by malo byť neoddeliteľnou súčasťou každého hodnotenia dopadov,
- Pri práci s údajmi zo sekundárnych zdrojov môže byť nedostatok informácií o kvalite zberu týchto dát faktorom obmedzujúcim možnosti ich analýzy a validitu výsledkov,
- Je potrebné byť si vedomý a zvoliť efektívny prístup riešenia obmedzení prítomných pri hodnotení dopadov (čas, dáta, zdroje).



cutting through complexity



Tento projekt bol spolufinancovaný Európskou úniou. Investícia do Vašej budúcnosti.

© 2015 KPMG Slovensko spol. s r.o., the Slovak member firm of KPMG International Cooperative ("KPMG International"), a Swiss entity. All rights reserved.

The KPMG name, logo and "cutting through complexity" are registered trademarks or trademarks of KPMG International.

10/02