## Příloha 8: Metodika kontrafaktuální analýzy

### Metodika realizace kontrafaktuální analýzy pro potřeby evaluační otázky 4

**Párování pomocí propensity skóre**

Pro konstrukci kontrolního vzorku jsme pro PP Technologie použili tři postupy:

1. Kontrolní skupina byla konstruovaná z nepodpořených subjektů, které žádaly o podporu.
2. Kontrolní skupina byla konstruovaná výhradně ze subjektů z vnější databáze (tj. nepodpořené subjekty žádající o podporu v této skupině nefigurují).
3. Kontrolní skupina byla vytvořena ze všech dostupných nepodpořených subjektů.

Tento postup byl zvolen z toho důvodu, jelikož subjekty žádající o podporu mohou být v nějakém smyslu odlišné od ostatních (nežádajících).

Pro PP Marketing jsme kontrolní vzorek konstruovali pouze přístupy 2. a 3., jelikož nepodpořených žádajících subjektů bylo méně než podpořených.

Spárování dat podpořených subjektů s kontrolními proběhlo na základě odhadnutého prospensity skóre (dále PS). Toto skóre je odhadnuto jako tzv. logit pravděpodobnosti zahrnutí do podpory pro subjekt s danou sadou vysvětlujících proměnných.

kde *Yi*  je identifikátor toho, zda je daný subjekt *i* zahrnut do podpory (tj. *Yi=1*) nebo není (*Yi=0*), ***Xi***je vector vysvětlujících proměnných *(Xi1, … , Xik) a*  jsou neznámé parametry modelu, které jsou odhadnuty metodou maximální věrohodnosti.

Do skupiny vysvětlujících proměnných ***Xi***byly zahrnuty následující proměnné charakterizující daný subjekt:

* Kraj, velikost obce
* Kód CZ NACE v modifikované formě (viz níže)
* Počet zaměstnanců, počet let existence
* Ekonomické ukazatele: celková aktiva, oběžná aktiva, vlastní kapitál, výkony, tržby z prodeje výrobků a služeb, hospodářský výsledek

Dvoumístný CZ NACE kód bohužel nemohl být využit přímo (nedostatek kontrolních subjektů v některých kategoriích), a proto bylo uvažováno jak kódování pomocí písmen, tak jednomístný CZ NACE. Výsledky párování byly uspokojivější pro druhou zvolenou variantu. Prezentované výsledky tedy odpovídají výsledkům párování pro model s jednomístným CZ NACE kódem.

Dále pak byla v modelu pro výpočtu PS zahrnuta interakce mezi počtem zaměstnanců a celkovými aktivy. Pro SC 2.2 navíc bylo vhodné zahrnout i interakci kategoriální veličiny určující kraj a celkových aktiv. V případě SC 2.1 nebylo zapojení této interakce vhodné.

Pro párování byl zvolen algoritmus nejbližšího souseda (nearest neighbor matching), kdy je jeden subjekt z podpořené skupiny spárován s jedním subjektem z kontrolní skupiny s nejbližší hodnotou propensity skóre. Pro projekty z roku 2015 jsme vybírali výhradně ze subjektů, pro které byla k dispozici data pro tento rok, podobně pak pro rok 2016. Jiné algoritmy párování byly také zváženy, ale vedly k horším výsledkům.

Po provedení párování byla provedena kontrola kvality párování. Na základě dvouvýběrového t-testu bylo otestováno, že se hladiny jednotlivých proměnných v uvažovaných dvou skupinách (podpoření a kontrolní) neliší. Pro kategoriální proměnné byl uvažován chí kvadrát test nezávislosti (dovoloval-li to rozsah dat). V případě, že párování nebylo uspokojivé, byl použit složitější model pro odhad PS. Výsledné párování bylo vyhodnoceno jako úspěšné.

**Metoda rozdílu v rozdílech**

Pro vyhodnocení výsledného efektu byla použita metodika rozdílu v rozdílech (tzv. Differences in differences DiD). Tento efekt lze (při uvažování podpořených subjektů pouze z jednoho roku) spočíst jako rozdíl rozdílu průměrných hodnot mezi podpořenými a nepodpořenými ve výsledném roce a rozdílu před započetím projektu. Zajímá-li nás DiD efekt veličiny Z, pak jej lze spočíst jako

kde je průměrná hodnota Z pro podpořené subjekty v čase T2 (výsledný rok), je průměr Z pro kontrolní subjekty v čase T2, je průmer Z pro podpořené v čase T1 a analogicky pro .

Graficky je tento efekt znázorněn na obrázku 1. Bod P1 odpovídá průměrné hodnotě ukazatele Z pro podpořené projekty v čase T1, bod P2 je hodnota ukazatele pro podpořené v čase T2. Podobně, K1 je průměrná hodnota Z pro kontrolní subjekty v čase T1 a K2 je hodnota téhož ukazatele pro tutéž skupinu v čase T2. Bod X odpovídá situaci, kdy by byl časový vývoj v obou skupinách (podpořené a kontrolní) stejný (tj. lineární vývoj se stejnou směrnicí). Odpovídá tedy hypotetické situaci, kdy by podpořené projekty podporu nedostaly a časový vývoj ukazatele v této skupině by byl stejný jako v kontrolní skupině. Efekt podpory DiD je pak znázorněn v grafu červenou barvou.

*Obrázek 1: Grafická ilustrace DiD efektu*



Obecně (a především pak v případě, že kombinujeme data pro subjekty s přidělením podpory v různých letech), je vhodnější použít metodu lineárního regresního modelu, která nám poskytne vedle odhadu DiD efektu i odhad rozdílu ukazatelů před započetím podpory a dále také statistickou významnost jednotlivých rozdílů.

**Výsledky párování pomocí techniky PSM**

Pro párování projektů PP Technologie bylo k dispozici 897 subjektů z externí databáze pro rok 2015 a 584 subjektů pro rok 2016. Dále byly k dispozici data o nepodpořených subjektech žádajících o podporu, a to 254 pro rok 2015 a 360 pro rok 2016. Soubor všech možných kontrolních subjektů byl tedy o rozsahu 1151 pro rok 2015 a 944 pro rok 2016.

Pro PP Marketing bylo k dispozici 921 subjektů z externí databáze s daty z roku 2015 a 575 subjektů s daty z roku 2016. Po zahrnutí i nepodpořených subjektů žádajících o podporu bylo k dispozici pro párování 946 subjektů pro rok 2015 a 588 pro rok 2016.

Výsledné modely zvolené pro výpočet PS pro párování pro jednotlivé programy podpory jsou uvedeny výše v metodické části. Připomeňme, že pro PP Technologie byly uvažovány tři různé párování (s výslednými třemi různými kontrolními vzorky), pro PP Marketing byla provedena dvě různá párování.

Grafy na obrázcích2 až 7 zobrazují rozložení PS pro jednotlivá výsledná párování pro PP Technologie. Grafy 2, 4 a 6 zobrazují PS pro nespárované podpořené (žádné nejsou), spárované podpořené, spárované kontrolní a nespárované kontrolní. Grafy 3, 5 a 7 pak zobrazují příslušné histogramy. Ze všech grafů je patrné, že výsledná párování byla úspěšná.

Tabulka 1 uvádí p-hodnoty dvouvýběrových t-testů (Welchův test), které porovnávají střední hodnotu jednotlivých veličin pro podpořené a kontrolní subjekty. Jelikož jsou všechny p-hodnoty vyšší než testovací hladina 0.05, je patrné, že úroveň všech veličin se statisticky významně neliší.

Tabulka 1: P-hodnoty dvouvýběrových t-testů (Welchův test) porovnávající střední hodnoty jednotlivých veličin v podpořeném a kontrolním vzorku, PP Technologie

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Párování | Velikost obce | Počet zam | Počet let existence | Celková aktiva | Oběžná aktiva | Kapitál | Výkony | Tržby | Hosp. výsledek | PS |
|  1. | 0,55 | 0,80 | 0,09 | 0,73 | 0,57 | 0,87 | 0,50 | 0,48 | 0,15 | 0,15 |
| 2. | 0,14 | 0,82 | 0,68 | 0,79 | 0,65 | 0,82 | 0,55 | 0,56 | 0,99 | 0,21 |
| 3. | 0,85 | 0,85 | 0,85 | 0,81 | 0,44 | 0,87 | 0,83 | 0,84 | 0,78 | 0,88 |

*Obrázek 2: Rozdělení PS pro párování 1, PP Technologie*



*Obrázek 3: Rozdělení PS pro párování 1, PP Technologie*



*Obrázek 4: Rozdělení PS pro párování 2, PP Technologie*



*Obrázek 5: Rozdělení PS pro párování 2, PP Technologie*



*Obrázek 6: PS pro párování 3, PP Technologie*



*Obrázek 7: Rozdělení PS pro párování 3, PP Technologie*



Pro program PP Marketing jsme provedli dvě možná párování (popsaná výše, označené jako 2 a 3). Výsledkem jsou dva kontrolní vzorky. Výsledky těchto párování (rozložení PS) jsou zobrazeny v grafech 8 - 11, p-hodnoty t-testů jsou uvedeny v Tabulce 2. Na základě všech těchto výsledků lze opět konstatovat, že výsledky párování jsou uspokojivé.

*Tabulka 2:* *P-hodnoty dvouvýběrových t-testů porovnávajících střední hodnoty jednotlivých veličin, PP Marketing*

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Párování | Velikost obce | Počet zam | Počet let existence | Celková aktiva | Oběžná aktiva | Kapitál | Výkony | Tržby  | Hosp. výsledek | PS |
| 2 | 0,69 | 0,08 | 0,21 | 0,67 | 0,76 | 0,76 | 0,40 | 0,36 | 0,66 | 0,73 |
| 3 | 0,65 | 0,11 | 0,34 | 0,93 | 0,43 | 0,60 | 0,74 | 0,71 | 0,93 | 0,85 |

*Obrázek 8: Rozložení PS pro párování 2 (výhradně externí subjekty), PP Marketing*



*Obrázek 9: Rozložení PS pro párování 2 (výhradně externí subjekty), PP Marketing*



*Obrázek 10: Rozložení PS pro porovnání 3 (všechny subjekty), PP Marketing*



*Obrázek 11: Rozložení PS pro porovnání 3 (všechny subjekty), PP Marketing*



### Metodika realizace kontrafaktuální analýzy pro potřeby evaluační otázky 5

Metodika realizace kontrafaktuální analýzy zaměstnanosti je totožná s metodikou uvedenou v evaluační otázce 4.

Pro proměnnou počtu zaměstnanců bohužel nebyla dostupná data z roku 2017. Tudíž, veškeré výsledky pro tuto proměnnou by mohly být založené pouze na projektech započatých v roce 2015.

V SC 2.1, kterého se analýza vývoje počtu zaměstnanců týká primárně, je ale velmi nízký počet subjektů, u kterých byla realizace zahájena v roce 2015 a zároveň mohou vstoupit do analýzy (jsou ukončené a jsou k nim dostupná data v externí databázi). Z tohoto důvodu byla pro zaměstnanost doplňkově analyzována kategoriální veličina udávající počet zaměstnanců. Tato data byla dostupná pro roky 2014, 2015, 2016 a 2018 (pro rok 2018 byla data stažena z databáze ARES, jde tedy o aktuální zařazení do kategorie počtu zaměstnanců). Na základě regresního modelu byl odhadnut DiD efekt odpovídající jednomu roku a z důvodu lepšího porovnání s ostatními výsledky byl tento efekt přepočten na dva roky (a tento údaj je uváděn v tabulkách).

Je ovšem nutné upozornit, že jde pouze o jakýsi proxy indikátor, který může napovědět, jakým způsobem se vyvíjí zaměstnanost v podpořených podnicích a u kontrolní skupiny. Analýza v zásadě poskytuje odpověď na to, zda u podpořené skupiny dochází častěji k přesunům podpořených subjektů do vyšší kategorie počtu zaměstnanců, než tomu je u kontrolní skupiny.

*Tabulka 3: Kódování kategoriální zaměstnanosti*

|  |  |
| --- | --- |
| 1 | 1 - 5 zaměstnanců |
| 2 | 6 - 9 zaměstnanců |
| 3 | 10 - 19 zaměstnanců |
| 4 | 20 - 24 zaměstnanci |
| 5 | 25 - 49 zaměstnanců |
| 6 | 50 - 99 zaměstnanců |
| 7 | 100 - 199 zaměstnanců |
| 8 | 200-249 zaměstnanců |
| 9 | 250 - 499 zaměstnanců |