



evropský  
sociální  
fond v ČR



EVROPSKÁ UNIE



OPERAČNÍ PROGRAM  
LIDSKÉ ZDROJE  
A ZAMĚSTNANOST



PODPORUJEME  
VAŠI BUDOUCNOST  
[www.esfcr.cz](http://www.esfcr.cz)

# **Návrh metodiky pro provádění kontrafaktuálních analýz s pilotním ověřením**

## AKRONYMA

<b>CIE</b>	<b>Counterfactual Impact Evaluation; Kontrafaktuální Analýza</b>
<b>ČSÚ</b>	<b>Český statistický úřad</b>
<b>IS VAV</b>	<b>Informační systém výzkumu, experimentálního vývoje a inovací</b>
<b>DiD</b>	<b>Difference in Differences</b>
<b>Firma</b>	<b>Obchodní korporace podle 90/2012 Sb.</b>
<b>OP LZZ</b>	<b>Operační program Lidské zdroje a zaměstnanost</b>
<b>OPPP</b>	<b>Operační program Průmysl a podnikání 2004-2006</b>
<b>Podnik</b>	<b>Obchodní korporace podle 90/2012 Sb.</b>
<b>PSM</b>	<b>Propensity Score Matching</b>
<b>ROA</b>	<b>Návratnost aktiv (Return on Assets)</b>
<b>ROE</b>	<b>Návratnost vlastního kapitálu (Return on Equity)</b>
<b>ROS</b>	<b>Rentabilita tržeb (Return on Sales)</b>
<b>R&amp;D</b>	<b>Research and Development</b>
<b>TA ČR</b>	<b>Technologická Agentura České Republiky</b>
<b>VaVal</b>	<b>Výzkum, vývoj a inovace</b>
<b>VS</b>	<b>Veřejná soutěž</b>

## Finanční údaje

<b>TA</b>	<b>Aktiva celkem</b>
<b>OA</b>	<b>Oběžná aktiva</b>
<b>VK</b>	<b>Vlastní kapitál</b>
<b>V</b>	<b>Výkony</b>
<b>OB</b>	<b>Obrat</b>
<b>PHV</b>	<b>Provozní hospodářský výsledek</b>
<b>HVpZ</b>	<b>Hospodářský výsledek před zdaněním</b>

## Obsah

<b>1. Manažerské shrnutí.....</b>	<b>6</b>
<b>2. Úvod.....</b>	<b>11</b>
2.1. Účel a cíle metodiky .....	12
<b>3. Metody a data aplikované v CIE .....</b>	<b>14</b>
3.1. Teoretický základ kontrafaktuální analýzy .....	14
3.3. Metoda Instrumental Variables .....	17
3.4. Metoda párování (Propensity Score Matching).....	19
3.5. Metoda rozdílu v rozdílech (Difference in Differences) .....	23
<b>4. Možnosti implementace CIE v podmínkách TAČR.....</b>	<b>28</b>
4.1. Vhodnost využití CIE vůči konkrétním programům TA ČR.....	28
4.1.1. Vhodnost využití CIE v programu ALFA .....	29
4.1.2. Vhodnost využití CIE v ostatních programech TA ČR.....	31
4.2. Metodika CIE pro program ALFA.....	31
4.3. Popis proměnných pro analýzu programu ALFA metodou PSM .....	36
4.4. Popis dostupných dat vhodných pro CIE .....	37
4.4.1. IS Patriot.....	37
4.4.2. IS VaV.....	37
4.4.3. Magnus databáze (bisnode).....	37
4.4.4. ČSÚ.....	38
4.4.5. Použité datové položky .....	39
<b>5. Praktická ukázka hodnocení programu ALFA metodou CIE.....</b>	<b>40</b>
5.1. Výsledky hodnocení programu ALFA metodou CIE.....	40
5.1.1. Porovnání konkurenceschopnosti první kontrolní skupina .....	41
5.1.2. Porovnání konkurenceschopnosti druhé kontrolní skupiny .....	42
5.1.3. Distribuce PSM, první kontrolní skupina.....	42
5.1.4. Distribuce PSM, druhá kontrolní skupina.....	43
5.1.5. Vypočet efektu, první kontrolní skupina .....	44

5.1.6.	Vypočet efektu, druhá kontrolní skupina .....	46
<b>6.</b>	<b>Poučení a závěry .....</b>	<b>49</b>
6.1.	Závěry z hlediska využití CIE .....	49
6.2.	Poučení.....	49
<b>7.</b>	<b>Realizace CIE v softwaru R .....</b>	<b>51</b>
7.1.	Požadovaná úprava datasetu .....	51
7.2.	RS studio, úvod .....	52
7.2.1.	Kde RStudio stáhnout a jak nainstalovat?.....	52
7.2.3.	Přesunutí datasetu do programu .....	54
7.2.4.	Úprava datasetu.....	55
7.3.	Interpretace výsledků.....	56
7.3.1.	Přehled dat, DataReview .....	56
7.3.2.	Kolmogorov Smirnov test, K-Stest, K-StestA .....	57
7.3.3.	Efekt vyčíslen pomocí metody Difference in Difference, DiD.....	58
7.3.4.	Poměrové finanční ukazatele .....	58
7.3.5.	Dataset.....	59
<b>8.</b>	<b>Použitá a doporučená literatura.....</b>	<b>62</b>
8.1.	Použitá literatura .....	62
8.2.	Konkurenceschopnost .....	64
8.3.	Kontrafaktuální dopadové evaluace obecně.....	64
8.4.	Mezi další literaturu řadíme následující tituly .....	65
8.5.	Párování.....	65
8.6.	Rozdíly v rozdílech.....	67
<b>9.</b>	<b>Přílohy .....</b>	<b>70</b>
9.1.	Příloha 1: Rešerše literatury a praktické užití CIE .....	70
9.2.	Příloha 2: Výběr finančních ukazatelů a typ přiřazení firem, České prostředí .....	73
9.3.	Příloha 3: Obsah IS VaV databáze .....	75

9.4.	Příloha 4: DiD OLS model .....	76
9.5.	Příloha 4: Program kontrafaktuální analýzy .....	78

# 1. Manažerské shrnutí

Hlavním účelem této metodiky je teoreticky i prakticky vyložit problematiku kontrafaktuální analýzy jako jedné z evaluačních metod pro měření dopadů VaVaI a pilotně ověřit její aplikaci v prostředí TA ČR pro vyhodnocení programu.

Cílem je připravit strukturovaný podklad pro práci analytického týmu TAČR a současně definovat nástroj, který umožní přispět k měření dopadu intervencí a který bude možné využívat automatizovaně a opakovaně.

Dílčím cílem je vytvořit podklady pro co nejsnazší implementaci využívání tohoto analytického nástroje do běžného provozu TA ČR. Součástí tohoto cíle je také nastavení standardů pro potřeby práce s daty pro provádění kontrafaktuální analýzy.

Tato metodika obsahuje tři hlavní části popisující kontrafaktuální analýzu (CIE).

- Teoretický popis CIE
- Metodologický popis CIE
- Pilotní studii programu ALFA pomocí CIE

Kontrafaktuální analýza (CIE) je nástrojem na měření dopadu veřejné intervence a vyjadřuje rozdíl mezi situací pozorovanou (observable; factual) po intervenci a situací, která by nastala, pokud by k intervenci nedošlo (non-observable; counterfactual). Výsledné hodnocení a výpočet efektu je založen na kvantitativních datech („tvrdých“ datech), jejichž dostupnost a kvalita určuje kvalitu samotné analýzy. Potenciální objektivita CIE převyšuje v užití většinu ostatních nástrojů hodnocení. Především z toho důvodu je v současné době tento způsob hodnocení stále více využíván.

## **Teoretický popis CIE**

Na základě vlastností programů TA ČR a dalších parametrů byla vybrána jedna z metod používaných pro zpracování CIE. Popsaná kontrafaktuální analýza se skládá z párování podniků pomocí Propensity Score Matching (PSM) a výpočtu efektu pomocí rozdílu DiD (Difference in Differences).

Smyslem párování je přiřadit k sobě podniky, u kterých je předpoklad vývoje bez intervence co nejpodobnější. Pro minimalizaci výběrového zkreslení je tedy nutné zajistit proměnné, které nejlépe vystihují situaci podniku. Pokud vysvětlujeme dopad intervence pomocí navýšení nákladů na VaVal (v případě TA ČR tomu tak je), je nutné, aby párované podniky měly před intervencí výdaje na VaVal s co nejvyšší možnou mírou podobnosti.

Výpočet efektu pomocí Difference in Differences (DiD) je výpočtem dvou rozdílů. Jeden rozdíl je určený v čase (před a po intervenci), druhý rozdíl je určen subjekty (podpořenými a nepodpořenými). Rozdíl je proveden mezi dvěma skupinami. První skupina je tvořena podpořenými podniky a druhá skupina zvaná kontrolní je tvořena podniky nepodpořenými. Před intervencí, kdy jsou podniky napárovány, jsou obě skupiny velmi podobné. Kontrolní skupina může být tvořena z náhodných podniků, nemusí tedy jít o žadatele do jednoho intervenčního programu.

Součástí teoretického popisu jsou i obecné předpoklady kontrafaktuální analýzy. Patří mezi ně velký počet sledovaných subjektů, dostatečná proporcionalita intervence a homogenita testovaného vzorku. V případě kontrafaktuální analýzy měřící dopad intervence v oblasti VaVal, je nutné přijmout předpoklad, že pro daný vzorek neexistují „R&D spillovers“<sup>1</sup>.

Předpokladem pro aplikaci párování také je, že existuje dostatečný přesah předintervenčních vlastností a charakteristik mezi podpořenými a nepodpořenými subjekty.

Důležitým předpokladem pro metodu Difference in Differences je paralelismus, tedy předpokládaný vývoj podpořené skupiny v případě, kdy by podpořena nebyla.

### **Metodologický popis CIE**

Metodologický popis se přesně soustředí na doporučené a zároveň realizovatelné provedení CIE programu ALFA za daných okolností. Je popsán model párování pomocí PSM za užití finančních ukazatelů.

---

<sup>1</sup> Aktivity R&D vytváří znalosti. Tyto znalosti mohou být využity i někým jiným. Volné šíření těchto znalostí se nazývá R&D spillovers. Z toho důvodu existuje duševní ochrana, která zabraňuje volnému šíření specifických znalostí.

$$\text{Logit}(\mathbb{E}[P_i | \text{Fin}_1 + \text{Fin}_2 + \dots + \text{Fin}_n]).$$

$P_i$  = Binární proměnná  $[0; 1]$ , 0 = Podnik nepodpořen 1 = Podnik podpořen

$\text{Fin}_n$  = finanční údaj podniku (př. aktiva celkem)

Stejně tak popsán postup výpočtu rozdílu DiD:  $(Y_{T2} - Y_{C2}) - (Y_{T1} - Y_{C1})$ .

$Y_{T2}$  = Daná proměnná u podpořeného podniku v čase 2

$Y_{T1}$  = Daná proměnná u podpořeného podniku v čase 1

$Y_{C2}$  = Daná proměnná u nepodpořeného podniku (z kontrolní skupiny) v čase 2

$Y_{C1}$  = Daná proměnná u nepodpořeného podniku (z kontrolní skupiny) v čase 1

### Pilotní evaluace programu ALFA pomocí CIE

Kontrafaktuální dopadová evaluace programu ALFA byla vypracována na základě výše uvedeného. V rámci evaluace byly použity dvě kontrolní skupiny. První byla tvořena neúspěšnými žadateli o podporu ve stejné veřejné soutěži jako podpoření. Druhá byla tvořena neúspěšnými žadateli v programu ALFA. Tato část také obsahuje vypracovaný program v softwaru R (RStudio), který do jisté míry výpočet CIE automatizuje. Je však stále nutné ad hoc projít pospané předpoklady a připravit vhodně upravený dataset (podle popisu, který je součástí této analýzy).

Efekt intervence TA ČR byl měřen na těchto finančních ukazatelích: aktiva celkem, oběžná aktiva, vlastní kapitál, výkony, obrat, provozní hospodářský výsledek a hospodářský výsledek před zdaněním. Byly vypočítány také poměrové ukazatele: návratnost aktiv (ROA), návratnost vlastního kapitálu (ROE) a návratnost tržeb (ROS). Studie využívá Sedláčkovu (2014) paralelu, že ukazatel ROA vyjadřuje konkurenceschopnost podniku. Ostatní finanční ukazatele poskytují poměrně plastický pohled na vývoj společnosti.

### Hlavní závěry a poučení

- Výpočet efektu a párování firem je možné provádět automatizovaně tak, aby to po krátkém zaškolení byl schopen provádět referent nebo analytik.



- Z provedené pilotní studie vyplývá, že efekt podpory je z hlediska celkových aktiv, vlastního kapitálu a zisku před zdaněním pozitivní. Návratnost vlastního kapitálu se však u podpořených podniků snížila. Z důvodu dostupnosti údajů jsou výsledky předběžné a kontrafaktuální analýzu bude nutné zopakovat po získání post intervenčních metrik.
- Přístup ke kontrafaktuální analýze je vhodné nadále vyvíjet, revidovat, testovat nové metody a přizpůsobovat je stále širší datové základně.
- Kvalita kontrafaktuální analýzy je určována kvalitním datasetem. Pro pilotní ověřování metodického postupu byla využita data dostupná v době realizace analýzy, nicméně pro zvýšení reálné vypovídací hodnoty samotné analýzy by bylo nutné získat potřebná data v dostatečné kvalitě a kompletnosti. Při práci s touto analýzou je nutné na tuto skutečnost brát ohled právě v souvislosti s omezeným přístupem k výchozím datům.
- Kvalita dat je nutnou podmínkou pro dobrou kontrafaktuální dopadovou studii. V případě problematiky VaVal je nutno zahrnout výdaje na R&D. V rámci českých účetních standardů nejsou tyto výdaje součástí finančních výkazů (na rozdíl od IFRS). Veřejná instituce má možnost získat tento údaj měřený prostřednictvím dotazníku VTR 5-01 v gesci ČSÚ. Tato analýza byla zpracována na základě dat dostupných v době jejího zpracování, tj. vycházelo se z finančních ukazatelů uváděných ve finančních výkazech dle platných právních předpisů.
- Účelová podpora VaVal udělovaná podnikům provádějící kontinuální R&D aktivity je v ČR de facto plošná. To činí vyhodnocení efektu podpory VaVal pomocí CIE problematickým. Podniky z kontrolní skupiny obdrželi podporu od jiných veřejných institucí. Z důvodu zajištění předpokladu, že daný podnik provádí R&D aktivity byla kontrolní skupina tvořena nepodpořenými žadateli do programu ALFA. Téměř všechny tyto podniky byly podpořeny z jiných veřejných zdrojů vymezených na podporu R&D. V důsledku toho tedy není možné v současné situaci porovnat podniky podpořené a vůbec nepodpořené. Porovnání je tudíž možné uskutečnit mezi podniky podpořenými TA ČR a podniky podpořenými jinými veřejnými institucemi, jejichž podpora však byla celkově nižší.

- Výsledky pilotního ověřování (2 skupiny podpořených podniků z hlediska míry konkurenceschopnosti):
  - První skupina (Podpořené podniky byly před podporou více konkurenceschopné)
- Efekt podpory je v absolutních hodnotách finančních ukazatelů pozitivní ve všech sledovaných ukazatelích
- Poměrové ukazatele u podpořených podniků se oproti nepodpořeným snížily ve všech případech.

Podpořené podniky mají vyšší finanční ukazatele, jako aktiva a zisk, ale jejich efektivita se oproti nepodpořeným snížila. Aktiva, vlastní kapitál a tržby rostly podpořeným podnikům méně, než zisk před zdaněním.

- Druhá skupina (Podpořené podniky byly před podporou méně konkurenceschopné)
- Efekt podpory je v absolutních hodnotách pozitivní (kromě obratu) ve všech sledovaných ukazatelích
- Efekt podpory v programu ALFA je z hlediska návratnost aktiv (ROA) a návratnost tržeb (ROS) pozitivní a návratnosti vlastního kapitálu (ROE) negativní

Podpořené podniky mají vyšší finanční ukazatele, jako aktiva a zisk. Jejich konkurenceschopnost je vyšší než konkurenceschopnost nepodpořených. Návratnost kapitálu (ziskovost) se však snížila.

## 2. Úvod

V posledních letech roste potřeba měřit výsledky a dopady různých intervencí pomocí kontrafaktuálních dopadových evaluací (Counterfactual impact evaluation; CIE). Na úrovni Evropské unie se zmínka o potřebě využití kontrafaktuálních dopadových evaluací objevuje například v Páté zprávě o hospodářské, sociální a územní soudržnosti (Evropská komise, 2010) nebo tzv. Barcově zprávě (Barca, 2009). Zároveň, se zvyšující se snahou o zjištění přínosů a efektů intervencí veřejných výdajových programů, se objevil i požadavek na dopadové evaluace v programovacím období 2014 – 2020. Zpracování kontrafaktuálních dopadových evaluací je u Kohezní politiky relativně nový požadavek ze strany Evropské komise, ale na příkladu mezinárodních rozvojových projektů jsou metody CIE již dlouhodobě známé a vyzkoušené. Obecně však v České republice existuje poměrně malé povědomí a schopnost zpracovat tento typ evaluací. To platí jak na straně orgánů veřejné správy, tak na straně zpracovatelů evaluací. Novost a přínos této metodiky pro případ Technologické Agentury České Republiky (TA ČR) spočívá také v její aplikační části.

Kontrafaktuální dopadové evaluace vychází z pojetí dopadu (impact), který vyjadřuje rozdíl mezi situací pozorovanou (observable; factual) po intervenci a situací, která by nastala, pokud by k intervenci nedošlo (non-observable; counterfactual)<sup>2</sup>. Cílem evaluace dopadu je rigorózním způsobem změřit efekt přiřítelný intervenci. Cílovou skupinou jsou podpořené podniky, u kterých jsou sledovány finanční údaje, jako např. vlastní kapitál, nebo zisk před zdaněním.

Tato metodika byla vytvořena za účelem popsání aplikace CIE na případu TA ČR, programu ALFA v ČR. Zároveň by tato metodika měla sloužit jako návod pro budoucí ex-post evaluace programů a intervencí podporovaných TA ČR formou CIE. Pro tento účel byl vytvořen program v RStudio<sup>3</sup>, který uživateli umožní

---

<sup>2</sup> Potluka, O., Špaček, M., (2013). "Postupy a metody kontrafaktuálních dopadových evaluací pro Operační program Zaměstnanost v období 2014-2020." Prague, 15. října 2013 Prague, September 19-21, 2013.

<sup>3</sup> RStudio je statistický a ekonometrický software. Formou kódu v něm lze zadávat úkony k provedení. Tyto kódy je možné uložit a následovně používat v celku.

aplikovat metodické přístupy CIE. Praktická ukázka úpravy dat a výpočtu je také součástí metodiky.

## 2.1. Účel a cíle metodiky

Dokument vzniká jako výstup projektu “Zefektivnění činnosti TA ČR v oblasti podpory VaVaI a podpory posilování odborných kapacit organizací veřejné správy v oblasti VaVaI” (dále také Projekt Zefektivnění nebo Projekt), resp. v rámci jeho klíčové aktivity 3 (KA3) “Příprava nových analytických metodik hodnocení VaVaI”. Projekt je kofinancován z Evropského sociálního fondu EU v rámci Operačního programu lidské zdroje a zaměstnanost.

Hlavním cílem projektu je zefektivnění poskytování podpory VaVaI ze strany TA ČR a dalších organizací veřejné správy, posílení odborných kapacit organizací veřejné správy v oblasti VaVaI, posílení chápání významu aplikovaného VaVaI a jeho výsledků pro další rozvoj ČR a do budoucna i sjednocení způsobů a podmínek poskytování podpory VaVaI.

Cíle Projektu Zefektivnění jsou naplňovány sedmi klíčovými aktivitami. Klíčová aktivita KA3 „Příprava nových analytických metodik hodnocení VaVaI“ je zaměřena na podporu posilování odborných kapacit organizací veřejné správy v oblasti VaVaI“. Cílem aktivity je zefektivnit a zkvalitnit činnosti TA ČR v oblasti analytických služeb jako základní metody zdůvodňování tvorby podkladů pro nastavení a hodnocení podpor. Toho má být dosaženo vytvořením nových analytických metodik pro podporu zajišťování hlavních činností a procesů realizovaných TA ČR.

Tato metodika byla vypracována na základě analytických zjištění a zkušeností získaných při realizaci KA3 a to v období 10/2014 – 11/2015.

Cílem této metodiky je teoreticky i prakticky vyložit problematiku kontrafaktuální analýzy tak, aby ji bylo možné aplikovat v prostředí TA ČR pro vyhodnocení programu.

Kromě výkladu kontrafaktuální analýzy je také hlavním účelem pilotně ověřit její aplikaci v prostředí TA ČR pro vyhodnocení programu.

Cílem je připravit strukturovaný podklad pro práci analytického týmu TAČR a současně definovat nástroj, který umožní přispět k měření dopadu intervencí a bude ho možné využívat automatizovaně a opakovaně.

Dílčím cílem je vytvořit podklady pro co nejsnazší implementaci využívání tohoto analytického nástroje do běžného provozu TA ČR. Součástí tohoto cíle je také nastavení standardů pro potřeby práce s daty a pro vlastní provádění kontrafaktuální analýzy.

V rámci této metodiky je podrobně popsán postup

- (1) jaké je nutno ověřit předpoklady
- (2) jak párovat podniky pomocí metody Propensity Score Matching
- (3) a jak vypočíst efekt pomocí metody Difference in Differences.

Dokument obsahuje praktickou ukázkou výpočtu kontrafaktuální analýzy a program pomocí kterého je možné kontrafaktuál vypočítat.

Zadavatelem projektu je Technologická agentura ČR. Zpracovatelem této metodiky je projektový tým KA3 vytvořený v rámci Projektu Zefektivnění.

## 3. Metody a data aplikované v CIE

### 3.1. Teoretický základ kontrafaktuální analýzy

Teoretický základ tohoto dokumentu ve velké míře čerpá z metodiky Potluky a Špačka (2013). V následující části jsou kriticky představeny metody, které lze použít při kontrafaktuálních dopadových evaluacích (včetně možností a limitů při aplikaci v praxi). Základ představovaných metod tvoří statistické kvantitativní metody Counterfactual Impact Evaluation (CIE), které, pokud se správně použijí, poskytují poměrně přesné odhady dopadů zkoumaných politik nebo programů. Pro tyto metody je však potřeba dostatečné množství homogenních případů a jasná specifikace cílů podpory, takže je nelze použít vždy.

Stejně jako metodika Potluky a Špačka (2013) se tato metodika kontrafaktuálních dopadových evaluací věnuje evaluacím s velkým počtem případů, které srovnávají situaci zkoumaného případu se situací bez podpory.

Pro aplikaci následujících metod je nutné splnění určitých podmínek:

#### ● Velký počet případů<sup>4</sup>

Primárně je tato podmínka uvažována pro zajištění toho, aby velký počet případů zajišťoval statistickou významnost odhadů a tím přispěl k důvěryhodnosti výsledků. Může jít o případy na úrovni jednotlivců, firem, obcí.

#### ● Homogenní případy

Aby bylo možné použít metod CIE, je nutné, aby zkoumané případy vystihovaly stejný typ podpory, tj. aby v jednom vzorku nebyly smíchávány různé druhy podpor, které se svým charakterem liší, a tudíž mechanismus změny sledovaných indikátorů se liší u cílové skupiny. Tj. metodami CIE nelze měřit dopady vysoce heterogenních programů či vágních intervencí. CIE se nezabývá zjišťováním sociální komplexity, ale vždy se zaměřuje na přímé otázky, kde byl jasně definován jeden očekávaný výsledek v kauzálním spojení na jednu intervenci. Jednou z možností jak testovat homogenitu případů je provést

<sup>4</sup> Ve statistické rovině se většinou považují vzorky s 80 a více subjekty za relevantní. V našem případě pro jistotu stavíme hranici na 100 sledovaných subjektů.

Kolmogorov Smirnov test<sup>5</sup> na údajích použitých pro srovnání. Tento test porovnává podobnost rozdělení. Tento test homogenity tedy nemá kvalitativní rozměr.

### ● Dostupnost dat

Bez důvěryhodných a ucelených dat není možné zpracovat kvalitní kontrafaktuální dopadovou evaluaci. S ohledem na význam této podmínky je jí následně věnována celá kapitola.

### ● Proporcionalita

Rozsah intervence, jejíž dopady mají být metodami CIE měřeny, musí mít potenciál nějakou změřitelnou změnu vytvořit. Vyhodnocovat lze jak regionální tak celonárodní formy podpory, vždy je však nutné zvážit, zda byl rozsah podpory dostatečný k tomu, aby ve sledovaném prostředí mohl způsobit změnu, která by byla odlišitelná od dalších obecných vlivů. Pro tento účel je možné porovnat přidělenou dotaci vůči obratu, nebo výkonům. Je také důležité, z jakých finančních zdrojů jsou dotace poskytovány. Není výjimkou, že jeden podnik čerpá veřejnou podporu na VaVal od více veřejných subjektů.

### ● Nulový spillover efekt

Pro účely použití CIE v prostředí podpory VaVal je nutno uvažovat nulový spillover efekt<sup>6</sup>. Výsledné zlepšení připisujeme zlepšení procesů, nebo využití daného vynálezu, který byl financován z veřejných prostředků. Pokud by se znalosti, tedy i schopnost využívat nejnovější vynálezy ostatními subjekty, volně šířily, nebylo by možné přičíst VaVal zlepšené výsledky konkrétních subjektů na kterých provádíme měření.

### ● Podmínka časového hlediska

Aby výsledky CIE byly relevantní, musí být splněna podmínka časového hlediska. Analýza musí být provedena na datech s dostatečným časovým odstupem, aby se zkoumaný vliv intervence mohl ve sledovaných datech projevit. U projektů s VaVal zaměřením lze očekávat, že zavedení nového vynálezu do provozu a prodeje bude časově nákladné přibližně jako projekt samotný.

### ● Nedochází k vytlačení investic

<sup>5</sup> Jedná se o statistický test homogenity. Výsledkem oboustranného testu je p hodnota. Pokud je tato hodnota vyšší než 0,05 jsou obě porovnávaná rozdělení považována za homogenní (dostatečně podobná).

<sup>6</sup> Volné šíření znalostí bezprostředně po dosažení nové znalostní úrovně (vynálezu)



Vytlačení investic<sup>7</sup> je jev popsáný v ekonomii. Jeho důsledkem je, že stát financuje soukromé aktivity, které by se odehrály, i kdyby podnik dotace neobdržel. K tomu aby bylo možné určit efekt veřejné podpory, je nutné, aby nedocházelo k vytlačení investic. Pokud k vytlačení investic nedochází, je veřejná podpora dodatečným financováním výzkumu, které by se bez veřejného zásahu neudálo. Tento předpoklad se testuje také pomocí kontrafaktuální analýzy. Pro otestování tohoto předpokladu je nutné disponovat podnikovými výdaji na R&D.

### ● Minimální možné výběrové zkreslení

V ideálním případě by měly být napárovány totožné podniky s tím, že jeden z nich by byl podpořen a druhý ne. Toho v reálných podmínkách není možné dosáhnout. V případě, že jsou podniky podpořeny s ohledem na charakteristiku, pro kterou se počítá kontrafaktuál, mohou být výsledky zavádějící. Podpořený podnik by ve výsledku byl lepší, protože lepší podniky jsou vybírány pro podporu. Nešlo by tedy o výsledek vyšších výdajů na VaVaI způsobené podporou samotnou. Jednou z možných kontrol je porovnat sledovanou proměnnou obou podniků, např. konkurenceschopnost podniků před podporou. Pro konkrétní metody jsou dále uváděny další poznámky k předpokladům.

## 3.2. Metoda Regression Discontinuity Design

Metoda regresní diskontinuity je použitelná, pokud je nárok na zapojení se do programu (na intervenci) určen pravidlem, které stanovuje, že subjekty nad/pod určitou hranicí jsou způsobilé pro program, zatímco subjekty pod/nad touto hranicí jsou nezpůsobilé<sup>8</sup>. Základní myšlenkou Regression Discontinuity Design (RDD) je vybrat co nejbližší podpořené a nepodpořené subjekty na základě počtu bodů ve výběrovém řízení. K tomu je využito cut-off criteria, které je předělem mezi podpořenými a nepodpořenými. Tento způsob párování není v této metodice použit z důvodu složitého výběrového procesu projektů v programu ALFA, a protože počet bodů nemá vypovídající hodnotu ohledně podobnosti podniku.

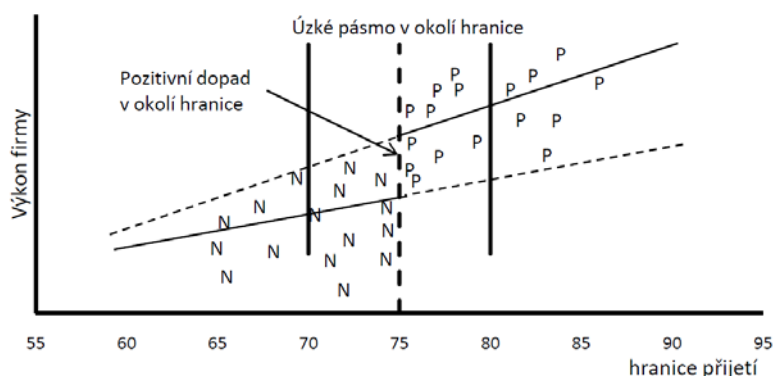
### Obrázek 1: Grafické vyjádření RDD

---

<sup>7</sup> Vytlačení investic nebo tzv. vytlačovací efekt, tedy proces, který má následek zvýšení úrokové míry, v důsledku čehož poklesnou investice soukromé a jehož výsledkem je tak paradox stejné, tedy nezvýšené míry investic. Z veřejných zdrojů se tak hradí soukromé aktivity, které by se odehrály tak jako tak, jen s rozdílnými zdroji financování.

<sup>8</sup> Tato podmínka není v případě TA ČR uspokojena.





Zdroj: Potluka a Špaček (2013), Evropská komise (2009)

### 3.3. Metoda Instrumental Variables

Tuto metodu lze využít v případech, kdy vystavení se politice nebo intervenci není určeno pouze rozhodnutím samotných oprávněných žadatelů, ale také do značné míry událostmi a procesy, které jsou mimo jejich kontrolu. Tato „nedobrovolná změna“ umožňuje odstranit výběrová zkreslení. Celý model se podobá experimentu. Aby mohla být tato metoda využita, musí být splněny dvě podmínky.

1. Instrumentální proměnná musí být signifikantním prediktorem pravděpodobnosti získání podpory.
2. Zároveň instrumentální proměnná nesmí ovlivňovat zkoumaný indikátor jinak, než přes získání podpory.

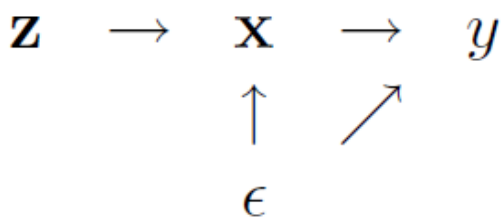
Příkladem instrumentálních proměnných mohou být náhlé změny právních předpisů vyvolané politickým procesem, geografické faktory, jako je vzdálenost od klientů poskytovaných služeb, neočekávané výpadky ve financování programu, změny v administrativně stanovených hranicích. Je také možné využít přístup, kdy je část oprávněných žadatelů přímo informována o možnosti se do programu zapojit a zbylá část nikoliv. Pro použití instrumentální proměnné pro odhad dopadu intervence existují dva základní přístupy. Jedná se o metodu Waldova estimátoru<sup>9</sup> a o dvoustupňovou metodu nejmenších čtverců<sup>10</sup> (dvoustupňovou regresi). Metoda Instrumental Variables (IV) umožňuje výpočet odhadů v případě endogenity. IV je metodou běžně používanou ve statistice a ekonometrii. Její použití je motivováno vlastností dat. Ve všech případech, kde není možné zajistit podmínku OLS 'nekorelovanost regresorů s reziduální složkou,' je doporučováno IV využít. V tomto případě je endogenita považována

<sup>9</sup> Způsob používaný k výpočtu odhadu v případě instrumentální proměnné

<sup>10</sup> Metoda nejmenších čtverců (MNČ; OLS) je vytvoření odhadu na základě sklonu křivky jejichž vzdálenost je v kvadratickém vyjádření nejmenší od všech bodů (data points). Dvoustupňová MNČ je vylepšením použitelným pro přesně identifikované i předidentifikované rovnice. Jedná se o opakovanou aplikaci MNČ pro dva odhady: (1) odhad neomezeného redukovaného tvaru, (2) odhad strukturních parametrů.

za silnou korelaci vysvětlované proměnné (regressor;  $x$ ) a statistické chyby (error term;  $e$ ). Abychom tento problém vyřešili, hledáme proměnnou ( $z$ ), která je silně korelována s proměnnou, kterou vysvětlujeme regresí ( $x$ ). Tato metoda nebude v metodice použita a vysvětlena, protože nebyla nalezena tato instrumentální proměnná (IV) a problém s endogenitou bude řešen minimalizací výběrového zkreslení.

## Obrázek 2 Instrumental Variables



### 3.4. Metoda párování (Propensity Score Matching)

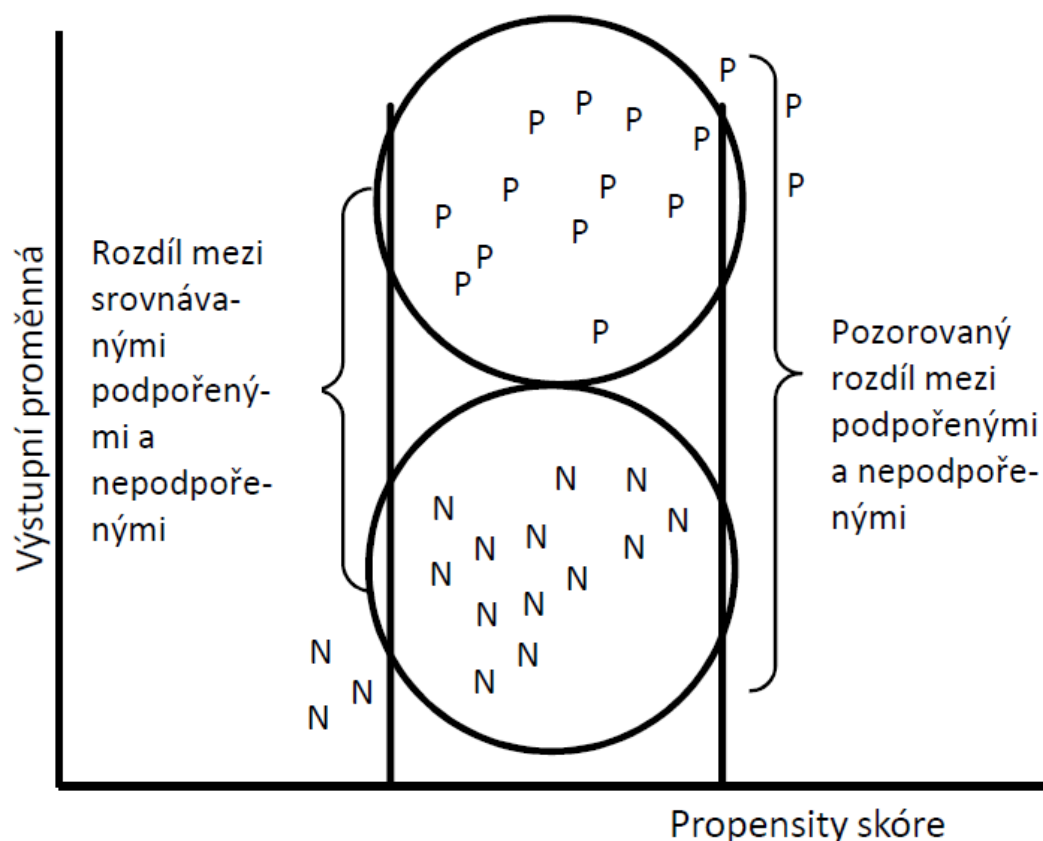
Základní myšlenkou párování (Matchingu) je snaha o sestavení takové kontrolní skupiny<sup>11</sup>, která se ve všech ohledech kromě obdržené podpory co nejvíce podobala skupině příjemců. V kontrolní skupině jsou zařazeny subjekty, které se shodují ve svých charakteristikách s podpořenými příjemci. Pokud je tato podoba dostatečná, lze výstupy kontrolní skupiny považovat za kontrafaktuál a dopad intervence lze získat jako rozdíl mezi průměrnými výsledky obou skupin. Metoda je tedy založena na předpokladu, že jsme schopni pozorovat všechny charakteristiky, které mají vliv na výsledek. Intuitivním způsobem konstrukce kontrolní skupiny a porovnáváním průměrů se tato metoda podobá náhodnému výběru. Avšak zásadní rozdíl mezi párováním (Matchingem) a náhodným výběrem je, že zatímco u náhodného výběru jsou mezi kontrolní a podpořenou skupinou shodné všechny charakteristiky, pozorovatelné i nepozorovatelné, u sebedokonalejšího párování lze brát v potaz pouze pozorovatelné charakteristiky. Základním předpokladem pro platnost párování je, že jsou shodné pozorovatelné charakteristiky mezi oběma skupinami a tyto skupiny jsou vyvážené s ohledem na všechny charakteristiky relevantní pro sledovaný výstup. Čím více dostupných předintervenčních vlastností do analýzy zahrneme, tím větší je šance na naplnění tohoto předpokladu. Dalším předpokladem pro aplikaci párování také je, že existuje dostatečný přesah těchto vlastností mezi podpořenými a nepodpořenými subjekty. Tento přesah se nazývá společná podpora (Common Support).

V následujícím grafu jde o situaci, kdy pro některé podpořené (P) a nepodpořené (N) jednotky platí to, že jejich propensity skóre je shodné (v grafu jde o situaci, kdy některá P a N leží „nad sebou“). Tento jev může nastat, protože PSM na základě vlastností těchto subjektů vyhodnotí pravděpodobnost obdržení podpory.

#### Graf 1: Metoda párování podle propensity skóre

---

<sup>11</sup> V případě, že pro vytvoření kontrolní skupiny používáme randomizaci, je anglicky nazývána „control group“, v případě, že kontrolní skupina nevznikla randomizovaně, je nazývána „comparison group“.



Zdroj: Potluka a Špaček (2013), Evropská komise (2009)

V ideálním případě by se v rámci párování měl pro každého příjemce podpory nalézt ve všech aspektech relevantních pro výběr identický nepodpořený subjekt. Avšak takového stupně podobnosti je v praxi velmi obtížné dosáhnout. Proto se užívá technika zvaná Propensity Score Matching (PSM). Tato metoda ve své podstatě převádí řadu charakteristik zkoumaných příjemců na jedno číslo. Zahrneme-li do analýzy velké množství charakteristik, je poté velice obtížné pro každého příjemce podpory nalézt přesně identický nepodpořený subjekt, který by měl shodné hodnoty u všech zkoumaných charakteristik. Tuto překážku lze překonat pomocí tzv. Propensity Score, které vyjadřuje pravděpodobnost získání podpory na základě pozorovaných charakteristik. PSM tak shrnuje všechny pozorované charakteristiky daného subjektu do jediného čísla (pokud pracujeme s jednotlivými charakteristikami, jde o metodu Multiple Regression). Tímto způsobem je možné redukovat vícerozměrný problém na jednorozměrný. Každý příjemce podpory je spárován s nepodpořeným subjektem, který má podobnou pravděpodobnost získání podpory. Tato pravděpodobnost je vypočtena na základě pozorovaných charakteristik a je vyjádřena právě pomocí

Propensity Score. Dopad je poté získán opět jako rozdíl průměrných výstupů obou skupin.

Klíčovou podmínkou pro užití párování je dostupnost pozorovatelných předintervenčních charakteristik, neboť charakteristiky pozorované až po intervenci již mohou být touto intervencí zkresleny. Další podmínkou je dostatečný přesah charakteristik mezi podpořenými a nepodpořenými subjekty. Matching bývá také velice často používán v kombinaci s metodou rozdílu v rozdílech (viz dále), kdy jsou nejdříve subjekty pomocí párování spárovány a následně je pomocí rozdílu v rozdílech odhadnut dopad intervence. Kombinací těchto dvou metod se snižuje riziko výběrového zkreslení výsledků.

Pro aplikaci párování existuje několik kroků (Evalsed 2009):

**1) Odhadnutí Propensity Score.** Tento krok se neskládá pouze ze samotného odhadu Propensity Score pro každý subjekt (pomocí logit, nebo probit regrese), ale obsahuje také analýzu proměnných (charakteristik), které je nutné do modelu zahrnout. Velká pozornost by měla být věnována výběru proměnných tak, aby se pomocí nich podařilo eliminovat výběrové zkreslení. Je proto nutné vybrat sadu proměnných, které věrohodně splní tuto podmínku. Vynechání důležitých proměnných může vážně zkreslit následný odhad dopadu politiky. Do modelu by měly být zahrnuty proměnné, které ovlivňují oprávněnost zapojit se do politiky a výstupní proměnnou.

**2) Rozdělení ve zkoumané a kontrolní skupině** musí být takové, aby se tyto dvě skupiny „překrývaly“, tj. že neexistuje systematické zkreslení mezi podpořenou a kontrolní skupinou.

**3) Párování subjektů na základě Propensity Score.** Je-li odhadnut celý propensity score model a je-li vypočítáno skóre pro každý subjekt, tím skóre je pravděpodobnost určená logit/probit modelem. Následuje párování na základě zvoleného algoritmu.

Existují čtyři základní algoritmy párování (další lze nalézt například v Khandker, Koolwal a Samand 2010):

**a) Párování s nejbližším sousedem (Nearest available PS matching).**

Jeden subjekt z podpořené skupiny je spárován s jedním subjektem z kontrolní skupiny s nejbližší hodnotou Propensity Score. Můžeme užít

variantu „s vrácením“ nebo „bez vrácení“ (with /without replacement). V prvním případě použijeme každý subjekt z kontrolní skupiny pouze jednou, v druhém případě lze subjekty z kontrolní skupiny použít opakovaně. Tento způsob párování je nejčastěji používaným.

b) **Kaliper/párování rádiusem** (Calliper/radius matching). Toto párování vychází z párování s nejbližším sousedem, ale navíc stanovuje maximální vzdálenost Propensity Score, do které je možné takového souseda hledat. Radius matching je variantou Calliper matchingu, kdy do párování zahrnuje všechny subjekty s Propensity Score ležící do dané vzdálenosti.

c) **Stratifikační či intervalové párování (Stratification or Interval Matching)** vychází z rozdělení všech subjektů do několika intervalů dle výše jejich Propensity Score. Intervaly se liší počtem podpořených a nepodpořených subjektů. Následně se vypočítá dopad pro každý interval zvlášť a celkový dopad se získá jako vážený průměr těchto dílčích dopadů, kde váhami jsou proporce podpořených subjektů v jednotlivých intervalech.

d) **Kernel matching** neporovnává pouze dva subjekty mezi sebou, ale pro sestavení kontrafaktuálu k výstupu podpořeného subjektu využívá všechny subjekty z kontrolní skupiny. Vypočítá se jako vážený průměr výstupu všech nepodpořených subjektů, kde váhami jsou vzdálenosti jejich Propensity Score od Propensity Score daného podpořeného subjektu.

**4) Posouzení kvality párování.** V tomto kroku se posuzuje společná podpora mezi podpořenými a nepodpořenými subjekty a také se ověřuje, zda po provedeném párování zůstalo zachováno rozdělení relevantních proměnných. Prediktorem přesnosti párování je také rozdíl průměru podpořené a kontrolní skupiny.

**5) Odhad průměrného dopadu a standardní chyby.** Po provedení párování lze dopad vyjádřit jako rozdíl průměrných hodnot výstupu podpořených a nepodpořených subjektů. Kontroluje se také statistická signifikance dopadu a standardní chyba. Párování má dvě nevýhody oproti experimentálním technikám. Jednou z nevýhod párování je nutný předpoklad podmíněné nezávislosti, což znamená předpokládat, že výběrové zkreslení lze odstranit



testování pozorovatelných proměnných. Druhou nevýhodou je, že lze odhadovat dopad intervence pouze tam, kde se překrývají charakteristiky příjemců a nepodpořených subjektů<sup>12</sup>.

Je vhodné upozornit na to, že metoda PSM není samospasitelná. Nevyžaduje předpoklad ohledně funkční závislosti, ale vyžaduje předpoklad o funkční formě při odhadu propensity skóre. Problém vícerozměrnosti<sup>13</sup> není tedy vyřešen, ale je přesunut do prvního kroku – do odhadu propensity skóre. Na druhou stranu však párování nevyžaduje ex-ante předpoklad o shodě typu a podobě proměnných, se kterými se pracuje při párování. Výhodou párování oproti randomizaci je odbourání etických problémů spojených s náhodným výběrem. Metoda párování je také mnohdy méně nákladná. Výhodou párování vůči ostatním regresním metodám je fakt, že se zabývá problematikou společné podpory, tedy rozdílů mezi podpořenými a nepodpořenými subjekty. Párování dále také nevyžaduje předpoklad funkční podoby rovnosti výstupů, jako je tomu u regresních metod, které předpokládají použití funkční závislosti (často lineární), což obvykle není odůvodněno daty ani teorií (Angrist, Pischke 2009).

### 3.5. Metoda rozdílu v rozdílech (Difference in Differences)

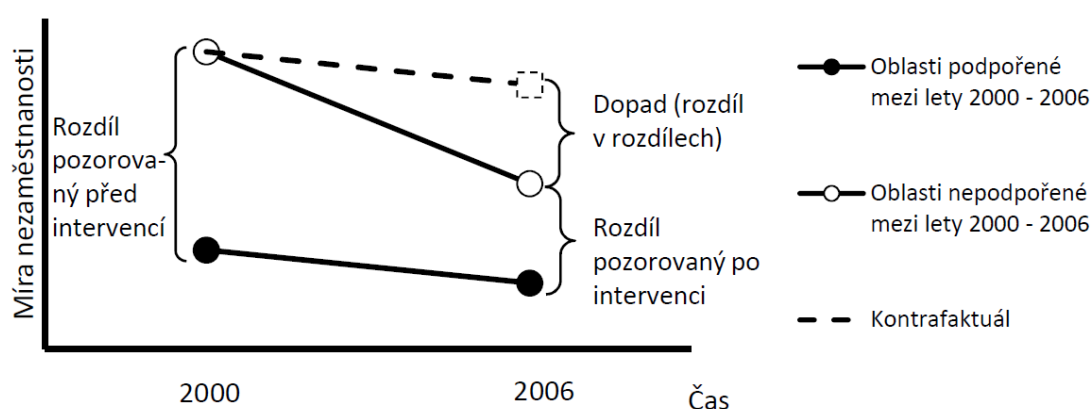
Metoda rozdílů v rozdílech (dále také DiD) vychází z toho, že dopad intervence na sledovaný výstup lze vypočítat pomocí dvou rozdílů. Jeden rozdíl je určený v čase (před a po intervenci), druhý rozdíl je určen subjekty (podpořenými a nepodpořenými). DiD je tedy užitečná metoda, pokud máme dvě skupiny se stejnou historií. Ve své nejjednodušší formě tato metoda vyžaduje pouze souhrnné údaje o sledované proměnné výstupu. Stačí mít pouze data za příjemce a nepodpořené subjekty po dobu nejméně dvou časových období a je možné pomocí metody rozdílů v rozdílech odhadnout dopad intervence. Základním předpokladem, který tato metoda považuje za daný, je paralelismus, tj.

<sup>12</sup> Podmíněné párování (Conditional matching) je využíváno v případě, že existuje odůvodněná obava, že nějaká nepozorovatelná, ale významná charakteristika může být nerovnoměrně rozložena mezi subjekty podpořené a nepodpořené skupiny a působit jako absolutní překážka párování (absolute constrain), např. příslušnost k regionu nebo k výrobnímu sektoru u porovnávaných firem, nebo věková kategorie u porovnávaných osob. Nicméně jde o pozorovatelnou veličinu.

<sup>13</sup> Při snaze postavit naproti sobě co nejpodobnější podniky se potýkáme s problémem vícerozměrnosti. Podnik není definován jedním rozměrem. Proto je přiřazení podniků netriviální záležitost. Jeden ze způsobů jak problém zjednodušit do jedné proměnné je přepočítání sledovaných charakteristik podniku na propensity score pomocí logit (nebo podobného) modelu.

předpokládaný vývoj podpořené skupiny v případě, kdy by podpořena nebyla. Nicméně i přesto je pro stanovení dopadu nutné uvažovat některé netestovatelné předpoklady. Následující schéma zobrazuje způsob, jakým se k odhadu dopadu přistupuje v metodě rozdílů v rozdílech. V grafu je znázorněn výše zmíněný paralelismus. Dopad je pak vypočítán jako rozdíl mezi situací ve dvou různých obdobích (před intervencí a po ní) a rozdílu mezi dvěma skupinami (podpoření vs. nepodpoření).

**Graf 2: Metodika rozdílů v rozdílech**



Zdroj: Potluka a Špaček (2013), Evropská komise (2009)

Existují dva různé způsoby pro vysvětlení toho, jak pomocí těchto dvou rozdílů odhadnout dopad intervence.

1) Intuitivní přístup spočívá v tom, že se nejprve změří rozdíl ve sledované proměnné mezi příjemci a nepodpořenými po poskytnutí intervence. Tento rozdíl však není způsoben pouze intervencí (existence tzv. natural deviation), ale i jinými rozdílnými charakteristikami mezi příjemci podpory a nepodpořenými. To se nazývá výběrové zkreslení (selection bias). Následně porovnáme hodnoty sledované proměnné mezi příjemci a nepodpořenými v době, kdy intervence ještě nebyla poskytnuta. Odečteme-li pak rozdíl ve sledované proměnné mezi příjemci a nepodpořenými před intervencí od jejich rozdílu po intervenci, eliminujeme tak část výběrového zkreslení, které je konstantní v čase. Pokud jsou tedy nepozorovatelné rozdíly mezi příjemci a nepodpořenými konstantní v čase<sup>14</sup>, jsme schopni touto metodou získat věrohodný odhad dopadu intervence.

<sup>14</sup> Jde o významnou podmínku při použití této metody.



2) Alternativní způsob vysvětlení, jak pomocí rozdílů v rozdílech odhadnout dopad politiky, je začít od změny pozorované v průběhu času u příjemců podpory. Ani tento rozdíl samotný však nemůže být vykládán jako vliv politiky na sledovanou proměnnou, protože v čase působí kromě intervence i mnoho dalších faktorů a procesů, které mnoho být příčinnou pozorované změny. Způsob, jak vzít tuto přirozenou dynamiku v úvahu, je výpočet změny sledované proměnné i pro nepodpořené subjekty. Následným odečtením změny naměřené u nepodpořených subjektů od změny pozorované u příjemců podpory získáme odhad dopadu intervence. Opět je nutné brát v úvahu předpoklad, že trend vývoje sledované proměnné by byl bez přítomnosti intervence stejný.

Oba způsoby vysvětlení nakonec dospějí ke stejným výsledkům. Používají pouze jinou argumentaci při interpretaci dat. Zatímco první způsob zdůrazňuje výběrové zkreslení, tedy rozdílné charakteristiky podpořených a nepodpořených subjektů a nutnost jeho odstranění, druhý způsob se zaměřuje na eliminaci zkreslení způsobeného přirozenou dynamikou u sledované proměnné. Pro použití metody rozdílu v rozdílech je vyžadováno, aby byla data opakovatelně k dispozici v čase. To znamená, že jsme schopni sledovanou proměnnou měřit opakovaně ve stejných jednotkách v průběhu času a nezávisle na existenci politiky. Data pro tuto metodu mohou být rutinně sbírána například statistickými úřady, ale v některých případech je nutné je získat ad hoc sběrem.

V druhém případě může nastat překážka použitelnosti metody rozdílu v rozdílech pramenící z opomenutí shromáždit potřebná data ještě před spuštěním intervence. Pokud nám tato data chybí, lze využít retrospektivní měření například pomocí dotazníků, které může být spojeno s měřením po intervenci. Zde však hrozí určité zkreslení, budou-li se data pro obě období (před a po intervenci) získávat v rámci jednoho šetření. Pro využití této metody je také nutné, aby byla intervence diskrétní povahy a jasně určovala, které jednotky jsou intervenci vystaveny a které nikoliv. Metoda je obtížně použitelná pro intervence kontinuálního charakteru.

Následuje několik doporučených kroků pro použití metody rozdílu v rozdílech (EVALSED, 2009a):

**1) Definování výstupní proměnné případně proměnných.** Analýza může být provedena pro co největší počet výsledných proměnných, pro které jsou dostupná data.

**2) Definování časového rozměru.** Důležité je stanovit rozsah sledovaných období před intervencí, v průběhu intervence a po intervenci a počet prováděných měření. Musí být zcela obhajitelné, že se vliv podpory projevil skrze aplikaci nově vyzkoumaných procesů a výrobků. Pro posouzení robustnosti výsledků je tedy nutné využít údaje o tom, kolik projektů je již ukončených, případně jak.

**3) Výpočet rozdílů.** Vypočítají se průměrné hodnoty sledované proměnné pro skupinu příjemců a nepodpořených minimálně ve dvou časových obdobích. Následným porovnáním buď v čase (před a po) nebo mezi skupinami (podpoření a nepodpoření) získáme jednotlivé dílčí rozdíly (viz výše). Odečtením těchto dílčích rozdílů získáme odhadovanou hodnotu dopadu intervence.

**4) Předpoklad paralelního vývoje.** Existuje několik způsobů, jak ověřit, že jsou vývojové trendy pro obě skupiny subjektů stejné. Jednou z možností je rozšíření časového rozměru o další měření před intervencí, na kterých se ověřuje paralelnost vývoje sledované proměnné bez intervence. Druhou možností je rozšíření modelu o další data, která ovlivňují výslednou proměnnou a jsou korelovány se získanou podporou. Model tímto však ztrácí svou jednoduchost. Dalším způsobem je testování na skupinách, o kterých víme, že intervencí nebyly ovlivněny.

**5) Použití regrese k replikaci výsledků.** Převedení modelu do regresní podoby sice poskytuje naprosto stejné výsledky, avšak výhodou je, že do pravé strany rovnice lze následně přidávat ostatní proměnné, což umožňuje uvolňovat přísný předpoklad paralelního vývoje.

**6) Zahrnutí kovariací do regrese.** Do modelu se přidávají interakce mezi jednotlivými proměnnými, což přispívá k zpřesňování odhadu dopadu intervence.

Přes svou širokou použitelnost není metoda rozdílů v rozdílech všelékem na hodnocení dopadu intervence. Výhodou této metody je, že k odhadu dopadu nevyžaduje komplexní datové struktury, ale postačují jí pouze souhrnné údaje o výstupech politiky shromážděné před a po intervenci. V praxi je však často získávání dat z období před intervencí nepřekonatelnou překážkou, nejčastěji z důvodu nedostatečného plánování sběru dat. Jednoduchost metody je také vykoupena silným předpokladem neměnnosti trendů sledovaného ukazatele u

podpořených a nepodpořených jednotek. Tento předpoklad lze navíc testovat pouze tehdy, máme-li k dispozici více údajů (tzn. dva a více sběrů dat ukazujících trend před intervencí).

## 4. Možnosti implementace CIE v podmínkách TAČR

### 4.1. Vhodnost využití CIE vůči konkrétním programům TA ČR

Aby mohla být CIE provedena, musí být splněny předpoklady uvedené v obecné části metodiky. Jde především o velký počet případů, homogenitu projektů, dostatečnou datovou specifikaci, dostatečnou proporcionalitu podpory a především volbu správného časového hlediska.

Dle § 5 zákona 130/2002 Sb. je účelová podpora TA ČR rozdělena do programů.

Technicky zaměřený program ALFA si klade za cíl zvýšit výkonnost ekonomických subjektů, konkurenceschopnost českého hospodářství a zlepšení celkové kvality životního prostředí. Toho dosahuje skrze kolaborativní podporu aplikovaného výzkumu a experimentálního vývoje progresivních technologií. Účastníky jsou tedy jak podniky, tak výzkumné organizace a vysoké školy.

Cílem programu GAMA je tyto výsledky připravit pro komercializaci formou „proof-of-concept“. Spojením podnikového a vědeckého know-how vzniká pomyslný most přes selhání trhu zvané „údolí smrti“<sup>15</sup>. Příjemci se tedy typově od programu ALFA neliší.

Cílem programu DELTA je skrze zvýšení hladiny know-how českých podniků dosáhnout vyšší konkurenceschopnosti české ekonomiky. Tento program se buduje na strategických partnerstvích se zahraničními výzkumnými ústavy např. z Vietnamu a Taiwanu na základě memoranda dojednaném Technologickou Agenturou ČR.

---

<sup>15</sup> Údolí smrti (Valley of Death) je pojem první fáze startupu (nebo vývoje výrobku či služby), kdy produkt ještě není v zisku.

Program EPSILON má velmi podobnou intervenční logiku jako program ALFA. „Centra Kompetence“ podporují dlouhodobou spolupráci podnikové a výzkumné sféry. Jde o malé množství projektů<sup>16</sup> s účastníky sloučených do větších klastrů.

Výzkumné potřeby vybraných ústředních orgánů státní správy jsou řešeny v programu BETA a zadávány formou veřejných zakázek.

Hlavním cílem programu OMEGA je posílení výzkumných aktivit v oblasti aplikovaných společenských věd a uplatnění výsledků těchto aktivit pro zvýšení konkurenceschopnosti České republiky, zvýšení kvality života jejích obyvatel a vyvážený socio-ekonomický rozvoj společnosti.

#### 4.1.1. Vhodnost využití CIE v programu ALFA

Nejvhodnější pro hodnocení kontrafaktuální metodou se zdá být program ALFA a to především z důvodu velkého počtu podpořených podniků s dostatečnou úrovní homogenity programu.

V tomto kolaborativním výzkumu tvoří participanty programu průmyslové podniky, výzkumné organizace a vysoké školy. Je nutno zmínit, že výsledná analýza nebere ohled na vytvořenou hodnotu ve výzkumných organizacích. Pro účely analýzy je efekt podpory měřen pouze pro podniky z důvodu jasné interpretace finančních údajů.

Ve všech projektech programu ALFA spolupracují určité kombinace podniků, výzkumných organizací a vysokých škol. Proto není jednoznačné, jak má být interpretována podpora podnikům. Nabízí se tedy rozdělení podpory na přímou a nepřímou. Výše podpory přímé je finanční obnos obdržený podpořenou organizací (obchodní korporací). Podpora nepřímá je finanční obnos vynaložený veřejnou sférou (státní sférou) na celý projekt rozdělený mezi obchodní korporace, které se v projektu účastnily.

Protože je efekt přičítán vyšším výdajům na VaVaI je nutné uplatnit předpoklad, že nedochází k vytlačení investic. Pro testování tohoto předpokladu je nutné disponovat výdaji podniků na VaVaI u obou skupin. Tento údaj prozatím nebyl k dispozici.

Vyšší výdaje na VaVaI mají produkovat nové znalosti. Předpokládáme, že se tyto znalosti volně nešíří od podpořených podniků k nepodpořeným. Jde tedy o

---

<sup>16</sup> První veřejná soutěž CK podpořila 22 projektů, druhá 11.

předpoklad nulového spillover efektu. Tento předpoklad je podepřen vznikem patentů, nebo užitných vzorů na konci projektu.

V programu ALFA je velký počet projektů<sup>17</sup>, ve kterých se účastní dostatečný počet obchodních korporací<sup>18</sup>. Z časového hlediska je možné hodnotit pouze omezený počet případů. V programu ALFA bylo v průběhu roku 2014 administrativně ukončeno a zhodnoceno<sup>19</sup> - 163 projektů. Kontrafaktuální analýza byla prováděna na podpořených podnicích. Ve finálním datasetu je celkem běžnou záležitostí, že se podpořené subjekty účastní ve více projektech. Stejně tak na jednom projektu se často podílí více podniků. Metrika, kolik je v celkovém datasetu podpořených podniků s ukončenými projekty a kolik projektů je probíhajících, je poměrně složitá. V ideálním případě je časový odstup používaný pro měření 3 až 4 roky po ukončení projektu. V takové fázi se nenachází ani jeden podnik podpořený TA ČR.

Z podniků, které se účastnily na větším počtu projektů, má jeden ze sledovaných podniků řádně ukončeny 3 projekty, šest ze sledovaných podniků má řádně ukončeny 2 projekty a dvacet pět ze sledovaných podniků má ukončen jeden projekt<sup>20</sup>.

Další předpoklad pro robustní CIE je relativní proporcionalita. Tedy předpoklad, že velikost intervence TA ČR je vůči velikosti obchodní korporace relevantní. Ve vybraném<sup>21</sup> vzorku 115ti obchodních korporací byla na základě výpočtu proporcionalita podpory z programu ALFA vůči ostatním podporám ze státních účelových programů 17%<sup>22</sup>. Splnění tohoto předpokladu se z tohoto důvodu jeví jako poněkud problematické. Celkový podíl dotací z programu ALFA vůči výkonům<sup>23</sup> provedený na stejném vzorku obchodních korporací byl roven 3,5%<sup>24</sup>.

<sup>17</sup> ALFA celkem 966 podpořených projektů, ALFA VS1 256, ALFA VS2 245, ALFA VS3 168, ALFA VS4 297; Zdroj: IS VAV

<sup>18</sup> ALFA celkem 815 podpořených obchodních korporací, ALFA VS1 252, ALFA VS2 265, ALFA VS3 188, ALFA VS4 320, očištěno od VO podle RVVI, které jsou zároveň soukromé PO

<sup>19</sup> Ukončení a zhodnocení probíhá formou Závěrečného Oponentního Řízení (ZOR)

<sup>20</sup> Platné v měsíci únoru 2015

<sup>21</sup> Pro pilotní CIE TA ČR, výběr na základě participace a plné datové specifikace

<sup>22</sup> Platí pro časové rozmezí 2008-2013, Zdroj: IS VAV, vlastní výpočet

<sup>23</sup> Finanční údaj uváděný na VZZ

<sup>24</sup> Tzn. Celková veřejná podpora na VaVaI vybraných obchodních korporací tvoří cca. 20,5% výkonů a to bez započítání podpory přidělené VŠ a VO na projekty s podporou související.



Závěr: Z hlediska předpokladů program ALFA pro hodnocení kontrafaktuální analýzou splňuje: (1) homogenitu (2) dostatečnou velikost pozorovaného vzorku (3) absenci spillover efektu. Čím kompletnější bude požadovaná datová základna, tím bude možné otestovat i ostatní předpoklady s větší relevancí.

#### 4.1.2. Vhodnost využití CIE v ostatních programech TA ČR

V Cetrech kompetence je podpořeno přibližně 219 podniků. Za předpokladu plné datové specifikace by bylo možné provést CIE i na tomto programu.

Princip intervence je podobný principu intervence programu ALFA, proto může být použito stejného způsobu výpočtu efektu podpory.

Program BETA je velmi heterogenní skupinou veřejných zakázek financovaných prostřednictvím TA ČR. Homogenita je jeden z nejdůležitějších předpokladů pro použití CIE, proto není program BETA pro hodnocení touto metodou vhodný.

V programu GAMA je pouze 28 projektů. Čím menší je vzorek pozorování, tím méně jsou výsledky kontrafaktuální analýzy robustní. Právě proto je i program GAMA pro hodnocení kontrafaktuální analýzou nevhodný. To také podtrhává fakt, že příjemci jsou pouze VO, které není možné hodnotit poměrovými ukazateli finanční analýzy.

Ze stejného důvodu není pro analýzu vhodný ani program DELTA a EPSILON.

Program OMEGA je zaměřen na sociálně vědní výzkum, velká heterogenita projektů znemožňuje měřit efekt podpory jednotlivce, jako např. měří fond dalšího vzdělání u projektu „Stáže ve firmách“. Proto bude hodnocení prováděno spíše nepřímým a pomocí kvalitativních metod.

Závěr: Pro hodnocení kontrafaktuální analýzou je v případě současných programů TAČR možné hodnotit program ALFA a Centra Kompetence. V případě budoucích programů, bude možné hodnotit i programy s cílem zapojit studenty. Pokud by však TA ČR chtěl kontrafaktuální analýzu pro tyto účely využívat, musí to být promítnuto do návrhu programu a způsobu sběru dat.

## 4.2. Metodika CIE pro program ALFA

Tato případová studie nabízí ukázkou možností, jakým způsobem lze aplikovat metodický přístup CIE v podmínkách TA ČR. Konkrétně byl na základě

posouzení základních předpokladů pro evaluaci zvolen program ALFA. Z hlediska metodologického přístupu bylo čerpáno z článků mnoha autorů, např. Abadie, Diamon, Hainmueller (2010), Evropská Komise (2013)<sup>25</sup> a české verze tohoto výzkumu, který na českých datech provedl, Čadil, Mirošík, Petkovová, Zeman, Hartman (2014), Sedláček (2014) a Potluka (2013).

Nejdříve je nutné otestovat zmíněné předpoklady.

### **Velký počet případů**

Velký počet případů je nutný z důvodu relevantnosti výpočtu. Může dojít k výběru, který nebude dostatečně charakterizovat zkoumanou skupinu, v tomto případě všechny účastníky programu ALFA. Hraniční velikostí je tedy přibližně 100 případů. Proporcionální rozdělení velikosti podniků, by mělo odpovídat proporcionálnímu rozdělení všem podnikům v programu ALFA.

### **Homogenní případy**

Pro vyhodnocení homogenity se používá K-S test. Ten je připraven v programu, který je přílohou této metodiky. K-S test porovnává rozdělení podpořené skupiny a kontrolní skupiny. Pokud p-hodnota bude nižší, než 0,05 jsou rozdělení shodná. Tímto je potvrzena, nebo vyvrácena podobnost a homogenita obou skupin.

### **Dostupnost dat**

Tento důležitý předpoklad není nutno testovat. Pokud nejsou dostupná data kontrafaktuální analýza má omezenou vypovídající hodnotu.

### **Proporcionalita**

Proporcionalita byla vypočítána v odstavci 4.1.1. Jde o poměr výkonů, nebo tržeb poměřených vůči dotaci. V tomto případě byly zvažovány dotace podnikům ve sledovaném období poměřené vůči roční položce. Podíl 3,5% vyjadřuje, jak velkou část výkonů tvoří ročně dotace. Za relevantní je možné považovat podíl na výkonech jedním procentem a vyšší. Ideální by bylo porovnat výši dotace vůči VaVaI nákladům. Pro relevanci by dotace měly činnit alespoň 10% nákladů na VaVaI.

### **Nulový spillover efekt**

Silný předpoklad nulového spillover efektu je testován pomocí binární proměnné. Pokud účastník projektu dosáhl nějaké formy duševní vlastnictví

<sup>25</sup> Design and commissioning of counterfactual impact evaluation



v důsledku projektu, je v tomto případě považován předpoklad nulového spillover efektu za uspokojený.

### **Podmínka časového hlediska**

V současnosti je tato podmínka testována pomocí binární proměnné. Pokud bylo provedeno závěrečné oponentní řízení (ZOŘ), je podmínka považována za splněnou. V budoucnosti bude jako proxy použit vyplněný implementační plán, dotazník, který přibližně tři roky po ukončení projektu vyplňují účastníci daného projektu.

### **Nedochází k vytlačení investic**

Tento předpoklad se testuje pomocí kontrafaktuální analýzy. Pro otestování tohoto předpokladu je nutné disponovat podnikovými výdaji na R&D. Vytlačení investic testuje Gonzáles a Pazó (2008). Porovnává podniky nepodpořené a kontrolní a sleduje nárůst VaVaI výdajů. Pokud by se výdaje oproti kontrolní skupině nezměnily, šlo by o stoprocentní vytlačení investic. Pokud by efekt podpory z hlediska VaVaI výdajů byl stejný, nebo větší, než podpora samotná nedocházelo by k vytlačení investic vůbec.

### **Minimální možné výběrové zkreslení**

Minimální možné výběrové zkreslení je zajištěno dobrým párováním. Tento předpoklad upozorňuje na nutnost pochopení výběrového procesu a případné porovnání konkurenceschopnosti podniků po párování. Párování zpřesňují kvantitativní ukazatele jako výdaje na R&D a počet zaměstnanců a kvalitativní údaje jako NACE odvětví podniků a inovační kapacita podniku.

Následujícím krokem je snaha přiřadit k podpořenému podniku co nejpodobnější podnik z kontrolní skupiny. Jedním z kritérií je, že podnik byl také žadatelem do programu ALFA, ale nebyl podpořen<sup>26</sup>. Tento předpoklad zajišťuje dodatečné rozměry podniku.

(1) Skutečnost, že se podnik hlásil o podporu ve stejném programu, signalizuje podobný charakter aplikovaného výzkumu.

(2) Skutečnost, že se podnik hlásil o podporu ve stejném programu, zajišťuje dodatečné informace, které k CIE mohou být využity.

---

<sup>26</sup> Není nezbytnou podmínkou, aby kontrolní skupina byla tvořena z žadatelů o podporu ve stejné veřejné soutěži, nebo programu.

Výsledný kontrafaktuál se musí dále vypořádat s rizikem endogenity. Podpořené podniky mohou být podpořeny, protože byly oproti nepodpořeným lepší<sup>27</sup>. To je v přímém rozporu s tvrzením, že podniky si vedou lépe v důsledku podpory. Tedy jinými slovy, pokud byly podniky vybrány k podpoře, protože byly lepší (nebo horší), nelze prohlásit výsledek kontrafaktuální analýzy za objektivní.

K výběru projektu k podpoření dochází na základě komplexní skupiny kritérií. Proces výběru projektů k podpoření není lehké rozklíčovat na základě kvantitativních proměnných. Z toho důvodu mluvíme o výběrové zkreslení (selection bias). Zjednodušeně proces probíhá tak, že na základě doporučení oponentů<sup>28</sup> rozhoduje PŘTA<sup>29</sup> o tom, které projekty budou podpořeny. Oponenti a PŘTA hodnotí i kvalitativní rozměry, jako například skutečnost, které nejsou zaznamenány a nejsou kvantifikovány. Je tedy nutno daný výběr a párování co nejvíce zpřesnit. Proto by pro párování podniků mělo být použito co nejvíce relevantních proměnných.

Toho je dosaženo prostřednictvím metody Propensity Score Matching. Pro tuto metodu je využit logit model na odhadnutí pravděpodobnosti, podle kterého jsou k sobě podniky přiřazeny.

Vyjádření má následující podobu:

$$\text{Logit}(\mathbb{E}[P_i | \text{Fin}_1 + \text{Fin}_2 + \dots + \text{Fin}_n]).$$

$P_i$  = Binární proměnná [0; 1], 0 = Podnik nepodpořen 1 = Podnik podpořen

$\text{Fin}_n$  = finanční údaj podniku (př. aktiva celkem)

K vyhodnocení efektu je použita metoda Difference in Differences (DiD), jde o následující odečet:

$$(Y_{T2} - C_{C2}) - (Y_{T1} - Y_{C1}).$$

$Y_{T2}$  = Daná proměnná u podpořeného podniku v čase 2

$Y_{T1}$  = Daná proměnná u podpořeného podniku v čase 1

$Y_{C2}$  = Daná proměnná u nepodpořeného podniku (z kontrolní skupiny) v čase 2

<sup>27</sup> Jejich finanční a ekonomické ukazatele byly lepší

<sup>28</sup> Oponent je zkušený expert se zázemím ve třetím sektoru, jehož oblast akademického zájmu je relevantní vůči dané problematice, kterou navrhovaný projekt řeší.

<sup>29</sup> Předsednictvo TA ČR

$Y_{c1}$  = Daná proměnná u nepodpořeného podniku (z kontrolní skupiny) v čase 1

Dále je využit lineární model:

$$(Y_{it} = b_0 + b_1(P_i)t + b_2(P_i) + b_3t + \varepsilon_{it})$$

$(Y)$  závislá proměnná

$i$  po podpoře ( $T2$ ), nebo před podporou ( $T1$ ),

$(P)$  je binární proměnná ( $1 =$  podpořen,  $0 =$  kontrolní), tedy informace zdali byl podnik podpořen

$t$  je časová binární proměnná

parametr  $b1$  vyčísluje efekt podpory vůči konkrétní finanční proměnné

Hodnoty těchto proměnných byly pro zjednodušení logaritmovány.

Pro relevantní výsledky CIE musí být splněny následující předpoklady.

Zprvé, soubor firem musí být homogenní. To je testováno Kolmogorov-Smirnov testem.

Zadruhé, neexistují R&D spillovers, nebo existují v omezené míře. Tento předpoklad je velmi silný, pokud však má být změřený efekt přičten právě R&D výzkumu, nesmí být tyto znalosti používány konkurencí bezprostředně po vynalezení. Tento předpoklad platí, protože před ukončením projektu je vytvořen prvek duševní ochrany (patent, nebo užitný vzor), který zabraňuje volnému šíření použití těchto znalostí.

### 4.3. Konstrukce kontrolní skupiny

Sestavení kontrolní skupiny není triviální a podléhá určitým předpokladům. První nutnost je kompletní datová specifikace kontrolní skupiny ve všech párovaných a dopadových proměnných. V případě ALFA, pokud není dostupný údaj „výdaje na výzkum a vývoj“ je nutné zajistit předpoklad, že subjekty z kontrolní skupiny provádějí výzkum a vývoj. To je v případě ALFA možné tím, že bude kontrolní skupina sestavena z neúspěšných (nepodpořených) žadatelů. Při zajištění potřebných údajů je však možné sestavit kontrolní skupinu z náhodných subjektů, které se shodují kromě finančních údajů i v údajích rozměrových (místo provozu, sídlo) a

v údajích zařazující podnik do daného odvětví (NACE kód). Pro ještě přesnější párování je možné použít dodatečné kvantitativní a kvalitativní údaje jako např. Z-Score a Inovační kapacitu (metodika výpočtu inovační kapacity je specifikována v projektu INKA a je dostupná TA ČR).

#### 4.4. Popis proměnných pro analýzu programu ALFA metodou PSM

Pro přiřazení podniků pomocí Propensity Score Matching domácí (češští) ekonomové Potluka (2013), Čadil (2014) používají především finanční ukazatele firem, na rozdíl od autorů zahraniční literatury. Gonzáles, Pazó (2008) pro vyhodnocení blízkosti podniků (pravděpodobnosti podpory) využívají proměnné, jako např. zda byl podnik v minulosti již podpořen, počet zaměstnanců, růst vlastního kapitálu, stáří podniku, zda je firma exportérem a typ daného průmyslu (NACE). Heterogenita v prováděném výzkumu minimálně podněcuje k zamyšlení. Do jisté míry výzkumná metoda může být určována dostupností dat.

Ačkoliv data využitá španělskými ekonomy Gonzáles, Pazó (2008) jsou v ČR sbírána ČSÚ, není jejich individuální použití vždy samozřejmostí. Podobná omezení byla bariérou i pro kontrafaktuální analýzu prováděnou na případu TA ČR ALFA. Pro jednotnost údajů bylo využito údajů z bisnode databáze (MAGNUS). Datová politika ČSÚ vycházející ze Zákona o státní službě neumožňuje využití výsledků VTR 5-01 dotazníků pro konkrétní subjekty (obchodní korporace) a dotazníku P 5-01. Proto srovnání postrádá nefinanční údaje a údaje o výdajích na VaVal.

Finálně po obdržení všech dostupných údajů byl logit model vytvořen takto:

$$\text{Logit}(E[P_i | TA_{t0} + OA_{t0} + VK_{t0} + V_{t0} + OB_{t0} + PHV_{t0} + HVpZ_{t0}]).$$

Tedy byly využity finanční údaje: Aktiva celkem (TA), oběžná aktiva (OA), vlastní kapitál (VK), výkony (V), obrat (OB), provozní hospodářský výsledek (PHV), hospodářský výsledek před zdaněním (HVpZ). Využití těchto ukazatelů může být při realizaci dalších analýz doplněno o další údaje, které by přesnost přirovnání obchodních korporací zvyšovalo. Popis použitých údajů využitých k provedení kontrafaktuální analýzy je uveden v Tabulka 12.

## 4.5. Popis dostupných dat vhodných pro CIE

Relevantní data pochází jak z interních, tak i z externích zdrojů TAČR. Mezi interní data se řadí IS Patriot a IS Beta. Z důvodu heterogenity projektů, není vhodné program BETA a IS Beta pro účely kontrafaktuální analýzy využívat. V IS Patriot se nachází podrobné údaje o projektech, podpořených projektech a subjektech a RIV výsledcích. Z externích zdrojů jsou data čerpána z databáze Magnus (bisnode). V této databázi jsou uvedeny finanční údaje pro konkrétní podniky za konkrétní rok. Následně je možné využít databáze IS VAV, Úřadu Průmyslového Vlastnictví a patentové databáze PATSTAT. Data z následujících zdrojů je nutno upravit podle popisu v části: „7.1 Požadovaná úprava datasetu“.

### 4.5.1. IS Patriot

Jako jeden ze zdrojů dat slouží finanční a kvalitativní údaje uvedené v podkladech pro hodnocení projektů. Tyto údaje mají sloužit k vyhodnocení finanční stability společnosti, která žádá o participaci v projektu. Finanční údaje dostupné v Patriotu jsou popsány v přílohách. Pokud by byla snaha přiřazování podniků provést pomocí Regression Discontinuity Design, mohou zde být čerpány údaje o počtu obdržených bodů od oponentů projektu před rozhodnutím projekt podpořit.

### 4.5.2. IS VaV

Tato národní databáze shrnuje údaje o všech dotačních titulech v ČR zaměřených na VaVaI. Je zde možné získat údaje o celkových výdajích státního rozpočtu vůči konkrétnímu subjektu. Tento údaj je důležitý pro zasazení výsledného čísla do kontextu. Pro konkrétní subjekty s dostupnými finančními údaji z Magnus databáze tvoří dotace obdržené od TA ČR přibližně 17%. K účelům kontrafaktuální analýzy by mohly být využity i CEPové kategorie výsledků. To však tvůrci této metodiky nedoporučují. Výsledky pro jednu konkrétní obchodní korporaci vykazují vysokou heterogenitu CEPových skupin a jejich kategorizace často neodpovídá jejich NACE zařazení<sup>30</sup>.

### 4.5.3. Magnus databáze (bisnode)

Tato databáze obsahuje údaje o soukromých podnicích, včetně základních finančních údajů. Je však obtížné získat kompletní datovou specifikaci ke každému podniku ke každému požadovanému roku. Plná datová specifikace byla

---

<sup>30</sup> CEPové skupiny jsou stanoveny RVVI a jejich cílem je popsat obor ve kterém se výzkum provádí, př: JE – Nejaderná Energetika. Skupiny NACE (Nomenclature statistique des Activités économiques dans la Communauté Européenne) popisují obory průmyslu podle logického členění 80. let 20. století.

dosažena přibližně u 40% podniků. V této databázi jsou obsaženy údaje i na nepodpořené podniky. Jde tedy o ideální zdroj dat pro kontrolní skupinu.

#### 4.5.4. ČSÚ

Veřejně Český statistický úřad poskytuje pouze agregované údaje. Z podstaty fungování statistického úřadu tato organizace disponuje mnoho informacemi o jednotlivých subjektech. Pro účely kontrafaktuální analýzy se nepodařilo zajistit potřebné informace a další možná spolupráce s ČSÚ je v jednání.

#### 4.5.5. Použité datové položky

Seznam použitých dat v případě programu ALFA je uveden v následující tabulce.

Datová položka	Zdroj dat	Výsledková proměnná
IČO	MAGNUS/IS PATRIOT	NE
Podpořen	IS PATRIOT	NE
Aktiva celkem	MAGNUS	ANO
Oběžná aktiva	MAGNUS	ANO
Vlastní kapitál	MAGNUS	ANO
Výkony	MAGNUS	ANO
Obrat, výnosy	MAGNUS	ANO
PV (Provozní výsledek)	MAGNUS	ANO
HVpZ (Hospodářský výsledek před zdaněním)	MAGNUS	ANO
Obrat	MAGNUS	ANO
Obdržená dotace	IS PATRIOT	NE
Celkové náklady projektu	IS PATRIOT	NE
Státní podpora podniku	IS VaV	NE
Celkové náklady podpory	IS VaV	NE
Státní podpora projektů, kde se PO podílí	IS VaV	NE
Celkové náklady projektů	IS VaV	NE



## 5. Praktická ukázka hodnocení programu ALFA metodou CIE

### 5.1. Výsledky hodnocení programu ALFA metodou CIE

Tato kapitola vyčísluje a interpretuje výsledky pilotní studie CIE. Z důvodu omezenosti interních datových zdrojů TA ČR byla využita data z databáze Magnus pro rok 2008. Rok 2008 je rokem před zahájením intervence. V programu ALFA bylo dosaženo plné datové specifikace u 121 podpořených obchodních korporací. Z toho 73 bylo podpořeno v první veřejné soutěži. Plná datová specifikace pro podpořené v první a druhé veřejné soutěži byla dostupná pro 117 subjektů. Kontrolní skupina sestavená z nepodpořených žadatelů do první veřejné soutěže obsahovala 115 vzorků s plnou datovou specifikací. Druhý kontrolní vzorek obsahující neúspěšné žadatele do programu ALFA obsahuje 672 vzorků s plnou datovou specifikací. Následující poměrové ukazatele byly vypočítány standardním způsobem:

$$ROA = \frac{HV \text{ před zdaněním}}{Aktiva celkem}$$

$$ROE = \frac{\text{Čistý zisk}}{\text{Vlastní kapitál}}$$

$$ROS = \frac{HV \text{ před zdaněním}}{\text{Celkové Tržby (obrat)}}$$

Tyto ukazatele slouží k porovnání výkonnosti podniků před podporou a po podpoře.

Pro otestování homogenity byl použit dvojstranný Kolmogorov-Smirnov test. Pokud je kombinovaná p-hodnota proměnné, vyšší než 0,05 je rozdělení obou porovnávaných skupin považováno za shodné. Výsledek testu je vyobrazen v Tabulce 1. **Předpoklad homogenity je uspokojen v obou případech.**



**Tabulka 1: KS test, p-hodnoty a homogenita vzorku**

K-S test	TA	OA	VK	V	OB	PHV	HVpZ
První kontrolní	0,15	0,26	0,47	0,63	0,44	0,39	0,42
Homogenita	ANO	ANO	ANO	ANO	ANO	ANO	ANO
Druhá kontrolní	0,085	0,52	0,015	0,88	0,38	0,87	0,32
Homogenita	ANO	ANO	NE	ANO	ANO	ANO	ANO

#### 5.1.1. Porovnání konkurenceschopnosti první kontrolní skupina

Z důvodu testování velikosti výběrového zkreslení jsou po párování podniků sledovány poměrové ukazatele. Výběrové zkreslení je těžké kvantifikovat, proto budou oba výsledky poměrových ukazatelů porovnány. Pokud by se prokázalo, že konkurenceschopnější podniky jsou systematicky vybírány pro podporu, nebylo by možné prohlásit vypočtený efekt za relevantní.

**Tabulka 2: Porovnání poměrových ukazatelů před udělením podpory**

Rok 2008	ROA	ROE	ROS
<b>Podpoření</b>	<b>6,56%</b>	<b>8,73%</b>	<b>16,70%</b>
<b>Kontrolní</b>	5,41%	8,43%	11,46%

Porovnání poměrových ukazatelů v roce 2008<sup>31</sup> nasvědčuje tomu, že skupina podpořených podniků byla z hlediska efektivity podniku lepší. Sedláček (2014) ustanovil paralelu mezi konkurenceschopností a návratností aktiv (ROA). **Ze srovnání tedy vyplývá, že skupina podpořených byla již před podporou více konkurenceschopná.**

**Závěr:** V případě první kontrolní skupiny byly podniky před podporou více konkurenceschopné. Srovnání z druhé skupiny však vyvrací, že by podniky byly vybrány k podpoře, protože více konkurenceschopné byly. Z tohoto hlediska je tedy možné považovat výsledky kontrafaktuální analýzy za relevantní.

<sup>31</sup> Rok 2008 je rok před udělení podpory.

### 5.1.2. Porovnání konkurenceschopnosti druhé kontrolní skupiny

Z důvodů uvedených v porovnání konkurenceschopnosti první skupiny byl výpočet poměrových ukazatelů proveden i u druhé skupiny. Protože počet prvků v první kontrolní skupině byl nižší, než počet prvků v skupině podpořených vykazují poměrové ukazatele jiné hodnoty. Výsledky párování mezi podpořenými a druhou kontrolní skupinou lze prohlásit za robustnější.

**Tabulka 3: Porovnání poměrových ukazatelů před udělením podpory**

2008	ROA	ROE	ROS
Podpoření	6,20%	11,07%	14,40%
Kontrolní	<b>6,78%</b>	<b>11,31%</b>	<b>17,06%</b>

Rozdílné výsledky mezi první a druhou kontrolní skupinou naznačují, že výběrová logika není příliš vázána na konkurenceschopnost podniků v projektech<sup>32</sup>. **Kontrolní skupina podniků je v tomto případě více konkurenceschopná.**

Závěr: V porovnání s druhou kontrolní skupinou, byly podpořené méně konkurenčněschopné. Porovnání v první skupině ukazuje, že podpoření vykazovali vyšší poměrové ukazatele. Z toho je možné vyvozovat, že poměrové a finanční ukazatele nebyly předmětem výběru podniků pro podporu. Z tohoto hlediska je možné výsledky výpočtu efektu prohlásit za relevantní.

### 5.1.3. Distribuce PSM, první kontrolní skupina

Při prvním srovnání s menší kontrolní skupinou byla splněna podmínka, že kontrolní skupina obsahuje subjekty, které neúspěšně žádali o podporu v ALFA1 a celkově nebyly podpořeni v jakémkoliv programu TAČR.

Vyčíslení PSM bylo provedeno pomocí binární regrese:

$$\text{Logit}(E[P_i|TA_{t0} + OA_{t0} + VK_{t0} + V_{t0} + OB_{t0} + PHV_{t0} + HVpZ_{t0}]).$$

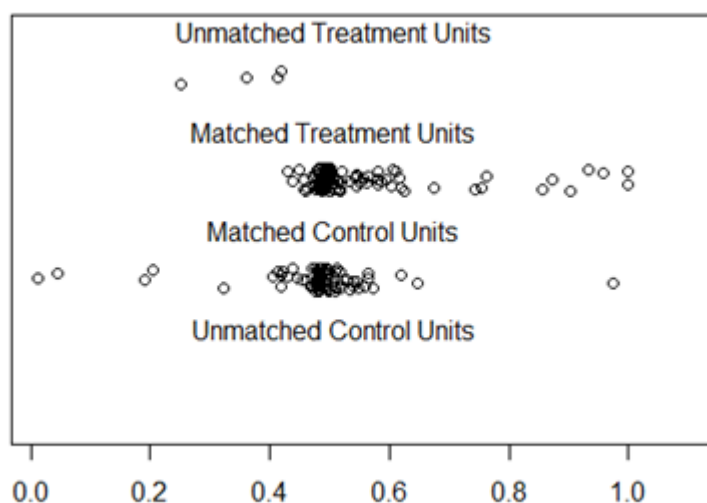
Pro přesnější vyčíslení Propensity Score by bylo vhodné použít i některé dodatečné údaje, které v současné chvíli nejsou dostupné. Jde například o počet

<sup>32</sup> Z finančního hlediska je zkoumána pouze solventnost podniku

zaměstnanců, výdaje na R&D, apod. Současná pilotní studie tedy nenabízí nejlepší možný výpočet Propensity Score. Nevhodné přiřazení podniků se může projevit nižší relevancí výpočtu efektů.

Následující obrázek popisuje párování podniků podle Propensity Score (osa x). Propensity Score je jedním číselným údajem, který v sobě obsahuje veškerá srovnávací kritéria. Jeden kroužek představuje jeden podnik a jeho relativní pozice vůči ose x vyjadřuje Propensity Score konkrétního podniku. Treatment Units vyjadřují podpořené podniky a Control Units kontrolní. Ze srovnání je vidět, že byly vynechány tři podpořené, protože pro ně nebyl nalezen podobný partner.

**Obrázek 3: První kontrolní skupina, přiřazení podpořených a kontrolních obchodních korporací podle Propensity Score**



Závěr: Z technického hlediska párování proběhlo úspěšně. Z důvodu omezené velikosti kontrolní skupiny nebyly však nalezeny vhodné kontrolní podniky pro rozmezí Propensity Score 0,75 – 1.

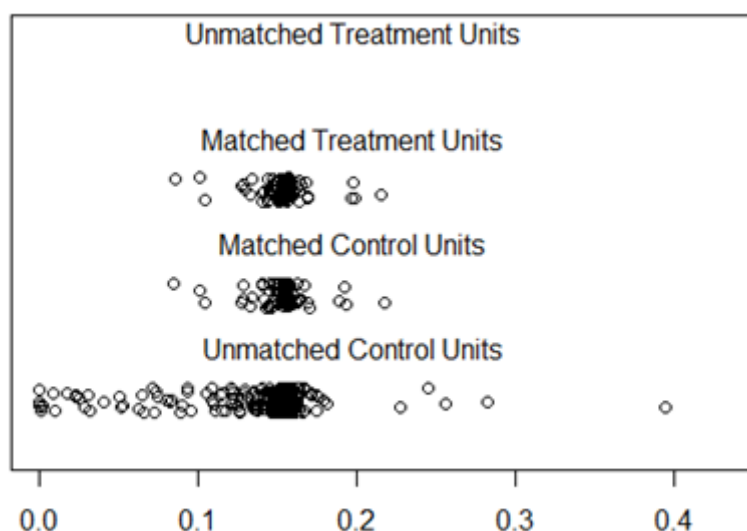
#### 5.1.4. Distribuce PSM, druhá kontrolní skupina

Protože druhá kontrolní skupina obsahovala 672, tedy 5,8 krát více prvků než první kontrolní skupina, Propensity Score poskytlo robustnější srovnání. To je

možné pozorovat na Obrázku 2. Kontrolní skupina obsahuje neúspěšné žadatele do jakékoliv veřejné soutěže ALFA<sup>33</sup>.

Podobně jako v případě předchozího obrázku je na následujícím obrázku možné pozorovat párování podniků. Zde vidíme, že kontrolní skupina podniků byla daleko větší a každý podpořený podnik byl spárován s podnikem kontrolním. Podniky z kontrolní skupiny, které nebyly spárovány, je možné pozorovat mezi „Unmatched Control Units“.

**Obrázek 4: Druhá kontrolní skupina, přiřazení obchodních korporací podle Propensity score**



Závěr: Větší kontrolní skupina poskytla možnost daleko přesnějšího párování z hlediska Propensity Score.

#### 5.1.5. Vypočet efektu, první kontrolní skupina

Metoda rozdílu v rozdílech (DiD) je porovnáním dvou časových období a konkrétních skupin. Na základě první kontrolní skupiny byly vypočteny poměrové ukazatele.

<sup>33</sup> Program ALFA je rozdělena do čtyř veřejných soutěží.

**Tabulka 4: Poměrové ukazatele před a po intervenci, první kontrolní skupina**

	ROA08	ROA13	ROE08	ROE13	ROS08	ROS13
Kontrolní	5,41%	5,43%	8,43%	8,46%	11,46%	11,37%
Podpořen	6,56%	6,27%	8,73%	8,3%	16,7%	16,16%
Rozdíl (P-K)	1,15%	0,85%	0,3%	-0,17%	5,24%	4,78%
<b>DiD</b>		-0,3%		-0,47%		-0,46%

Výsledný efekt DiD je negativní v každém ukazateli podpořených podniků. Z hlediska absolutních hodnot finančních ukazatelů nebyl efekt výrazný. Vývoj poměrových ukazatelů byl sledován v letech 2008 až 2013. Z časového hlediska tedy výzkum nemohl být zpracován. Oproti nepodpořeným se snížila návratnost aktiv o 0,3%, návratnost vlastního kapitálu o 0,47% a návratnost tržeb o 0,46%. **Konkurenceschopnost podpořených podniků se tedy snížila.**

**Tabulka 5: Výpočet efektu, první kontrolní skupina**

	TA	VK	Ob	V	PHV	HVpZ
DiD	1,94E+11	1,24E+10	6,67E+09	8,18E+10	5,58E+09	9,5E+8

Největšího efektu podpory bylo zaznamenáno u celkových aktiv (TA), zatímco nejnižší vliv byl zaznamenán u hospodářského výsledku před zdaněním (HVpZ). I přesto, že celkový vliv podpory je kladný je nutno poznamenat, že výsledky aplikovaného výzkumu jsou středně až dlouhodobou záležitostí, proto plný efekt lze očekávat později než v roce skončení projektu.

**Tabulka 6: Vyjádření růstu podle konkrétních finančních ukazatelů**

	TA	VK	Ob	V	PHV	HvpZ
Podpořená	22,05%	22,90%	20,60%	3,67%	5,68%	16,73%
Kontrolní	3,72%	3,51%	4,77%	-7,28%	0,34%	4,01%
<b>Efekt</b>	<b>18,34%</b>	<b>19,39%</b>	<b>15,83%</b>	<b>10,95%</b>	<b>5,34%</b>	<b>12,73%</b>

Předchozí tabulka porovnává růst vybraných finančních ukazatelů u obou skupin. První kontrolní skupina nevykazovala přílišný růst finančních ukazatelů, v celkovém součtu se této kontrolní skupině výkony propadly o 7%, je nutno poznamenat, že roky 2009 až 2013 jsou v případě ČR krizovým obdobím. Je tedy

možné spekulovat, že vyšší agilita podniků ve schopnosti získat veřejnou podporu vyřešila některé jejich krátkodobé finanční problémy.

**Tabulka 7: Vypočtený efekt růstu poměrových ukazatelů (Očištěný)**

	TA	VK	Ob	V	PHV	HVpZ
DiD	22,81%	24,70%	25,22%	18,87%	9,85%	7,07%

Řádek DiD vyjadřuje poměr “efektu podpory” vůči danému ukazateli (TA = Celková Aktiva; VK = Vlastní kapitál; Ob = Obrat; V = Výkony; PHV = Provozní hospodářský výsledek; HVpZ = Hospodářský výsledek před zdaněním. Jde tedy o vyjádření velikosti daného efektu. V případě Obratu (Ob) efekt podpory dosahuje 25% původního obratu. Základ tohoto podílu tvoří průměr obou skupin.

Závěr: Efekt podpory je v absolutních hodnotách finančních ukazatelů pozitivní ve všech sledovaných ukazatelích, ale poměrové ukazatele u podpořených podniků se oproti nepodpořeným snížily ve všech případech. To ve výsledku znamená, že podpořené podniky jsou na tom v absolutních číslech lépe, než nepodpořené, ale jejich efektivita se snížila. Celkový efekt podpory je v případě obratu roven 25% a v případě zisku před zdaněním 7% těchto finančních ukazatelů. Podpořené podniky po sledované období 2009 až 2013 vykazovaly daleko vyšší růst než podniky nepodpořené. Konec tohoto období, rok 2013, je rokem ukončení projektů, proto nelze výsledky považovat za finální.

#### 5.1.6. Vypočet efektu, druhá kontrolní skupina

V druhém případě je kontrolní skupina větší a předpoklad žádosti ve stejné veřejné soutěži neplatí. Předpoklad homogenity byl zamítnut v případě jednoho finančního ukazatele, tedy vlastního kapitálu. Z toho důvodu poměrový ukazatel návratnost kapitálu (ROE) nebude nejspíš relevantní.

**Tabulka 8: Poměrové ukazatele před a po intervenci, druhá kontrolní skupina**

	ROA8	ROA13	ROE8	ROE13	ROS8	ROS13
Kontrolní	6,78%	5,91%	11,31%	11,20%	17,06%	12,61%
Podpořen	6,20%	5,72%	11,07%	9,37%	14,40%	15,14%
Rozdíl	-0,58%	-0,19%	-0,24%	-1,84%	-2,66%	2,53%

DiD		0,40%		-1,60%		5,18%
-----	--	-------	--	--------	--	-------

Tabulka 8 obsahuje výsledný efekt vypočítaný metodou Difference in Differences v absolutních hodnotách. Z tabulky je patrný pozitivní vliv intervence na některé finanční ukazatele celého vzorku. Poměrové ukazatele vykazují rozdílný trend. Návratnost aktiv u podpořené skupiny klesla. Tento pokles byl však nižší než u skupiny kontrolní. Z toho důvodu je výsledný výpočet DiD v ukazateli návratnosti aktiv kladný. Vnitřní dynamika nepodpořených subjektů (obchodních korporací) nemusí odpovídat dynamice podpořených. V případě vlastního kapitálu, není splněna podmínka homogenity (testovaná K-S testem), proto výsledky této proměnné nelze považovat za směrodatné.

**Tabulka 9: Výpočet efektu, druhá kontrolní skupina**

	TA	VK	Ob	V	PhV	HVpZ
DiD	6E+09	1,63E+10	-1E+10	-2,9E+09	-5,9E+09	1,05E+09

Největším efektem podpory v absolutních hodnotách byl vlastní kapitál (VK). V tomto rozměru vyšel K-S test negativně a proto není možné považovat vypočtený efekt vlastního kapitálu za relevantní. Efekt podpory vyšel negativně v obratu, výkonech a provozním zisku. To může být způsobeno měřením v roce ukončení projektu. Intervence v oblasti VaVaI se projeví s 3-4 letým odstupem.

**Tabulka 10: Vyjádření růstu podle konkrétních finančních ukazatelů**

	TA	VK	Ob	V	PHV	HvpZ
Podpořená	17,54%	28,28%	3,25%	3,95%	5,63%	8,53%
Kontrolní	5,01%	-7,63%	23,77%	5,63%	10,34%	-8,49%
Efekt	12,53%	35,91%	-20,53%	-1,67%	-4,71%	17,02%

Největším efektem podpory v absolutních hodnotách byl vlastní kapitál (VK). V tomto rozměru vyšel K-S test negativně a proto není možné považovat vypočtený efekt vlastního kapitálu za relevantní. Efekt podpory vyšel negativně v obratu, výkonech a provozním zisku. To může být způsobeno měřením v roce ukončení projektu. Intervence v oblasti VaVaI se projeví s 3-4 letým odstupem.



Tabulka 10 vyjadřuje růst finančních ukazatelů souboru firem. Podpořená skupina rostla rychleji z hlediska celkových aktiv. Vlastní kapitál z důvodu testu homogenity nelze považovat za směrodatný. Z hlediska růstu obrátu podpořené podniky rostly o dvacet procent méně. Zisk před zdaněním rostl však rychleji o 17%.

**Tabulka 11: Vypočtený efekt růstu poměrových ukazatelů (Očištěný)**

	TA	VK	Ob	V	PHV	HVpZ
DiD	5,8%	28,3%	-23,2%	-3,3%	-6,9%	17,5%

Závěr: Efekt podpory v absolutních hodnotách je pozitivní a to s výjimkou obrátu, výkonů a provozního zisku. Efekt podpory v programu ALFA je z hlediska návratnosti aktiv (ROA) a návratnosti tržeb (ROS) pozitivní a návratnosti vlastního kapitálu (ROE) negativní. Ve srovnání s více širokou skupinou, kde je zajištěno přesnější párování se podpořeným podnikům zlepšila konkurenceschopnost, ale snížila se jim ziskovost vůči vlastnímu kapitálu. Z hlediska absolutních údajů se podpořeným zisk zvýšil. Z porovnání tedy vyplývá, že podniky považují za důležitější přístup ke zdrojům, které zprostředkovává TAČR oproti návratnosti vlastního kapitálu

## 6. Poučení a závěry

### 6.1. Závěry z hlediska využití CIE

Kontrafaktuální analýzu je možné v případě TA ČR technicky provést, je však nutné zajistit dodatečná primární data společností, a to především výdaje na R&D s co nejvyšší objektivností, aspirace podniku<sup>34</sup>, počet výzkumných pracovníků apod. Nejen z tohoto důvodu je vhodné analýzu stále vyvíjet a uzpůsobovat rozšířené datové základně.

Na základě této metodiky bude možné, aby výpočet efektu prováděl zaškolený pracovník. Příprava dat a kontrola předpokladů, kromě homogenity, bude muset být provedena manuálně.

Kdyby byly vybírány k podpoře podniky, které by byly na základě sledovaných kritérií lepší, nebylo by možné tvrdit, že jsou podniky lepší v důsledku dané intervence. V rámci pilotní studie vyplynulo najevo, že konkurenceschopnost podniků nehraje důležitou roli ve výběru podniku pro podporu. Z tohoto hlediska je tedy možné prohlásit vypočtený efekt za relevantní.

Z realizované pilotní studie vyplývá, že efekt podpory je z hlediska celkových aktiv, vlastního kapitálu a zisku před zdaněním pozitivní. Návratnost vlastního kapitálu se však u podpořených podniků snížila. Tento jev byl způsoben vyšším růstem vlastního kapitálu oproti zisku.

V rámci vývoje kontrafaktuální analýzy by bylo vhodné stále rozšiřovat v tomto ohledu znalostní základnu TA ČR, revidovat výsledky na nových datech, zkoušet jiné metody a kontinuálně upravovat automatizovaný systém pro širší datovou základnu.

### 6.2. Poučení

Hlavní poučení: Kvalita a dostupnost dat je základním předpokladem pro kvalitní CIE. Je třeba sbírat data o podpořených podnicích průběžně, a to tak, aby byla analytikům přístupná.

---

<sup>34</sup> Aspirace podniku být inovačním lídrem, následovníkem, nebo optimalizátorem

Data, která musí být dostupná každoročně: Vlastní Kapitál, Zisk před zdaněním, Přidaná hodnota, Aktiva celkem, Výdaje na R&D (interní/externí), počet zaměstnanců, počet R&D zaměstnanců. Bylo by také možné používat některé kvalitativní ukazatele, např. zda je společnost z hlediska inovační kapacity světový leader, nebo následovník atd. V rámci českých účetních standardů nejsou tyto výdaje na R&D součástí finančních výkazů (na rozdíl od IFRS). Veřejná instituce má možnost získat tento údaj měřený prostřednictvím dotazníku VTR 5-01 v gesci ČSÚ. Z hlediska sběru dat by bylo ideální získat tyto údaje, případně požadovat finanční výkazy podle IFRS a tyto datové požadavky promítnout do implementačního plánu, který TA ČR zasílá účastníkům projektů přibližně rok po jeho ukončení.

Hlavní poučení: Je třeba rozvíjet spolupráci s ČSÚ pro získání dat pro hodnocení, plus je nutná harmonizace sběru dat s ČSÚ, aby data mohla být použita pro výpočet efektu.

Hlavní poučení: V ČR je v oblasti VaVaI podpořeno cca 2410 subjektů z 2500 kontinuálně provádějících VaVaI (ČSÚ), tzn. podpora je udělena plošně téměř každému, kdo kontinuálně provádí VaVaI. Vliv veřejné podpory VaVaI skrze podniky podpořené v programu xy vůči nepodpořeným v programu xy v zásadě poskytuje srovnání podniků podpořených v programu xy vůči podnikům podpořeným v programu xz. Efekt podpory TA ČR v této pilotní studii tedy neodpovídá na otázku, zda je veřejná podpora VaVaI výhodná pro růst HDP, zaměstnanost atd. protože vždy porovnáváme podpořené s podpořenými.

Vedlejší poučení: Stoprocentní automatizace výpočtu je možná především v kombinaci výpočtu PSM a DiD, ostatní přístupy by mohly být problémové. Je však nutné, aby předpoklady a dataset byl sestaven analytikem, který rozumí kontextu problematiky.

## 7. Realizace CIE v softwaru R

Tato část slouží jako praktický popis, jak získat požadovaná data, jak kontrafaktuální analýzu zhotovit a jak interpretovat výsledky. Praktická část se omezuje pouze na kombinovanou metodu Propensity Score Matching a Difference in Differences. Shrnutí postupu v bodech:

- 1) Úprava dat
- 2) Stažení a Instalace R
- 3) Instalace balíčků v R
- 4) Spuštění programu a import dat
- 5) Zpětná úprava dat na základě přehledu ve výsledku
- 6) Finální výpočet a interpretace

### 7.1. Požadovaná úprava datasetu

Tento úsek popisuje jak data upravit a vložit do programu (v přílohách), současně byla po jeho spuštění vyexportována kontrafaktuální analýza. Potřebná struktura dat závisí na napsání programu v Rstudiu. V tomto konkrétním případě program očekává vstupy těchto finančních proměnných: Aktiva celkem (TA), oběžná aktiva (OA), vlastní kapitál (VK), výkony (V), obrat (OB), provozní hospodářský výsledek (PHV), hospodářský výsledek před zdaněním (HVpZ). Ve sloupci (P) musí být uvedeno, zda byl subjekt podpořen Technologickou Agenturou. Místo navrhovaných údajů můžou být použity jiné proměnné, např počet zaměstnanců apod.

- Pro aktiva celkem bude počáteční rok označen „TA0“ a poslední rok „TA1“. Záleží na pojmenování sloupce, ne na jejich pořadí.
- Po vytvoření těchto sloupců musí být tabulkový soubor převeden do csv.
- Protože se software R problematicky vyrovnává s čárkami (které v českém pojetí fungují jako desetinná čárka) z české verze MS Office Excel, je vhodné údaje zaokrouhlit na nejbližší celé číslo. V tomto je důležitá důslednost, protože soubory \*.csv někdy desetinnou stránku skrývají apod. K tomu je vhodné použít funkci v Excelu ROUND (aj Excel), nebo ZAOKROUHLIT (čj. Excel).
- Název s obsahem nesouvisí, proto v TA0 a TA1 může být místo celkových aktiv klidně počet zaměstnanců.
- Software R je “key sensitive“, to znamená, že registruje rozdíl v malém a velkém písmenu, proměnné apod. musí tedy být naprosto shodné.

**Obrázek 5: Příklad tabulky v softwaru R**

P	TA0	TA1	OA0	OA1
0	104000000000	77064000000	17361000000	12046000000
0	102000000000	76920892000	9276191000	3390994000

## 7.2. RS studio, úvod

Tento úsek by měl sloužit jako popis a seznámení se softwarem, v kterém bude kontrafaktuální analýza zhotovena. RS studio je rozšířená verze statistického softwaru R. Jde o freeware, proto program v této metodologii (v její příloze) bude moc využít kdokoli bez jakýchkoliv poplatků. Pro vytvoření nějakého relevantního (vlastního) ekonometrického programu je nutno mít základní (nebo mírně pokročilou) znalost ekonometrie. Pro plné porozumění programu R je nutná znalost angličtiny. Na jakýkoliv případ programování v R lze nalézt návod (v angličtině).

### 7.2.1. Kde RStudio stáhnout a jak nainstalovat?

Pro instalaci RStudio je také třeba nainstalovat samotný software R. Ten je možné stáhnout v odkazu 2). Autoři metodiky doporučují nejdříve nainstalovat RStudio, které je možné stáhnout v odkazu 1). Pro hladké fungování je ještě vhodné přinstalovat Java, tak aby se shodoval s operačním systémem. (Windows x64, Java x64; defaultně je Java nainstalovaná x86)

1) RStudio: <https://www.rstudio.com/products/rstudio/download/>

2) R: <https://cran.rstudio.com/bin/windows/base/release.htm>

3) Java x64bit: <http://www.oracle.com/technetwork/java/javase/downloads/jre8-downloads-2133155.html> (default je x86 a to někdy dělá problémy)

**Instalace Java se musí shodovat s operačním systémem. Pokud je tedy windows x64 Java musí být nainstalována také ve vydání x64. Automaticky se instaluje Java x86 a to může činit určité problémy pro vytvoření výstupu do excelu.**

### Jak pustit program v prostředí R?

Na základě programu v prostředí R se provádí konkrétní analýza, regrese, atp. Pro tento případ kontrafaktuální analýzy není třeba psát žádný program, stačí

pouze spustit program přiložený, nebo do prázdného pole (po otevření RStudio) zkopírovat text v přílohách označený jako „Program kontrafaktuální analýzy“.

Pokud by však analytik toužil po vlastním sestavení kontrafaktuální analýzy v softwaru R, je toto nejvýstižnější a nejpochopitelnější návod na internetu: <http://pareonline.net/getvn.asp?v=19&n=18>

**Výpočet je spuštěn označením celého textu programu (Ctrl+A) a stisknutím spuštění programu (Ctrl+Enter). Celý program je také možno spustit pomocí označení celého textu a stiskem „Run“ na horní liště okna programu vpravo.**

### 7.2.2. Instalace balíčků

R jako software využívá ke svému fungování balíčky, které rozšiřují funkční pole softwaru. Pro fungování programu kontrafaktuální analýzy je nutné nainstalovat následující balíčky. Tyto balíčky jsou: ("Matching"; "MatchIt"; "rJava"; "xlsxjars"; "xlsx"). Instalování balíčku je možné v rozhraní RStudio pomocí označení textu pouze „install.packages ("Matching")“ a stisknutím Ctrl+Enter. Je doporučováno každý balíček instalovat zvlášť a dosledovat zda instalace proběhla v pořádku. To lze zjistit z konzole (Obrázek 7: Potvrzení úspěšného nainstalování balíčku v konzoli).

#### **Obrázek 6: Instalace balíčku v rozhraní RStudio**

```

RStudio
File Edit Code View Plots Session Build Debug Tools Help
Go to file/function

TA CR CIE.R * TACRCIE * Dataset * DataReview * DiD * KStest *
Source on Save

1 # Name: TACR CIE
2 # Author: Petr Horak
3 # Date: 26 May 2015
4 # Goal:
5 # 1) Dataset review
6 # 2) K-S test of homogeneity
7 # 3) Propensity Score Matching 1:1
8 # 4) After-peering K-S test of homogeneity
9 # 5) Contrafactual analysis
10 # - take weighted averages
11 # - Average causal effect = Sum (Supported Yx - UnsupportedYx)
12 # (compare difference in means in certain strata = estimate in certain value)
13 # 6) Difference in Difference
14 # 7) Outcome summary to be printed into excel
15 #
16 #-----
17 INPUTS:
18 TAX=0.19 #Corporate tax is 19%
19
20 # Match packages: install.packages ("Matching"); install.packages ("zeig"); https://www
21 # install.packages ("MatchIt")
22 # install.packages ("rJava")
23 # install.packages ("xlsxjars")
24 # install.packages ("xlsx")
~

Console ~/
Error in install.packages : updating loaded packages
> install.packages("MatchIt")
trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.2/MatchIt_2.4-21.zip'
Content type 'application/zip' length 78265 bytes (76 KB)
downloaded 76 KB

package 'MatchIt' successfully unpacked and MD5 sums checked

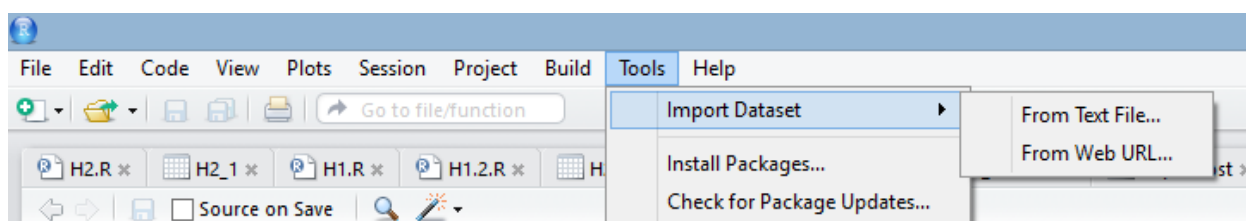
The downloaded binary packages are in
C:\Users\petr.horak\AppData\Local\Temp\RtmpMTAVQA\downloaded_packages

```

**Obrázek 7: Potvrzení úspěšného nainstalování balíčku v konzoli**

### 7.2.3. Přesunutí datasetu do programu

Po zdařilé instalaci program je třeba nahrát dataset v požadované podobě.



Data musí mít specifickou (textovou) podobu, proto musí být importována v csv formátu (tabulka oddělovaná středníkem, nebo čárkou). Druhá možnost importu dat je možná prostřednictvím příkazu přímo v programu: `mydata<-read.csv("C:\xx\data.csv")`, data ovšem musí být také strukturována v podobě csv. Příklad:



```
TACRCIE <- read.csv("C:/.../TACRCIE.csv", sep=";", dec=",")
```

TACRCIE, je místo kam se dataset uloží a tím pádem je možné s ním v R pracovat.

<- je příkaz, “ulož do”; read.csv, je příkaz “načti CSV”.

Podrobnosti v následujícím odkazu: <http://www.r-tutor.com/r-introduction/data-frame/data-import>

Program bude nastaven na předdefinovanou složku, proto není potřeba s programem nijak manipulovat.

### 7.2.4. Úprava datasetu

Pro početně správný výpočet je třeba zajistit správnou úpravu datasetu. Protože tato analýza byla vypočítána na základě dat z MAGNUS databáze bude použit tento příklad.

#### 1) Propojení datasetu

Nejdříve je nutné si opatřit data vhodná pro kontrafaktuální analýzu. Nezbytným údajem je, zda byl podnik podpořený (vyjádřeno binárně 1 = podpořený, 0 = nepodpořený), to je možné ověřit v interních datových zdrojích TA ČR (IS PATRIOT), nebo ve veřejném datovém zdroji pro ČR (ISVAV). Podrobný popis požadované podoby datasetu je v následujícím obrázku:

**Obrázek 8: Podoba datasetu pro výpočet**

	A	B	C	D		O	P
1	IČO	P	TA0	TA1		HVpZ0	HVpZ1
2		n	48000	2254000		-130000	-91000

První řádek musí obsahovat popisy sloupců

P; TA0;TA1;OA0;OA1;VK0;VK1;V0;V1;OB0;OB1;PHV0;PHV1;HVpZ0; HVpZ1

Významově jde o Aktiva celkem (TA), oběžná aktiva (OA), vlastní kapitál (VK), výkony (V), obrat (OB), provozní hospodářský výsledek (PHV), hospodářský výsledek před zdaněním (HVpZ), přiřazená 0 znamená v čase před podporou a přiřazená 1 znamená v čase po ukončení programu (TA1 = Aktiva celkem po podpoře). Jeden řádek je jedním sledovaným podnikem.

#### 2) Desetinné čárky

Protože R je poměrně archaický program a kvůli faktu, že desetinné čárky jsou jinak zaznamenávány v US (0.1) a v ČR (0,1) je nutné dané hodnoty zaokrouhlit. To může být provedeno pomocí jednoduché funkce v Excelu „Zaokrouhlit“.

### 3) Vlastnosti datasetu, outliers, nesmyslné údaje

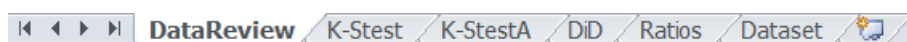
Z hlediska obsahového je nutno zajistit smysluplnost dat. K tomu napomáhá 7.3.1 Přehled dat. Je nutné zajistit, aby v daném datasetu nebyly nesmyslně záporné hodnoty u veličin u kterých to není možné (všechny kromě zisku před zdaněním a provozním zisku). Pokud jsou např. Aktiva celkem záporná je nutné podnik vyřadit. V přehledu dat jsou uvedeny minimální hodnoty. Pokud je tedy program spuštěn jsou po jeho dokončení záporné hodnoty odhaleny a je třeba zpětně o tyto podniky dataset očistit. (Vymazání nesmyslných outlierů je v souladu se statistickou metodologií).

### 4) Subjekty podpořené z jiných veřejných zdrojů než TA ČR

Velmi často dochází k tomu, že podniky jsou podpořeny i z jiných veřejných zdrojů než pouze z TA ČR. Tyto podniky není možné vynechat (omittovat), protože jev je příliš častý. Pro tento případ je v této metodice zmíněn předpoklad „proporcionalita“. Pro určitou míru relevance by dotace TA ČR měla tvořit alespoň 1% výkonů.

## 7.3. Interpretace výsledků

Výsledky budou z programu exportovány do xls souboru, kde bude více záložek. V každé záložce se nachází jeden z předpokládaných výstupů. Ty budou v následující části popsány a interpretovány. Výsledek je výstup programu v podobě xls souboru, který obsahuje více listů (DataReview.xls) a vytvoří se v adrese uvedené v proměnné **OUTfile**. **Nutno nastavit**. „C:/.../DataReview.xls“. Výstup obsahuje 6 listů.



### 7.3.1. Přehled dat, DataReview

Tabulka popisující dataset se nachází na listu výsledného dokumentu jménem DataReview, Mean = Průměr, StD = Standard deviation, Min = Minimální hodnota, Max = Maximální hodnota, TA, OA, VK = proměnné které jsou

popisovány, jako aktiva celkem atd.,  $P = 0$  je nepodpořený podnik,  $P = 1$  je podpořený podnik. Následující obrázek obsahuje tabulku.

**Obrázek 9: Popisná tabulka vloženého datasetu do R**

	A	B	C	D	E	F	G	H	I
1		MeanP0	MinP0	MaxP0	SDP0	MeanP1	MinP1	MaxP1	SDP1
2	TA	1,11E+09	48000	1,04E+11	6,89E+09	5,81E+08	261000	6,6E+09	1,24E+09
3	OA	3,81E+08	48000	2,28E+10	1,7E+09	2,98E+08	191000	4,48E+09	6,83E+08
4	VK	6,31E+08	-4,7E+07	7,82E+10	4,63E+09	3,25E+08	-1192000	6,09E+09	8,41E+08
5	V	1,03E+09	56000	1,02E+11	5,44E+09	6,03E+08	169000	6,62E+09	1,16E+09
6	OB	4,63E+08	-928000	4,04E+10	2,5E+09	2,5E+08	190000	4,75E+09	6,44E+08
7	PHV	9,26E+08	22000	9,82E+10	5,13E+09	5,55E+08	139000	6,16E+09	1,07E+09
8	HVpZ	86818112	-9E+08	1,54E+10	7,28E+08	35998271	-1,9E+08	4,75E+08	86415104

### 7.3.2. Kolmogorov Smirnov test, K-Stest, K-StestA

K-S test porovnává rozložení dvou datasetů. V tomto případě byl porovnáván dataset podpořených a nepodpořených. Pokud je výsledná p-hodnota K-S testu pro oba datasety vyšší než 0,05 můžeme považovat rozdělení za shodná, tudíž předpokládáme homogenitu. Na následujícím obrázku vidíme, že v případě proměnné VK nebylo homogenity dosaženo, protože daná hodnota je nižší než 0,05 ve sloupci K-S test value (p-hodnota testu). Program tuto vlastnost vyhodnotil a výsledný verdikt vypsál do následujícího sloupce. Budou provedeny dva K-S testy, první před přirovnání pomocí PSM metody a druhý po přirovnání. Pro vyhodnocení je homogenita velmi důležitá, pouze homogenní veličiny (po přirovnání PSM) lze považovat za relevantní. Pro ověření správnosti párování je zařazen K-S test před (K-Stest) i po (K-StestA) párování. Pokud není některá proměnná homogenní, není možné výstup této proměnné dále interpretovat.

**Obrázek 10: Kolmogorov Smirnov test, p-hodnoty pro oba datasety a verdikt**

	A	B	C
1		K-S test value	Homogeneity
2	TA	0,084970846	YES
3	OA	0,519608376	YES
4	VK	0,014557789	NO
5	V	0,875145807	YES
6	OB	0,376722267	YES
7	PHV	0,836118375	YES
8	HVpZ	0,315698784	YES

### 7.3.3. Efekt vyčíslen pomocí metody Difference in Difference, DiD

V této části se nachází velmi očekávané výsledky, totiž odpověď na otázku zdali si vybraná skupina podpořených vede lépe oproti skupině nepodpořených (v konkrétním programu TA ČR, ALFA). Kvantifikace efektů je vypočtena pro srovnávací proměnné. Protože tento postup je někdy nevhodný, viz. Gonzáles, Pazó (2008), kde srovnávací proměnné jsou jiné od pozorovaných, na kterých se vypočítává efekt DiD, je možno na následných datech vypočítat efekt ručně podle návodu v části 7.3.5 Dataset. Sloupec „Effect“ popisuje velikost v absolutních číslech vypočítaného efektu pro celý dataset. Podpořené si tedy v případě celkových aktiv vedli o  $7,61 \cdot 10^9$  lépe, než skupina nepodpořená. To je v případě datasetu přibližně o 7% Celkových aktiv (sloupec ratio). Pokud jsou předpoklady dobře zajištěny, je možné výsledek interpretovat tak, že v důsledku podpory si podpořené podniky vedli o 7% lépe v aktivech a o 25% lépe v zisku.

**Obrázek 11: Vyčíslení efektu pomocí metody difference in differences**

	A	B	C
1		Effect	Ratio
2	TA	7,61E+09	0,073463
3	OA	2,24E+09	0,04473
4	VK	1,75E+10	0,302758
5	V	-2,9E+09	-0,03176
6	OB	-1E+10	-0,2223
7	PHV	-6,3E+09	-0,07159
8	HVpZ	1,56E+09	0,255488

### 7.3.4. Poměrové finanční ukazatele

Výsledný efekt poměrových finančních ukazatelů je popsán v záložce „Ratios“.

**Obrázek 12: Výsledný efekt poměrových ukazatelů**

	A	B	C	D
1		DiDROE	DiDROA	DiDROS
2	1	-0,009	0,007234	0,060923

Nachází se zde výpočet efekt poměrových ukazatelů. Obrázek 12 zachycuje výsledný výstup. Jednou z interpretací by mohlo být, že podpořené obchodní korporace v programu ALFA oproti nepodpořeným (odmítnutým) vykázaly pozitivní změnu v návratnosti aktiv o 0,7% a návratnosti tržeb o 6,1%, naopak vykázaly propad ve vývoji návratnosti vlastního kapitálu. Tyto tři veličiny jsou považovány akademiky Viaene (1998), Skandalis, K. (2010) a Meric et al (2011)

za **konkurenceschopnost podniku**. Ve výsledku tedy vidíme, zda došlo ke zvýšení konkurenceschopnosti a v jakém přibližně měřítku.

### 7.3.5. Dataset

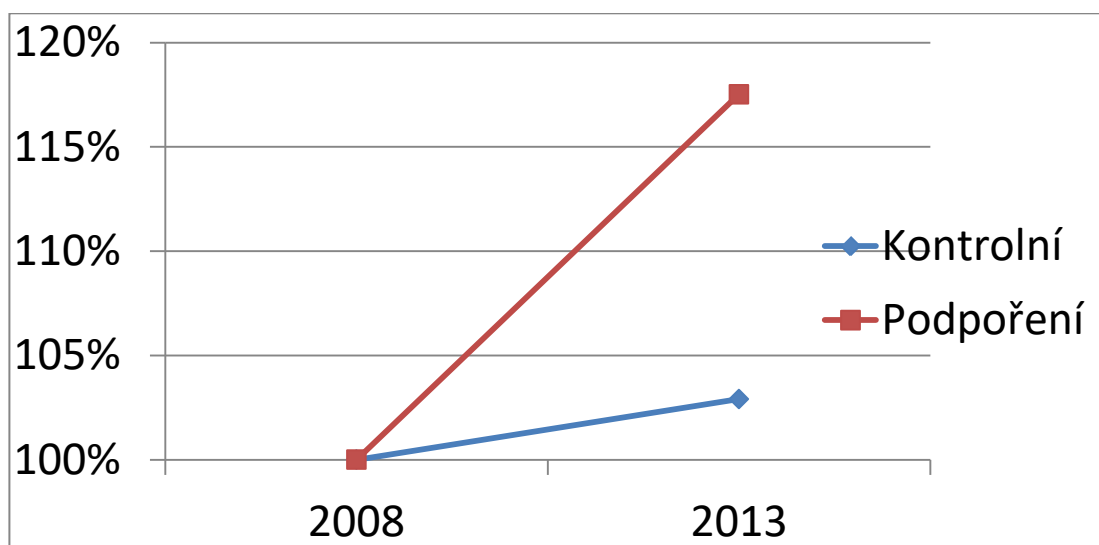
Pro případ již uvedený, tedy kdyby bylo nutno srovnávat obchodní korporace podle jiných proměnných, než podle těch, u kterých chceme vyčíslit výsledný efekt, je ve výstupech programu také výsledný srovnávací dataset. V původním datasetu mohou být i další sloupce kromě zmíněných povinných sloupců. Každá podpořená obchodní korporace je přirovnána ke svému nejbližšímu nepodpořenému protějšku. Výsledkem je tedy dataset na kterém může uživatel programu provádět samostatné analýzy, které nejsou částí programu. Proto za proměnné TA,OA,VK... mohou být použity proměnné, které nenabývají hodnot celkových aktiv, oběžných aktiv a vlastního kapitálu, obchodních korporací, ale uživatelem vybraných rozměrů třeba po vzoru Gonzáles, Pazó (2008). Výsledek může být následně vypočten podle metodiky DiD, jak je popsáno v části 2.4. Použitá metodologie pro případ TA ČR ALFA.

**Obrázek 13: Ukázka datasetu**

	A	B	C	D	E	F	G	H	I
1		ICO	P	TA0	TA1	OA0	OA1	VK0	VK1
2	3		0	6,24E+10	5,76E+10	1,08E+10	7,41E+09	5,21E+10	3,98E+10
3	7		0	2,36E+10	2,74E+10	2,28E+10	2,72E+10	2,37E+09	3,46E+09
4	15		1	6,6E+09	8,35E+09	4,27E+08	6,76E+08	6,09E+09	7,73E+09
5	16		1	6,34E+09	8,17E+09	4,48E+09	6,75E+09	1,59E+09	1,34E+09

Pomocí datasetu je také možné vytvořit jednoduché přehledy, které mohou být převedeny do grafického vyjádření. Převedením do kontingenční tabulky v excelu je následně možné vypočítat rozdíly pro podpořené a nepodpořené. Pomocí toho je možné vytvářet jednoduché, ale informativní grafy (viz. níže).

**Graf 3: Rozdílný vývoj v aktivech podniků**



### Dopsat interpretaci na základě výpočtu datasetu

Dále jsou poskytnuty návrhy na výpočet vzhledných statistik, které by mohly sloužit jako způsob komunikace s veřejností a jako obhajoba podpory aplikovaného výzkumu kolaborativním způsobem. Program tyto údaje nepočítá, protože je třeba dohledat výši finanční státní podpory<sup>35</sup> konkrétních subjektů.

#### Za 1kč dotací Xkč v HDP

Pro tento výpočet je nutno disponovat plnou datovou specifikací přidané hodnoty (nalezneme na VZZ dané obchodní korporace). Pokud sumu vypočteného efektu přidané hodnoty metodou Difference in Differences vydělíme obdržanou podporou, dostáváme efekt přímé podpory podnikům, tedy kolik jedna koruna v dotacích vytvoří korun v HDP<sup>36</sup>. Přímý efekt podpory lze tedy pozorovat na přidané Hodnotě (Obchodní marže + výroba - výrobní spotřeba). Efekt přidané hodnoty je to, o kolik podpořené podniky vytvořily větší přidanou hodnotu (průměrně a mediánově). Proto jde o část HDP, která je díky podpoře vyšší. Je nutno také vyčíslit, jak velká přidaná hodnota v podnicích vznikla a za jaké náklady (dotace + adicionalita). Podpora podnikům musí být vztažena k efektu přidané hodnoty. Podniky ovšem přijímaly i podporu nepřímou v podobě výsledků veřejně financovaných výzkumných organizací. Pokud chceme daný problém pozorovat komplexněji, popsání výpočet nebude stačit. V kolaborativním výzkumu jsou podpořeny i výzkumné ústavy, které své

<sup>35</sup> Případně z programu ALFA

<sup>36</sup> Podle definice je HDP součtem všech přidaných hodnot ve společnosti. CAHLÍK (2010)



poznatky s podniky sdílí<sup>37</sup>. Proto pokud chceme vyčíslit efekt přímé i nepřímé podpory, musíme výslednou hodnotu z výpočtu typu difference in difference vydělit dotacemi, které obdrželi jak konkrétní obchodní korporace, tak ostatní subjekty. Vytvořená hodnota ve výzkumných organizacích apod. však zůstane ignorována.

### **Za 1kč dotací Xkč daní**

Další z návrhů, který nabízí možnost výpočtu návratnosti je obecný přístup k daním. Pokud budeme uvažovat, že právně stanovená míra zdanění právnických osob je rovna míře efektivní<sup>38</sup>, je možno tvrdit, že Hospodářský výsledek před zdaněním \* míra daně je rovna dani vybrané<sup>39</sup>. Pokud je tedy vybraná daň vydělena výší podpory konkrétním podnikům naše výsledné číslo může být interpretováno, jako kolik jedna koruna v dotacích vytvořila příjem státního aparátu formou daní.

### **Návratnost podpory IRR (%)**

Internal Rate of Return je často používaný indikátor ke stanovení návratnosti investice. Pro svou jednoduchost je používán pro srovnání různých investičních příležitostí. Pokud je tedy podpora kolaborativního výzkumu investicí s pozitivním IRR, bylo by vhodné ho vyčíslit.

$$NPV = 0 = \frac{CF_1}{(1 + IRR)^1} + \frac{CF_2}{(1 + IRR)^2} + \frac{CF_3}{(1 + IRR)^3} + \frac{\sum CF_t}{(1 + IRR)^t}$$

To by bylo možné pomocí složitějších nástrojů. Pokud vytvoříme kontrafaktuální analýzu pro každý rok od počátku definovaného období<sup>40</sup>,  $CF_1$  odpovídá *Hospodářský výsledek před zdaněním*<sub>1</sub> \* *Míra daně*  $PO_1$  a tak dále. Pro tyto účely je vhodné hodnotit s delším odstupem od ukončení podpory, což v případě TA ČR není možné, a proto tato metoda nebyla použita. Pro konkrétní výpočet IRR je vhodná funkce IRR v Excelu.

<sup>37</sup> Na základě smlouvy o spolupráci a § 16 zákona 130/2002 Sb.

<sup>38</sup> DEMARZO, 2014. P. Fundamentals of Corporate Finance Student Value Edition.

<sup>39</sup> Pro zpřesnění je možno využít efektivní míru zdanění právnických osob z minulého roku

<sup>40</sup> Období, kdy podpořené projekty prokazatelně vytváří vyšší přidanou hodnotu firmě, tedy rok, nebo více po úspěšném ukončení projektu



## 8. Použitá a doporučená literatura

### 8.1. Použitá literatura

Abadie, A., Diamond, A. and Hainmueller, J. (2014), "Comparative Politics and the Synthetic Control Method". American Journal of Political Science. 2014, Forthcoming; Formerly MIT Political Science Department Research Paper No. 2011-25.

Abadie, A. and G. W. Imbens (2011), "Bias-Corrected Matching Estimators for Average Treatment Effects," Journal of Business and Economic Statistics, vol. 29(1), 1-11.

Abadie, A., Diamond A., and Hainmueller, J. (2010), "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." Journal of the American Statistical Association 105 (2010): 493-505.

Angrist, J. D., Pischke, J. (2009) "Mostly harmless econometrics: an empiricist's companion." Princeton: Princeton University Press, c2009, xiii, 373 p. ISBN 978-0691120355.

Barca, F. (2009), "An Agenda for a Reformed Cohesion Policy: A Place-Based Approach to Meeting European Union Challenges and Expectations, Independent Report Prepared at the Request of the European Commissioner for Regional Policy", Danuta Hübner, Bruxelles: European Commission.

Bertrand M., Duflo E. and Mullainathan S. (2004), "How Much Should We Trust Differences-in-Differences Estimates?," The Quarterly Journal of Economics, 2004, vol. 119, n. 1, pp. 249-275.

Bondonio, D. (2010). "Counterfactual Impact evaluation of Enterprise Support Programmes. Evidence from a Decade of Subsidies to Italian Firms."

Bondonio, D. (2014). "Counterfactual Impact Evaluation of Enterprise Support Programmes. Evidence from a Decade of Subsidies to Italian Firm. European D.G.Regio Research, Grant: Counterfactual Impact Evaluation of Cohesion Policy. Work Package 1: Examples from Enterprise Support."

CAHLÍK, T., HLAVÁČEK, M., SEIDLER, J. (2010). "Makroekonomie." 2., upr. vyd. Praha: Karolinum, 228 s. ISBN 978-80-246-1906-4.

Caliendo, Marco and Kopeinig, Sabine, (2008). "Some Practical Guidance for the Implementation of Propensity Score Matching." IZA Discussion Paper No. 1588.

Čadil, Mirošík, Petkovová, Zeman, Hartman (2014). "Hodnocení efektivnosti podpory R&D v ČR", pracovní studie.

DEMARZO, P. (2014). "Fundamentals of Corporate Finance Student Value Edition."  
Pearson College Div.

EC; EVALSED (2009a). "Counterfactual impact evaluation: Methods and Techniques:  
Difference-in-Difference." Evropska komise, SOURCEBOOKS, 2009.

EC; Evropská komise (2013). "Design and commissioning of counterfactual impact  
evaluations guidance to help employers and workers to manage the transition to the  
new classification, labelling and packaging system." Luxembourg: Publications Office,  
2013. ISBN 978-92-79-28238-6.

González, X., Pazó, C. (2008) "Do public subsidies stimulate private R&D spending?"  
*Research Policy*. 2008, roč. 37, č. 3, s. 371-389.

Hausman, Jerry A. & Taylor, William E., (1981). "Panel data and unobservable individual  
effects," *Journal of Econometrics*, Elsevier, vol. 16(1), pages 155-155, May.

Heckman, J., Ichimura, H., & Todd, P.E. (1997). "Matching as an econometric evaluation  
estimator: Evidence from evaluating a job training programme." *The Review of  
Economic Studies*, 64, 605-654. doi 10.2307/2971733

Hottenrott, H. and Lopes, B., Cindy. (2014). "R&D Partnerships and Innovation  
Performance: Can There Be Too Much of a Good Thing?" (July 15, 2014). ZEW - Centre  
for European Economic Research Discussion Paper No. 14-108.

Khandker S. R., Koolwal G. B. and H. A. Samad (2010), " Handbook on Impact Evaluation:  
Quantitative Methods and Practices," The International Bank for Reconstruction and  
Development, The World Bank, Washington, DC

Potluka, O., Brůha J., Špaček M., (2013). "Does the ESF assistance contribute to sales in  
private firms?" Working paper: The 7th International Days of Statistics and Economics

Potluka, O., Špaček, M., (2013). "Postupy a metody kontrafaktuálních dopadových  
evaluací pro Operační program Zaměstnanost v období 2014-2020." Prague, 15. října  
2013 Prague, September 19-21, 2013.

Rosenbaum, P. R. and RUBIN, D. B. (1983) "The central role of the propensity score in  
observational studies for causal effects." *Biometrika* (1983) 70 (1): 41-55  
doi:10.1093/biomet/70.1.41

Sedláček (2014), Vliv dotace na konkurenceschopnost podniku. Disertační práce,  
školicel: doc. Ing. Petr Suchánek, Ph.D.

Wooldridge, J.M., (2010), *Econometric Analysis of Cross Section and Panel Data*,  
2nd Edition. The MIT Press. (Chapter 21)

Doporučený seznam literatury sestavený O. Potlukou pro zájemce o kontrafaktuální dopadové evaluace:

## 8.2. Konkurenceschopnost

Liargovas, P. and Skandalis, K. (2010), „Factors Affecting Firm Competitiveness: The Case of Greek Industry“. Working Papers 0012, University of Peloponnese, Department of Economics.

Meric, Gulser; Welsh, Carol; Weidman, Stephanie; Marmon, Richard. (2011), “The Effects of the 2008 Stock Market Crash on the Managerial Behavior, Financial Characteristics and Competitiveness of Large U.S. Corporations“. Journal of Global Business Issues 5.1 (Spring 2011): 11-20.

Viaene, J.; Gellynck, X. (1998), “Small firms, old traditions equal low profit: pigmeat processing in Belgium.” E. Pitts a B. Trail. Competitiveness in the Food Industry. London: Blackie and Professional. 075-14-043-14.

## 8.3. Kontrafaktuální dopadové evaluace obecně

Mezi základní literaturu řadíme tyto tři publikace. První z nich je dokonce psána stylem, který se věnuje intuitivně pouze principům a vynechává ekonometrické zálady.

Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B. and C. M. J. Vermeersch (2011), "Impact Evaluation in Practice," The International Bank for Reconstruction and Development, The World Bank, Washington, DC

Khandker S. R., Koolwal G. B. and H. A. Samad (2010), " Handbook on Impact Evaluation: Quantitative Methods and Practices," The International Bank for Reconstruction and Development, The World Bank, Washington, DC

Angrist, J. D. and J. S. Pischke (2009), Mostly Harmless Econometrics: An Empiricist Companion. Princeton University Press.

## 8.4. Mezi další literaturu řadíme následující tituly

Angrist, J. D. and A. B. Krueger (2000), "Empirical Strategies in Labor Economics," in A. Ashenfelter and D. Card eds. Handbook of Labor Economics, vol. 3. New York: Elsevier Science. Sections 1 and 2.

Duflo, E., R. Glennerster and M. Kremer (2008), "Using Randomization in Development Economics Research: A Toolkit," in T.P. Schultz and J.A. Strauss eds. Handbook of Development Economics, vol. 4. New York: Elsevier Science.

Imbens, G.W. and J.M. Wooldridge (2009) "Recent Developments in the Econometrics of Program Evaluation," Journal of Economic Literature, vol. 47(1), 5-86.

Wooldridge, J.M., (2010), Econometric Analysis of Cross Section and Panel Data, 2nd Edition. The MIT Press. (Chapter 21)

## 8.5. Párování

Abadie, A. and G. W. Imbens (2006), "Large Sample Properties of Matching Estimators for Average Treatment Effects," Econometrica, vol. 74, 235-267.

Abadie, A. and G. W. Imbens (2008), "On the Failure of the Bootstrap for Matching Estimators," Econometrica, vol. 76, 1537-1557.

Abadie, A. and G. W. Imbens (2011), "Bias-Corrected Matching Estimators for Average Treatment Effects," Journal of Business and Economic Statistics, vol. 29(1), 1-11.

Angrist, J. D. and J. S. Pischke (2009), Mostly Harmless Econometrics: An Empiricist Companion. Princeton University Press.

Becker, S. O. and A. Ichino (2002), "Estimation of average treatment effects based on propensity scores," Stata Journal, vol. 2 (4), 358-377.

Bryson A., Dorsett R. and Purdon S. (2002), "The Use of Propensity Score Matching in the Evaluation of Active Labour Market Policies," UK Department for Work and Pensions, London.

Caliendo, M. and S. Kopeinig (2008) "Some Practical Guidance for the Implementation of Propensity Score Matching," *Journal of Economic Surveys*, vol. 22 (1), 31–72.

Cochran, W. G., (1968), "The Effectiveness of Adjustment by Subclassification in Removing Bias in Observational Studies," *Biometrics*, vol. 24, 295-313.

Cochrane W., Chambers S. (1965), "The Planning of Observational Studies of Human Populations", *Journal of the Royal Statistical Society, Series A*, 128, pp. 234-266.

Dehejia, R. H. and S. Wahba (1999), "Causal Effects in Non-Experimental Studies: Re-Evaluating the Evaluation of Training Programs," *Journal of the American Statistical Association*, vol. 94, 1053-1062.

Gadd H., Hansson G. and Mansson J. (2008), *Evaluating the Impact of Firm Subsidy Using a Multilevel Propensity Score Approach*, Centre for Labour Market Policy Research (CAFO), School of Management and Economics, Vaxjo University.

Hardle, W and O. Linton (1994), "Applied Nonparametric Methods," in R. F. Engle and D.L. McFadden eds. *Handbook of Econometrics*, vol. 4. New York: Elsevier Science.

Heckman, J. J., H. Ichimura and P. E. Todd (1997), "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme," *Review of Economic Studies*, vol. 64, 605-654.

Hirano K., Imbens G. (2004), "The Propensity Score with Continuous Treatments, in *Missing Data and Bayesian Methods in Practice: Contributions by Donald Rubin's Statistical Family*," Wiley, NYC

Imbens, G.W. (2003), "Sensitivity to Exogeneity Assumptions in Program Evaluation," *American Economic Review (Papers & Proceedings)*, vol. 93(2), 126-132.

Imbens, G.W. (2004), "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review," *Review of Economics and Statistics*, vol. 86(1), 4-29.

Lechner M. (2002), "Some practical issues in the evaluation of heterogeneous labour market programmes by matching methods," *Journal of the Royal Statistical Society, Series A*, 165, pp. 59-82.

Pearl, J. (2009), *Causality* (second edition). Cambridge University Press.

Rosenbaum, P. R. (1995), *Observational Studies*. New York: Springer-Verlag. Chapter 3.

Rosenbaum, P. R., and D. B. Rubin (1983), "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, vol. 70, 41-55.

Rubin, D. B. (1977), "Assignment to Treatment Group on the Basis of a Covariate," *Journal of Educational Statistics*, vol. 2, 1-26.

White, H. (1980), "Using Least Squares to Approximate Unknown Regression Functions," *International Economic Review*, vol. 21, 149-170.

## 8.6. Rozdíly v rozdílech

Abadie, A. (2005), "Semiparametric Difference-in-Differences Estimators," *Review of Economic Studies*, vol. 72, 1-19.

Abadie, A., A. Diamond and J. Hainmueller (2010), "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program," *Journal of the American Statistical Association*, vol. 105, 493-505.

Abadie, A. and J. Gardeazabal (2003), "The Economic Costs of Conflict: A Case Study of the Basque Country," *American Economic Review*, vol. 93(1), 113-132.

Ashenfelter, O. and D. Card (1985), "Using the Longitudinal Structure of Earnings to Estimate the Effects of Training Programs," *Review of Economics and Statistics*, vol. 67, 648-660.

Bertrand M., Duflo E. and Mullainathan S. (2004), "How Much Should We Trust Differences-in-Differences Estimates?," *The Quarterly Journal of Economics*, 2004, vol. 119, n. 1, pp. 249-275.



Card, D. (1990), "The Impact of the Mariel Boatlift on the Miami Labor Market," Industrial and Labor Relations Review, vol. 44, 245-257.

Card, D. and A. B. Krueger (1994), "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania," American Economic Review, vol. 84, 772-793.

Card D., Krueger A. (1997), "Myth and Measurement: The New Economics of the Minimum Wage," Princeton, NJ, Princeton University Press.

Duflo E. (2001), "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment," American Economic Review, vol. 91, 795-813.

ITPS (2004), "The EC Regional Structural Funds impact in Sweden 1995-1999: A quantitative analysis," Swedish Institute for Growth Policy Studies, University of Umea, Department of Geography.

Meyer, B. D. (1995), "Natural and Quasi-Experiments in Economics," Journal of Business & Economic Statistics, vol. 13, 151-161.

## Seznam tabulek

<b>Tabulka 1: KS test, p-hodnoty a homogenita vzorku .....</b>	<b>41</b>
<b>Tabulka 2: Porovnání poměrových ukazatelů před udělením podpory .....</b>	<b>41</b>
<b>Tabulka 3: Porovnání poměrových ukazatelů před udělením podpory .....</b>	<b>42</b>
<b>Tabulka 4: Poměrové ukazatele před a po intervenci, první kontrolní skupina</b>	<b>45</b>
<b>Tabulka 5: Výpočet efektu, první kontrolní skupina .....</b>	<b>45</b>
<b>Tabulka 6: Vyjádření růstu podle konkrétních finančních ukazatelů .....</b>	<b>45</b>
<b>Tabulka 7: Vypočtený efekt růstu poměrových ukazatelů (Očištěný .....</b>	<b>46</b>





evropský  
sociální  
fond v ČR



EVROPSKÁ UNIE



OPERAČNÍ PROGRAM  
LIDSKÉ ZDROJE  
A ZAMĚSTNANOST



PODPORUJEME  
VAŠI BUDOUCNOST  
[www.esfcr.cz](http://www.esfcr.cz)

**Tabulka 8: Poměrové ukazatele před a po intervenci, druhá kontrolní skupina**46

**Tabulka 9: Výpočet efektu, druhá kontrolní skupina** .....47

**Tabulka 10: Vyjádření růstu podle konkrétních finančních ukazatelů** .....47

**Tabulka 11: Vypočtený efekt růstu poměrových ukazatelů (Očištěný)**.....48

## 9. Přílohy

### 9.1. Příloha 1: Rešerše literatury a praktické užití CIE

Jak je naznačeno v úvodu, kontrafaktuální analýza je aktuálně zdůrazňována, jako jeden z možných nástrojů hodnocení projektů podpořených v rámci různých dotačních titulů (velký důraz je kladen především na úrovni Evropské komise). Způsob aplikace CIE v každé konkrétní zemi i organizaci se může mírně měnit a to především kvůli datové dostupnosti, jež může významně limitovat možnosti implementace této metody. V českém prostředí byla CIE využita především na evaluaci evropských dotačních titulů. Potluka (2013) je jedním z prvních, kdo v českém prostředí aplikuje kontrafaktuální analýzu v širším měřítku. Na toto téma zpracoval metodologii a vykázal ji v rámci projektu TAČR OMEGA TD010083. Primárně se zaměřil na hodnocení vlivu intervencí z OP LZZ na zvyšování konkurenceschopnosti firem v České republice (evaluace vlivu podpory na tržby, zaměstnanost a mnoho dalších kritérií).

Metodiku kontrafaktuální analýzy, velmi podobnou té od Potluky (2013), vytvořila také Evropské Komise<sup>41</sup> (2013). Ad hoc, mimo ex-post evaluaci, se CIE v českém prostředí zabývali: Čadil (2014) a Sedláček (2014). Čadilovy (2014) výsledky ukazovaly na nejednoznačnost efektu podpory TA ČR. Výzkum byl prováděn na neukončených projektech. Sedláček (2014) na základě dotazníkového šetření rozdělil podniky z „Operačního programu Průmysl a podnikání“ do tří skupin dle jejich produktivity (ROA a ROE). Dle této studie je patrné, že více produktivní firmy jsou častěji podpořené. Hypotéza, že „*podniky, které čerpaly dotaci, jsou po jejím využití více konkurenceschopné (vyšší ROA), než byly před jejím využitím*“, nebyla potvrzena. Období výzkumu bylo prováděno na datech 2001 – 2010, finální rok byl tedy dnem finanční krize.

Na mezinárodní scéně není kontrafaktuální analýza žádnou novinkou. Na toto téma je vedena poměrně pestrá debata, a to jak z hlediska používaných metod,

<sup>41</sup> EUROPEAN COMMISSION. *Design and commissioning of counterfactual impact evaluations guidance to help employers and workers to manage the transition to the new classification, labelling and packaging system*. Luxembourg: Publications Office, 2013. ISBN 9789279282386.

tak vymýcení případných nedostatků dat. Přirovnání subjektů pomocí Propensity Score (popis metody viz dále) popsal vyčerpávajícím způsobem Rosenbaum a Rubin (1983). S více než dvacetiletým odstupem Caliendo and Kopeinig (2008) poskytli shrnutí v totožné problematice. Samotná metoda přirovnání subjektů se také dočkala mírných variací ve své statistické podobě a to především pro zajištění robustnosti modelu Heckman, Ichimura, and Todd (1997). S další metodou přirovnání subjektů pro výpočet průměrného efektu přichází Abadie, A. and G. W. Imbens (2011).

Statistická aplikace rozdílů v rozdílech (DiD – viz dále) je technicky poměrně náročná. Porozumění ekonometrické teorii je nutným předpokladem. Tomu se věnuje Wooldridge (2010) ve své učebnici ekonometrie v příslušně nazvané kapitole. Praktickou ukázkou vyhodnocení efektu na datech v podobě časových řad a cross-sectional data nabízí článek autorů Hausman a Taylor (1981). Velká přidaná hodnota článku spočívá také v tom, že popisuje jak naložit s nepozorovanými proměnnými, které činí lineární regresi nepřesnou (biased). Eliminace nepozorovaných proměnných je pro vyčíslení efektu důležité, jde totiž o jeden z předpokladů pro OLS BLUE. Jednou z překážek správné lineární regrese je také autokorelace. Jak tento problém vyřešit popisuje Bertrand, Duflo, and Mullainathan (2004). Protože se tato metodika zaměřuje především na co nejpřesnější přirovnání subjektů pomocí Propensity Score, nebude těmto statistickým metodám věnován další prostor. Pokud nastane situace, že by uživatel byl nucen použít podobné přístupy, doporučená literatura bude citována v části

#### Použitá a doporučená literatura.

Příkladem využití kontrafaktuální analýzy v účelové podpoře Výzkumu Vývoje a Inovací je např. článek autorů González a Pazó (2008). Použitím metody Propensity Score Matching na příkladu veřejné účelové podpory VaVaI ve španělských firmách vyvrátili efekt vytlačení investic<sup>42</sup>. Výsledky naopak hovoří o absenci „crowding out“ efektu. Dle jejich zjištění veřejná podpora naopak indikuje další soukromé investice skrze adicionalitu<sup>43</sup>. Výsledně je vypočten

<sup>42</sup> Proces, který má následek zvýšení úrokové míry, v důsledku čehož poklesnou investice soukromé a jehož výsledkem je tak paradox stejné, tedy nezvýšené míry investic. Laicky řešeno, z veřejných zdrojů se tak hradí soukromé aktivity, které by se odehrály tak jako tak, jen s rozdílnými zdroji financování.

<sup>43</sup> Veřejné financování soukromým podnikům je pouze částečné. Jak uvádí Rámec pro státní podporu výzkumu, vývoje a inovací pod číslem (2014/C 198/01) uveřejněný v Evropském

efekt přičítaný veřejné podpoře. Většího efektu je dosaženo u malých podniků pohybujících se v odvětví s menším používáním vyspělých technologií.

Bondonio (2010) provádí kontrafaktuální analýzu v Italském prostředí. Za použití metody Propensity Score Matching dochází k závěru, že k vytlačení investic nedochází. Je také schopen vyčíslit, kolik jedno euro veřejné podpory „vytvořilo“ eur v tržbách a investicích do hmotného majetku. Tento výzkum prováděl na programu Law488 a několika heterogenních programech v regionu Piemonte. Program Law488 je cílen na průmyslové firmy. V rámci tohoto programu je poskytována nevratná investiční podpora hmotného kapitálu. Dataset pro podporu regionu Piemonte tvoří heterogenní programy a zahrnuje pouze malé a střední podniky. Jedním ze závěrů je, že pro případ Piemonte jedno euro veřejné podpory vytvořilo čtyři eura v celkových tržbách podpořených podniků. V případě podpory v programu Law488 jedno euro veřejné podpory vytvořilo osm eur v celkových tržbách. Z hlediska zaměstnanosti jedno pracovní místo bylo vytvořeno za 33tis euro pro podporu regionu Piemonte a v programu law 488 jedno pracovní místo vzniklo za cenu 230tis euro. Díky této evaluaci byly porovnány různé typy podpory. Forma alternativních finančních nástrojů byla z hlediska využitých veřejných prostředků nejvíce rentabilní.

Hottenrott (2014) zkoumá veřejnou podporu kolaborace podniků a veřejných institucí z Flander. Nulová hypotéza, že dochází k vytlačení investic, byla zamítnuta. Veřejné programy tedy zvyšují investice soukromých podniků do VaVaI skrze adicionalitu. Z výzkumu vychází najevo, že malé a střední podniky spolupracující na mezinárodní úrovni dosahují vyšší úrovně pozitivního efektu podpory oproti malým a středním podnikům nespolupracujícím na mezinárodní úrovni. Tyto závěry mohou být využity pro úpravu rámce Evropské komise 198/01<sup>44</sup>, kde by výše podpory nebyla determinována pouze velikostí podniku, ale také tím, zdali podnik bude využívat mezinárodní spolupráci s oborově stejným partnerem.

---

věstníku dne 27. 6. 2014. Např. velký podnik, který v projektu provádí aplikovaný výzkum, může obdržet maximálně 70% vykázaných nákladů.

<sup>44</sup> Rámec pro státní podporu výzkumu, vývoje a inovací pod číslem (2014/C 198/01) uveřejněný v Evropském věstníku dne 27. 6. 2014.

## 9.2. Příloha 2: Výběr finančních ukazatelů a typ přiřazení firem, České prostředí

**Tabulka 12: výběr finančních ukazatelů a typ přiřazení firem**

Položka	Milan Sedláček	Jan Čadil	TACR, současnost	Srholec	Potluka
Aktiva celkem	ANO	ANO	ANO		
Oběžná aktiva	ANO				ANO
Krátkodobé pohledávky	ANO				
Finanční majetek krátkodobý	ANO		Fin. Majetek		
Vlastní kapitál	ANO		ANO		ANO
Krátkodobé závazky	ANO				
Krátkodobé bankovní úvěry	ANO				
Tžby za prodej zboží	ANO	Tržby			Tržby
Výkony	ANO		ANO		
Přidaná hodnota	ANO	ANO			
Osobní náklady	ANO	ANO			
Provozní hospodářský výsledek	ANO	ANO			
Nákladové úroky	ANO				
Výsledek hospodaření před daň.	ANO	ANO			
Daň (dopočet)	ANO	ANO			
Hmotné investiční výdaje		ANO			
Obchodní Marže		ANO			
Cizí zdroje			ANO		ANO
Dlouhodobé závazky			ANO		

Počet zaměstnanců			ANO		ANO
NACE			ANO	ANO	ANO
Počet bodů Hodnotitelé			ANO	ANO	ANO
Počet bodů Předsednictvo			ANO	ANO	ANO
Identifikace hodnotitele			ANO		ANO
Metoda	NACE	PSM	PSM	IV <sup>45</sup>	PSM,IV

***Použité veličiny Propensity Score Matching od Gonzáles, Pazó (2008):***

Subsidy dummy, Number of workers, Capital growth, Age, Technological sophistication, Foreign capital dummy, Domestic exporter dummy, Firm with market power dummy, Abnormal subsidy dummy, Industry, region and time dummies

***Použité proměnné pro výpočet efektu Gonzáles, Pazó (2008):***

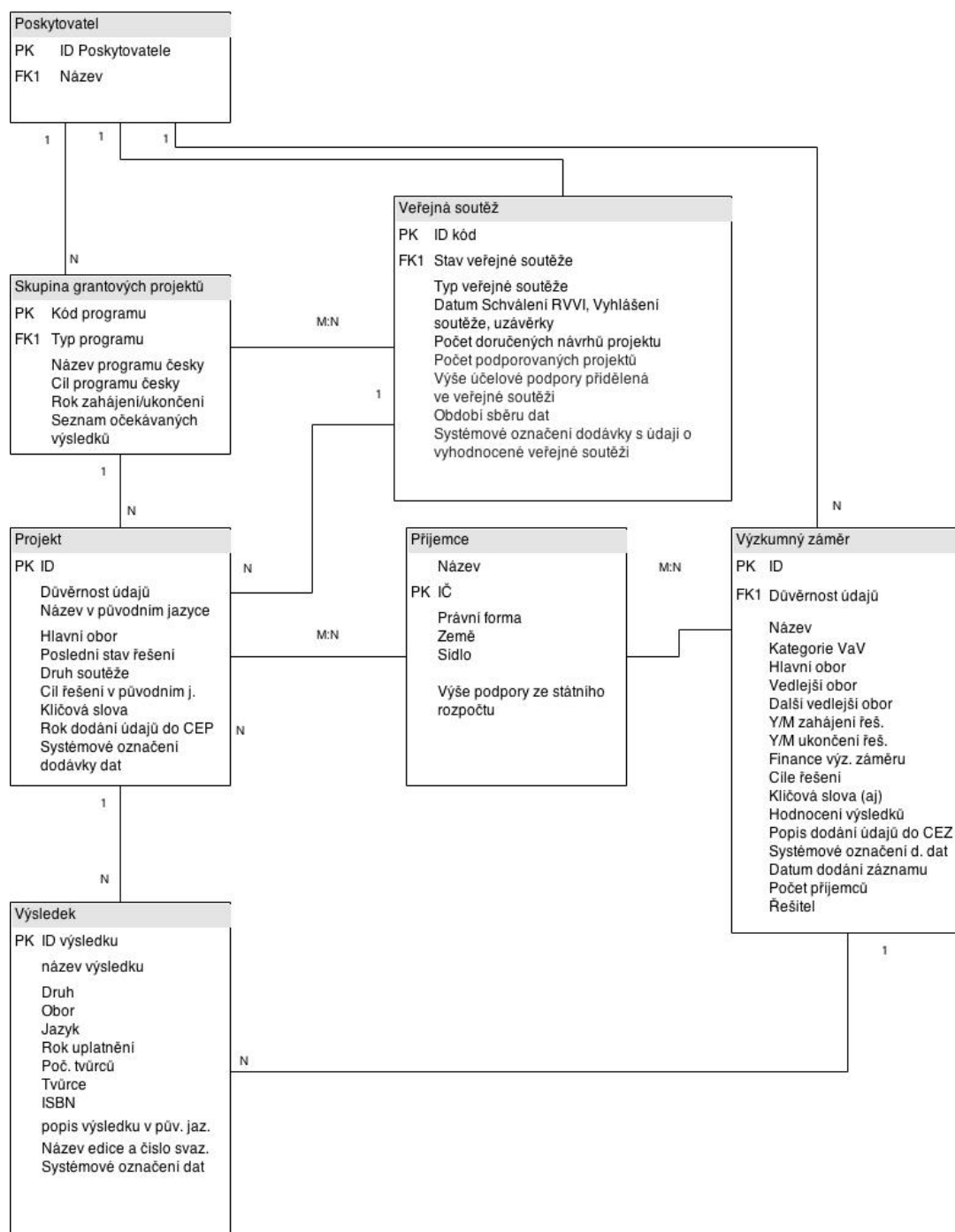
Total R&D expenditure (effort, high-tech etc.), Payments for licences, Public R&D funding, firm size

---

<sup>45</sup> Instrumental Variables

## 9.3. Příloha 3: Obsah IS VaV databáze

Obrázek 14: Mapování IS VaV databáze





## 9.4. Příloha 4: DiD OLS model

Pro následující zhodnocení byl použit zmíněný OLS model. Z výsledku tohoto modelu je patrné, že koeficienty u všech proměnných jsou nesignifikantní. Malá signifikance může být také vysvětlena tím, že vzorek je malý a předpokladem, že podpora R&D projeví až v dlouhodobém časovém horizontu.

**Tabulka 13: Aktiva celkem**

Estimate	Std. Error	t value	Pr(> t )	
Intercept	17.79867	0.19872	89.565	<2e-16 ***
P	0.72240	0.28104	2.570	0.0105 *
T	-0.00112	0.28104	-0.004	0.9968
P:t	0.37264	0.39745	0.938	0.3489

Residual standard error: 2.14 on 460 degrees of freedom  
Multiple R-squared: 0.04694, Adjusted R-squared: 0.04073  
F-statistic: 7.553 on 3 and 460 DF, p-value: 6.096e-05

**Tabulka 14: Vlastní kapitál**

	Estimate	Std. Error	t value	Pr(> t )
Intercept	16.1040	0.4061	39.655	<2e-16 ***
P	1.3486	0.5743	2.348	0.0193 *
T	-0.5508	0.5743	-0.959	0.3380
P:t	0.9930	0.8122	1.223	0.2221

Residual standard error: 4.374 on 460 degrees of freedom  
Multiple R-squared: 0.04595, Adjusted R-squared: 0.03973  
F-statistic: 7.385 on 3 and 460 DF, p-value: 7.668e-05

**Tabulka 15: Výkony**

	Estimate	Std. Error	t value	Pr(> t )
Intercept	18.03341	0.19131	94.261	<2e-16 ***
P	0.83303	0.27056	3.079	0.0022 **
T	-0.09739	0.27056	-0.360	0.7190
P:t	0.27397	0.38263	0.716	0.4743

Residual standard error: 2.061 on 460 degrees of freedom  
Multiple R-squared: 0.05401, Adjusted R-squared: 0.04784  
F-statistic: 8.755 on 3 and 460 DF, p-value: 1.174e-05

**Tabulka 16: Hospodářský výsledek před zdaněním**

	Estimate	Std. Error	t value	Pr(> t )
Intercept	15.34006	0.19782	77.547	<2e-16 ***
P	0.97466	0.27443	3.552	0.00043 ***
T	0.05652	0.29118	0.194	0.84620
P:t	-0.02066	0.39641	-0.052	0.95846

Residual standard error: 1.958 on 390 degrees of freedom (70 observations deleted due to missingness) Multiple R-squared: 0.05777, Adjusted R-squared: 0.05052, F-statistic: 7.97 on 3 and 390 DF, p-value: 3.608e-05

**Tabulka 17: Aktiva celkem**

Estimate	Std. Error	t value	Pr(> t )	
Intercept	18.18439	0.18066	100.654	<2e-16 ***
P	0.34149	0.25550	1.337	0.182
T	0.28976	0.25550	1.134	0.257
P:t	0.06958	0.36133	0.193	0.847

Residual standard error: 1.963 on 468 degrees of freedom

Multiple R-squared: 0.01598, Adjusted R-squared: 0.009676

F-statistic: 2.534 on 3 and 468 DF, p-value: 0.05631

**Tabulka 18: Vlastní kapitál**

	Estimate	Std. Error	t value	Pr(> t )
Intercept	17.2928	0.1979	87.361	<2e-16 ***
P	0.6177	0.2775	2.226	0.0265 *
T	0.3476	0.2787	1.247	0.2130
P:t	0.1072	0.3916	0.274	0.7844

Residual standard error: 2.086 on 450 degrees of freedom (18 observations deleted due to missingness) Multiple R-squared: 0.03432, Adjusted R-squared: 0.02788, F-statistic: 5.331 on 3 and 450 DF, p-value: 0.001286

**Tabulka 19: Výkony**

	Estimate	Std. Error	t value	Pr(> t )
Intercept	18.40969	0.17170	107.219	<2e-16 ***
P	0.45804	0.24282	1.886	0.0599 .
T	0.15051	0.24282	0.620	0.5357
P:t	0.01167	0.34340	0.034	0.9729

Residual standard error: 1.865 on 468 degrees of freedom Multiple R-squared: 0.01707, Adjusted R-squared: 0.01077 F-statistic: 2.71 on 3 and 468 DF, p-value: 0.04462

**Tabulka 20: Hospodářský výsledek před zdaněním**

	Estimate	Std. Error	t value	Pr(> t )
Intercept	15.74676	0.19423	81.071	<2e-16 ***
P	0.53076	0.26879	1.975	0.049 *
T	0.15190	0.27540	0.552	0.582
P:t	-0.09746	0.38022	-0.256	0.798

Residual standard error: 1.958 on 390 degrees of freedom (70 observations deleted due to missingness) Multiple R-squared: 0.05777, Adjusted R-squared: 0.05052, F-statistic: 7.97 on 3 and 390 DF, p-value: 3.608e-05

## 9.5. Příloha 4: Program kontrafaktuální analýzy

Program bude přiložen tak, že budu možno jej otevřít v softwaru R. Pro jistotu je sem vložen v textové podobě, aby bylo možné ho překopírovat a používat.

```
# Name: TACR CIE
# Author: Petr Horak
# Date: 26 May 2015
# Goal:
# 1) Dataset review
# 2) K-S test of homogeneity
# 3) Propensity Score Matching 1:1
# 4) After-peering K-S test of homogeneity
# 5) Countrafactual analysis
#    - take weighted avarages
#    - Avarage causal effect = Sum (Supported Yx - UnsupportedYx)
#    (compare difference in means in certain strata = estimate in certain
value)
# 6) Difference in Difference
# 7) Outcome summary to be printed into excel
#-----#
INPUTS:
      TAX=0.19 #Corporate tax is 19%

# Match packages: install.packages ("Matching"); install.packages ("Zelig");
https://www.java.com/en/download/manual.jsp
# install.packages ("MatchIt")
# install.packages ("rJava")
# install.packages ("xlsxjars")
# install.packages ("xlsx")

TACRCIE <- read.csv("C:/TACR_petr_horak/KA3/.../Data/H3/TACRCIE.csv", sep=";",
dec=",")
dataH3_2 <- TACRCIE

# 1) Dataset review
#-----#
---#
# Review of control group (P=0)
sdTA0<-sd((dataH3_2[which(dataH3_2$P=="0"),])$TA0,na.rm = FALSE) # Standard Deviation
for TA0
meanTA0<-mean((dataH3_2[which(dataH3_2$P=="0"),])$TA0,na.rm = FALSE) # mean of TA0
minTA0<-min((dataH3_2[which(dataH3_2$P=="0"),])$TA0,na.rm = FALSE) # min of TA0
maxTA0<-max((dataH3_2[which(dataH3_2$P=="0"),])$TA0,na.rm = FALSE) # max of TA0

sdOA0<-sd((dataH3_2[which(dataH3_2$P=="0"),])$OA0,na.rm = FALSE) # Standard Deviation
for OA0
meanOA0<-mean((dataH3_2[which(dataH3_2$P=="0"),])$OA0,na.rm = FALSE)
minOA0<-min((dataH3_2[which(dataH3_2$P=="0"),])$OA0,na.rm = FALSE)
maxOA0<-max((dataH3_2[which(dataH3_2$P=="0"),])$OA0,na.rm = FALSE)

sdVK0<-sd((dataH3_2[which(dataH3_2$P=="0"),])$VK0,na.rm = FALSE)
meanVK0<-mean((dataH3_2[which(dataH3_2$P=="0"),])$VK0,na.rm = FALSE)
minVK0<-min((dataH3_2[which(dataH3_2$P=="0"),])$VK0,na.rm = FALSE)
maxVK0<-max((dataH3_2[which(dataH3_2$P=="0"),])$VK0,na.rm = FALSE)

sdV0<-sd((dataH3_2[which(dataH3_2$P=="0"),])$V0,na.rm = FALSE)
meanV0<-mean((dataH3_2[which(dataH3_2$P=="0"),])$V0,na.rm = FALSE)
minV0<-min((dataH3_2[which(dataH3_2$P=="0"),])$V0,na.rm = FALSE)
maxV0<-max((dataH3_2[which(dataH3_2$P=="0"),])$V0,na.rm = FALSE)

sdOB0<-sd((dataH3_2[which(dataH3_2$P=="0"),])$OB0,na.rm = FALSE)
meanOB0<-mean((dataH3_2[which(dataH3_2$P=="0"),])$OB0,na.rm = FALSE)
minOB0<-min((dataH3_2[which(dataH3_2$P=="0"),])$OB0,na.rm = FALSE)
maxOB0<-max((dataH3_2[which(dataH3_2$P=="0"),])$OB0,na.rm = FALSE)

sdPHV0<-sd((dataH3_2[which(dataH3_2$P=="0"),])$PHV0,na.rm = FALSE)
meanPHV0<-mean((dataH3_2[which(dataH3_2$P=="0"),])$PHV0,na.rm = FALSE)
```

```

minPHV0<-min((dataH3_2[which(dataH3_2$P=='0'),])$PHV0,na.rm = FALSE)
maxPHV0<-max((dataH3_2[which(dataH3_2$P=='0'),])$PHV0,na.rm = FALSE)

sdHVpZ0<-sd((dataH3_2[which(dataH3_2$P=='0'),])$HVpZ0,na.rm = FALSE)
meanHVpZ0<-mean((dataH3_2[which(dataH3_2$P=='0'),])$HVpZ0,na.rm = FALSE)
minHVpZ0<-min((dataH3_2[which(dataH3_2$P=='0'),])$HVpZ0,na.rm = FALSE)
maxHVpZ0<-max((dataH3_2[which(dataH3_2$P=='0'),])$HVpZ0,na.rm = FALSE)

MeanP0 <- c(meanTA0, meanOA0, meanVK0, meanV0, meanOB0, meanPHV0, meanHVpZ0)
MinP0 <- c(minTA0, minOA0, minVK0, minV0, minOB0, minPHV0, minHVpZ0)
MaxP0 <- c(maxTA0, maxOA0, maxVK0, maxV0, maxOB0, maxPHV0, maxHVpZ0)
SDP0 <- c(sdTA0, sdOA0, sdVK0, sdV0, sdOB0, sdPHV0, sdHVpZ0)
#-----#
---#
# Review of supported group (P=1)
sdTA0<-sd((dataH3_2[which(dataH3_2$P=='1'),])$TA0,na.rm = FALSE) # Standard Deviation
for TA0
meanTA0<-mean((dataH3_2[which(dataH3_2$P=='1'),])$TA0,na.rm = FALSE) # mean of TA0
minTA0<-min((dataH3_2[which(dataH3_2$P=='1'),])$TA0,na.rm = FALSE) # min of TA0
maxTA0<-max((dataH3_2[which(dataH3_2$P=='1'),])$TA0,na.rm = FALSE) # max of TA0

sdOA0<-sd((dataH3_2[which(dataH3_2$P=='1'),])$OA0,na.rm = FALSE) # Standard Deviation
for OA0
meanOA0<-mean((dataH3_2[which(dataH3_2$P=='1'),])$OA0,na.rm = FALSE)
minOA0<-min((dataH3_2[which(dataH3_2$P=='1'),])$OA0,na.rm = FALSE)
maxOA0<-max((dataH3_2[which(dataH3_2$P=='1'),])$OA0,na.rm = FALSE)

sdVK0<-sd((dataH3_2[which(dataH3_2$P=='1'),])$VK0,na.rm = FALSE)
meanVK0<-mean((dataH3_2[which(dataH3_2$P=='1'),])$VK0,na.rm = FALSE)
minVK0<-min((dataH3_2[which(dataH3_2$P=='1'),])$VK0,na.rm = FALSE)
maxVK0<-max((dataH3_2[which(dataH3_2$P=='1'),])$VK0,na.rm = FALSE)

sdV0<-sd((dataH3_2[which(dataH3_2$P=='1'),])$V0,na.rm = FALSE)
meanV0<-mean((dataH3_2[which(dataH3_2$P=='1'),])$V0,na.rm = FALSE)
minV0<-min((dataH3_2[which(dataH3_2$P=='1'),])$V0,na.rm = FALSE)
maxV0<-max((dataH3_2[which(dataH3_2$P=='1'),])$V0,na.rm = FALSE)

sdOB0<-sd((dataH3_2[which(dataH3_2$P=='1'),])$OB0,na.rm = FALSE)
meanOB0<-mean((dataH3_2[which(dataH3_2$P=='1'),])$OB0,na.rm = FALSE)
minOB0<-min((dataH3_2[which(dataH3_2$P=='1'),])$OB0,na.rm = FALSE)
maxOB0<-max((dataH3_2[which(dataH3_2$P=='1'),])$OB0,na.rm = FALSE)

sdPHV0<-sd((dataH3_2[which(dataH3_2$P=='1'),])$PHV0,na.rm = FALSE)
meanPHV0<-mean((dataH3_2[which(dataH3_2$P=='1'),])$PHV0,na.rm = FALSE)
minPHV0<-min((dataH3_2[which(dataH3_2$P=='1'),])$PHV0,na.rm = FALSE)
maxPHV0<-max((dataH3_2[which(dataH3_2$P=='1'),])$PHV0,na.rm = FALSE)

sdHVpZ0<-sd((dataH3_2[which(dataH3_2$P=='1'),])$HVpZ0,na.rm = FALSE)
meanHVpZ0<-mean((dataH3_2[which(dataH3_2$P=='1'),])$HVpZ0,na.rm = FALSE)
minHVpZ0<-min((dataH3_2[which(dataH3_2$P=='1'),])$HVpZ0,na.rm = FALSE)
maxHVpZ0<-max((dataH3_2[which(dataH3_2$P=='1'),])$HVpZ0,na.rm = FALSE)

MeanP1 <- c(meanTA0, meanOA0, meanVK0, meanV0, meanOB0, meanPHV0, meanHVpZ0)
MinP1 <- c(minTA0, minOA0, minVK0, minV0, minOB0, minPHV0, minHVpZ0)
MaxP1 <- c(maxTA0, maxOA0, maxVK0, maxV0, maxOB0, maxPHV0, maxHVpZ0)
SDP1 <- c(sdTA0, sdOA0, sdVK0, sdV0, sdOB0, sdPHV0, sdHVpZ0)
#-----#
---#
# Combination of Reviews, final outcome

DataReview = data.frame(MeanP0,MinP0,MaxP0,SDP0,MeanP1,MinP1,MaxP1,SDP1)
row.names(DataReview) <- c("TA","OA","VK","V","OB","PHV","HVpZ")

write.csv2(DataReview, file = "C:/Skola/IES/Diplomka/Data/H3/DataReview.csv")

#-----#
---#
# 2) Start of the first part K-S test of homogeneity
summary(dataH3_2[which(dataH3_2$P=='1'),]) # Summary of supported ALL (boxplot etc
would be better etc.)
summary(dataH3_2[which(dataH3_2$P=='0'),]) # Summary of non-supported ALL (boxplot
etc would be better etc.)
# Ideas: B) The whole outcome would be printet into some .txt / .csv file nicely
ordered

```

```
# 2) K-S test of financials TA,TF,SO,PM, GVA, PE, OP, GP

Suppl <- dataH3_2[which(dataH3_2$P=='1'), ] # picks just supported companies
x <-Suppl$TA0
Supp2 <- dataH3_2[which(dataH3_2$P=='0'), ] # picks just non-supported companies
y <-Supp2$TA0
Test1 <- ks.test(x, y)$p.value #Kolmogorov - Smirnov test of Total Assets t0, P-Value:
The smaller this number is, the less likely that x=y is true.
# More technically, it is the probability of finding a situation more extreme than
what you have in your data, assuming that x=y
# The homogeneity holds if p-value > 0.05
x <-Suppl$OA0
y <-Supp2$OA0
Test2 <- ks.test(x, y)$p.value #Kolmogorov - Smirnov test of Obezna Aktiva(Current
Assets) t0
x <-Suppl$VK0
y <-Supp2$VK0
Test3 <- ks.test(x, y)$p.value #Kolmogorov - Smirnov test of Vlastni Kapital(equity)
t0
x <-Suppl$V0
y <-Supp2$V0
Test4 <- ks.test(x, y)$p.value #Kolmogorov - Smirnov test of Vykony (Performance,
czech accounting standards) t0
x <-Suppl$OB0
y <-Supp2$OB0
Test5 <- ks.test(x, y)$p.value #Kolmogorov - Smirnov test of Vynosy, Obrat (Revenues)
t0
x <-Suppl$PHV0
y <-Supp2$PHV0
Test6 <- ks.test(x, y)$p.value #Kolmogorov - Smirnov test Provozni Hospodarsky
Vysledek (Operating profit) t0
x <-Suppl$HVpZ0
y <-Supp2$HVpZ0
Test7 <- ks.test(x, y)$p.value #Kolmogorov - Smirnov test of Phv Pred zdanenim
KSTESTs <- c(Test1,Test2,Test3,Test4,Test5,Test6,Test7) #All the tests as vector (can
be used for an output)
H0TA <- ifelse(Test1>0.05,print("YES"),print("NO")) #If the test is bigger than 0.05
there is homogeneity
H0OA <- ifelse(Test2>0.05,print("YES"),print("NO")) #outcome
H0VK <- ifelse(Test3>0.05,print("YES"),print("NO")) #outcome
H0V <- ifelse(Test4>0.05,print("YES"),print("NO")) #outcome
H0Ob <- ifelse(Test5>0.05,print("YES"),print("NO")) #outcome
H0PHV <- ifelse(Test6>0.05,print("YES"),print("NO")) #outcome
H0PhvPz <- ifelse(Test7>0.05,print("YES"),print("NO")) #outcome
#-----#
---#
# 2) K-S test outcome to print to excel
KStest <- data.frame( c(Test1,Test2,Test3,Test4,Test5,Test6,Test7),
c(H0TA,H0OA,H0VK,H0V,H0Ob,H0PHV,H0PhvPz)) #Creates table, col1 = twotailed KStest,
col2 homogeneity (Y/N)
names(KStest) <- c("K-S test value", "Homogeneity") #Names columns
row.names(KStest) <- c("TA","OA","VK","V","OB","PHV","HVpZ") #Names Rows
#-----#
---#
# 3) Matching of companies & 5) Stratification of choosen group - Propensity Score
# propensity score matching
library(MatchIt)
#m.out <-
matchit(P~log(TA0)+log(OA0)+log(VK0)+log(V0)+log(OB0)+log(PHV0)+log(HVpZ0),data =
dataH3_2, method = "nearest", distance = 'logit', ratio = 1) #PSM by all financials
m.out <- matchit(P~TA0+OA0+VK0+V0+OB0+PHV0+HVpZ0,data = dataH3_2, method = "nearest",
distance = 'logit', ratio = 1) #PSM by all financials
Dataset <- match.data(m.out)
plot(m.out, type = "hist")
PSM=summary(m.out)
#plot(m.out, type = "jitter") # Plots distance of Matched Treated and Control Units
#-----#
---#
# X) Do countrafactual analysis by difference in difference (subtracting the
financials and obtaining final effect)
m.data1 <- match.data(m.out)
#write.table(m.data1, "C:/Skola/IES/Diplomka/Data/H3/match.txt", sep="\t")
# X) Matching outcome 1:1 to print to excel
#write.table(m.data1, "C:/Skola/IES/Diplomka/Data/H3/match.txt", sep="\t")
```

```
#-----#
---#
#-----#
---#
# 4) After peering K-S test of homogeneity
summary(Dataset[which(Dataset$P=='1'),]) # Summary of supported ALL (boxplot etc would
be better etc.)
summary(Dataset[which(Dataset$P=='0'), ]) # Summary of non-supported ALL (boxplot etc
would be better etc.)
# Ideas: B) The whole outcome would be printet into some .txt / .csv file nicely
ordered
# 4) K-S test of financials TA,TF,SO,PM, GVA, PE, OP, GP

SupplA <- Dataset[which(Dataset$P=='1'), ] # picks just supported companies
x <-SupplA$TA0
Supp2A <- Dataset[which(Dataset$P=='0'), ] # picks just non-supported companies
y <-Supp2A$TA0
Test1A <- ks.test(x, y)$p.value #Kolmogorov - Smirnov test of Total Assets t0, P-
Value: The smaller this number is, the less likely that x=y is true.
# More technically, it is the probability of finding a situation more extreme than
what you have in your data, assuming that x=y
# The homogeneity holds if p-value > 0.05
x <-SupplA$OA0
y <-Supp2A$OA0
Test2A <- ks.test(x, y)$p.value #Kolmogorov - Smirnov test of Obezna Aktiva(Current
Assets) t0
x <-SupplA$VK0
y <-Supp2A$VK0
Test3A <- ks.test(x, y)$p.value #Kolmogorov - Smirnov test of Vlastni Kapital(equity)
t0
x <-SupplA$V0
y <-Supp2A$V0
Test4A <- ks.test(x, y)$p.value #Kolmogorov - Smirnov test of Vykony (Performance,
czech accounting standards) t0
x <-SupplA$OB0
y <-Supp2A$OB0
Test5A <- ks.test(x, y)$p.value #Kolmogorov - Smirnov test of Vynosy, Obrat (Revenues)
t0
x <-SupplA$PHV0
y <-Supp2A$PHV0
Test6A <- ks.test(x, y)$p.value #Kolmogorov - Smirnov test Provozni Hospodarsky
Vysledek (Operating profit) t0
x <-SupplA$HVpZ0
y <-Supp2A$HVpZ0
Test7A <- ks.test(x, y)$p.value #Kolmogorov - Smirnov test of Phv Pred zdanenim
KSTESTsA <- c(Test1A,Test2A,Test3A,Test4A,Test5A,Test6A,Test7A) #All the tests as
vector (can be used for an output)
AH0TA <- ifelse(Test1A>0.05,print("YES"),print("NO")) #If the test is bigger than 0.05
there is homogeneity
AH0OA <- ifelse(Test2A>0.05,print("YES"),print("NO")) #outcome
AH0VK <- ifelse(Test3A>0.05,print("YES"),print("NO")) #outcome
AH0V <- ifelse(Test4A>0.05,print("YES"),print("NO")) #outcome
AH0Ob <- ifelse(Test5A>0.05,print("YES"),print("NO")) #outcome
AH0PHV <- ifelse(Test6A>0.05,print("YES"),print("NO")) #outcome
AH0PhvPz <- ifelse(Test7A>0.05,print("YES"),print("NO")) #outcome
#-----#
---#
# 4) K-S test outcome to print to excel
KStestA <- data.frame( c(Test1A,Test2A,Test3A,Test4A,Test5A,Test6A,Test7A),
c(AH0TA,AH0OA,AH0VK,AH0V,AH0Ob,AH0PHV,AH0PhvPz)) #Creates table, col1 = twotailed
KStest, col2 homogeneity (Y/N)
names(KStestA) <- c("K-S test value", "Homogeneity") #Names columns
row.names(KStestA) <- c("TA", "OA", "VK", "V", "OB", "PHV", "HVpZ") #Names Rows
#-----#
---#
# The calculation of the effect using Difference in Difference methodology
# 5) Difference in Difference (DiD), financials
Final <- NULL
# Effect of TA1 in sample (total assets)
# DiD = (YT2 - CT2) - (YT1-YC1); Y = outcome of financial variable for supported
companies; T = time; 1 or 2 = position at time
# C = Control group variable outcome, e.g. sum of Total Assets for group P=0
```



```
Final$TA2013effect <- sum(as.numeric((m.data1$TA1[which(m.data1$P=='1')])) -
sum(as.numeric((m.data1$TA1[which(m.data1$P=='0')])))) -
sum(as.numeric((m.data1$TA0[which(m.data1$P=='1')])) +
sum(as.numeric((m.data1$TA0[which(m.data1$P=='0')]))))
Final$OA2013effect <- sum(as.numeric((m.data1$OA1[which(m.data1$P=='1')])) -
sum(as.numeric((m.data1$OA1[which(m.data1$P=='0')])))) -
sum(as.numeric((m.data1$OA0[which(m.data1$P=='1')])) +
sum(as.numeric((m.data1$OA0[which(m.data1$P=='0')]))))
Final$VK2013effect <- sum(as.numeric((m.data1$VK1[which(m.data1$P=='1')])) -
sum(as.numeric((m.data1$VK1[which(m.data1$P=='0')])))) -
sum(as.numeric((m.data1$VK0[which(m.data1$P=='1')])) +
sum(as.numeric((m.data1$VK0[which(m.data1$P=='0')]))))
Final$V2013effect <- sum(as.numeric((m.data1$V1[which(m.data1$P=='1')])) -
sum(as.numeric((m.data1$V1[which(m.data1$P=='0')])))) -
sum(as.numeric((m.data1$V0[which(m.data1$P=='1')])) +
sum(as.numeric((m.data1$V0[which(m.data1$P=='0')]))))
Final$Ob2013effect <- sum(as.numeric((m.data1$OB1[which(m.data1$P=='1')])) -
sum(as.numeric((m.data1$OB1[which(m.data1$P=='0')])))) -
sum(as.numeric((m.data1$OB0[which(m.data1$P=='1')])) +
sum(as.numeric((m.data1$OB0[which(m.data1$P=='0')]))))
Final$PHV1 <- sum(as.numeric((m.data1$PHV1[which(m.data1$P=='1')])) -
sum(as.numeric((m.data1$PHV1[which(m.data1$P=='0')])))) -
sum(as.numeric((m.data1$PHV0[which(m.data1$P=='1')])) +
sum(as.numeric((m.data1$PHV0[which(m.data1$P=='0')]))))
Final$HVPZ1 <- sum(as.numeric((m.data1$HVPZ1[which(m.data1$P=='1')])) -
sum(as.numeric((m.data1$HVPZ1[which(m.data1$P=='0')])))) -
sum(as.numeric((m.data1$HVPZ0[which(m.data1$P=='1')])) +
sum(as.numeric((m.data1$HVPZ0[which(m.data1$P=='0')]))))
# 6) Calculation of ratio DiD/avrg.Financial measure
TAe<-Final$TA2013effect/((sum(as.numeric(m.data1$TA1)))/2)
Oae<-Final$OA2013effect/((sum(as.numeric(m.data1$OA1)))/2)
VKe<-Final$VK2013effect/((sum(as.numeric(m.data1$VK1)))/2)
Ve<-Final$V2013effect/((sum(as.numeric(m.data1$V1)))/2)
OBe<-Final$Ob2013effect/((sum(as.numeric(m.data1$OB1)))/2)
PHVe<-Final$PHV1/((sum(as.numeric(m.data1$PHV1)))/2)
HVPZe<-Final$HVPZ1/((sum(as.numeric(m.data1$HVPZ1)))/2)
# 6) Difference in Difference (DiD), outcome of financial and ratio into a table
DiD <-
data.frame(c(Final$TA2013effect,Final$OA2013effect,Final$VK2013effect,Final$V2013effect,
Final$Ob2013effect,
Final$PHV1,Final$HVPZ1), c(TAe,Oae,VKe,Ve,OBe,PHVe,HVPZe))
#Creates table, with effect of the support absolutely and relatively
row.names(DiD) <- c("TA","OA","VK","V","OB","PHV","HVPZ") #Names Rows
names(DiD) <- c("Effect","Ratio") #Names columns
# 6) Difference in Difference (DiD), financial ratios
# ROE = EAT / Equity
ROE_YT2 = sum(as.numeric(m.data1$HVPZ1[which(m.data1$P=='1')]))*(1-TAX) /
sum(as.numeric(m.data1$VK1[which(m.data1$P=='1')]))
ROE_CT2 = sum(as.numeric(m.data1$HVPZ1[which(m.data1$P=='0')]))*(1-TAX) /
sum(as.numeric(m.data1$VK1[which(m.data1$P=='0')]))
ROE_YT1 = sum(as.numeric(m.data1$HVPZ0[which(m.data1$P=='1')]))*(1-TAX) /
sum(as.numeric(m.data1$VK0[which(m.data1$P=='1')]))
ROE_CT1 = sum(as.numeric(m.data1$HVPZ0[which(m.data1$P=='0')]))*(1-TAX) /
sum(as.numeric(m.data1$VK0[which(m.data1$P=='0')]))
# Outcome to print to excel
DiDROE = (ROE_YT2 - ROE_CT2) - (ROE_YT1 - ROE_CT1)
# ROA = EBIT / Total Assets
ROA_YT2 = sum(as.numeric(m.data1$HVPZ1[which(m.data1$P=='1')])) /
sum(as.numeric(m.data1$TA1[which(m.data1$P=='1')]))
ROA_CT2 = sum(as.numeric(m.data1$HVPZ1[which(m.data1$P=='0')])) /
sum(as.numeric(m.data1$TA1[which(m.data1$P=='0')]))
ROA_YT1 = sum(as.numeric(m.data1$HVPZ0[which(m.data1$P=='1')])) /
sum(as.numeric(m.data1$TA0[which(m.data1$P=='1')]))
ROA_CT1 = sum(as.numeric(m.data1$HVPZ0[which(m.data1$P=='0')])) /
sum(as.numeric(m.data1$TA0[which(m.data1$P=='0')]))
# Outcome to print to excel
DiDROA = (ROA_YT2 - ROA_CT2) - (ROA_YT1 - ROA_CT1)
# ROS = EBIT / Total Sales
ROS_YT2 = sum(as.numeric(m.data1$HVPZ1[which(m.data1$P=='1')])) /
sum(as.numeric(m.data1$OB1[which(m.data1$P=='1')]))
ROS_CT2 = sum(as.numeric(m.data1$HVPZ1[which(m.data1$P=='0')])) /
sum(as.numeric(m.data1$OB1[which(m.data1$P=='0')]))
ROS_YT1 = sum(as.numeric(m.data1$HVPZ0[which(m.data1$P=='1')])) /
sum(as.numeric(m.data1$OB0[which(m.data1$P=='1')]))
```



```

ROS_CT1 = sum(as.numeric(m.data1$HVpZ0[which(m.data1$P=='0')])) /
sum(as.numeric(m.data1$OB0[which(m.data1$P=='0')]))
DiDROS = (ROS_YT2 - ROS_CT2) - (ROS_YT1 - ROS_CT1)
Ratios <- data.frame(DiDROE,DiDROA,DiDROS)
# Outcome to print to excel
#-----#
---#
# DataReview, KStest, PSM, DiD, Ratios, Dataset

library("xlsx") #calls libraby to print xls
OUTfile = "C:/TACR_petr_horak/KA3/CIE/DataReview.xls"
write.xlsx2(DataReview, file = OUTfile,
            sheetName="DataReview",col.names=TRUE, row.names=TRUE, append=FALSE)
#Creates file DataReview.xls and fills it with table "DataReview"
write.xlsx2(KStest, file = OUTfile,
            sheetName="K-Stest",col.names=TRUE, row.names=TRUE, append=TRUE) #Adds new
sheet "K-Stest" to existing file DataReview.xls
write.xlsx2(PSM, file = OUTfile,
            sheetName="PSM",col.names=TRUE, row.names=TRUE, append=TRUE)
write.xlsx2(KStestA, file = OUTfile,
            sheetName="K-StestA",col.names=TRUE, row.names=TRUE, append=TRUE) #Adds
new sheet "K-StestA" to existing file DataReview.xls
write.xlsx2(DiD, file = OUTfile,
            sheetName="DiD",col.names=TRUE, row.names=TRUE, append=TRUE) #Adds new
sheet "DiD" to existing file DataReview.xls
write.xlsx2(Ratios, file = OUTfile,
            sheetName="Ratios",col.names=TRUE, row.names=TRUE, append=TRUE) #Adds new
sheet "Ratios" to existing file DataReview.xls
write.xlsx2(Dataset, file = OUTfile,
            sheetName="Dataset",col.names=TRUE, row.names=TRUE, append=TRUE) #Adds new
sheet "Ratios" to existing file DataReview.xls

```