

Diskuze metodických postupů při evaluaci dopadu aktivní politiky zaměstnanosti ČR na datech OKpráce

Discussion of Methods and Approaches towards the Impact Evaluation of Czech Active Labor Market Policy Using „OKpráce“ Data

Ondřej Hora, Miroslav Suchanec*

Abstrakt

Evaluace dopadu je spojena s mnoha metodologickými volbami ze strany výzkumníka. Výsledky evaluace jsou pak značně závislé na definici konkrétního evaluačního modelu. Cílem tohoto článku je diskutovat základní oblasti, principy a postupy hodnocení výsledků a dopadů programů aktivní politiky zaměstnanosti (APZ) s využitím administrativních dat systému OKpráce. Metodologickým postupem přitom máme na mysli sekvenci voleb a dílčích procedur evaluace, vedoucí k validním výsledkům. Článek se zabývá prezentací postupů, které jsou vhodné pro individualizovaná data OKpráce a opomíjí postupy, které nelze na těchto datech uspokojivým způsobem využít. Článek uplatňuje perspektivu tzv. částečného ekvilibria, kdy hodnotí dopady programů na jejich účastníky a nikoli na trh práce jako celek. Jádrem je diskuze postupů s ohledem na interní validitu naměřeného dopadu. Jako prostředek pro dodatečné vybalancování předprogramových rozdílů mezi skupinou intervence a kontrolní skupinou diskutuje proces párování, přičemž zmiňuje důležitost zahrnutí faktorů předprogramová perspektiva a historie nezaměstnaných na trhu práce a fáze hospodářského cyklu. Článek uzavírá, že výběr konkrétní techniky analýzy dat závisí na tom, zda chce evaluátor zjistit velikost absolutního nebo relativního rizika, jeho průběh v čase anebo pouze přítomnost nezaměstnaného v registru.

Abstract

Realization of impact evaluation brings possibility of various methodological choices that are limited to distinct assumptions concerning the context of the study, nature of the data etc. Results of the impact evaluation are then dependent considerably on definition of particular evaluation model. Aim of this article is to discuss main areas, principles and approaches to evaluation of Czech active labor policy programs (ALMP) outcomes and impacts, using administrative data of OKpráce. The article presents approaches that are suitable for this kind of data and is based on partial equilibrium perspective, thus omitting impacts on labor market as whole. The core of the article is discussion of diverse methodical steps in context of internal validity of measured impact. The matching is discussed as a method of balancing pre-program differences between program participants and control group. Inclusion of factors of labor market perspective and history, as well as phase of economic cycle, during this process is of high importance. Article concludes that choice of particular statistical method of data analysis hinges on whether evaluator wants to find absolute or relative risks of leaving unemployment, their progress in time or merely the presence of unemployed in register.

Klíčová slova

Aktivní politika zaměstnanosti, dopadová evaluace, kvazi-experimentální design, OKpráce

Keywords

Active labor market policy, impact evaluation, quasi-experimental design, OKpráce

Úvod

Tato studie se zabývá metodologickým postupem programového hodnocení programů¹ aktivní politiky zaměstnanosti (APZ) na individualizovaných datech OKpráce. Zabýváme se evaluací v perspektivě tzv. částečného ekvilibria (mikro-ekonomickou evaluací), tedy efektem intervence na její účastníky, nikoli na trh práce jako celek. Metodologickým postupem myslíme sekvenci voleb a dílčích procedur evaluace, vedoucí k validním výsledkům. Vycházíme z předpokladu, že zde rozpracované kvazi-experimentální postupy mohou být vhodnou „druhou nejlepší volbou“ (výsledkově přiměřeně srovnatelnou s provedením experimentu)². Zároveň v případě dat OKpráce lze v těchto kvazi-experimentálních postupech nalézt nesporné výhody:

- značný rozsah sledovaných dat umožňuje bohatost zjištění v oblasti dopadů programů na různé skupiny nezaměstnaných za různých podmínek,
- k dispozici je úplná populace nezaměstnaných,
- evaluace je mnohem levnější ve srovnání s pravým experimentem.

Evaluace dopadu je založena na řadě předpokladů spojených s jednotlivými postupy a na mnoha metodologických volbách ze strany výzkumníka. S určitou nadsázkou lze říci, že samotný evaluační postup je u řady evaluací druhou pomyslnou černou skříňkou s řadou nevyjasněných evaluačních míst (první černou skříňkou jsou neznámé podmínky působící uvnitř programu). Výsledky evaluace jsou pak značně závislé na definici konkrétního modelu (Betcherman, Olivas a Dar 2004).

Cílem studie je definovat základní oblasti, principy a postupy hodnocení výsledků a dopadů programů APZ s využitím administrativních dat systému OKpráce. Důraz je kladen zejména na zhodnocení postupů tak,

¹ Jedná se o postup hodnotící následné výsledky a dopady jednotlivých programů a nikoli o procesní evaluaci programů.

² Card, Kluve a Weber (2009) např. v meta-analýze nezjistili mezi výsledky experimentálních a kvazi-experimentálních studií významné rozdíly. Pro srovnání zjištění rozdílu v odhadu dopadu mezi experimentálním a kvazi-experimentálním přístupem na shodných individuálních datech srovnaj LaLonde (1986), Dehejia a Wahba (2002) a Smith a Todd (2005). O velikosti rozdílu zřejmě rozhoduje především kvalita párování případů (viz níže).

aby výsledkem jejich použití byl interně validní indikátor dopadu programu, tedy takový indikátor, jež by se maximálně blížil skutečnému kauzálnímu efektu programu na výsledek. Cílem však není zpracování metodiky evaluace dopadu, nýbrž diskuze základních oblastí, principů a postupů, na kterých bude budoucí metodika založena. Studie je určena zejména zadavatelům dopadových evaluací, analytikům ÚP a ostatním realizátorům dopadových evaluací.

Systém OKpráce je využíván ke správě administrativních dat vkládaných pracovníky jednotlivých poboček Úřadu práce ČR. Hlavním účelem tohoto systému je především vést evidenci o jednotlivých nezaměstnaných, nárocích na dávky sociálního zabezpečení či spravovat informace o APZ. Zároveň je ovšem vhodným zdrojem dat pro sekundární analýzu APZ³.

Pro hledání vhodného postupu analýzy využíváme jako podklad především zahraniční metodologickou literaturu. Rozhodující pro posouzení jednotlivých dopadů je ale především interpretace společenské hodnoty uvedeného postupu a dosaženého dopadu ze strany uživatelů. Studie je koncipována tak, aby vyhovovala způsobu definice dat v systému OKpráce či podobném systému, neboť je pravděpodobné, že základní požadavky na funkcionalitu a formu systému administrace dávek budou i při případné budoucí změně dodavatele systému (po vyhlášení výběrového řízení) zachovány. Tato studie se tedy zabývá prezentací určitého postupu či postupů, které jsou vhodné pro dostupná data, a pomíjí tak postupy, které nelze na využívaných datech uspokojivým způsobem využít.

Díky povaze dat OKpráce se diskuze metod a postupů dopadové evaluace odehrává v rámci kvantitativního kontrafaktuálního paradigmatu, jež je založeno na jednofaktorovém pojetí kauzality (kdy je naší snahou prostřednictvím kontrafaktuálu izolovat efekt programu na odchod z neza-

³ Pro potřeby jednotlivých statistických analýz jsou z mnohorozměrné tabulkové databáze dat OKpráce generovány anonymizované výtahy užší skupiny dat do dvourozměrného formátu (datové matice), kde každý uchazeč o zaměstnání je veden na jednom řádku, a jednotlivé údaje o uchazečích jsou vedeny ve sloupcích. Velikost datového souboru může být podle potřeb konkrétní analýzy od několika stovek či tisíců případů až po desítky tisíc účastníků APZ a stovky tisíc ostatních uchazečů o zaměstnání (pro potřeby vytvoření kontrolní skupiny) v jednom konkrétním roce. Z databáze jsou využívány především informace o jednotlivých uchazečích včetně informací o jejich historii nezaměstnanosti a dále informace o jednotlivých programech aktivní politiky zaměstnanosti.

městnanosti od efektu jiných faktorů) a zároveň dopad měříme kvantitativně prostřednictvím statistické analýzy. Z jiných přístupů zde zmíníme pouze konfigurační přístup, např. Booleho analýzu (Suchanec 2011), založený na komplexním (vícefaktorovém a interakčním) pojetí kauzality⁴.

V textu se dále zabýváme následujícími základními oblastmi evaluačního postupu: a) definicí hodnocených osob a programů, b) postupem stanovení rozsahu a cílenosti programu, c) definicí výsledku programu, d) hodnocením dopadu, e) problémem selekce a jeho důsledky, f) párováním případů (matching), g) problémem přeběhnutí a řešením nedokončené účasti v programu, h) definicí časových bodů pro měření, i) diskuzí působení času a jeho vlivu na evaluaci programů a j) technikami vyhodnocení výsledku. Uvedené oblasti představují podle našeho názoru klíčová metodologická místa důležitá pro vhodnou volbu konkrétního evaluačního postupu. Aplikační potenciál textu spatřujeme především v jeho využití při tvorbě konkrétního evaluačního postupu, konkrétně při přípravě dat pro hodnocení, volbě vhodného způsobu párování případů a volbě vhodné techniky pro vlastní analýzu.

Definice hodnocených osob a programů

Definice programu v první řadě zahrnuje vymezení jednotlivých programů pro hodnocení a jejich pro evaluaci významných aspektů. Vymezení programů podle jednotlivých typů a sub-typů bylo již v odborné literatuře mnohokrát prezentováno jak v ČR, tak v zahraničí (viz např. Bonoli 2010,

⁴ Hlavní devizou tohoto nestatistického přístupu je možnost identifikace nutných a dostačujících podmínek ke vzniku požadované události, a tím i identifikace potenciálních klíčových součástí programů nutných k dosažení pozitivního dopadu při relativně velmi malé velikosti zkoumaného vzorku. V případě vyčerpávajícího šetření a velké velikosti souboru však přístup není obecně vhodný, neboť se úměrně velikosti vzorku zvyšuje pravděpodobnost zastoupení všech potenciálních kombinací nezávislých a závislého faktoru, a analýzu tak není možné provést z důvodu ambivalence působení jednotlivých faktorů. V případě OKpráce tedy lze konfigurační přístup doporučit: a) pokud je naším cílem identifikace klíčových součástí lokálních programů malého rozsahu u specifických populací v konkrétních podmínkách, b) po agregaci a úpravě původního datového souboru, kdy jednotkou analýzy nebudou nezaměstnaní, ale programy a naším cílem bude např. identifikace klíčových elementů programů a vnějších podmínek skrze srovnání celkových dopadů jednotlivých programových cyklů.

Hora a Suchanec 2014) a nebudeme je zde opakovat. Upozorníme jen na některé významné specifické aspekty hodnocení.

- V evaluaci na datech OKpráce se zaměřujeme na osoby, které byly ve sledovaném období v evidenci Úřadu práce.
- Protože máme k dispozici údaje o všech nezaměstnaných, všech programech a současně i informace o regionech a poskytovatelích programů, je možné identifikovat regionální či jiné rozdíly mezi programy a vybrat vhodné neúčastníky pro kontrolní skupinu. Tato informace může být významná, pokud se dopady programů liší regionálně či u jednotlivých poskytovatelů.
- Data nám umožňují hodnotit všechny programy bez nutnosti definice sub-vzorků pro analýzu, pokud je nechceme definovat vzhledem ke specifickým cílům analýzy⁵.
- Jednotlivé programy musí mít definován skutečný počátek a konec intervence.

Specifickým problémem pro evaluaci je opakovaná (vícenásobná) účast nezaměstnaných v programech APZ ve sledovaném období, která byla popsána v ČR i v zahraničí (Hujer a Caliendo 2000, Richardson a van den Berg 2006, Hora a Suchanec 2014). V některých případech je obdobným problémem zapojení hodnoceného programu do širší skupiny intervencí – balíčku aktivit/služeb (Fay 1996).

Metodologický postup stanovení rozsahu a cílenosti programu

Z hlediska dopadové evaluace je významné především stanovení cílenosti (tedy výsledného zaměření programů na jednotlivé nezaměstnané v závislosti na jejich charakteristikách). I když je problém cílenosti implicitně řešen párováním případů (viz níže), má explicitní vyjádření cílenosti další přínosy. Za prvé může být pro politické činitele vodítkem ohledně zaměření jednotlivých programů, za druhé pak vytváří základní vodítko pro reflektovanou interpretaci dopadu programu (k problematice cílenosti viz např. Soukup 2006, Soukup, Michalička a Kotíková 2009, Hora et al. 2009).

⁵ Jedním ze základních evaluačních dilemat je výběr hodnocených účastníků programu. Častou situací je značná redukce dostupných dat (ořezání vzorku tak, aby splnil metodologická kritéria). Evaluátor tak získá nenáhodně vybraný sub-vzorek populace, pro který jsou naplněny jeho evaluační předpoklady, ale za cenu pochybnosti o nulovém vlivu takové selekce respektive za cenu snížení vnější validity výsledků (viz také Bryson, Dorsett a Purdon 2002).

Definice výsledku programu

V této evaluaci je základním hodnotícím kritériem posouzení dopadu jednotlivých programů. Východiskem hodnocení je shoda na definici očekávaného výsledku programů APZ. Mohr (1992) definuje výsledek následujícím způsobem:

- reaguje na nějaký společenský problém,
- je nějak společensky ceněný,
- jedná se zpravidla o stav a nikoliv o aktivitu.

Podle Wimera (2006) je možné výsledky dělit na pozitivní, neutrální a negativní, a to jak z matematického, tak z normativního hlediska. Tento přístup je hojně využíván v meta-analytických studiích (např. Calmfors, Forslund a Hemström 2002, de Koning a Peers 2007, Card, Kluve a Weber 2009). Výsledek programu poskytuje první orientační informaci o úspěšnosti programu a je základním předpokladem pro hodnocení dopadu programu⁶.

Výsledky můžeme rozlišit také na individuální, skupinové a společenské. Vliv programu není v jednotlivých případech možné uspokojivě předpovědět (Kluve et al. 2005), neboť je kontra-faktuální. Zároveň je možné předpokládat, že efekt programu se u jednotlivých účastníků odlišuje, a je proto závislý na zařazení konkrétních osob (Bryson, Dorsett a Purdon 2002). Hlavním předmětem zájmu v této dopadové perspektivě je proto průměrný efekt jednotlivých programů na účastníky těchto programů (viz např. Borland, Tseng a Wilkins 2005, Dias, Ichimura a van den Berg 2008, Suchanec 2014a)⁷.

⁶ Jakkoli je nejvyšší důraz kladen na zhodnocení dopadu programů, nelze při politickém rozhodování ani při interpretaci dopadu opomíjet meritorní význam dosaženého výsledku – např. kolik nezaměstnaných si po programu našlo zaměstnání.

⁷ Další potencionálně zajímavou možností pro některé typy výzkumných analýz je sledování distribuce výsledků u nezaměstnaných (distribuční efekty) – viz např. Heckman a Smith 1996, Hujer, Caliendo a Radić 2004. Toto rozlišení pomáhá mj. odpovědět na otázku, zda mezi nezaměstnanými neexistují významné rozdíly v dopadu z hlediska sledovaných charakteristik (někomu mohl program pomoci, zatímco jinému uškodit). Zahraniční zkušenosti z řady studií např. ukazují, že dopady programů APZ se odlišují u mladých nezaměstnaných do 25 let a dalších věkových skupin nezaměstnaných.

V evaluačních studiích (Betcherman, Olivas a Dar 2004, Kluve et al. 2005, Hujer, Thomsen a Zeiss 2006, de Koning a Peers 2007) můžeme nalézt např. následující druhy výsledků APZ:

- odchod z evidence/nepřítomnost v evidenci úřadu práce,
- zkrácení nezaměstnanosti,
- nalezení zaměstnání,
- udržení (stabilita) zaměstnání,
- zvýšení počtu odpracovaných hodin,
- zvýšení produktivity,
- zvýšení mzdy (příjmů),
- pracovní podmínky vytvořených zaměstnání,
- sociální participace,
- zdravotní stav.

Některé evaluační studie se zaměřily na dílčí výsledky, které jsou mezikroky k dosažení konečných výsledků, např. na hledačskou aktivitu nebo na reakce zaměstnavatelů na účast nezaměstnaných v programu (Calmfors, Forslund a Hemström 2002). Výsledek je zpravidla dále definován v rámci vzájemně se vylučujících kategorií (např. zaměstnaný, nezaměstnaný, ekonomicky neaktivní). Kvalita jeho měření pak závisí na schopnosti „správného“ přiřazení jednotlivých případů do těchto kategorií (viz Kluve et al. 2005). V dynamických modelech nemusí být výsledek vyjádřen jako stav. Vermunt (1996) upozorňuje, že tyto modely spíše pracují s pojmy změna (change), přechod (transition) nebo událost (event). Výsledek je pak např. míra výskytu události (viz níže způsob hodnocení výsledku).

Důležitým aspektem evaluací zaměřených na odchod z evidence je možnost rozpoznat typ odchodu: do zaměstnání, do podnikání, do vzdělávání, do ekonomické neaktivity, sankčním vyřazením či dokonce ztrátou kontaktu mezi nezaměstnaným a pracovníkem úřadu práce (viz např. Sianesi 2003). Podle Card, Kluve a Weber (2009) se zde pracuje s (implicitním a normativním) předpokladem, že odchod do zaměstnanosti je „dob-

rým“ typem výsledku, zatímco odchod z jiných důvodů je „špatný“ výsledek. Do této úvahy vstupuje např. následující:

- cíle politik,
- ti, kdo odešli z registru Úřadu práce, nejsou příjemci sociálních dávek (Sianesi 2003),
- širší hlediska ekonomického vlivu na veřejné rozpočty,
- důraz na etiku práce.

Card, Kluve a Weber (2009) uvádějí, že evaluační studie, které jsou zaměřeny na výsledek „délku registrované nezaměstnanosti“, přinášejí pozitivnější krátkodobé výsledky než studie, které se zaměřují na výsledek „zaměstnanost“ nebo „příjem“ (pro konkrétní příklad rozdílnosti výsledku viz Dias, Ichimura a van den Berg 2008).

Důsledkem chybějících informací o výsledku některých nezaměstnaných je riziko špatného přiřazení a tím i odhadu dopadu programu (Sianesi 2003). Současně hrozí riziko, že pravděpodobnost chybějící informace se systematicky odlišuje mezi skupinou intervence a kontrolní skupinou (Sianesi 2003). V některých evaluačních studiích autoři pracují s chybějícími výsledky jako s cenzorovanými případy, jindy se pokoušejí o odhad míry odchodu do zaměstnání. Bring a Carling (2001) řešili otázku neznámého důvodu odchodu nezaměstnaných z evidence provedením dodatečného telefonního šetření a zjistili, že více než polovina těchto osob odešla do ekonomické neaktivity. Dále nepochybně záleží na tom, jak velká proporce ekonomicky neaktivních se později vrátí do zaměstnání (v případě Bring a Carling 2001 byla tato proporce velmi malá)⁸.

Je třeba brát v úvahu, že účast v programu může mít očekávané i neočekávané vedlejší výsledky. Calmfors, Forslund a Hemström (2002) a Hujer,

⁸ V přechodných studiích v ČR měl výsledek vždy dichotomizovanou podobu („je v evidenci“/„není v evidenci“). Výhodou tohoto přístupu je přesné vymezení obou stavů, ovšem při značné nejistotě ohledně podstaty situace „není v evidenci“. V zahraničí, kde se zpravidla pracuje s konkrétní znalostí výsledků (jako je zaměstnání či odchod na rodičovskou dovolenou) někdy přetrvává problém s vysokým podílem chybějících údajů (tento problém existuje i v ČR). Z tohoto hlediska je významná diskuze, jak s těmito údaji pracovat metodologicky korektním způsobem, např. které stavy (situace) lze chápat jako trvání cenzorované zprava (viz např. Bring a Carling 2001, Richardson a van den Berg 2006).

Thomsen a Zeiss (2006) upozorňují na to, že programy mohou současně vytvářet protichůdné efekty (např. podporovat i potlačovat motivaci k hledání zaměstnání). Častým takovým důsledkem u některých typů programů je efekt uzamčení (locking-in) nebo automatické participace. Účastníci vzdělávacích programů mohou mít během trvání programu menší aktivitu při hledání zaměstnání a menší reálnou míru zaměstnanosti než neúčastníci, naopak účastníci programů tvorby míst jsou často automaticky chápáni jako zaměstnaní či nezaměstnaní (de Koning a Peers 2007, Card, Ibarra a Villa 2011, Forslund, Fredriksson a Vinkström 2011). Z tohoto důvodu je nutné dosáhnout shody na definici výsledku a odlišení jednotlivých stavů, a to především u programů tvorby míst.

Richardson a van den Berg (2006) na konkrétním programu testovali vliv efektu uzamčení. Zjistili, že i když byl při nezapočtení délky programu efekt dlouhodobějších programů větší než efekt programů krátkodobých, tento efekt byl zcela potlačen v případě, že se započítala délka programu (viz ibid.). Hora a Suchanec (2014) ukázali, že u účastníků více než jednoho programu byly dopady programů menší než u účastníků jediného programu. Posouzení efektu programu v takovém případě vychází z normativních hledisek⁹.

Dalším příkladem efektu, který je možné podle některých autorů (Suchanec 2014a) identifikovat na základě dat v perspektivě částečného ekvilibria, je tzv. efekt mrtvé váhy. Tito autoři odhadují efekt mrtvé váhy jako podíl osob v kontrolní skupině, které si našly zaměstnání i bez účasti v programu tvorby míst.

Je třeba poznamenat, že analýza v perspektivě částečného ekvilibria může vést ke zkreslenému odhadu, pokud je neadekvátně zobecněna. Tento přístup totiž není schopen odhalit část popsanych vedlejších efektů APZ na celý trh práce (srovnej např. Calmfors 1994).

V sumativní analýze není výsledek (na rozdíl od některých jiných na procesní evaluaci založených přístupů) dostatečným cílem evaluace, ale autoři usilují o zjištění dopadu (viz níže).

⁹ Např. potřebnost nezaměstnaných, vzdělávací efekt programu, (ne)vhodnost hodnocení efektu, který nemohl dosud nastat.

Hodnocení dopadu – způsob řešení kontra-faktuálního problému

Evaluační vychází z očekávání kauzálního vztahu mezi účastí v programu APZ a definovaným výsledkem. Dopad programu je podle Borland, Tseng a Wilkins (2005) definován takto: nakolik se změní měřený výsledek participací v programu...¹⁰. Základním evaluačním problémem je ovšem odlišení vlivu programu od vlivu dalších vnějších faktorů (Hujer, Caliendo a Radić 2004). Nejistota panuje především ohledně toho:

- zda je výsledek skutečně způsoben programem a ne něčím jiným, např. kompetencemi jednotlivých nezaměstnaných,
- zda by se nezaměstnaným nevedlo lépe bez programu (zda by si práci nenalezli, i pokud by se programu nezúčastnili).

Z tohoto důvodu by bylo žádoucí porovnat situaci stejného účastníka při účasti v programu se situací bez účasti v programu, což ovšem nelze. Efekt či dopad programu tedy představuje pomyslný rozdíl mezi dosaženým výsledkem účastníků programu a jejich kontra-faktuálním (rozuměj nenastalým) výsledkem, pokud by se programu nebyli účastnili (viz Hujer a Wellner 2000, de Koning a Peers 2007, Card, Ibarrán a Villa 2011). Protože přímé pozorování obou stavů současně není možné, je efekt programu odhadován (Hujer a Wellner 2000, Betcherman, Olivas a Dar 2004, Card, Ibarrán a Villa 2011). K odhadu (proxy) výsledku skupiny intervence v situaci bez programu se nejčastěji využívá výsledku, kterého dosáhli neúčastníci programu (Hujer, Caliendo a Radić 2004). Tato situace vyžaduje tzv. definování identifikačních předpokladů (Bryson, Dorsett a Purdon 2002), tedy podmínek, které musí být splněny, aby byly výsledky evaluace platné (viz též níže).

Ve většině evaluačních studií dochází k odhadu parametru populace z menšího výběrového vzorku s rizikem nepřesnosti odhadu (viz Kluve et al. 2005). Mohr (1992) definuje citlivost designu jako měřítko toho, jak malý může být rozdíl ve výsledku mezi experimentální skupinou a kontrafaktuální skupinou, abychom stále věřili, že program má pozitivní kauzální dopad. Rozdíl, který jsme schopni detekovat, záleží na přesnosti našeho odhadu a řadě dalších faktorů (mj. velikost vzorku). Citlivost designu je ovšem významná pouze tam, kde účastníky programu vybíráme

¹⁰ Respektive ve srovnání se situací neúčasti v programu.

z populace nezaměstnaných náhodně při snaze zobecnit na tuto populaci, což v případě administrativních dat OKpráce neplatí.

Jelikož v případě analýzy na datech OKpráce nelze účastníky a neúčastníky rozdělit do obou skupin náhodně (jak je tomu v klasickém experimentu), vychází konstrukce obou skupin pro měření dopadu dodatečně z rozhodnutí výzkumníků na základě skutečných dat. Tento postup je některými autory označován za kvazi-experimentální design výzkumu, zatímco jiní autoři jej chápou jako design studie na základě pozorování. Hlavní výhodou takového přístupu je nízká míra rizika narušení prostředí výzkumem. Zásadní nevýhodou je ovšem potřeba vypořádat se s problémem selekce a dalšími problémy, které vyplývají z povahy dostupných dat.

Problém selekce a jeho důsledky

Dopad programu je validně odhadnut pouze, pokud nedochází k tzv. zkreslení na základě selekce (Hujer a Wellner 2000). Problém zkreslení na základě selekce vzniká (při nepřítomnosti náhodného výběru účastníků a neúčastníků programu) v důsledku existence různých okolností vedoucích k účasti či neúčasti jednotlivých nezaměstnaných v programech APZ. Takovými okolnostmi mohou být např. podmínky nároku pro vstup do programu, kdy jsou preferováni obtížně umístitelní uchazeči, samo-výběr účastníků do programu, výběr na základě tzv. efektu slízávání smetany (cream-skimming)¹¹ či předpoklad nástupu do zaměstnání (Bryson, Dorsett a Purdon 2002, Lechner a Wunsch 2009, Card, Ibarrán a Villa 2011).

Jako základní kritéria posouzení rizika problému selekce jsou brány např. (ne)jednotnost a (ne)centralita zařazování či (ne)zařazování podle jasných kritérií¹². Při nedodržení těchto hledisek roste nejistota ohledně skrytých faktorů působících na různost obou skupin a tím i pochybnost

¹¹ K efektu slízávání smetany dochází, pokud jsou do programu vybráni účastníci, kteří by nejpravděpodobněji uspěli i bez programu na úkor nezaměstnaných, kterým by program mohl nejvíce pomoci (viz např. Blache 2011).

¹² Jednotlivá zařazení do programu se tedy pravděpodobněji budou odlišovat, pokud osoby do programu zařazují různé osoby, a každá z nich využívá jiná kritéria pro zařazení nebo tato kritéria jinak interpretuje, anebo pokud sami nezaměstnaní mohou rozhodovat, zda se programu zúčastní.

o shodnosti obou skupin. V důsledku selekce je rozdíl mezi výsledky účastníků a neúčastníků programu způsoben jak samotným programem, tak současně či výhradně řadou dalších potencionálních faktorů (viz Mohr 1992, Card, Ibarrán a Villa 2011). Praktický příklad důsledku nepravého efektu v důsledku selekce uvádí Richardson a van den Berg (2006). Pokud účastník po programu poměrně rychle opustí evidenci, může to být v důsledku a) pozitivního kauzálního efektu programu nebo b) skrytých faktorů, které by vedly k rychlému odchodu z evidence i bez programu. Skutečný rozsah problému selekce přitom často není výzkumníkovi znám (Fay 1996).

Výzkumník tedy nemůže z výše uvedených důvodů realisticky předpokládat, že by k problému selekce nedocházelo (mj. proto, že je dlouhodobě zdokumentována cílenost programů APZ na určité skupiny, a tudíž šance na vstup do programu rozhodně není pro všechny stejná). V této metodice zároveň vycházíme z předpokladu, že výzkumník nemůže v neexperimentálním výzkumu proces selekce efektivně ovlivnit¹³, a proto musí očekávat existenci před-programových rozdílů mezi skupinou intervence a kontrolní skupinou a volit adekvátní řešení za pomoci korekce volbou vhodného designu a estimátoru (Hujer a Caliendo 2000).

Kluve et al. (2005) dělí faktory s možným vlivem na výsledky programu na: (a) současné charakteristiky a (b) předchozí výsledky na trhu práce. Kluve et al. (2005) uvádí, že je důležité, aby byly tyto faktory vnějšími faktory, tedy jde o to nepodmiňovat na faktorech, jejichž změna může být výsledkem participace v programu. Faktory, které mohou způsobovat odlišnost skupin lze dále rozdělit na snadno zjištělné (takové, které lze ve formě proměnných zahrnout do výzkumného souboru a lze pro ně kontrolovat) a obtížně zjištělné či nezjištělné. Toto rozdělení má zásadní význam pro identifikaci vhodného způsobu řešení problému selekce. Problém nepozorované heterogenity v důsledku nezařazení proměnných do modelu jsou závažným rizikem i u modelů hazardu (Vermunt 1996). Bryson, Dorsett a Purdon (2002) poukazují na to, že zahrnutím pozorovaných proměnných můžeme částečně zahrnout i vliv nepozorovaných proměnných, pokud jsou tyto vzájemně korelovány (např. před-programová historie a motivace).

¹³ I v designu pravého experimentu může nastat selekce z důvodu randomizace a/nebo substituce (Suchanec 2014b).

Řešení problému selekce za pomoci párování případů (matching)

Párování případů (často v zahraničí označováno jako tzv. matching), je jedním z často používaných postupů rozdělení vzorku na dvě srovnatelné skupiny: účastníky programu a neúčastníky programu¹⁴. Podstatou procesu párování je výběr osob do kontrolní skupiny (z širšího okruhu osob) tak, aby před-programové charakteristiky těchto osob co nejvíce odpovídaly charakteristikám osob, které se programu účastnily (de Koning a Peers 2007, Card, Ibarrán a Villa 2011). Jinými slovy, cílem párování je vybalancovat rozložení všech relevantních před-programových charakteristik v obou skupinách, a tím dosáhnout nezávislosti mezi potencionálními výsledky a zařazením do programu (Hujer, Caliendo a Radić 2004). Identifikační předpoklad podmíněné nezávislosti (conditional independence assumption – CIA) totiž vyžaduje, aby individuální účast v programu byla nezávislá na potenciálním výsledku programu v situaci bez programu (Sianesi 2003, Borland, Tseng a Wilkins 2005, Dias, Ichimura a van den Berg 2008). CIA je splněna pouze pokud všechny proměnné, které ovlivňují jak proces selekce, tak potenciální výsledek programu, jsou zahrnuty do párování (Bryson, Dorsett a Purdon 2002, Reinowski a Schultz 2006). Neidentifikované faktory jsou též potenciálně nebezpečné pro zkreslení výsledků pouze tehdy, pokud současně souvisejí jak s participací v programu, tak s potenciálním výsledkem. Z výše uvedeného vyplývá, že ke zkvalitnění párování dochází zahrnutím co nejširší skupiny z výše uvedeného hlediska relevantních proměnných.

Párování je ovšem založeno též na předpokladu společné podpory (common support assumption – CSA), tedy že ke každému účastníkovi získáme shodného či obdobného neúčastníka (Sianesi 2003, Borland, Tseng a Wilkins 2005, Reinowski a Schultz 2006). Rizikem procesu párování je proto narůstající počet proměnných zahrnutých do procesu párování, neboť s každou další proměnnou exponenciálně narůstá i počet buněk a tím možností, kde se párované osoby mohou odlišovat (viz např. Dehejia a Wahba 2002, Hujer, Caliendo a Radić 2004). CSA může být pro-

¹⁴ Jedním z alternativních postupů pro řešení problému selekce může být nespojitá regrese (regression discontinuity design). Její využití vyžaduje ovšem spíše specifické podmínky, které je obtížné naplnit. Pro vysvětlení způsobu, jak se nespojitá regrese vypořádává s problémem selekce, viz Suchanec (2014b), pro konkrétní využití při evaluaci dopadu viz např. Dias, Ichimura a van den Berg (2008).

blémem také v případě povinných programů, neboť není možné najít odpovídající neúčastníky (Bryson, Dorsett a Purdon 2002).

Při párování případů se jako proměnná pro párování používá tzv. propensity score tedy podmíněná pravděpodobnost pro účast v programu (Hujer a Wellner 2000, Dehejia a Wahba 2002). Velkou výhodou tohoto postupu je možnost párovat na jediné proměnné (Hujer, Caliendo a Radić 2004). Další značnou výhodou je, že díky propensity score můžeme párovat na velkém množství proměnných bez rizika zkreslení výsledku (Dehejia a Wahba 2002). Párování je spojeno s rizikem zkreslení především v těch případech, kdy v kontrolní skupině nenalézáme vhodné případy pro spárování (Hujer, Caliendo a Radić 2004). Z toho vyplývá, že je výhodou mít dostatečný počet osob, ze kterých je možné do kontrolní skupiny vybírat. Případy, pro které nemáme odpovídající protějšky v kontrolní skupině, nemůžeme zahrnout do procesu párování. V některých případech autoři kombinovali využití propensity score s přesným párováním na konkrétní proměnné historie před programem (Puhani 1998, Hora a Suchanec 2014) – k důvodům pro využití tohoto přístupu viz níže definice časových bodů.

Podle možnosti je ke každému účastníku programu vyhledán jeden či více neúčastníků anebo je jeden neúčastník využit i vícekrát (Hujer, Caliendo a Radić 2004). Existuje více způsobů využití propensity score pro párování na základě dostupnosti vhodných případů pro párování (viz poznámka)¹⁵. Při využití procesu párování podle skupin jsou někdy použity váhy, které zajišťují stejnou distribuci charakteristik ve skupině intervence a v kontrolní skupině podle počtu případů (Kluve et al. 2005).

Vhodným způsobem párování na datech OKpráce může být párování bez náhrady s náhodným přiřazením osob z kontrolní skupiny (viz např. Dehejia a Wahba 2002, Reinowski a Schultz 2006). Párování bez náhrady v praxi znamená, že každý případ je zařazen do párování pouze jednou,

¹⁵ Základními způsoby jsou přesné párování (exact matching), párování „nejbližšího souseda“ (nearest neighbour matching), párování nejbližšího souseda ve vymezeném intervalu (caliper matching) či vážení případů na základě jejich podobnosti s párovaným případem (kernel matching) – viz např. Hujer a Wellner 2000, Bryson, Dorsett a Purdon 2002, Hujer, Caliendo a Radić 2004. V případě OKpráce dosud vždy bylo možné párovat na přesném propensity score, což je ve srovnání s párováním v intervalu přístup, který by měl nejvíce redukovat potenciální zkreslení.

tj. je-li vytvořen pár, oba případy jsou z dalšího párování vyřazeny (Dehejia a Wahba 2002). To je pravděpodobně vhodnější přístup, pokud máme vysokou podobnost respektive možnost podobnosti kontrolní skupiny a skupiny intervence (viz Dehejia a Wahba 2002)¹⁶. Nevýhodou tohoto přístupu je, že může být citlivý na pořadí řazených případů (Dehejia a Wahba 2002). Velkou výhodou dat OKpráce je vysoký počet dostupných případů pro párování. V minulých analýzách bylo možné spárovat zpravidla 80–95 procent případů. Zpravidla pro téměř všechny případy ve skupině intervence máme vyšší počet potencionálních partnerů v kontrolní skupině. Tyto případy lze přiřadit ke skupině intervence pomocí náhodných čísel. Alternativním přístupem by mohlo být zaměření se na párování dosud nespárovaných případů¹⁷.

Nevýhodou procesu párování (matching) je potenciální neznalost skrytých charakteristik (např. motivace, aktivita při hledání zaměstnání, kompetence nezaměstnaných), které mohou mít vliv jak na vstup do programu tak na šance nezaměstnaných na trhu práce (de Koning a Peers 2007, Dias, Ichimura a van den Berg 2008, Forslund, Fredriksson a Vinkström 2011). Richardson a van den Berg (2006) předpokládají, že skrytým faktorem s vlivem na efekty programu je zvýšená hledačská aktivita zprostředkovatelů zaměstnání bezprostředně po skončení vzdělávacího programu.

Řešení problému potenciální existence nepozorovaných proměnných je obtížné. Puhani (1998) doporučuje provádět test nepozorované heterogeneity. De Koning a Peers (2007) navrhuji porovnávat šance nezaměstnaných na nalezení zaměstnání před programem a po programu, což by

¹⁶ V praxi jsou pro párování významná dvě základní hlediska: a) velikost skupiny potenciálních partnerů pro párování do kontrolní skupiny, b) definovaná kritéria pro potřebnou blízkost hodnot propensity score v rámci jednoho páru. V našem případě je množství případů využitelných pro párování velmi vysoké, a proto nemusíme případy do párování zařazovat opakovaně a přesto získáváme většinu párů s přesnou shodou na propensity score.

¹⁷ V předchozí analýze (Hora a Suchanec 2014) jsme při poměrně komplexním způsobu párování nespárovali asi 15 procent případů. Dodatečná analýza ukázala, že tyto případy dosahovaly horších průměrných místních výsledků než 85 procent spárovaných případů. U těchto případů zpravidla není možné najít srovnatelný případ v kontrolní skupině. Alternativním přístupem by proto mohlo být: a) upřednostnit tyto případy při párování, b) uvolnit u nich kritérium přesnosti odhadu v propensity score, c) povolit vícenásobné použití případu. Základní evaluační otázkou je, zda chceme přesný odhad dopadu o 85 procentech případů či méně přesný odhad u většího podílu případů.

mohlo pomoci, pokud se nepozorované faktory nemění v čase či různě u obou srovnávaných skupin. Lze předpokládat, že zahrnutí proměnné historie před programem, která indikuje šance nezaměstnaného na trhu práce, do párování může významně pomoci snížit riziko působení nepozorovaných faktorů¹⁸.

Nyní se zaměříme na otázku identifikace vhodných faktorů pro párování. Bryson, Dorsett a Purdon (2002) upozorňují, že volba těchto faktorů by měla být založena na dobré znalosti teorie a výsledků předchozích výzkumů. Zpravidla jsou využívány např. věk, pohlaví, vzdělání, zdravotní stav, region úřadu práce, předchozí obor činnosti atd. Zde se krátce zaměříme na diskuzi především některých těchto faktorů.

Jedním z klíčových faktorů úspěšnosti programů je před-programová perspektiva účastníků programu z hlediska nalezení zaměstnání. Výzkumy (např. Lechner a Wunsch 2009, Hora a Suchanec 2014) ukázaly, že dopady programů APZ jsou horší u nezaměstnaných s lepšími šancemi na nalezení zaměstnání. Dalším významným faktorem, který je nutné zahrnout do zjišťování efektů programu je před-programová historie účastníků a neúčastníků programu vzhledem k okolnosti, u níž očekáváme efekt programu např. délka nezaměstnanosti (Bryson, Dorsett a Purdon 2002, Card, Ibarra a Villa 2011, Forslund, Fredriksson a Vinkström 2011). V předchozí analýze (Hora a Suchanec 2014) jsme prokázali, že předchozí délka nezaměstnanosti má významný vliv na šance nezaměstnaných odejít z evidence Úřadu práce. Kromě této skutečnosti mají uvedené informace význam též pro posouzení cílenosti programu z hlediska načasování.

Výzkumník se může s problémem selekce vyrovnat především volbou vhodného způsobu odhadu dopadu (estimátoru) za pomoci věrohodné identifikační podmínky. V našem případě ke kontra-faktuálnímu odhadu využíváme párování s pomocí tzv. propensity score matching.

¹⁸ Někteří autoři interpretují tyto informace v jiném kontextu, a to jako načasování intervence (viz např. Hujer, Thomsen a Zeiss 2006).

Problém přeběhnutí (crossover, spillover) a řešení nedokončené účasti v programu

Rozdělení nezaměstnaných na účastníky a neúčastníky je dále komplikováno rizikem špatného zařazení nezaměstnaných do skupiny intervence a do kontrolní skupiny (např. při účasti neúčastníků v programu APZ či v jiném programu) – tedy problémem tzv. „cross-over“ (Card, Ibarra a Villa 2011).

První z výše uvedených situací (špatné zařazení do skupiny intervence) nastává, pokud panuje nejistota ohledně správnosti zařazení nezaměstnaných do skupiny intervence. Card, Ibarra a Villa (2011) identifikují dvě skupiny:

- a) nezaměstnané, kteří se do programu přihlásili, ale pak do něj nenastoupili,
- b) nezaměstnané, kteří v programu participovali, ale program nedokončili.

Card, Ibarra a Villa (2011) se domnívají, že obě tyto skupiny by měly být chápány jako účastníci skupiny intervence. Účast v programu může být též definována jako absolvování alespoň určité části programu (Card, Ibarra a Villa 2011).

Při druhé situaci (špatné zařazení do kontrolní skupiny) můžeme (někdy nevědomě) porovnávat účastníky sledovaného programu s účastníky jiných programů (např. při chybném zařazení nezaměstnaných do kontrolní skupiny na základě chybějící informace o jejich účasti v programu). Dále Calmfors, Forslund a Hemström (2002) uvádí, že při vyšším rozsahu APZ je pravděpodobné, že každý nezaměstnaný se zúčastnil, zúčastňuje nebo bude účastnit APZ, a proto je obtížné nalézt kontrolní skupinu bez jakékoliv účasti. Z tohoto hlediska se také může snadno stát, že evaluátor srovnává účastníky jednoho programu s účastníky jiného programu realizovaného později (ibid.). Tento problém byl popsán v řadě evaluačních studií (viz např. Heckman a Smith 1996, Sianesi 2003).

Praktický význam problému cross-over při hodnocení APZ lze identifikovat v následujících situacích.

- K účasti nezaměstnaného z kontrolní skupiny v programu mohlo dojít před programem, jehož výsledky plánujeme hodnotit, v době trvání tohoto programu, i době po skončení tohoto programu.
- Programy mohou být vzájemnými substituty (Heckman a Smith 1996, Richardson a van den Berg 2006).
- Nezaměstnaní mohli být účastníky jiných programů APZ již před začátkem sledovaného programu což může mít vliv jak na jejich pravděpodobnost odchodu do zaměstnání (v přechodném programu, po předchozím programu), tak na pravděpodobnost zařazení do sledovaného programu (Richardson a van den Berg 2006).
- To, že nezaměstnaní získali jiný program, může a nemusí být evidováno.

Někteří autoři se přiklánějí k názoru, že není správné, potřebné nebo nezbytné klást omezení ohledně statusu účastníků kontrolní skupiny (Sianesi 2003, Dias, Ichimura a van den Berg 2008). Tato úvaha může být též modifikována různými časovými hledisky. Specifickým řešením tohoto problému je záměrné srovnávání různých skupin účastníků programu (zajímavá je v tomto ohledu např. studie Sianesi 2003 o vlivu časování vstupu do programu). Další možností je záměrné srovnávání jednotlivých typů programů mezi sebou (Bryson, Dorsett a Purdon 2002).

Řešení problému cross-over v sobě zahrnuje dvě roviny: (a) normativní, tedy učinění rozhodnutí, zda osoby účastníci jiných programů zahrnovat do kontrolní skupiny, a (b) praktickou, kdy jsou autoři v některých zemích (zejména ve Švédsku) k tomuto postupu „nuceni“ neboť nelze nalézt nezaměstnané bez účasti v programu. V ČR je nalezení vhodných nezaměstnaných do kontrolní skupiny poměrně snadné, proto se jedná spíše o normativní rozhodnutí.

Definice časových bodů pro měření výsledku a dopadu

Základem pro vymezení časových bodů je vymezení sledovaného období. Jedná se o takové období, ke kterému lze logicky vztahovat prezentovaná data a které odpovídá záměrům hodnocení. Především se jedná o:

- informace o nezaměstnaných (účastnících a neúčastnících programů),
- informace o programech,
- informace o evidencích.

V rámci jedné evaluace je možné mít sledované období definováno různě pro různé okruhy údajů (neměnné údaje např. pohlaví se zpravidla sledují pouze jednou).

Významným metodologickým problémem je práce s údaji, které jsou užitečné pro hodnocení, ale nejsou částečně dostupné nebo se nacházejí vně sledovaného období (tzv. cenzory). Jedná se tedy o formu částečně chybějících údajů (viz Vermunt 1996). Důvodem existence cenzorů je neznalost historie před sledovaným obdobím (levý cenzor), konečnost sledovaného období (ty, kdo přišli později, je možné sledovat po kratší dobu – pravý cenzor) a nastání jiné události či zmizení z datového souboru (nevíme, co se s osobami stalo) během sledovaného období (Vermunt 1996).

Problém nastává především tehdy, pokud můžeme očekávat, že chybějící údaje mohou zkreslovat výsledky programů. Vermunt (1996) uvádí, že vyřazení případů s chybějícími (cenzorovanými) hodnotami může závažně zkreslit výsledky, pokud není distribuce těchto případů náhodná, a proto doporučuje zachovat maximální možné množství informací a snažit se zabránit zkreslení selekce. K řešení cenzorů zprava jsou využívány techniky vycházející z předpokladu, že míra cenzorování je nezávislá na sledovaném výsledku. Jedním z řešení je kontrolovat pro data vstupu do analýzy a další proměnné, které by mohly být příčinou systematického cenzorování (viz ibid.). Vermunt (1996) dále píše, že v případě existence cenzorů zleva je možné situaci řešit odstraněním těchto případů (doporučuje se, jen pokud jich je jen malý počet), zatímco odhadování cenzorovaných dat je obtížné. Je totiž možné předpokládat, že případy cenzorované zleva neodpovídají případům, o kterých máme kompletní informace (dochází ke zkreslení selekce). Běžným řešením levých cenzorů je tedy vyřa-

zení těchto případů z analýzy (např. Puhani 1998). V případě dat OKpráce není problémem přítomnost levých cenzorů, zatímco přítomnost pravých cenzorů je z logiky věci často nevyhnutelná.

Evaluace vyžaduje definice některých přesných časových bodů. Podle konkrétního typu hodnocení potřebujeme identifikovat situaci před programem, v programu a po programu.

Pro identifikaci před-programového stavu je podle Card, Ibarrán a Villa (2011) jako základní bod sledována situace těsně před začátkem programu. Řada autorů upozorňuje, že situace těsně před programem se může výrazně odlišovat od dřívější situace nezaměstnaných. Hujer, Thomsen a Zeiss (2006) např. předpokládají, že nezaměstnaní, kteří se dozvědí o své účasti v programu, mohou změnit své strategie hledání zaměstnání. Vliv tohoto efektu (popsaného poprvé Ashelferterem) se pravděpodobně bude odlišovat též podle zvoleného výsledku. V případě využití dat OKpráce a volby výsledku (ne)přítomnost v evidenci jsou zpravidla před programem všichni účastníci programu chápáni jako nezaměstnaní (implicitní pre-test) a sledován je jen podíl v evidenci/zaměstnaných po programu.

Velmi významnou otázkou je, zda měřit dopady programů od začátku nebo od konce programu, a jak v takovém případě definovat časové body pro začátek sledování u kontrolní skupiny. Některé studie prokázaly, že efekty programů se mohou odlišovat v závislosti na zvolení jednoho nebo druhého z těchto přístupů (srovnej Calmfors, Forslund a Hemström 2002).

U skupiny intervence můžeme uvažovat o následujících časových bodech: a) začátek evidence před programem (případně další před-programové evidence), b) datum začátku programu (data začátku více programů), c) datum konce programu (konců více programů).

Pro odvození základního bodu výpočtu pro kontrolní skupinu jsou využívány různé přístupy (viz např. Larson 2001 podle Calmfors, Forslund a Hemström 2002, Sianesi 2003, Lechner a Wunsch 2009).

Larson (2001 podle Calmfors, Forslund a Hemström 2002) uvádí, že pokud sledujeme výsledky programu po programu, a ne od počátku pro-

gramu, můžeme stanovit startovní bod pro kontrolní skupinu dvěma způsoby:

- a) zařazením členů kontrolní skupiny, kteří se stali nezaměstnanými ve stejném čase jako účastníci skupiny intervence a v době ukončení programu byli stále nezaměstnaní,
- b) zařazením členů kontrolní skupiny, kteří byli v době ukončení programu stejně dlouho nezaměstnaní jako účastníci skupiny intervence před programem.

Podle Larsona oba tyto způsoby vedou ke zkreslení (zvýšení) dopadu programu. V prvním případě proto, že v kontrolní skupině budou nadměrně zastoupeni nezaměstnaní s horšími šancemi na trhu práce. Ve druhém případě získávají účastníci programu delší čas na nalezení zaměstnání, byť lze očekávat menší intenzitu jejich hledání v průběhu realizace programu.

Pravděpodobně z těchto důvodů Puhani (1998) a Forslund, Fredriksson a Vinkström (2011) sledují výsledky programu od počátku programu. Někteří autoři dále rozlišují efekt intervence na část během programu a na část po programu. Puhani (1998), Sianesi (2003) a Hora a Suchanec (2014) stanovují podmínku, že jedinci v kontrolní skupině strávili v nezaměstnanosti alespoň takovou dobu, jako účastníci programu před vstupem do programu. Cílem tohoto přístupu je postavit účastníky na stejnou „startovní čáru“. Puhani (1998) tento postup zdůvodňuje tím, že historie před programem je známá pouze pro skupinu intervence, a proto nemůže být zahrnuta do propensity score.

Zásadním poznatkem z hlediska hodnocení dopadů APZ je možnost proměnlivosti dopadu APZ v čase, která může nastat např. tehdy, pokud je k projevení efektu potřeba určitý čas či naopak, pokud k efektu programu dochází nejvíce bezprostředně po skončení programu (Richardson a van den Berg 2006, Hujer, Thomsen a Zeiss 2006). Evaluační literatura zpravidla rozlišuje krátkodobé, střednědobé a dlouhodobé výsledky programů. Jako krátkodobé výsledky jsou označovány výsledky do jednoho roku, jako střednědobé a dlouhodobé výsledky pak výsledky v období více než dvou let (viz např. Card, Kluve a Weber 2009, Card, Ibarrán a Villa 2011)¹⁹. Pre-

¹⁹ Předchozí kvazi-experimentální studie ukázaly, že krátkodobé a dlouhodobé dopady programů APZ se mohou odlišovat (krátkodobé dopady byly méně příznivé než dlouhodobé dopady) (viz např. Calmfors, Forslund a Hemström 2002, Lechner a Wunsch 2009, Card, Kluve a Weber 2009). Opačná situace může nastat u programů tvorby míst

zentace dlouhodobých výsledků je často v ČR i v zahraničí omezena nedostupností vhodných dat a snahou přinést aktuální údaje (viz např. Lechner a Wunsch 2009). V modelech sledujících riziko spíše hovoříme o osobách vystavených riziku, o rizikovém období, tedy období, kdy je osoba vystavena riziku a trvání, než došlo ke sledované události (viz Vermunt 1996).

Diskuze působení času a jeho vlivu na evaluaci programů APZ

Výsledky programů APZ mohou být ovlivněny cyklickými a sezónními vlivy (Card, Ibarrán a Villa 2011, Forslund, Fredriksson a Vinkström 2011). Rozsah programů APZ má často pro-cyklický charakter (viz např. Calmfors, Forslund a Hemström 2002). Forslund, Fredriksson a Vinkström (2011) předpokládají (a empiricky potvrdili), že hlavním faktorem je zde rozdílná šance nezaměstnaných nalézt si zaměstnání v konjunktúře a v recesi, která je v interakci s participací v programu. Srovnání efektů v různých částech hospodářského cyklu vyžaduje naplnění řady dalších předpokladů o shodnosti dalších okolností ve srovnávaných obdobích (viz Forslund, Fredriksson a Vinkström 2011).

Kluve et al. (2005) a Borland, Tseng a Wilkins (2005) uvádějí, že kvazi-experimentální postupy jsou závislé na předpokladu stabilní hodnoty intervence (SUTVA). Tento předpoklad zahrnuje očekávání, že: (a) efekt intervence na každého jednotlivce není ovlivněn rozhodnutím o participaci dalších jednotlivců (Bryson, Dorsett a Purdon 2002, Kluve et al. 2005) či také (b) efekt účasti v programu je v čase stabilní a (c) výsledky neúčastníků nejsou programem ovlivněny (Bryson, Dorsett a Purdon 2002, Borland, Tseng a Wilkins 2005). Richardson a van den Berg (2006) uvádějí, že míra odchodu z evidence po programu se v čase odlišuje (klesá). Autoři upozorňují, že tento jev může být spojen jak s klesajícím efektem programu, tak s vlivem nepozorovaných proměnných.

Richardson a van den Berg (2006) např. uvádějí, že jejich model vychází z předpokladu neexistence vlivu programu před programem. K takovému vlivu může docházet, pokud kandidáti účasti v programu anticipují svou účast a mění na základě toho své chování.

(viz Reinowski a Schultz 2006). Z tohoto důvodu je doporučováno měřit výsledky ve více časových bodech včetně bodů minimálně dva roky po absolvování programu (např. Card, Kluve a Weber 2011).

Techniky vyhodnocení výsledku

Z hlediska pojetí času můžeme v závislosti na povaze dat a metodologických záměrech evaluátorů rozlišit přístupy založené na nespojitém čase (discrete time), událost může nastat v určitý určený čas, a přístupy založené na spojitém čase (continuous time), kde událost může nastat kdykoliv (Vermunt 1996). Nezaměstnaný může mít během sledovaného období jednu či více evidencí, které jsou ohraničeny přesnými časovými body. Data OKpráce mají klíčové údaje uvedeny s přesností na jednotlivé dny, a proto umožňují využití oběma základními způsoby. V evaluační literatuře (de Koning a Peers 2007, Card, Kluve a Weber 2009) se dále setkáváme s následujícími dvěma přístupy k hodnocení.

V prvním případě je výsledek definován jako stav v určitém časovém bodě, tedy např. kolik osob nebylo v určitém časovém bodě v evidenci. Ačkoliv je tento výsledek někdy též interpretován jako „kolik nezaměstnaných do určité doby opustilo evidenci Úřadu práce“, je třeba vzít v úvahu, že záleží na tom, zda sledujeme situaci i v jiných časových bodech. Vermunt (1996) upozorňuje, že změnu zjistíme v tomto případě pouze tehdy, pokud se obě měření odlišují. Na druhé straně, pokud se obě hodnoty shodují, mohlo či nemuselo mezi jednotlivými měřeními dojít ke změně. Např. osoba, která je v evidenci po roce od ukončení programu mohla být během tohoto roku 11 měsíců mimo evidenci, a přesto bychom mohli mylně předpokládat, že byla v evidenci ÚP po celých 12 měsících. Totéž platí pro období následující po sledovaném časovém bodě. Je ovšem možné sledovat situaci ve více časových bodech. Pro tento typ analýzy je možné využít např. logistickou regresi (Bryson, Dorsett a Purdon 2002). Příkladem evaluace provedené tímto způsobem je např. Reinowski a Schultz (2006).

Ve druhém případě hovoříme o časování událostí (timing of events) ve kterém jsou využívána srovnávání rizik na odchod z evidence v různých časových obdobích např. před programem a po programu (Richardson a van den Berg 2006, de Koning a Peers 2007). V těchto modelech je sledováno, zda události nastaly, případně v jakém pořadí, základní jednotkou pro zjišťování dopadu je trvání času, než poprvé či opakovaně nastane či nenastane sledovaná událost (Vermunt 1996). Vermunt (1996) rozděluje modely na parametrické modely založené na spojitém čase (kam

zařazuje analýzu přežití²⁰), semi-parametrické modely založené na spojitém čase (Coxova regrese) a analýzy založené na nespojitém čase (především různé varianty „discrete time hazard rate model“). Modely předpokládají různou (nikoliv stejnou) délku individuální účasti v nezaměstnanosti před programem (Richardson a van den Berg 2006). Některé modely umožňují časově se odlišující hodnoty nezávislých proměnných (viz Vermunt 1996). V takovém případě tedy můžeme např. zachytit, že se během sledovaného období změnila úroveň formálního vzdělání účastníka programu. Práce s časově proměnlivými hodnotami nezávislých proměnných může být významná především tam, kde existuje teoretické opodstatnění pro tento přístup. Příklady evaluací provedených tímto způsobem jsou např. van Ours (2002) či Richardson a van den Berg (2006).

Při rozhodování o volbě technik analýzy zvažujeme zejména odpovědi na čtyři otázky související s vytyčeným cílem evaluace:

- Chceme znát průběh velikosti rizika odchodu z nezaměstnanosti v čase nebo si vystačíme s velikostí rizika v určitém časovém bodě (např. půl roku po programu)?
- Chceme znát velikost rizika odchodu z registru v určitém čase nebo informaci o tom, jaká část nezaměstnaných již registr opustila?
- Chceme znát velikost skutečného (absolutního) rizika odchodu nebo se spokojíme s velikostí rizika ve srovnání s jinou skupinou nezaměstnaných?
- Požadujeme dodatečnou kontrolu efektů jiných proměnných?

Modelování pomocí logistické regrese neumožňuje posouzení rizika a tím i dopadu programu v průběhu času, nýbrž je zapotřebí, aby evaluátor zakotvil svou analýzu v určitém výběrovém okamžiku (často po jednotlivých měsících). Logistická regrese také volbou určitého časového bodu neumožňuje posouzení rizika odchodu, ale pouze posuzuje přítomnost či nepřítomnost nezaměstnaných v registru. Silnou stránkou je možnost posouzení absolutních rizik opuštění rizik. Slabinou je naopak možnost

²⁰ Podle Vermunta (1996) pravděpodobnost přežití (či funkce přežití) indikuje pravděpodobnost, že jev nenastane v čase t . Míra hazardu (či funkce hazardu) vyjadřuje nynější riziko, že událost nastane v čase t , pokud již nenastala v dřívějším období. Míra hazardu může dosahovat hodnot vyšších než jedna. Obdobná míra je využívána i v modelech s nespojitým časem pod názvem podmíněná pravděpodobnost (ibid.).

zkreslení dopadu nejen z důvodu zakotvenosti analýzy pouze ve svévolně vybraných časových bodech, ale i omezená možnost vypořádat se ve sledovaném období s chybějícími informacemi o odchodu nezaměstnaných (tzv. cenzory).

Modelování pomocí analýzy přežití poskytuje zřejmě nejpřehlednější obrazy o průběhu rizik odchodu nezaměstnaných z registru (a srovnáním těchto rizik u skupiny účastníků a neúčastníků i o dopadu programu) v čase. Navíc umožňuje posouzení absolutních rizik v čase i zjištění velikosti proporce účastníků, kteří již registr opustili v jakémkoli časovém bodě. Kontrola efektu jiných proměnných je možná skrze zastoupení těchto proměnných ve zkoumaných kombinacích, nicméně při kombinacích vyššího řádu velmi pracná. Metoda je také vhodná v případě, kdy chceme znát průběh odchodu z registru (nebo i průběh velikosti dopadu programu) u konkrétní podskupiny zaměstnaných za konkrétních podmínek.

Modelování dopadu pomocí Coxových modelů proporcionálních rizik přináší výhodu snadné kontroly efektu rušivých proměnných, nicméně má řadu nevýhod spočívajících zejména v zacházení s časem. V základní podobě Coxův model informací o čase nepodává – efekty jsou v čase „zprůměrovány“ a rozdíly v rizicích mezi posuzovanými skupinami nezaměstnaných jsou tak proporcionálně stejné bez ohledu na běh času. Předpoklad proportionality je možné uvolnit rozšířením modelu o lineární interakci velikosti dopadu s časem, nicméně narůstá tak pracnost a díky lineárnímu předpokladu nezískáváme reálný průběh dopadu. Druhá slabina Coxova modelu spočívá v proporcionálním a nikoli absolutním zhodnocení rizik – velikost rizik určité skupiny je vždy posuzována relativně vzhledem k jiné skupině, přičemž možnost zhodnotit reálnou velikost rizika odchodu určité skupiny v určitém časovém bodě je nemožná.

Vypořádání se s těmito problémy nabízí metoda modelování rizik v diskrétním čase, která umožňuje uvolnit jak předpoklad proportionality rizik, tak předpoklad lineární změny dopadu v čase. Výměnou za to však umožňuje posouzení změny dopadu v čase pouze v určitých časových bodech, např. po měsících. Je tedy zřejmé, že každá metoda má své silné a slabé stránky a výběr metody závisí na evaluátorových odpovědích na výše uvedené otázky.

Závěr

Cílem této studie bylo definovat a diskutovat základní oblasti, principy a postupy hodnocení výsledků a dopadů programů aktivní politiky zaměstnanosti s využitím administrativních dat systému OKpráce. Důraz byl kladen zejména na diskuzi postupů tak, aby výsledkem jejich použití byl validní indikátor dopadu. Zabývali jsme se zejména interní validitou naměřeného dopadu, tedy do jaké míry různé postupy při evaluaci přinášejí výsledky, jež je prostý různých zdrojů zkreslení, a utváří tak adekvátní představu o dopadu programu. Jako prostředek pro dodatečně vybalancování před-programových rozdílů mezi skupinou intervence a kontrolní skupinou, a tudíž i způsob zvýšení validity diskutujeme proces párování, přičemž zmiňujeme důležitost zahrnutí faktorů před-programová perspektiva a historie nezaměstnaných na trhu práce a fáze hospodářského cyklu do tohoto procesu. V poslední části jsme diskutovali volbu konkrétní techniky analýzy dat. Uzavíráme, že výběr techniky závisí na tom, zda chce evaluátor zjistit velikost absolutního nebo relativního rizika, jeho průběh v čase anebo pouze přítomnost nezaměstnaného v registru.

Zdroje

- [1] Betcherman, G., Olivas, K., Dar, A. (2004) *Impacts of Active Labor Market Programs: New Evidence from Evaluations with Particular Attention to Developing and Transition Countries*. Washington: The World Bank [online]. [cit. 2015-02-20]. Dostupné z: <http://siteresources.worldbank.org/SOCIALPROTECTION/Resources/SP-Discussion-papers/Labor-Market-DP/0402.pdf>
- [2] Bonoli, G. (2010) *The political economy of active labour market policy*. Edinburgh: RECOWE Publication, Dissemination and Dialogue Centre [online]. [cit. 2015-02-20]. Dostupné z: http://www.sps.ed.ac.uk/_data/assets/pdf_file/0010/39268/REC-WP_0110_Bonoli.pdf
- [3] Blache, G. (2011) *Active Labour Market Policies in Denmark: A Comparative Analysis of Post-Program Effects*. [online]. [cit. 2015-02-20]. Paris: CES. Dostupné z: <ftp://mse.univ-paris1.fr/pub/mse/CES2011/11071.pdf>
- [4] Borland, J., Tseng, Y., Wilkins, R. (2005) *Experimental and quasi-experimental methods of microeconomic program and policy evaluation*. Melbourne: Melbourne Institute of Applied Economic and Social Research [online]. [cit. 2015-02-20]. Dostupné z: http://cf.fbe.unimelb.edu.au/staff/jib/documents/Rog_YiP_Jeff_WPJan05.pdf
- [5] Bring, J., Kenneth Carling, K. (2001) Attrition and misclassification of drop-outs in the analysis of unemployment duration. Uppsala: IFAU. [online]. [cit. 2015-02-20]. Dostupné z: http://www.researchgate.net/publication/5095711_Attrition_and_misclassification_of_drop-outs_in_the_analysis_of_unemployment_duration
- [6] Bryson, A., Dorsett, R., Purdon, S. (2002) *The use of propensity score matching in the evaluation of active labour market policies*. London: Policy Studies Institute and National Centre for Social Research. [online]. [cit. 2015-02-20]. Dostupné z: http://eprints.lse.ac.uk/4993/1/The_use_of_propensity_score_matching_in_the_evaluation_of_active_labour_market_policies.pdf
- [7] Calmfors, L. (1994) *Active labour market policy and unemployment - a framework for the analysis of crucial design features*. Paris: OECD. [online]. [cit. 2015-02-20]. Dostupné z: <https://search.oecd.org/eco/growth/33936463.pdf>
- [8] Calmfors, L., Forslund, A., Hemström, M. (2002) *Does active labour market policy work? Lessons from the Swedish experiences*. Uppsala: IFAU. [online]. [cit. 2015-02-20]. Dostupné z: <http://www.ifau.se/upload/pdf/se/2002/wp02-04.pdf>
- [9] Card, D., Kluve, J., Weber, A. (2009) *Active Labor Market Policy Evaluations: A Meta-Analysis*. Bonn: Institute for the Study of Labor. [online]. [cit. 2015-02-20]. Dostupné z: <http://www.nber.org/papers/w16173>

- [10] Card, D., Ibarrán, P., Villa, J. (2011) *Building in an Evaluation Component for Active Labour Market Programs: a Practitioner's Guide*. Washington: Inter-American Development Bank. [online]. [cit. 2015-02-20]. Dostupné z: <http://publications.iadb.org/bitstream/handle/11319/5349/Building%20in%20an%20Evaluation%20Component%20for%20Active%20Labor%20Market%20Programs.pdf?sequence=1>
- [11] de Koning, J., Peers, Y. (2007) *Evaluating Active Labour Market Policies Evaluations*. Rotterdam: SEOR.
- [12] Dehejia, R. H., Wahba, S. (2002) „Propensity score matching methods for nonexperimental causal studies“ [online]. *The review of Economics and Statistics* 84(1): 151-161. [cit. 2015-02-20]. Dostupné z: <https://wagner.nyu.edu/files/faculty/publications/matching.pdf>
- [13] Dias, M., Ichimura, H., van den Berg, G. (2008) *The Matching Method for Treatment Evaluation with Selective Participation and Ineligibles*. Bonn: Institute for the Study of Labor. [online]. [cit. 2015-02-20]. Dostupné z: <http://ftp.iza.org/dp3280.pdf>
- [14] Fay, R. (1996). *Enhancing the effectiveness of active labour market policies: Evidence from Programme Evaluation in OECD countries*. OECD: Paris. [online]. [cit. 2015-02-20]. Dostupné z: <http://www.oecd-ilibrary.org/docserver/download/5lgsjhvj7tjl.pdf?expires=1424684458&id=id&accnam=e=guest&checksum=2CC5D447CE302083BB12D1CF43C8BA6E>
- [15] Forslund, A., Fredriksson, P., Vinkström, J. (2011) *What active labour market policy works in recession?* Uppsala: IFAU [online]. [cit. 2015-02-20]. Dostupné z: <http://www.ifau.se/Upload/pdf/se/2011/wp11-02-What-active-labor-market-policy-works-in-a-recession.pdf>
- [16] Heckman, J. J. and J. A. Smith (1996) „Experimental and Non-Experimental Evaluation“. In: Schmid, G. O'Reilly, J., Schomann K. (eds), *International Handbook of Labour Market Policy and Evaluation*. Cheltenham: Edward Elgar.
- [17] Hora et al. (2009) *Hodnocení programů aktivní politiky zaměstnanosti realizovaných v roce 2007 se zaměřením na rekvalifikace (struktura, cílenost, krátkodobé a střednědobé efekty na opuštění evidence)* Praha: VÚPSV [online]. [cit. 2015-02-20]. Dostupné z: http://praha.vupsv.cz/Fulltext/vz_300.pdf
- [18] Hora, O., Suchanec, M. (2014). „Zhodnocení programů aktivní politiky zaměstnanosti realizovaných v České republice v období krize“. In: Sirovátka T., Horáková M., Horák P. (eds), *Česká politika zaměstnanosti v době krize a po krizi*. Brno/Boskovice: Masarykova Univerzita/Albert. s. 143-182.
- [19] Hujer, R., Caliendo, M. (2000) *Evaluation of Labour Market Policy: Methodological Concepts and Empirical Estimates*. Bonn: Institute for the Study of Labor [online]. [cit. 2015-02-20]. Dostupné z: <http://ftp.iza.org/dp236.pdf>

- [20] Hujer, R., Wellner, M. (2000) *The Effects of Public Sector Sponsored Training on Individual Employment Performance in East Germany*. Bonn: Institute for the Study of Labor [online]. [cit. 2015-02-20]. Dostupné z: <http://ftp.iza.org/dp141.pdf>
- [21] Hujer, R., Caliendo, M., Radić, D. (2004) „Methods and Limitations of Evaluation and Impact Research“ In: Descy, P., Tessaring, M. (eds), *The Foundations of Evaluation and Impact Research*. Luxembourg: Cedefop/Office for Official Publications of the European Communities. s. 131-190.
- [22] Hujer, R., Thomsen, S., Zeiss, Ch. (2006) *The Effects of Short-Term Training Measures on the Individual Unemployment Duration in West Germany*. Mannheim: ZEW [online]. [cit. 2015-02-20]. Dostupné z: <ftp://ftp.zew.de/pub/zew-docs/dp/dp06065.pdf>
- [23] Kluve, J. et al. (2005). *Study on the effectiveness of ALMPs*. Essen: RWI [online]. [cit. 2015-02-20]. Dostupné z: http://www.rwi-essen.de/media/content/pages/publikationen/rwi-projektberichte/PB_ALMP.pdf
- [24] LaLonde, R. J. (1986) „Evaluating the Econometric Evaluations of Training Programs with Experimental Data“ [online]. *The American Economic Review* 76(4): 604-620. [cit. 2015-02-20]. Dostupné z: <http://www.jstor.org/stable/pdf/1806062.pdf?acceptTC=true>
- [25] Lechner, M., Wunsch, C. (2009) *Active Labour Market Policy in East Germany: Waiting for the Economy to Take Off*. St. Gallen: University of St. Gallen. [online]. [cit. 2015-02-20]. Dostupné z: http://www1.vwa.unisg.ch/RePEc/usg/dp2006/DP24_Le.pdf
- [26] Mohr, L. (1992) *Impact analysis for program Evaluation*. Newbury park, London, New Delhi: Sage publications.
- [27] Puhani, P. (1998) *Advantage through training? A microeconomic evaluation of the employment effects of active labour market programmes in Poland*. Leibniz: ZEW [online]. [cit. 2015-02-20]. Dostupné z: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=141462
- [28] Reinowski, E., Schultz, B. (2006). *Microeconomic Evaluation of Selected ESF-funded ALMP-Programmes*. Halle: Halle Institute for Economic Research [online]. [cit. 2015-02-20]. Dostupné z: <http://www.iwh-halle.de/d/publik/disc/17-06.pdf>
- [29] Richardson, K., van den Berg, G. (2006) *Swedish Labor Market Training and the Duration of Unemployment*. Bonn: Institute for the Study of Labor [online]. [cit. 2015-02-20]. Dostupné z: <http://ftp.iza.org/dp2314.pdf>
- [30] Sianesi, Barbara (2003) *An evaluation of the Swedish system of Active Labour Market Programmes in the 1990s*. London: The Institute for Fiscal Studies.

- [31] Smith, J., Todd P. (2005) „Does matching overcome LaLonde’s critique of nonexperimental estimators?“ [online]. *Journal of Econometrics*. 125(1-2): 305–353. [cit. 2015-02-20]. Dostupné z: http://ac.els-cdn.com/S030440760400082X/1-s2.0-S030440760400082X-main.pdf?_tid=e557ccd2-bb59-11e4-8b56-00000aabb0f02&acdnat=1424695730_7b4c09dad70db27e20bd76d106ccbdbf
- [32] Soukup, T. (2006) *Early assessment a profiling ve službách zaměstnanosti. Zahraniční zkušenosti a spolehlivost odhadu v ČR*. Praha: VÚPSV [online]. [cit. 2015-02-20]. Dostupné z: http://praha.vupsv.cz/Fulltext/vz_202.pdf
- [33] Soukup, T., Michalička, L., Kotíková, J. (2009) *Třídění uchazečů na úřadech práce – řešení problematiky cílení APZ a poradenství*. Praha: VÚPSV [online]. [cit. 2015-02-20]. Dostupné z: http://praha.vupsv.cz/Fulltext/vz_287.pdf
- [34] Suchanec, M. (2011) „Využití Booleho přístupu v komparativní analýze: příklad srovnání 27 zemí EU z hlediska zaměstnanosti žen“. *Fórum sociální politiky*5(4):17-19
- [35] Suchanec, M. (2014a) *Jak se metodologické volby v procesu evaluace dopadu projevují při evaluaci dopadů české APZ prostřednictvím analýzy datové matice OK práce?* Brno: FSS MU.
- [36] Suchanec, M. (2014b) „Evaluace dopadu“. In: Hora, O., Suchanec, M., Žížlavský, M. *Evaluační výzkum*. Brno: MUNI. s. 85-108.
- [37] van Ours, J. (2002) *The Locking-in Effect of Subsidized Jobs*. Michigan: The William Davidson Institute. [online]. [cit. 2015-02-20]. Dostupné z: <http://wdi.umich.edu/files/publications/workingpapers/wp474.pdf>
- [38] Vermunt, J. (1996) *Log-linear event history analysis: a general approach with missing data, latent variables, and unobserved heterogeneity*. Tilburg: Tilburg University Press [cit. 2015-02-20]. Dostupné z: <http://members.home.nl/jeroenvermunt/thesis.pdf>
- [39] Wimer, Ch. (2006) *Learning from Small-Scale Experimental Evaluations of After School Programs*. Cambridge: Harvard Graduate School of Education. [online]. [cit. 2015-02-20]. Dostupné z: <http://www.hfrp.org/publications-resources/browse-our-publications/learning-from-small-scale-experimental-evaluations-of-after-school-programs>